أســس
التحليل الإحصائي متعدد المتغيرات
Principles Of
Multivariate Statistical Analysis

الدكتور

إبراهيم محمد العلي

أستاذ في قسم الإحصاء والبرمجة

كلية الاقتصاد- جامعة تشرين- سورية

اللاذقية – سورية

2020 م – 1441 هـ

الفهرس العام

المقدمة :

إن الأبحاث الحديثة في مجالات العلوم المختلفة، أصبحت تتناول عدة متحولات كمية أو نوعية ، وينتج عنها بيانات متعددة . وإن معالجة هذه البيانات المتعددة أدت إلى ظهور فرع جديد في الإحصاء هو التحليل الإحصائي متعدد المتغيرات (Multivariate Statistical Analysis)، وإن هذا التحليل يقدم للباحثين طرائق إحصائية متطورة لتوصيف وتحليل واختبار تلك البيانات المتعددة، وتحديد العلاقات بينها والتنبؤ بسلوكها من أجل العمل على إدارتها والتحكم بها .

ويمكن القول أن التحليل الإحصائي متعدد المتغيرات هو تعميم للتحليل العادي لمتحول واحد أو لمتحولين، وعندما يعمل الإنسان بهذا التحليل ينقل من السير على الطريق إلى الطيران في الفضاء .

وإني عندما اطلعت على بعض جوانب هذا الفضاء ترددت في الإقدام عليه، تخوفاً من الضياع فيه ، ولكني قررت أخيراً أن أدخل إليه متسلحاً بمعارفي الرياضية اللازمة ومعتمداً على وسائلي العلمية المتوفرة، وأن أقوم بتأليف كتاب مرجعي عنه، لأضعه بين أيدي الباحثين العرب للاستفادة منه في أبحاثهم العلمية ، من خلال تطبيق طرائقه المختلفة في جميع المجالات الحياتية والاقتصادية والاجتماعية والبيولوجية والطبيعية.. الخ، وآمل أن أكون قد وفقت في ذلك.

ورأيت أن أصنف موضوعات هذا الكتاب ضمن خمسة أجزاء تتضمن 23 فصلاً، وهي كما يلي:

الجزء الأول: قضايا المصفوفات (فصلان) .

الجزء الثاني: الاستدلال الإحصائي (7 فصول) .

الجزء الثالث: الانحدار الخطي المتعدد والارتباط القانوني (3 فصول) .

الجزء الرابع: التحليل العاملي (العواملي) (4 فصول) .

الجزء الخامس: قضايا التمييز والتصنيف (7 فصول) .

وهنا أود أن أطمئن القارئ الكريم بأنه يستطيع أن يقوم بدراسة هذا التحليل وفهمه وتطبيقه، إذا كانت لديه خلفية رياضية بسيطة، ومعرفة عامة بنظرية الاحتمالات والإحصاء الرياضي، ودراية عامة بقضايا المصفوفات وقيمها الذاتية...الخ. لذلك رأيت أن أخصص الجزء الأول منه لقضايا المصفوفات، لاستعراض أنواعها والتعريف بخواص العمليات المعرفة عليها وكيفية الحصول على قيمها وأشعتها الذاتية. وأنصح القارئ الكريم أن يبدأ بمراجعة هذا الجزء قبل الدخول إلى فصول التحليل المتتالية، والعودة اليه عند كل حاجة. ولتسهيل فهم الأفكار الواردة في الفصول المختلفة وضعت في مستهل كل فصل تمهيداً خاصاً لشرح بعض المفاهيم اللازمة له، ثم دعمته بعدد من الأمثلة التطبيقية لتوضيح معاني الطرائق الرياضية المستخدمة فيه وشرحت كيفية تطبيقها على المشكلات الحياتية الفعلية .

وإني إذ أقدم هذا الكتاب للباحثين العرب وأنشره لهم على عدة مواقع الكترونية، لا أدعي فيه كمالاً ولا شمولاً، وأعتذر من القراء الكرام عن أي خطأ لم ألاحظه، وعن أي سهو لم أتداركه، وعن أي نقص لم

أعالجه، وآمل منهم أن يوافوني بآرائهم فيه وبملاحظاتهم عليه، حتى أتمكن من تصحيح تلك الهفوات المتوقعة . كما أرجو منهم الإشارة إليه عند الاستفادة منه .

وختاماً أتقدم بالشكر الجزيل للزملاء الذين ساعدوني في تدقيق ونشر هذا الكتاب وهم السادة الدكاترة: عبد الهادي الرفاعي، محمد عكروش، أيمن العشعوش، عز الدين حيدر، طالب أحمد، فادي خليل، محمود حسين، وسيم أحمد، ياسر علوش، رسلان العلي، ولطلاب الدكتوراة: لمى منلا، علا خدوج ، خضر عكاري، كما أخص بالشكر الآنسة سوزان صقر التي قامت على طباعته بكل صبر وإتقان .

والله من وراء القصد

اللاذقية في 2020/02/15

المؤلف

أ. د. إبراهيم محمد العلي

أسس

التحليل الإحصائي متعدد المتغيرات

الجزء الأول

قضايا المصفوفات

الفصل الأول

المصفوفات والعمليات المعرفة عليها

1-1 : الأشعة :

تلعب الأشعة والمصفوفات دوراً كبيراً في التحليل متعدد المتغيرات. وبدون أن نشرح أنواعها وخواصها نعرضها وفق التالي:

- تعريف الشعاع: يتألف الشعاع المتعدد X من عدة متحولات كمية نرمز لها بـ $X_1 , X_2 , \dots , X_P$،

ونكتبها على شكل عمود X كما يلي:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_P \end{bmatrix}$$

كما نرمز لمنقوله بـX' ،الذي يكتب على شكل سطر كما يلي:

$$X' = [X_1 , X_2 , X_3 , \dots , X_P] \qquad (1-1)$$

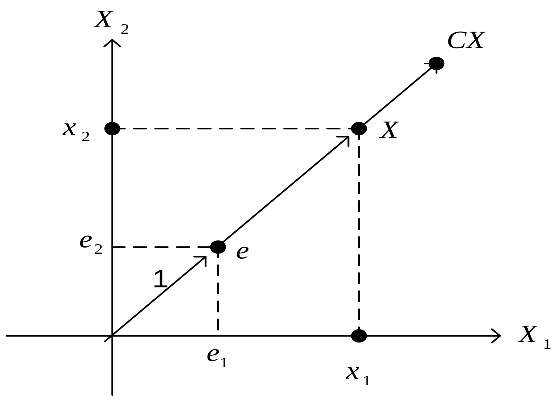
ويعرف الشعاع البسيط(العددي) X في الفضاء R^P بعلاقة مشابهة وبرموز مماثلة (وسنرمز لمركباته العددية بـ x_j) . وبناءً على ذلك نعرف مربع طوله من خلال الجداء الشعاعي التالي:

$$\|X\|^2 = X' * X = (x_1 , x_2 , \dots , x_P) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} = x_1^2 + x^2 + \cdots + x_p^2$$

وبذلك يكون طوله $\|X\|$ مساوياً لـ :

$$\|X\| = \sqrt{X' * X} = \sqrt{x_1^2 + x^2 + \cdots + x_p^2} \qquad (2-1)$$

وإن القيم العددية لهذه المركبات تمثل إحداثيات النقطة X في الفضاء R^P، فإذا كان $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ فإنه يرسم في المستوى كما يلي :



الشكل (1-1) تمثيل وتمديد وتقليص الأشعة في المستوى

ويمكن تمديد وتقليص الشعاع X بضربه بثابت C، فنحصل على الشعاع CX، كما يمكن تحويل الشعاع X إلى الشعاع الواحدي الممعير $\boldsymbol{e}$، الذي يكون طوله مساوياً للواحد ويكون اتجاهه باتجاه X ، كما في الشكل (1−1)، ونحصل على مركبات e بتقسيم مركبات X على طوله ℓ_x .

وللحصول على مركبات الشعاع الواحدي e من الشعاع X نقوم بما يلي:

1−نحسب طول الشعاع X من العلاقة (1−2) كما يلي:

$$\ell_x = \sqrt{X' * X} = \sqrt{x_1^2 + x_2^2 + x_3^2 + \cdots + x_p^2} = \|X\|$$

2− ثم نقسم كل مركبة من مركبات X على طوله ℓ_x فنحصل على الشعاع الواحدي الممعير التالي:

$$\boldsymbol{e} = \begin{bmatrix} \frac{x_1}{\ell_x} \\ \frac{x_2}{\ell_x} \\ \vdots \\ \frac{x_p^2}{\ell_x} \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_P \end{bmatrix} \qquad (3-1)$$

وللتأكد من أن طول هذا الشعاع يساوي الواحد نجد أن :

$$\|\boldsymbol{e}\| = \ell_e = \sqrt{\boldsymbol{e}' * \boldsymbol{e}} = \sqrt{\left(\frac{x_1}{\ell_x}\right)^2 + \left(\frac{x_2}{\ell_x}\right)^2 + \cdots + \left(\frac{x_P}{\ell_x}\right)^2} = 1$$

ويستخدم الشعاع الواحدي $\boldsymbol{e}$ في التعبير عن الحلول المختلفة في التحليل متعدد المتغيرات ويدخل في حساب الأشعة الذاتية للمصفوفات.

- خاصة التعامد: إذا كان X وY شعاعان متعامدان (في نفس الفضاء) فإن جداءهما السلمي يساوي الصفر، أي إذا كان X وY متعامدان فإن:

$$X' * Y = Y' * X = 0 \qquad (4-1)$$

ويمكن كتابة هذا الجداء على الشكل التالي:

$$X' * Y = x_1y_1 + x_2y_2 + x_3y_3 + \cdots + x_py_p = 0 \qquad (5-1)$$

أما إذا كان X وY غير متعامدين فإن جداءهما يساوي:

$$X' * Y = Y' * X = \ell_x\ell_y \cos\theta \qquad (6-1)$$

حيث أن θ هي الزاوية المحصورة بينهما .

1−2 : أنواع المصفوفات وخواصها:

- تعريف المصفوفة: تعرف المصفوفة بأنها جملة من الأعداد الحقيقية أو المركبة، الموضوعة على شكل أسطر وأعمدة ضمن مستطيل (أو مربع) مؤلف من p سطراً و n عموداً، ونرمز لها بالرمز A_{P*n}، (ويسمى $p * n$ مرتبة (order) المصفوفة A) ونكتبها كما يلي:

$$A_{P*n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix} \qquad (7-1)$$

- منقول المصفوفة A: ونرمز له بـ A' ويتم الحصول عليه باستبدال أسطرها بأعمدتها أو أعمدتها بأسطرها. فنحصل على مصفوفة جديدة من المرتبة $n * p$ ونرمز لها بـ A'_{n*p}
- الجمع والطرح: يمكن أن نجمع أو نطرح أي مصفوفتين إذا كانتا من نفس المرتبة ، وذلك بجمع أو طرح عناصرهما المتقابلة، فنحصل على مصفوفة ثالثة من نفس المرتبة كما يلي :

$$A_{P*n} \pm B_{P*n} = C_{P*n} \qquad (8-1)$$

- الضرب: يمكن ضرب أي مصفوفتين إذا كان عدد أعمدة الأولى (اليسرى) يساوي عدد أسطر اليمنى، فنحصل على مصفوفة جديدة ، ويتم حساب عناصرها على الترتيب من جراء ضرب كل سطر من الأولى بعمود من الثانية (ضرب أشعة) كما يلي:

$$A_{P*n} * C_{n*k} = B_{P*k} \qquad (9-1)$$

- القسمة:إن عملية القسمة على المصفوفات غير معرفة وتستبدل بالضرب في مقلوب المصفوفة.

1-2-1-المصفوفات المربعة: ولها عدة أنواع هي:

- المصفوفة النظامية (غير الشاذة unsingular): وهي المصفوفة المربعة A التي تكون قيمة محددها $|A|$ غير معدومة . أي التي يكون محددها (انظر المحددات لاحقاً):

$$|A| \neq 0 \qquad (10-1)$$

وهذا يعني أن أعمدة (أو أسطر) المصفوفة تكون مستقلة عن بعضها البعض .

- المصفوفة الشاذة (singular): وهي المصفوفة المربعة A التي تكون قيمة محددها $|A|$ معدومة . أي التي يكون محددها (انظر المحددات لاحقاً):

$$|A| = 0 \qquad (10a-1)$$

وهذا يعني أن أعمدة (أو أسطر) المصفوفة تكون غير مستقلة عن بعضها البعض .

- مقلوب المصفوفة المربعة A: ونرمز له بـ A^{-1} وهو مصفوفة مربعة تحقق العلاقة التالية:

$$A * A^{-1} = A^{-1} * A = I \qquad (11-1)$$

حيث I هي المصفوفة الواحدية من مرتبة A (انظر لاحقاً) . ويتم حساب A^{-1} من العلاقة:

$$A^{-1} = \frac{1}{|A|} * M'$$

حيث أن M' هي منقول المصفوفة المساعدة المؤلفة من مصغرات المصفوفة A مع تناوب الإشارة .

- المصفوفة المتناظرة: هي مصفوفة مربعة تكون عناصرها المتناظرة بالنسبة لقطرها الرئيسي متساوية، أي يكون فيها :

$$a_{ij} = a_{ji} \qquad (i \neq j) \qquad (12-1)$$

ومن التعريف نستنتج أنه إذا كانت A متناظرة فإن $A' = A$.

- المصفوفة القطرية: وهي مصفوفة مربعة تكون جميع عناصرها الواقعة خارج قطرها الرئيسي معدومة. ونرمز لها بـ D_{iag} حيث أن :

$$D_{iag}=\begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \quad : \; a_{ij}=0 \quad i\neq j \qquad (13-1)$$

ونشير هنا إلى أن مقلوب هذه المصفوفة القطرية يساوي مصفوفة قطرية أيضاً، وإن عناصر قطرها الرئيسي تساوي مقاليب عناصر المصفوفة القطرية الأصلية . أي أن:

$$D_{iag}^{-1}=\begin{bmatrix} \frac{1}{a_{11}} & 0 & 0 & 0 \\ 0 & \frac{1}{a_{22}} & 0 & 0 \\ 0 & 0 & \frac{1}{a_{33}} & 0 \\ 0 & 0 & 0 & \frac{1}{a_{44}} \end{bmatrix} \quad : \; a_{ij}=0 \quad i\neq j \quad a_{ii}\neq 0 \qquad (14-1)$$

- المصفوفة الواحدية: ونرمز لها بـ I وهي مصفوفة قطرية وتكون عناصرها القطرية تساوي الواحد، ونكتب ذلك على الشكل :

$$I=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (15-1)$$

ومن (1−15) نستنتج أن: $I^{-1}=I$ وأن: $I'=I$.

- المصفوفة المتعامدة : ونرمز لها بـ Q وهي التي تكون الأشعة المؤلفة لأعمدتها متعامدة، وهي تحقق العلاقة التالية:

$$Q' * Q = Q * Q' = I \qquad (16-1)$$

ومنها ومن العلاقة (1−11) نستنتج أن:

$$Q' = Q^{-1} \qquad (17-1)$$

أي أن Q^{-1} مقلوب المصفوفة المتعامدة Q يساوي منقولها Q' .

وتتميز هذه المصفوفات بأن جداءات الأعمدة فيها معدومة (لأنها متعامدة) أي أن:

$$q_i' * q_j = 0 \qquad for \;\; i\neq j \qquad (18-1)$$

- المصفوفة المتعارضة: نقول عن المصفوفة C إنها متعارضة إذا كانت عناصرها مؤلفة من 1 و 0 وكان مجموع عناصر كل سطر فيها يساوي الواحد، وكانت أسطرها مستقلة مثل:

$$C=\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad \mathrm{C}=\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

3-1 : خواص العمليات على المصفوفات :

أ- خواص عمليات الجمع والطرح: إذا كانت A وB وC مصفوفات من نفس المرتبة $m * n$ وكان e و d عددين حقيقيين، فإنها تحقق الخواص التالية:

$$1 - A \pm B = B \pm A$$
$$2 - (A \pm B) \pm C = A \pm (B \pm C) \qquad (19-1)$$
$$3 - C(A \pm B) = CA \pm CB$$
$$4 - (e + d)A = eA + dA$$
$$5 - (ed)A = e(dA)$$
$$6 - (A + B)' = A' + B'$$

ب- خواص الضرب: إذا كانت A وB وC متوافقة في المراتب لإجراء الضرب، وكان e عدداً، فإنها تحقق الخواص التالية(إن عملية القسمة على مصفوفة A غير معرفة وتستبدل بالضرب بمقلوبها):

$$1 - e(A * B) = (eA) * B = A * (eB)$$
$$2 - A(B * C) = (AB) * C \qquad (20-1)$$
$$3 - A(B + C) = AB + AC$$
$$4 - (B + C) * A = B * A + C * A$$
$$5 - A * B \neq B * A \qquad \left(\text{ولكن بصورة عامة}\right)$$
$$6 - e * A * B = A * (e * B) \qquad : \quad (e = \text{عدد})$$
$$7 - (A * B)' = B' * A'$$
$$8 - (A * B)^{-1} = B^{-1} * A^{-1}$$

ج-خواص المنقول : إذا كانت A وB مصفوفتان متوافقتان في الجمع والضرب ، أي إنهما تحققان الشروط اللازمة لذلك، فإن :

$$1 - (A + B)' = A' + B' \qquad (21-1)$$
$$2 - (e * A)' = eA' \qquad \left(\boldsymbol{e} = \text{عدد}\right)$$
$$3 - (A * B)' = B' * A' \qquad \left(\text{الجداء معكوس}\right)$$
$$4 - (A')' = A$$

د- خواص المقلوب : وهي خواص مشروطة بالمصفوفات المربعة:

إذا كانت A وB مربعتين ونظاميتين فيكون لهما مقلوبان A^{-1} وB^{-1}، وإذا كانتا متوافقتان من حيث الضرب ، فإنهما تحققان الخواص التالية :

$$`1 - A * A^{-1} = A^{-1} * A = I \qquad (22-1)$$
$$2 - (A^{-1})^{-1} = A$$
$$3 - (A * B)^{-1} = B^{-1} * A^{-1} \qquad \left(\text{الجداء معكوس}\right)$$
$$4 - (A^{-1})' = (A')^{-1}$$
$$5 - (eA)^{-1} = \frac{1}{e} * A^{-1} \qquad \left(e = \text{عدد}\right)$$

4-1 : المحددات (المعينات):

- تعريف المحدد: يعرف المحدد على المصفوفات المربعة فقط ، وهو عدد يعبر عن قيمة كامل المصفوفة، ويرمز له بـ $|A|$، ويمكن حساب قيمته بنشر المحدد حسب الأسطر أو الأعمدة وحساب الجداءات اللازمة حسب الطرائق المعروفة.
- خواص المحددات (المعينات) : إذا كانت A وB مصفوفتين مربعتين من المرتبة $n * n$ فإن محدديهما يحققان الخواص التالية:

$$1-\ |A| = |A'| \qquad (23-1)$$
$$2-|A * B| = |A| * |B|$$
$$3-|eA| = e^n|A|$$ (عدد حقيقي e)
$$4-|A * A^{-1}| = 1$$
$$5-|A * A^{-1}| = |A| * |A^{-1}| = 1$$
$$6-|A^{-1}| = \frac{1}{|A|} \quad \Rightarrow |A| = \frac{1}{|A^{-1}|}$$

إذا كانت عناصر أحد الأسطر أو الأعمدة معدومة فإن: $7-|A| = 0$

إذا كانت عناصر سطرين(أو عمودين) متناسبة فإن: $8-|A| = 0$

وعندها تكون المصفوفة A شاذة .

5-1 : رتبة المصفوفة المربعة أو المستطيلة :

إن رتبة المصفوفة A حسب الأسطر هي أكبر عدد من الأسطر المستقلة خطياً فيها (باعتبارها أشعة)، وإن رتبة المصفوفة A حسب الأعمدة هي أكبر عدد من الأعمدة المستقلة خطياً فيها (باعتبارها أشعة) إن رتبة المصفوفة A حسب الأسطر = رتبة المصفوفة حسب الأعمدة. ويدخل مفهوم الرتبة في تعريف المصفوفات النظامية وفق التالي: تكون المصفوفة المربعة نظامية (unsingular) إذا كانت رتبتها تساوي عدد أسطرها (أو عدد أعمدتها) .

- تعريف: إننا نقول عن المصفوفة المربعة A بأنها نظامية إذا كانت المعادلة $AX = 0$، تقتضي أن يكون $X = 0$.

وإذا لم تنجح المصفوفة A في أن تكون نظامية فإنها تسمى مصفوفة شاذة (singular) .

وهنا نلاحظ أن المعادلة $AX = 0$ يمكن كتابتها كما يلي:

$$\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}$$
$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_p & a_{p2} & \cdots & a_{pn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + a_2x_2 + \cdots a_nx_n = 0 \quad (24-1)$$

حيث a_1 وa_2 و..a_n هي الأشعة العمودية في A .

وهذا يعني أنه إذا كانت المعادلة (1−24) لا تتحقق إلا إذا كانت جميع $x_i = 0$، أي ان الأشعة a_i تكون مستقلة عن بعضها البعض وإن المصفوفة تكون من المرتبة n .

إذا كانت A مصفوفة مربعة نظامية من المرتبة $n * n$ فإنه يوجد مصفوفة وحيدة B من المرتبة $n * n$ أيضاً بحيث يكون لدينا:

$$A * B = B * A = I \qquad (82-2)$$

حيث I من المرتبة $n * n$ ، وعندها تسمى المصفوفة B بمقلوب المصفوفة A ونرمز لها بـA^{-1} ونكتب ذلك كما يلي:

$$B = A^{-1} \qquad (83-2)$$

- خواص رتبة المصفوفة (rank) : وهي تختلف عن مرتبة المصفوفة (order)، وتعرف الرتبة بأنها تساوي أكبر عدد من الأعمدة (أو الأسطر) المستقلة في A .

وتكون رتبة المصفوفة من المرتبة $n * n$ كاملة، إذا كانت رتبتها $r = n$، وفي هذه الحالة يكون محدد المصفوفة $|A| \neq 0$، ويكون المقلوب A^{-1} موجوداً . أما بالنسبة للمصفوفات المستطيلة A من المرتبة $m * n$ ، فإن رتبتها تكون أقل أو تساوي أصغر العددين n أو m ، أي أن:

$$rank(A) \leq min(n\,, m) \qquad (24-1)$$

ومن خواص الرتبة إنها لا تتغير إذا تم ضرب المصفوفة A من اليسار أو اليمين بأية مصفوفة نظامية B أو بأي عدد حقيقي e .

ومن أهم خواصها ما يلي:

$$1 - rank(A) = rank(A') = rank(A' * A) = rank(A * A')$$

$$2 - rank(I) = n \qquad (25-1)$$

$$3 - rank(A * B) \leq (B) \qquad \left(\text{موجود } A * B\right)$$

$$4 - rank(A) = rank(A * B) = r(B) \qquad \left(\text{نظامية } B\right)$$

إذا كانت A وB متكافئتين فإن:

$$5 - rank(A) = rank(B)$$

إذا كانت B_{n*n} مربعة ونظامية وC_{m*n} مربعة ونظامية فإن :

$$6 - rank(B * A * C) = rank(A)$$

7--- إذا كانت A مصفوفة قطرية فإن رتبتها تساوي عدد العناصر القطرية غير المعدومة فيها

8- إذا كانت A نظامية من المرتبة $n * n$ فإن رتبتها تكون كاملة وتساوي n لأن $|A| \neq 0$ ولأن A^{-1} يكون موجوداً .

6−1 : أثر المصفوفة: (trace)

إن أثر المصفوفة المربعة A هو مجموع عناصرها القطرية (a_{ii}) ونرمز له بـ $tr(A)$ وهو يساوي:

$$tr(A) = \sum_{i=1}^{k} a_{ii}$$

- خواص أثر المصفوفة: إذا كانت A و B من المرتبة $n * n$ وكان e عدد حقيقي فإن:

$$1 - tr(e * A) = e * tr(A)$$
$$2 - tr(A \pm B) = tr(A) \pm tr(B)$$
$$3 - tr(A * B) = tr(B * A) \qquad : \left(\text{الجداء معكوس}\right)$$
$$4 - tr(B^{-1} * A * B) = \mathrm{tr}(A) \qquad : \left(B \text{ نظامية}\right) \qquad (26-1)$$
$$5 - tr(A * A') = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}^2 \qquad : \left(\text{مجموع مربعات عناصر } A\right)$$

7-1 :القيم الذاتية والمصفوفات الشاذة (singular):

لنفترض أنه لدينا المعادلة النموذجية التالية :

$$A_{k*k} * X_{k*1} = \lambda * X \qquad : X \neq 0 \qquad (28-1)$$

لإيجاد حلول لهذه المعادلة نكتبها كما يلي:

$$[A - \lambda I]X = 0 \qquad : X \neq 0 \qquad (29-1)$$

وحتى تتحقق هذه المعادلة (بفرض أن $X \neq 0$) فإنه يجب أن تكون المصفوفة $[AX - \lambda I]$ مصفوفة شاذة وأن يكون محددها مساوياً للصفر، أي يجب أن يكون لدينا :

$$|A - \lambda I| = 0 \qquad (30-1)$$

وتسمى هذه المعادلة بالمعادلة المميزة للمصفوفة A .

وبعد نشر هذا المحدد نحصل على معادلة مؤلفة من كثير حدود بالنسبة لـ λ، نقوم بإيجاد جذورها الجبرية ونرتبها تنازلياً، فنحصل على ما يسمى بالجذور الكامنة للمصفوفة A أو ما يسمى بالقيم الذاتية لـ A ، ونكتبها كما يلي:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \qquad (31-1)$$

ثم نعوض هذه القيم الذاتية واحدة بعد الأخرى في المعادلة الأساسية (1-29) فنحصل على أن :

$$[A - \lambda I]X = 0 \qquad (32-1)$$

نقوم بحلها فنحصل منها على الحلول الخاصة X، المقابلة لتلك القيم الذاتية ونرمز لها بما يلي:

$$\ell_1, \ell_2, \ell_3, \dots, \ell_n$$

ثم نقوم بمعيرتها فنحصل على الحلول الممعيرة (التي طول كل منها يساوي الواحد) التالية:

$$e_1, e_2, e_3, \dots, e_n$$

وعندها نجد أن كل من هذه الحلول يحقق المعادلة النموذجية الأولى (1-28)، لنأخذ أحد تلك الحلول وليكن الحل ℓ_i ونعوضه في العلاقة 01-28) ثم نعالجه كما يلي: $A * \ell_i = \lambda_i * \ell_i$

نضرب الطرفين من اليسار بـ ℓ_i' ثم نقسم على $||\ell_i||^2$ فنحصل على أن:

$$\ell_i' * A * \ell_i = \lambda_i * \ell_i' * \ell_i = \lambda_i * ||\ell_i||^2 \qquad (33-1)$$

$$\frac{\ell_i'}{||\ell_i||} * A * \frac{\ell_i}{||\ell_i||} = \lambda_i$$

أما بالنسبة للحلول المميزة $\boldsymbol{e_i}$ فنجد من المعادلة (1−28) أن:

$$A * \boldsymbol{e_i} = \lambda_i * \boldsymbol{e_i}$$
$$\boldsymbol{e_i'} * A * \boldsymbol{e_i} = \lambda_i * \boldsymbol{e_i'} * \boldsymbol{e_i} = \lambda_i * 1 = \lambda_i \qquad (34-1)$$

وبمقارنة العلاقتين (1−33) و(1−34) نجد أن:

$$\boldsymbol{e_i} = \frac{\ell_i}{\|\ell_i\|}$$

8−1 : معكوس المصفوفة (شبه المقلوب Pseud Inverce) :

نعلم أن القيم الذاتية للمصفوفة A هي الجذور التربيعية للقيم الذاتية غير المعدومة للمصفوفة $A * A'$ أو للمصفوفة $A' * A$، ويمكننا أن نكتب أية مصفوفة A من المرتبة $m * n$ وفق الصيغة التالية :

$$A = U * \sigma * V' = \sum_{i=1}^{r} \sigma_i * u_i * v_i \qquad (35-1)$$

حيث أن r هي رتبة المصفوفة A، وبذلك يكون $r \le (m\,, n)$، وإن U هي مصفوفة من المرتبة $m * r$، والتي أعمدتها مؤلفة من الأشعة $u_1\,, u_2\,, \dots u_r$، التي هي الأشعة الذاتية اليسرى، وإنها تحقق العلاقة $u' * u = I$ ، حيث I من المرتبة $r * r$، وأن V هي مصفوفة من المرتبة $n * r$، والتي أعمدتها مؤلفة من الأشعة $v_1\,, v_2\,, \dots v_r$، التي هي الأشعة الذاتية اليمينية، وإن V تحقق العلاقة $V' * V = I$، حيث I من المرتبة $r * r$، وأنσ هي المصفوفة القطرية المؤلفة من القيم الذاتية σ_i للمصفوفة A ، أي أن:

$$\sigma_{r*r} = diag(\sigma_1\,, \sigma_2\,, \dots \sigma_r)$$

وبناءً على ذلك تم تعريف معكوس المصفوفة A (شبه المقلوب Pseud – Inverce) على شكل مصفوفة من المرتبة المعكوسة $n * m$، ويُرمز لها بالرمز الخاص التالي $A^{\dagger}$ وهو يساوي :

$$A^{\dagger} = V * \sigma^{-1} * U' = \sum \frac{1}{\sigma_i} v_i * u_i' \qquad (36-1)$$

وتستخدم هذه المصفوفة لإيجاد الحل المثالي X الذي يجعل مربع الخطأ التالي: $\|AX - B\|^2$ أصغر ما يمكن . وإن هذا الحل يساوي :

$$X = A^{\dagger} * B \qquad (37-1)$$

وإذا كانت رتبة A أقل من n، فليس لهذه المسألة حل وحيد بالنسبة لـ X ، وإن تحليل القيمة الذاتية سيعطينا الحل بأصغر قيمة ممكنة للخطأ المذكور.

إن معكوس المصفوفة A يتمتع بالخواص التالية:

$$\begin{aligned} &1 - A * A^{\dagger} * A = A \\ &2 - A^{\dagger} * A * A^{\dagger} = A^{\dagger} \\ &3 - \left(A * A^{\dagger}\right)' = A * A^{\dagger} \\ &4 - \left(A^{\dagger} * A\right)' = A^{\dagger} * A \end{aligned} \qquad (38-1)$$

9-1 : قواعد اشتقاق المصفوفات :

تعتمد عمليات اشتقاق المصفوفات على القواعد التالية :

- قواعد الاشتقاق بالنسبة للمتحول t :

$$1-\frac{dA(t)}{dt}=\left[\frac{da_{ij}(t)}{dt}\right] \qquad (39-1)$$

$$2-\frac{d}{dt}(A\pm B)=\frac{dA}{dt}\pm\frac{dB}{dt} \qquad (40-1)$$

$$3-\frac{d}{dt}(\propto * A)=\propto\frac{dA}{dt} \qquad (41-1)$$

$$4-\frac{d(A*B)}{dt}=\frac{dA}{dt}*B+A*\frac{dB}{dt} \qquad (42-1)$$

- قواعد الاشتقاق بالنسبة للمصفوفة ذاتها فهي:

$$5-\frac{\partial}{\partial X}(C' * X)=C \qquad (43-1)$$

$$6-\ \frac{\partial(X'X)}{\partial X}=2X$$

وإذا كانت A متناظرة فإن:

$$7-\frac{\partial}{\partial X}(X' * A * X)=2AX \qquad (44-1)$$

$$8-\frac{\partial|A|}{\partial A}=(adjount\ of\ A)=|A| * (A^{-1})' \qquad (45-1)$$

وذلك بشرط وجود A^{-1} :

$$9-\frac{\partial}{\partial A}[tr|A'MA|]=MA+M'A \qquad (46-1)$$

10-1 : الصيغة التربيعية والمسافات :

الصيغة التربيعية :

إذا كانت A مربعة ومتناظرة وكان $X'(x_1x_2\ \dots x_n)$ شعاع من مرتبة A، فإننا نقول عن الجداء $X' * A * X$ بأنه الصيغة التربيعية لـ X ونرمز له بـ :

$$Q(X)=X'_{1n} * A_{nn} * X_{n1}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij} * x_i * x_j\ =\ \text{عدد} \qquad (47-1)$$

وكمثال على ذلك نأخذ المثال البسيط التالي :

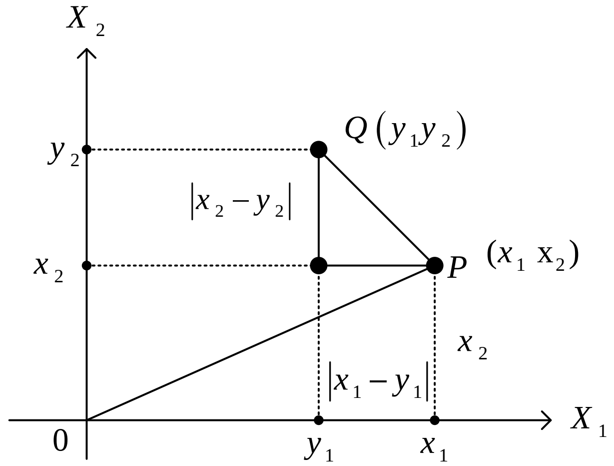
$$Q(X)=[x_1\ x_2]\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=[x_1\ x_2]\begin{bmatrix}a_{11}x_1+a_{12}x_2\\ a_{21}x_1+a_{22}x_2\end{bmatrix}$$

$$Q(X)=a_{11}^2x_1^2+a_{22}x_2^2+2a_{12}^2x_1x_2 \qquad (48-1)$$

وهي تعبر عن مربع المسافة الممعيرة من النقطة $X(x_1x_2)$ حتى مركز الإحداثيات 0.

وبعد تعريف الصيغة التربيعية نستعرض تعريف المسافة بين نقطتين في المستوى أو في الفضاء R^P، وذلك من خلال إحداثيات الشعاع الواصل بينهما .

لنفترض أنه لدينا نقطة $P(x_1 , x_2)$ في المستوى $X_1 0\, X_2$ كما في الشكل (2−1) التالي :



الشكل (2−1): المسافات

فإن المسافة $d(P\,,0)$ من النقطة $P(x_1 , x_2)$ حتى مركز الإحداثيات $O(0\,,0)$ تساوي حسب نظرية فيثاغورث ما يلي:

$$d(P\,,0) = \sqrt{x_1^2 + x_2^2} \qquad (49-1)$$

ولكن المسافة $d(P\,,\mathrm{Q})$ من النقطة $P(x_1 , x_2)$ إلى النقطة $Q(y_1 , y_2)$ تساوي :

$$d(P\,,Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \qquad (50-1)$$

وتسمى هذه المسافات بالمسافات الإقليدية (Euclidean)، لأنها تفترض أن X_1 و X_2 لهما نفس وحدة القياس (متر أو غيره)، وبالتالي فإن المسافة $d(P\,,Q)$ تحسب بنفس تلك الواحدة القياسية (متر)، أما إذا كانت واحدة قياس X_1 (متر) تختلف عن واحد قياس X_2 (كغ)، فإن المقدار الذي تحت الجذر يصبح غير قابل للجمع . وبالتالي تصبح المسافة غير معرفة .

وللتخلص من هذه المشكلة نقوم بمعيرة المتحولين X_1 و X_2 وتحويلها إلى متحولين ممعيرين، وذلك بتقسيم كل منها على انحرافه المعياري $\sqrt{S_{11}}$ و $\sqrt{S_{22}}$، فنحصل على متحولين جديدين مجردين من واحدات القياس نرمز لهما بـ : $Z_2 = \frac{X_2}{\sqrt{S_{22}}}$ $\qquad Z_1 = \frac{X_1}{\sqrt{S_{11}}}$

وبناء على ذلك نقوم بوضع تعريف جديد للمسافة من النقطة $P(Z_1 , Z_2)$ حتى المركز (O) بالعلاقة التالية:

$$d^*(P\,,0) = \sqrt{Z_1^2 + Z_2^2} = \sqrt{\frac{X_1^2}{S_{11}} + \frac{X_2^2}{S_{22}}} \qquad (51-1)$$

وتسمى هذه المسافة $d^*(P\,,0)$ بالمسافة الإحصائية من P حتى المركز O .

وبناء على ذلك يمكننا تعريف المسافة الإحصائية بين أي نقطتين $P(x_1\,, x_2)$ و $Q(y_1\,, y_2)$ في المستوى وحسابها من العلاقة :

$$d^*(P\,, Q) = \sqrt{\frac{(x_1 - y_1)^2}{S_{11}} + \frac{(x_2 - y_2)^2}{S_{22}}} \qquad (52-1)$$

وتسمى هذه المسافة بالمسافة الإحصائية من النقطة P حتى النقطة Q، وبصورة عامة يمكننا تعريف المسافة الإحصائية في الفضاء R^P بين اي نقطتين $P\left(x_1\,, x_2, \ldots . x_p\right)$ و $Q\left(y_1\,, y_2, \ldots . y_p\right)$ بالعلاقة :

$$d^*(P\,, Q) = \sqrt{\frac{(x_1 - y_1)^2}{S_{11}} + \frac{(x_2 - y_2)^2}{S_{22}} + \cdots + \frac{\left(x_p - y_p\right)^2}{S_{pp}}} \qquad (53-1)$$

حيث أن $S_{11}, S_{22}\,, S_{33}, \ldots, S_{pp}$ **هي تباينات المتحولات الإحداثية** $\left(X_1\,, X_2\,, X_3 \ldots . X_p\right)$ **على الترتيب**

وهنا نلاحظ ما يلي :

1- إذا وضعنا الإحداثيات $y_1 = y_2 = y_3 = \cdots = y_p = 0$ في المعادلة (53-1) فإننا نحصل على المسافة من النقطة P حتى المركز (O) في الفضاء R^P ونحصل على المعادلة:

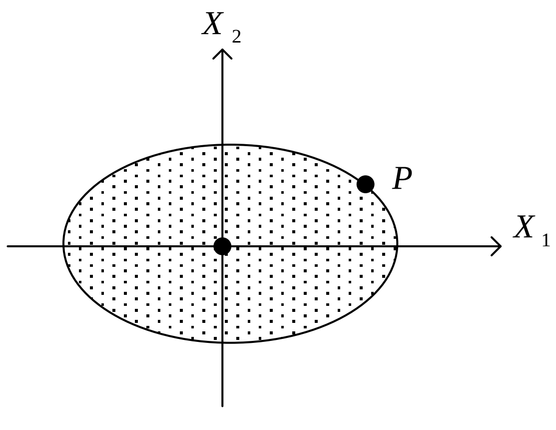
$$d^*(\mathrm{P}\,, 0) = \sqrt{\frac{x_1^2}{S_{11}} + \frac{x_2^2}{S_{22}} + \cdots + \frac{x_p^2}{S_{pp}}} \qquad (53a-1)$$

2- إذا كانت التباينات متساوية وتساوي الواحد: $S_{11} = S_{22} = \cdots = S_{pp} = 1$، فإننا نحصل على المسافة (الإقليدية) المعرفة في (50-1) .

3-إذا وضعنا المسافة $d^*(P\,, Q)$ مساوية لعدد ثابت C في (52-1)، وربعنا الطرفين فإننا سنحصل على المعادلة التالية (في حالة متحولين):

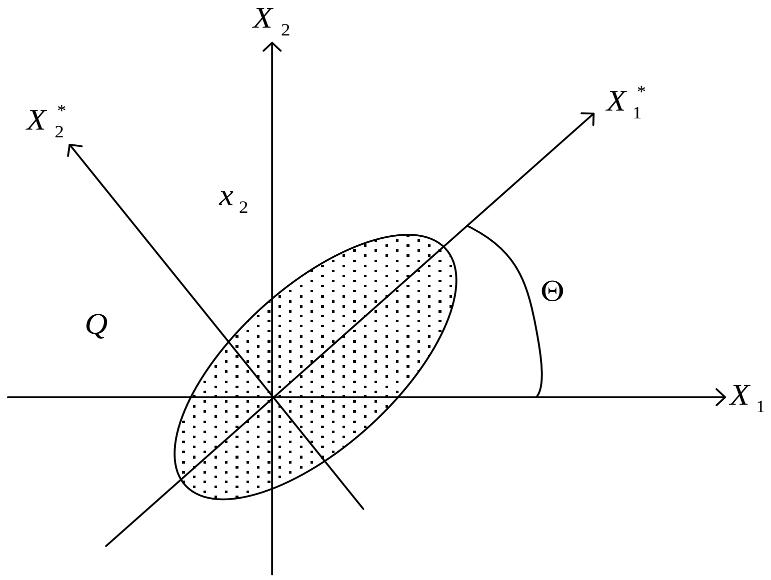
$$\frac{x_1^2}{S_{11}} + \frac{x_2^2}{S_{22}} = C^2 \qquad (54-1)$$

وهي معادلة قطع ناقص مركزه المبدأ (O) ومحوراه يوازيان المحورين الإحداثيين OX_1 و OX_2، كما هو موضح على الشكل التالي:



الشكل (1- 3) شكل القطع الناقص

وإذا كانت قياسات X_1 مرتبطة مع قياسات X_2 فإن شكل الانتشار لهما يكون له الشكل التالي:



الشكل (1−4) قطع ناقص

وعندها فإن المسافة من النقطة P حتى المركز (O) تعطى بالعلاقة $(53a-1)$، وإذا وضعنا تلك المسافة مساوية لعدد ثابت C فإننا نحصل على قطع ناقص كالقطع المبين على الشكل (1−4)، وحتى نحصل على قطع ناقص مشابه لشكل الانتشار (1−3) ، فإننا نقوم بتدوير المحاور الإحداثية OX_1 وOX_2 إلى اليسار بزاوية قدرها (θ)، فنحصل على محاور جديدة نرمز لها بـ OX_1^* وOX_2^*،

ولإيجاد الإحداثيات الجديدة بدلالة الإحداثيات القديمة نستخدم معادلتي التحويل التاليتين :

$$x_1^* = x_1 \cos\theta + x_2 \sin\theta$$
$$x_2^* = -x_1 \sin\theta + x_2 \cos\theta \qquad (55-1)$$

ومنهما نحسب تبايني X_1^* وX_2^* الجديدين ونرمز لهما بـ S_{11}^* وS_{22}^*، وعندها نجد أن المسافة من نفس النقطة $P(x_1^*, x_2^*)$ ذات الإحداثيات الجديدة إلى المركز (O) تعطى بالعلاقة $(53a-1)$ التي تأخذ الشكل التالي :

$$d^*(P,O) = \sqrt{\frac{x_1^{*2}}{S_{11}^*} + \frac{x_2^{*2}}{S_{22}^*}} \qquad (56-1)$$

وإذا قمنا بتعويض (x_1^*, x_2^*) وكذلك S_{11}^* وS_{22}^* من (1− 55) في (1− 56)، فإننا سنحصل على المسافة الجديدة $d^{**}(P,O)$ بدلالة الإحداثيات القديمة x_1, x_2، وبعد إجراء الإصلاحات المطولة نحصل على المسافة الجديدة x_1^*, x_2^* بدلالة الإحداثيات القديمة من العلاقة التالية :

$$d^{**}(P,O) = \sqrt{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2} \qquad (57-1)$$

أن المقدار الذي تحت الجذر يأخذ صيغة تسمى بالصيغة التربيعية، لأنه يتضمن حدود لـ X من الدرجة الثانية هي $(x_1^2), (x_2^2)$ و $(x_1 * x_2)$ مضروبة بالأعداد a_{11}, a_{22} و $2a_{12}$ على الترتيب .

حيث أن a_{ij} هي الأعداد المناسبة التي تجعل المسافة $d^{**}(P\ ,O)$ غير سالبة لجميع قيم x_2, x_1، وهي تحدد وتحسب من العلاقات التالية :

$$a_{11} = \frac{cos^2\,\theta}{cos^2\,\theta * S_{11} + 2\,sin\,\theta\,cos\,\theta * S_{12} + sin^2\,\theta * S_{22}} + \frac{sin^2\,\theta}{cos^2\,\theta * S_{22} - 2\,sin\,\theta\,cos\,\theta * S_{12} + sin^2\,\theta * S_{11}}$$

$$a_{22} = \frac{sin^2\,\theta}{cos^2\,\theta * S_{11} + 2\,sin\,\theta\,cos\,\theta * S_{12} + sin^2\,\theta * S_{22}} + \frac{cos^2\,\theta}{cos^2\,\theta * S_{22} - 2\,sin\,\theta\,cos\,\theta * S_{12} + sin^2\,\theta * S_{11}}$$

$$a_{12} = \frac{cos\,\theta\,sin\,\theta}{cos^2\,\theta\,S_{11} + 2\,sin\,\theta\,cos\,\theta * S_{12} + sin^2\,\theta * S_{22}} - \frac{sin\,\theta\,cos\,\theta}{cos^2\,\theta * S_{22} - 2\,sin\,\theta\,cos\,\theta * S_{12} + sin\,\theta * S_{11}}$$

وإذا ربعنا طرفي العلاقة (1−57) نحصل على مربع المسافة $d^{**}(P\ ,O)$ ويكون لدينا :

$$d^{**2}(P\ ,O) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \qquad (58-1)$$

وهنا نلاحظ أنه يمكن كتابة (1-58) على شكل مصفوفة كما يلي:

$$d^{**2}(P\ ,O) = (x_1\ ,x_2)\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad (59-1)$$

$$d^{**2}(P\ ,O) = X' * A * X \qquad (60-1)$$: وهي الصيغة التربيعية المطلوبة

وإذا وضعنا تلك المسافة مساوية لعدد ثابت C فإن المعادلة (1-56) تعطينا أن :

$$\frac{x_1^{*2}}{S_{11}^*} + \frac{x_2^{*2}}{S_{22}^*} = C^2$$

وهي معادلة قطع ناقص بدلالة الإحداثيات الجديدة x_1^* وx_2^*، ومحوراه يوازيان المحورين OX_1^* وOX_2^* ، اللذين يصنعان مع المحورين القديمين OX_1 وOX_2 زاوية قدرها θ، وإن هذا القطع يأخذ نفس وضعية القطع في الشكل (1−4) بالنسبة للمحورين OX_1^* وOX_2^* .

ولكن إذا جعلنا تلك المسافة مساوية لعدد ثابت C في المعادلة (1−58) فإننا نحصل على أن:

$$d^{**2}(P\ ,O) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = C^2 \qquad (61-1)$$

وهي معادلة القطع المذكور بدلالة الإحداثيات القديمة OX_1 وOX_2 ، كما هو مبين على الشكل (1−4) السابق .

وبصورة عامة فإنه يمكننا تعريف المسافة من نقطة P إلى مركز الإحداثيات (O) في الفضاء R^P بواسطة العلاقة التالية:

$$d^{**}(P\ ,O) = \sqrt{\mathrm{a}_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + \cdots + a_{PP}x_P^2 + 2a_{12}x_1x_2 + +2a_{13}x_1x_3 + \cdots + 2a_{P-1\quad P}x_{P-1}x_P} \qquad (62-1)$$

وبتربيع الطرفين يمكن كتابة هذه العلاقة مصفوفياً كما يلي:

$$d^{**2}(P\,,O) = (x_1 x_2 x_3 \dots x_P)\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1P} \\ a_{21} & a_{22} & \cdots & a_{2P} \\ \vdots & \vdots & \dots & \vdots \\ a_{P1} & a_{P2} & \cdots & a_{PP} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} = X' * A * X$$

ولحساب المسافة بين نقطتين $P(x_1 x_2 x_3 \dots x_P)$ و $Q(y_1 y_2 y_3 \dots y_P)$ في الفضاء R^P نعمم العلاقة (62−1) فنجد أن:

$$d(P\,,O) = \sqrt{a_{11}(x_1 - y_1)^2 + a_{22}(x_2 - y_2)^2 + \cdots a_{PP}(x_P - y_P)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2) + 2a_{13}(x_1 - y_1)(x_3 - y_3) + \cdots + 2a_{P-1}P(x_{P-1} - y_{P-1})(x_P - y_P)} \quad \mathbf{(63-1)}$$

وإذا وضعنا هذه المسافة مساوية لعدد ثابت C، ثم ربعنا الطرفين فإننا سنحصل على معادلة قطع ناقص كما يلي:

$$d^{**2}(P\,,Q) = a_{11}(x_1 - y_1)^2 + a_{22}(x_2 - y_2)^2 +..\, a_{PP}(x_P - y_P)^2 + 2a_{12}(x_1 - y_1)(\mathrm{x}_2 - y_2)$$
$$+2a_{13}(x_1 - y_1)(x_3 - y_3) + \cdots + 2a_{P-1}P(x_{P-1} - y_{P-1})(x_P - y_P) = C^2 \qquad (64-1)$$

وهي معادلة قطع ناقص مركزه O ومحوراه يوازيان المحورين الجديدين X_1^* و X_2^*، اللذين يصنعان مع المحورين القديمين OX_1 و OX_2 الزاوية θ .

وأخيراً يمكننا كتابة مربع المسافة من P حتى Q مصفوفياً كما يلي:

$$d^{**2}(P\,,Q) = [(x_1 - y_1)(x_2 - y_2) \dots (x_P - y_P)]\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1P} \\ a_{21} & a_{22} & \cdots & a_{2P} \\ a_{31} & a_{32} & \cdots & a_{3P} \\ \vdots & \vdots & \dots & \vdots \\ a_{P1} & a_{P2} & \cdots & a_{PP} \end{bmatrix}\begin{bmatrix} (x_1 - y_1) \\ (x_2 - y_2) \\ x_2 - y_2 \\ \vdots \\ (x_P - y_P) \end{bmatrix} \quad \mathbf{(65-1)}$$

كما يمكن كتابتها على شكل الصيغة التربيعية المعممة كما يلي:

$$d^{**2}(P\,,Q) = (X - Y)' * A * (X - Y) \qquad (66-1)$$

الفصل الثاني

قضايا القيم الذاتية والمصفوفات الشاذة

2-1 تمهيد :

لقد ظهرت مشكلة القيم الذاتية للمصفوفات عند معالجة المعادلات الخطية المتجانسة. ولكي نوضح جوهر هذه القضية دعنا ننطلق من المعادلة الخطية التالية:

$$a * x = \lambda * x \qquad (1-2)$$

والتي يمكن كتابتها على الشكل التالي:

$$(a-\lambda) * x = 0 \qquad (2-2)$$

ومن الواضح أن الحل البسيط لهذه المعادلة بالنسبة لـ x هو الحل الذي يكون فيه $x = 0$. ويسمى هذا الحل بالحل التافه (Trivial solution). لأنه لا يفيدنا في معالجة المسائل العلمية المختلفة. ويجعلنا نخسر جميع المعلومات المتوفرة في العلاقة (2-1) أو في العلاقة (2-2) .

وحتى نمنع هذا الحل التافه (x = 0) من الحدوث نستبعده من الحلول الممكنة، ونفترض أو نشترط أن يكون المضروب الأول $(a-\lambda)$ معدوماً، وعندها نجد أن λ يجب أن تأخذ قيمة معينة هي a لأن:

$$(a-\lambda) = 0 \quad \Rightarrow \quad \lambda = a \qquad (3-2)$$

وتسمى هذه القيمة لــ λ بالقيمة الذاتية (الكامنة) للمعادلة (2-2) (eigenvalue) . وعند تعويض هذه القيمة في المعادلة (2-2) فإنها تعطينا لانهاية من الحلول المقبولة (المحدودة وغير المعدومة) بالنسبة للمتحول المجهول x . وذلك لأن جداء أية قيمة محدودة لـ x في $(a-\lambda)$ يساوي الصفر عندما $(\lambda = a)$.

وكذلك الأمر في حالة المصفوفات المربعة، حيث نجد أنه لو كانت لدينا جملة من المعادلات الخطية كما يلي:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= \lambda x_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= \lambda x_3 \end{aligned} \qquad (4-2)$$

فإنه يمكننا كتابتها بدلالة المصفوفات والأشعة كمايلي:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad (5-2)$$

والتي يمكن كتابتها باختصار على الشكل التالي:

$$A_{3.3} * X_{3.1} = \lambda * X_{3.1} \qquad (6-2)$$

وهذه العلاقة تعبر عن عملية تحويل الشعاع X إلى شعاع آخر هو λX . وهذه العملية تصادفنا كثيراً في الدراسات والبحوث العلمية .

وبصورة عامة إذا كان لدينا n معالة خطية وتتضمن n متحولاً، ومشابهة للمعادلات (2-4) فإنه يمكننا كتابتها بدلالة مصفوفة A وشعاع X كما يلي:

$$A_{n.n}X_{n.1} = \lambda X_{n.1} \qquad (7-2)$$

حيث أن: A هي مصفوفة مربعة من المرتبة $n.n$ (وتسمى بمصفوفة التحويل)، وتتضمن الأمثال العددية a_{ij} المصاحبة للمتحول X_j في المعادلة الخطية i، وإن $X_{n.1}$ هو الشعاع العمود المؤلف من المتحولات X_j ، أي أن $X_{n.1}$ هو الشعاع العمود الذي يرمز للمتحولات X_j نفسها، أما λ فهو عدد حقيقي (سلمي) يستخدم في عملية التحويل من X إلى λX، ويسمى شكل المعادلة (2-7) بالشكل النموذجي أو المعياري لمشكلة القيم الذاتية للمصفوفة A . (standard eigenvalue problem) وذلك لتميزها عن الشكل العام لمشكلة لقيم الذاتية للمصفوفة A (Generalized eigenvalue problem) والذي يكون على الشكل التالي:

$$A_{n.n}X_{n.1} = \lambda B_{n.n}X_{n.1} \qquad (8-2)$$

حيث أن B هي مصفوفة مربعة من المرتبة $n.n$ أيضاً ومحددة إيجابياً (positive definite) وأن A مصفوفة متناظرة، ويمكن تحويل المشكلة العامة للقيم الذاتية المبينة في (2-8) إلى المشكلة النموذجية المعروضة في العلاقة (2-7)، وذلك باستخدام تحويل وتحليل تشوليسكي (cholesky) [انظر ذلك في الطويل- المصفوفات- ص 138). لذلك فإننا سنركز اهتمامنا في هذه الورقة على المشكلة النموذجية للقيم الذاتية وسنعرضها من خلال الفقرة التالية .

2-2 المشكلة النموذجية للقيم الذاتية:

إن الشكل النموذجي لهذه المشكلة يكون على شكل المعادلة (2-7) السابقة وهي:

$$A * \mathrm{X} = \lambda * X \qquad (9-2)$$

والتي يمكن كتابتها على الشكل التالي:

$$[A - \lambda I] * X = 0 \qquad (10-2)$$

حيث أن: I هي مصفوفة أحادية من المرتبة n .

إن حل المعادلة (2-10) بالنسبة للشعاع X يصطدم بالحل التافه (الذي هو X = 0).

وحتى نمنع حدوث ذلك الحل التافه. فإننا نفترض أو نشترط أن تكون المصفوفة اليسارية $[A - \lambda I]$ شاذة. وهذا يعني أنه يجب أن يكون محددها معدوماً. أي نشترط أن يحقق محددها الشرط التالي .

$$|A - \lambda I| = 0 \qquad (11-2)$$

وبما أن هذا المحدد هو من المرتبة $n.n$ وان عناصر قطره الرئيسي تساوي $(a_{ii} - \lambda)$ فإن العلاقة (2-11) تأخذ الشكل التالي:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{1n} \\ a_{21} & a_{22}-\lambda & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{nn}-\lambda \end{vmatrix} = 0 \qquad (12-2)$$

وعندما نقوم بحساب منشور أو مفكوك هذا المحدد، فإننا سنحصل منه على كثير حدود من المرتبة n بالنسبة للوسيط المجهول λ ويكون له الشكل التالي:

$$a_0 + a_1\lambda + a_2\lambda^2 + a_3\lambda^3 + \cdots + a_n\lambda^n = 0 \qquad (13-2)$$

وتسمى المعادلة (13-2) بالمعادلة المميزة للمصفوفة A (Characteristics equation)، وعندما نقوم بحل هذه المعادلة بالنسبة لـ λ فإننا سنحصل على n جذراً لها: أي إننا سنحصل منها على n قيمة لـ λ وسنرمز لها بالرموز :

$$\lambda_1\ ,\lambda_2\ ,\lambda_3\ \ldots\ldots\ldots\ \lambda_n \qquad (14-2)$$

وتسمى هذه الجذور بالقيم الذاتية أو القيم المميزة للمصفوفة A، كما تسمى بالجذور الكامنة للمصفوفة A، وإن هذه الجذور هي قيم λ ، التي تحول المصفوفة A إلى المصفوفة الشاذة $[A-\lambda I]$ ، وهي تستخدم كثيراً في المسائل العلمية. وإن هذه القيم الذاتية قد لا تكون مختلفة عن بعضها البعض بل قد يكون بعضها مكرراً عدة مرات (عدة جذور مضاعفة). كما يمكنها أن تكون حقيقية أو عقدية (مركبة). ولكن إذا ظهرت جذور عقدية فإنها تكون على شكل أزواج مترافقة .

وهنا نشير إلى أنه إذا كانت المصفوفة A نظامية $(|A| \neq 0)$ فإن عدد قيمها الذاتية يساوي مرتبتها n. وإذا كانت A شاذة، فإن عدد القيم الذاتية لها يكون مساوياً لرتبتها r، حيث أن $r \leq n$.

والآن نعود إلى مسألة إيجاد الشعاع X الذي يحقق العلاقة (10-2) مقابل كل قيمة من القيم الذاتيةλ_i.

ولذلك نأخذ أحدى القيم الذاتية ولتكن $\lambda = \lambda_i$ ونعوضها في المعادلة (10-2) فإننا سنحصل على المعادلة المحددة التالية:

$$[A - \lambda_i I] * X = 0 \qquad (15-2)$$

وبما أن قيمة λ_i تجعل المصفوفة $[A - \lambda_i I]$ مصفوفة شاذة، فإنه يمكننا الحصول من (15-2) على عدد لانهائي من الحلول المقبولة لـ X . وإذا أخذنا أحد تلك الحلول وليكن X_i المقابل للقيمة الذاتية λ_i فإنه سيكون على شكل شعاع عمود كما يلي:

$$X_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_i \qquad (16-2)$$

ويسمى هذا الشعاع X_i بالشعاع الذاتي (المميز) للمصفوفة A والمقابلة للقيمة الذاتية λ_i

ولكن بما أن جملة المعادلات (10-2) هي جملة معادلات خطية متجانسة، فإن أي حل مقبول لها مثل X_i، يشكل مع مضاعفاته حزمة من الحلول المقبولة (غير التافهة) ونرمز لها بالرموز:

$$X_i,\quad C_1X_i,\quad C_2X_i,\quad C_3X_i, \ldots\ldots \qquad (17-2)$$

حيث C_k أي عدد حقيقي (غير الصفر) يضاعف أو يقلص الحل X_i .

وبذلك نحصل على لانهاية أخرى من الحلول المرتبطة بالحل X_i المقابلة للقيمة الذاتية λ_i وحدها، ولكن هذه الحلول المتضاعفة تحدد لنا فقط الاتجاه العام لشعاع الحل X_i، ولا تحدد لنا عناصر الحل المطلوب (مركبات X_i). ولذلك تبقى أطوال هذه الحلول كيفية (غير محددة). وهذا أمر لا يساعدنا على التوصل إلى حلول محددة الطول بل محددة الاتجاه فقط .

وللتخلص من هذه المشكلة يمكن أن نقوم بتحويل هذه الأشعة المتضاعفة إلى أشعة ممعيرة واحدية بحيث يكون طول كل منها مساوياً للواحد الصحيح .

لذلك نأخذ أي شعاع منها مثل X_i ونحسب مربع طوله ℓ_i كما يلي:

$$\|X_i\|^2 = X_i' * X_i = x_1^2 + x_2^2 + x_3^2 + \cdots = \ell_i^2 = \text{مربع الطول} \qquad (18-2)$$

ثم نحسب طوله $\boldsymbol{\ell_i}$ فنجد أن :

$$\|X_i\| = \sqrt{\ell_i^2} = +\ell_i$$

ثم نقسم عناصر ذلك الشعاع على طوله $(+\ell_i)$ فنحصل على الشعاع الواحدي للشعاع X_i ونرمز له بـ e_i وهو يساوي

$$\boldsymbol{e_i} = \frac{X_i}{\|X_i\|} = \frac{X_i}{\ell_i} \qquad (19-2)$$

وبالتالي نحصل على شعاع جديد يكون له نفس الاتجاه ويكون طوله الجديد مساوياً للواحد لأن: $\boldsymbol{e_i'} * \boldsymbol{e_i} = 1$، ويسمى الشعاع $\boldsymbol{e_i}$ بالشعاع الذاتي الممعير المقابل للقيمة الذاتية λ_i، ونحصل على نفس النتيجة إذا حولنا الأشعة الأخرى في (2–17) إلى أشعة واحدية . وإذا كان للمصفوفة n قيمة ذاتية مختلفة (جذراً مختلفاً) فإنه سيكون لها n شعاعاً ذاتياً واحدياً مختلفاً مقابلاً لها. وللحصول على هذه الأشعة نعود ونعوض كل من القيم الذاتية التي حصلنا عليها في (2-14) في المعادلة (2-10)، واحدة تلو الأخرى حتى نحصل على جميع الأشعة الذاتية الواحدية المقابل لكل منها.

مثال (2–1): أوجد القيم الذاتية والأشعة الذاتية للمصفوفة $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$

الحل: نكتب المعادلة الذاتية لـ A كما يلي:

$$AX = \lambda X \qquad \Longrightarrow [A - \lambda I] * X = 0$$

ثم نضع محدد المصفوفة اليسرى مساويا للصفر فنحصل على أن :

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} = 0$$

ثم نقوم بحساب مفكوكه فنحصل على كثير حدود من المرتبة الثانية بالنسبة لـ λ وهو:

$$(1-\lambda)(1-\lambda) - 3 * 3 = 0$$

$$\lambda^2 - 2\lambda - 8 = 0$$

وبحل المعادلة الأخيرة بالنسبة لـ λ نحصل على الجذرين التاليين:

$$\lambda_1 = 4 \qquad \lambda_2 = -2$$

ولإيجاد الأشعة الذاتية المقابلة لهما، نأخذ كل قيمة على حدة، فنجد أنه عندما نأخذ $\lambda_1 = 4$ فإن مصفوفة المعادلة المميزة تساوي :

$$[A-\lambda_1 I] = \begin{bmatrix} 1-4 & 3 \\ 3 & 1-4 \end{bmatrix} = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix}$$

وللحصول على الأشعة الذاتية المقابلة لهذه القيمة نكتب المعادلة المميزة كما يلي :

$$[A-\lambda_1 I] * X = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

ومنها نحصل على المعادلتين غير المستقلتين التاليتين:

$$-3x_1 + 3x_2 = 0$$
$$3x_1 - 3x_2 = 0$$

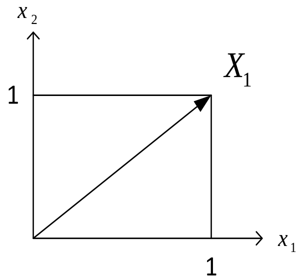
وهما عبارة عن معادلة واحدة (الثانية تنتج من الأولى بضربها بـ (1-))

نأخذ أحدهما ولتكن الأولى: $-3x_1 + 3x_2 = 0$

فنحصل منها على أن: $x_1 = x_2$

أي أن العنصر x_1 يساوي x_2 وله نفس الإشارة

وبإعطاء x_1 أية قيمة كيفية . يمكننا أن نحصل من المعادلة الأخيرة على لانهاية من الحلول المقبولة ويكون لها نفس الاتجاه .



وحتى نحدد أحد الحلول الخاصة نضع كيفياً $x_1 = 1$

فنحصل على أن: $x_2 = 1$

وبالتالي نحصل على الشعاع الذاتي الأول، والذي نرمز له بـ X_1 والمبين على الشكل المقابل وهو $X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ وهو يقابل القيمة الذاتية $\lambda_1 = 4$.

ومنه نحصل على حزمة الأشعة الذاتية الأخرى المقابلة لنفس القيمة الذاتية $\lambda_1 = 4$، وذلك بأخذ مضاعفات X_1 كما يلي:

$$X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad X_{c_1} = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad X_{c_2} = c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \dots \dots \dots \dots \dots$$

حيث c_i أي عدد حقيقي (غير الصفر)

أما عندما نأخذ القيمة الذاتية الثانية $\lambda_2 = -2$ فإننا نجد أن المصفوفة:

$$[A-\lambda_2 I] = \begin{bmatrix} 1+2 & 3 \\ 3 & 1+2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$$

ثم نكتب المعادلة الذاتية كما يلي:

$$[A-\lambda_2 I]X = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

ومنها نحصل على المعادلتين غير المستقلتين التاليتين:

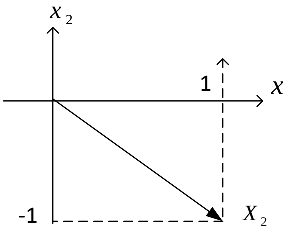
$$3x_1 + 3x_2 = 0$$
$$3x_1 + 3x_2 = 0$$

وهما عبارة عن معادلة واحدة، وإذا أخذنا أحدهما نجد أن:

$$x_1 = -x_2$$

أي أن العنصر x_1 يساوي $(-x_2)$ وأن العنصران مختلفين بالإشارة .

وبإعطاء x_1 أية قيمة كيفية، فإنه يمكننا أن نحصل من المعادلة الأخيرة على لانهاية من الحلول المقبولة كمايلي:



نضع بشكل كيفي $x_1 = 1$ فنحصل على أن $x_2 = -1$، وبذلك نحصل على الشعاع الذاتي الخاص X_2 المقابل للقيمة الذاتية الثانية $\lambda_2 = -2$ وهو $X_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$، كما في الشكل المقابل :

ومنه نحصل على حزمة الأشعة الذاتية الأخرى المقابلة لنفس القيمة الذاتية $\lambda_2 = -2$ وهي:

$$X_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, X_{c_1} = c_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}, X_c = c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \ldots \ldots \ldots$$

حيث c أي عدد حقيقي (غير الصفر)

وللتحقق من أن هذه الأشعة تشكل حلول مقبولة للمعادلة (2-9) أو للمعادلة (2-10) نقوم بتعويضها في المعادلة الذاتية $AX = \lambda X$ لنتأكد من تحققها فنجد أن :

$$AX_1 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda_1 X_1$$

$$AX_2 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \lambda_2 X_2$$

أي أن الشعاعين الذاتين X_2, X_1 يحققان المعادلة الذاتية، وبذلك يشكلان حلين مقبولين لها. علماً بأن مضاعفات كل منهما تشكل حلول مقبولة لها أيضاً، ونشير هنا أيضاً إلى أن الحلين X_1 و X_2 متعامدان لأنهما يحققان شرط التعامد وهو أن يكون الجداء السلمي لهما مساوياً للصفر:

$$\langle X_1 \, , X_2 \rangle = X_1^* * X_2 = (1 \, ,1) \begin{bmatrix} +1 \\ -1 \end{bmatrix} = 0$$

ولكن هذين الحلين محددان بالاتجاه فقط ، وغير محددين بالطول لأن جميع مضاعفاتهما تشكل حلول مقبولة أيضاً .

لذلك يفضل أن نقوم بحساب الأشعة الذاتية الممعيرة الواحدية لهما، ولهذا نقوم بحساب طول كل منها فنجد أن:

$$\ell_1^2 = \|X_1\|^2 = x_1^2 + x_2^2 = 1^2 + 1^2 = 2 \qquad \Longrightarrow \qquad \ell_1 = \sqrt{2}$$

$$\ell_2^2 = \|X_2\|^2 = x_1^2 + x_2^2 = 1^2 + (-1)^2 = 2 \qquad \Longrightarrow \qquad \ell_2 = \sqrt{2}$$

وبذلك نجد أن الأشعة الذاتية الواحدية هي:

$$\boldsymbol{e_1} = \frac{X_1}{\|X_1\|} = \frac{X_1}{\ell_1} = \frac{X_1}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$e_2 = \frac{X_2}{\|X_2\|} = \frac{X_2}{\ell_2} = \frac{X_2}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

ويمكننا التأكد من أن هذين الشعاعين يحققان المعادلة النموذجية (2-9) أو (2-10) فنجد أن:

$$A * e_1 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} \end{bmatrix} = 4 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \lambda_1 e_1$$

$$A * e_2 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{bmatrix} = -2 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \lambda_2 e_2$$

أي أن الشعاعين e_2, e_1 يحققان المعادلة الذاتية المفروضة ويشكلان الحلين المميزين لها .

2-3 نظريات مختلفة حول القيم والأشعة الذاتية :

لقد أشرنا سابقاً إلى أنه إذا كانت المصفوفة A نظامية ومن المرتبة $n * n$، فإنه عدد قيمها الذاتية سيكون مساوياً لـ n قيمة ذاتية. ونرمز لها $\lambda_1 \ \lambda_2 \ \lambda_3 \ \dots \ \lambda_n$، وبناء على ذلك نقدم النظريات التالية (للبرهان أنظر المصفوفات للطويل):

ن1: إن مجموع القيم الذاتية للمصفوفة A يساوي أثر trace تلك المصفوفة أي أن:

$$\sum_{k=1}^{n} \lambda_i = tr(A) = \sum_{i=1}^{n} a_{ii} \qquad (20-2)$$

ن2: إن قيمة محدد المصفوفة A يساوي جداء قيمها الذاتية أي أن:

$$|A| = \prod_{\text{k}=1}^{n} \lambda_i \qquad (21-2)$$

وكتطبيق على ذلك نأخذ القيم الذاتية للمصفوفة A في المثال(2−1) السابق فنجد أن:

$$\left.\begin{aligned} \sum_{k=1}^{n} \lambda_i &= 4 + (-2) = 2 \\ tr(A) &= 1 + 1 = 2 \end{aligned}\right| \quad \Rightarrow \quad \sum_{k=1}^{n} \lambda_i = tr(A)$$

$$\left.\begin{aligned} |A| &= 1 - 9 = -8 \\ \prod_{k=1}^{n} \lambda_i &= 4(-2) = -8 \end{aligned}\right| \quad \Rightarrow \quad |A| = \prod_{k=1}^{n} \lambda_i$$

ن3: إذا كانت المصفوفة A مصفوفة صفرية فإن جميع قيمها الذاتية تكون أصفاراً أي أن:

$$A = [0] \Rightarrow \qquad \lambda_i = 0: \qquad \text{i} = 1,2, \dots \dots n \qquad (22-2)$$

ن4: إذا كانت المصفوفة A هي المصفوفة الأحادية I فإن جميع قيمها الذاتية تكون متساوية وتساوي الواحد الصحيح أي أن:

$$A = I \Longrightarrow \qquad \lambda_i = 1 \qquad \mathrm{i} = 1,2, \ldots .. n \qquad (23-2)$$

ن5: إذا كانت المصفوفة A مصفوفة قطرية D فإن قيمها الذاتية تساوي عناصرها القطرية. أي أن:

$$A = D = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \Longrightarrow \quad \lambda_i = a_{ii} \quad i = 1,2, \ldots .. n \qquad (24-2)$$

ويستفاد من هذه النظرية في إيجاد القيم الذاتية λ_i لأي مصفوفة وذلك بتحويلها إلى مصفوفة قطرية.

ن6: تكون أي مصفوفة A مصفوفة شاذة $(|A| = 0)$ إذا وفقط إذا كانت إحدى قيمها الذاتية مساوية للصفر (لأن قيمة $|A| = \prod \lambda_i$) .

أما إذا كانت المصفوفة قطرية D فإنها تكون شاذة إذا وفقط إذا كان أحد عناصرها القطرية معدوماً .

ن7: إذا كانت جميع القيم الذاتية للمصفوفة A مختلفة، أي إذا كانت: $\lambda_1 \neq \lambda_2 \neq .. \lambda_i .. \neq \lambda_n$ ، فإن الأشعة الذاتية المقابلة لها تكون مختلفة (من حيث الاتجاه) .

ن8: إذا كانت جميع القيم الذاتية للمصفوفة A مختلفة فإن الأشعة الذاتية المقابلة لها تكون مستقلة خطياً (متعامدة) .

ن9: يمكن أن يكون لمصفوفتين مختلفتين A , B ولهما نفس المرتبة n.n نفس القيم الذاتية: وعندها تكون المصفوفتان A وB متشابهتين. (يستفاد منها في التعرف على المصفوفات المتشابهة) .

ن10: إن للمصفوفة A ولمنقولها A' نفس القيم الذاتية وذلك لأن لهما نفس كثير الحدود بالنسبة لـ λ .

ن11: إذا كانت المصفوفة A حقيقية ومتناظرة فإن قيمها الذاتية تكون حقيقية. وإذا كانت هذه القيم مختلفة فإن الأشعة الذاتية المقابلة لها تكون متعامدة، أي يكون الجداء السلمي لأي شعاعين مساوياً الصفر $< U_k * U_n >= U_k' * U_n = 0$، أما إذا لم تكن مختلفة فيمكن تحويل الأشعة الذاتية المقابلة لها إلى أشعة متعامدة (حسب نظرية غرام).

ن12: إذا كانت المصفوفة A حقيقية ومتناظرة سلبياً فإن قيمها الذاتية تكون تخيلية صرفة (الجزء الحقيقي معدوم). وإذا كانت هذه القيم مختلفة فإن الأشعة الذاتية المقابلة لها تكون متعامدة. أو يمكن تحويلها إلى متعامدة (نظرية غرام) .

ن13: إذا كان الشعاع U هو الشعاع الذاتي للمصفوفة A والمقابل للقيمة الذاتية λ، وكان a أي عدد حقيقي غير الصفر، فإن الشعاع $(\alpha * u)$ يكون أيضاً شعاعاً ذاتياً للمصفوفة A ومقابلاً للقيمة الذاتية λ نفسها .

ن14: إذا كانت A مصفوفة مربعة ومتناظرة، وكانت L أي مصفوفة نظامية $(|L| \neq 0)$ ومن مرتبة A ، فإن المصفوفة B المعرفة بالعلاقة $B = L^{-1} * A * L$ تكون مشابهة لـ A ويكون لها نفس القيم الذاتية التي لـ A (ولكن الأشعة الذاتية تبقى غير متطابقة) .

وإن الأشعة الذاتية لـ B تساوي حاصل ضرب L^{-1} في الأشعة الذاتية لـ A. أي أن: $V = L^{-1} * U$

حيث: U هي الأشعة الذاتية لـ A و V هي الأشعة الذاتية لـ B

وبالتالي فإن المصفوفتين B,A تكونان متشابهتين. وتسمى المصفوفة L بتحويلة التشابه .

ن15: إذا شكلنا من أعمدة الأشعة الذاتية للمصفوفة A بحسب ترتيبها مصفوفة T، فإننا سنحصل على مصفوفة مربعة من مرتبة A، وتسمى المصفوفة T بالمصفوفة الظاهرية لـ A(medal matrix)، وإذا كانت المصفوفة T نظامية $(|T| \neq 0)$ ، فإن المصفوفة D المعرفة بالعلاقة:

$D = T^{-1} * A * T$ تكون مصفوفة قطرية، وإن عناصر قطرها الرئيسي هي القيم الذاتية لـ D وتساوي القيم الذاتية لـA حسب الترتيب السابق . أي أن:

$$D = T^{-1} * A * T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} \qquad (25-2)$$

أي أن المصفوفتين D,A تكونان متشابهتين، وتسمى المصفوفة T بتحويلة الاستقطار والتشابه Similarity diagonalization transformation، ويستفاد من هذه النظرية في عملية تحويل المصفوفات إلى مصفوفات قطرية، وفي عمليات التحقق من صحة الأشعة الذاتية.

وهناك شكل آخر لهذه النظرية وهو أن يكون:

$A * T = T * \Lambda$ (وتطبق عندما يكون T^{-1} غير موجود)

حيث Λ: هي مصفوفة قطرية مؤلفة من القيم الذاتية λ_i على القطر الرئيسي .

ن16: إذا كانت λ قيمة ذاتية للمصفوفة A ومكررة m مرة وكانت الأشعة $U_1, U_2, U_3 \dots . U_m$ هي الأشعة الذاتية المقابلة لـ λ المكررة، فإن التركيب الخطي لها $\sum_{i=1}^{m} \alpha_i * U_i$ يكون أيضاً شعاعاً ذاتياً للمصفوفة A ومقابلاً للقيمة الذاتية λ نفسها حيث أن a_i أية أعداد حقيقية غير الصفر.

ن17: إذا كانت λ قيمة ذاتية للمصفوفة النظامية A $(|A| \neq 0)$ ويقابلها الشعاع الذاتي U فإن λ^k تكون قيمة ذاتية للمصفوفة الأسية A^k ويقابلها نفس الشعاع الذاتي U . حيث أن k عدد صحيح بشكل عام.

نتيجة هامة: إذا كان k=-1 فإنه يكون لدينا المصفوفة A^{-1} ويكون λ^{-1} قيمة ذاتية لها، أما شعاعها الذاتي فيبقى U هو نفسه الذي لـ A . أي القيم الذاتية لـ A^{-1} هي مقاليب القيم الذاتية لـ A، ولكن الأشعة الذاتية تبقى نفسها .

ن18: إذا كان للمصفوفة A القيمة والشعاع الذاتيين(λ , U) وكان لدينا التابع $f(A) = \sum_{i=1}^{m} \alpha_i * A^i$، فإن $f(A)$ يكون له القيمة والشعاع الذاتيين $[f(\lambda), u]$.

ن19: نظرية (كايلي– هاملتون) .

إذا كانت A مصفوفة مربعة من المرتبة n وكانت معادلتها المميزة بالنسبة لـ λ هي :

$$|A - \lambda I| = a_0 + a_1\lambda + a_2\lambda^2 + \cdots a_n\lambda^n = 0 \qquad (26-2)$$

فإن A نفسها تحقق تلك المعادلة الذاتية أي يكون لدينا:

$$a_0 I + a_1 A + a_2 A^2 + \cdots a_n A^n = 0 \qquad (27-2)$$

ن20: يمكن الحصول على المقلوب A^{-1} من العلاقة (27−2) $\sum_{i=0}^{n} \alpha_i * A^i = 0$، وذلك بضرب تلك الصيغة بـ A^{-1} من اليسار فنجد أن :

$$a_0 A^{-1} + a_1 I + a_2 A + a_3 A^2 + \cdots a_n A^{n-1} = 0$$

وبالتالي نحصل على المقلوب A^{-1} من العلاقة:

$$A^{-1} = \frac{-1}{a_0}[a_1 I + a_2 A + a_3 A^2 + \cdots a_n A^{n-1}] \qquad (28-2)$$

ن21: يمكن فك أو نشر أي مصفوفة أسية A_{n*n}^k (حيث $k < n$) على شكل متسلسلة قوى من A حتى القوة (n-1) أي يمكن كتابة A_n^k على الشكل التالي:

$$A_n^k = B_0 + B_1 A^1 + B_2 A^2 + \cdots B_{n-1} A^{n-1} \qquad (29-2)$$

ن22: يمكن فك واقتطاع أي دالة مصفوفة $f(A)$ إلى الحد A^{n-1} كحد أعلى. أي يمكن كتابتها كما يلي:

$$f(A) = \sum_{i=0}^{n-1} a_i A^i \qquad \text{بشرط تقارب المتسلسلة} \qquad (30-2)$$

وهناك نظريات أخرى يمكن البحث عنها في المراجع المختصة .

مثال (2−2): لنأخذ المصفوفة المتناظرة التالية : $A = \begin{bmatrix} 5 & -2 & -4 \\ -2 & 2 & 2 \\ -4 & 2 & 5 \end{bmatrix}$

والمطلوب: إيجاد القيم والأشعة الذاتية لـ A ، ثم حساب المصفوفة T والجداء $T^{-1} * A * T$

ثم التأكد من أن: $T^{-1} * A * T = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$

الحل: لإيجاد القيم الذاتية لهذه المصفوفة نشكل المعادلة الذاتية لها وهي:

$$A * X = \lambda X$$
$$[A - \lambda I] * X = 0$$

وحتى لا نحصل على الحل التافه للشعاع العمودي (X=0) نضع محدد المصفوفة مساوياً للصفر، أي نجعل: $|A - \lambda I| = 0$ فنجد أن:

$$\begin{vmatrix} 5-\lambda & -2 & -4 \\ -2 & 2-\lambda & 2 \\ -4 & 2 & 5-\lambda \end{vmatrix} = 0$$

وبأخذ مفكوك هذا المحدد حسب الأقطار نجد أن:

$$(5-\lambda)(2-\lambda)(5-\lambda) + 16 + 16 - 16(2-\lambda) - 4(5-\lambda) - 4(5-\lambda) = 0$$
$$(10 - 5\lambda - 2\lambda + \lambda^2)(5-\lambda) + 32 - 32 + 16\lambda - 20 + 4\lambda - 20 + 4\lambda = 0$$
$$50 - 35\lambda + 5\lambda^2 - 10\lambda + 7\lambda^2 + \lambda^3 + 24\lambda - 40 = 0$$

ومنه نحصل على المعادلة المميزة التالية :

$$10 - 21\lambda + 12\lambda^2 - \lambda^3 = 0$$

نحل هذه المعادلة بالنسبة لـ λ حسب خوارزمية معادلات الدرجة الثالثة، فنحصل على (3) جذور كامنة أو على (3) قيم ذاتية للمصفوفة السابقة A وهي: $\lambda_1 = 1\ ,\ \lambda_2 = 1\ ,\ \lambda_3 = 10$

وهنا نلاحظ أن: مجموع هذه القيم لذاتية = مجموع عناصر القطر الرئيسي للمصفوفة A= trace(A)

$$1 + 1 + 10 = 5 + 2 + 5 = 12$$

وللحصول على الأشعة الذاتية المقابلة لكل من هذه القيم الذاتية نعوض قيم λ على التوالي في مصفوفة المعادلة الذاتية: فعندما نضع $\lambda_1 = 1$ فإنه يكون لدينا:

$$\begin{bmatrix} 5-1 & -2 & -4 \\ -2 & 2-1 & 2 \\ -4 & 2 & 5-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

ومنها نحصل على المعادلات الخطية المتجانسة التالية:

$$4x_1 - 2x_2 - 4x_3 = 0$$
$$-2x_1 + 1X_2 + 2x_3 = 0$$
$$-4x_1 2 + 2x_2 + 4x_3 = 0$$

وهي معادلات غير مستقلة عن بعضها البعض، بل هي عبارة عن معادلة واحدة (لأن الثانية تنتج من الأولى بضربها بـ $\left(-\frac{1}{2}\right)$ والثالثة من الأولى بضربها بـ (1-)) .

ولذلك نأخذ معادلة واحدة منها ولتكن المعادلة الثانية فنجد أن:

$$-2x_1 + x_2 + 2x_3 = 0 \qquad (1)$$

وهي تعطينا لانهاية من الحلول المقبولة، لذلك نبدأ بالحل البسيط الذي يكون فيه $x_3 = 0$ فنحصل على أن: $x_2 = 2x_1$ ثم نضع $x_1 = -1$ فنحصل على أن $x_2 = -2$

وبذلك نحصل على الشعاع الذاتي الخاص الأول وهو: $u_1 = \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix}$

ومنها نحصل على حزمة الأشعة المقابلة للقيمة الذاتية: $\lambda = 1$ من العلاقة:

$$U_1 = C * \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix}$$

حيث C هو أي عدد حقيقي (غير الصفر) .

ولكن إذا وضعنا $x_2 = 0$ في المعادلة (1) فإننا سنحصل على أن $x_3 = x_1$، لذلك يمكن أن نضع $x_1 = 1$ فنحصل على $x_3 = 1$، وبالتالي نحصل على الشعاع الذاتي الخاص الآخر المقابل أيضاً للقيمة الذاتية $\lambda = 1$ وهو: $u_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

ومنها نحصل على حزمة أخرى من الأشعة المقابلة للقيمة الذاتية $\lambda = 1$ من العلاقة الآتية :

$$U_2 = C * \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

ولكن إذا وضعنا $x_1 = 0$ في المعادلة (1) فإننا سنحصل على أن : $x_2 = -2x_3$ ، وعندها يمكن أن نضع $x_3 = 1$ فنحصل على أن $x_2 = -2$ ، وبالتالي نحصل على الشعاع الذاتي الخاص الثالث المقابل أيضاً لـ $\lambda_2 = 1$ وهو :

$$u_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

ومنه نحصل على حزمة ثالثة من الأشعة المقابلة أيضاً للجذر الذاتي $\lambda = 1$ من العلاقة الآتية:

$$U_3 = cu_3$$

وحسب النظرية (16) نلاحظ أن: $u_3 = u_1 + u_2$ أي أن u_3 غير مستقل عن u_1, u_2 لأنها تقابل نفس القيمة الذاتية $\lambda = 1$.

وللتحقق من أن هذه الأشعة u_1, u_2, u_3 صحيحة نتأكد من أن قيمها تحقق المعادلة الذاتية التالية:

$$AU = \lambda * U$$

وكمثال على ذلك نأخذ أحد أشعة الحزمة الثالثة فنجد أنها محققة :

$$\begin{bmatrix} 5 & -2 & -4 \\ -2 & 2 & 2 \\ -4 & 2 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} = 1 * \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

وأخيراً لنأخذ القيمة الذاتية الثالثة $\lambda_2 = 10$ ونعوضها في المعادلة $[A - \lambda I]X = 0$ فنجد أن:

$$\begin{bmatrix} (5-10) & -2 & -4 \\ -2 & (2-10) & 2 \\ -4 & 2 & (5-10) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -5 & -2 & -4 \\ -2 & -8 & 2 \\ -4 & 2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

ومنها نحصل على المعادلات الخطية التالية:

$$-5x_1 - 2x_2 - 4x_3 = 0$$
$$-2x_1 - 8x_2 + 2x_3 = 0$$
$$-4x_1 + 2x_2 - 5x_3 = 0$$

وهي معادلات غير مستقلة عن بعضها البعض (لأن محددها معدوم)، ولحلها نقوم بضرب الثانية بـ (2)، ثم نطرحها من الثالثة فنحصل على المعادلة التالية:

$$0 + 18x_2 - 9x_3 \Rightarrow x_3 = 2x_2$$

فإذا وضعنا $x_2 = 1$ فإن $x_3 = 2$، وبالتعويض في المعادلة الثالثة أو غيرها نجد أن $x_1 = -2$

وبذلك نحصل على الحل الخاص الأول المقابل للقيمة الذاتية: $\lambda_3 = 10$ وهو:

$$u_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} \quad \Rightarrow \quad U_4 = C \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$$

وللتأكد من صحته نعوضه في المعادلة $A * U = \lambda U$ فنجد أنها محققة :

$$\begin{bmatrix} 5 & -2 & -4 \\ -2 & 2 & 2 \\ -4 & 2 & 5 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -20 \\ 10 \\ 20 \end{bmatrix} = 10 \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$$

ثم نقوم بتشكيل المصفوفة T من الأشعة الذاتية المقابلة للقيم الذاتية وهي: U_1, U_2, U_4 فنحصل على أن:

$$T = \begin{bmatrix} -1 & 1 & -2 \\ -2 & 0 & 1 \\ 0 & 1 & +2 \end{bmatrix} \Rightarrow T^{-1} = \frac{1}{9}\begin{bmatrix} 1 & 4 & -1 \\ 4 & -2 & 5 \\ -2 & 1 & 2 \end{bmatrix}$$

وبذلك نجد أن:

$$T^{-1} * A * T = \frac{1}{9}\begin{bmatrix} 1 & 4 & -1 \\ 4 & -2 & 5 \\ -2 & 1 & 2 \end{bmatrix} * \begin{bmatrix} 5 & -2 & -4 \\ -2 & 2 & 2 \\ -4 & 2 & 5 \end{bmatrix} * \begin{bmatrix} -1 & 1 & -2 \\ -2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$T^{-1} * A * T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

وإذا أردنا أن نحصل على الأشعة الذاتية الواحدية لهذه المصفوفة، فإننا نقوم بحساب أطوال الأشعة الذاتية السابقة ونقسم عناصر كل شعاع على طوله كما يلي:

$$\|u_1\|^2 = 1^2 + 2^2 + 0^2 = 5 \quad \Rightarrow \quad \|u_1\| = \sqrt{5} = \ell_1$$

$$\|u_2\|^2 = 1^2 + 0^2 + 1^2 = 2 \quad \Rightarrow \quad \|u_2\| = \sqrt{2} = \ell_2$$

$$\|u_3\|^2 = (-2)^2 + 1^2 + 2^2 = 9 \quad \Rightarrow \quad \|u_3\| = \sqrt{9} = 3 = \ell_3$$

وهكذا نجد أن الأشعة الذاتية الممعيرة الواحدية هي:

وهو يقابل $\lambda_1 = 1$: $\boldsymbol{e_1} = \frac{U_1}{\ell_1} = \begin{bmatrix} \frac{-1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \\ 0 \end{bmatrix}$

وهو يقابل $\lambda_2 = 1$: $\boldsymbol{e_2} = \frac{U_2}{\ell_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

وهو يقابل $\lambda_3 = 10$: $\boldsymbol{e_3} = \frac{U_3}{\ell_3} = \begin{bmatrix} \frac{-2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$

وهي تحقق العلاقة السابقة: $T^{-1} * A * T = \Lambda$

كما أنها تحقق المعادلة الذاتية :

$$A * E = \lambda * E$$

حيث أن : E هي مصفوفة الأشعة المتعامدة $\boldsymbol{e_1}\ \boldsymbol{e_2}\ \boldsymbol{e_3} \dots \boldsymbol{e_n}$.

2-4 نظريات مكملة(أنظر المراجع المرفقة)

1- نظرية النشر الطيفي (Spectral decomposition)

إذا كانت A مصفوفة مربعة ونظامية وحقيقية من المرتبة $n*n$ ، فإنه يكون لها عدداً من القيم الذاتية الحقيقية يساوي n قيمة هي:

$$\lambda_1 , \lambda_2 , \lambda_3 \ \dots \lambda_n \qquad (31-2)$$

ويقابلها عدداً n من الأشعة الذاتية المميزة (الواحدية) هي:

$$\boldsymbol{e_1} , \boldsymbol{e_2} , \boldsymbol{e_3} \ \dots \boldsymbol{e_n} \qquad (32-2)$$

وإذا كانت القيم الذاتية مختلفة فإن الأشعة الذاتية المقابلة لها تكون متعامدة أي أنها تحقق العلاقتين :

$$\boldsymbol{e'_i}.\boldsymbol{e_i} = 1 \qquad \boldsymbol{e'_i}.\boldsymbol{e_j} = 0 \qquad : i \neq j$$

وإذا كانت E مصفوفة الأشعة المتعامدة $\boldsymbol{e_1} \ \boldsymbol{e_2} \ \boldsymbol{e_3} \dots \boldsymbol{e_n}$

فإنه يمكننا نشر المصفوفة A بدلالة تلك الأشعة على الشكل التالي:

$$\boxed{A = \lambda_1\boldsymbol{e_1}\boldsymbol{e'_1} + \lambda_2\boldsymbol{e_2}\boldsymbol{e'_2} + \cdots \lambda_n\boldsymbol{e_n}\boldsymbol{e'_n} = E * \Lambda * E' \qquad \mathbf{(33-2)}}$$

وللتأكد من ذلك نقوم بتطبيقها على المصفوفة $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ من المثال (2-1)، حيث كان لها القيم الذاتية التالية:

$$\lambda_1 = 4 \qquad \lambda_2 = -2$$

وكان لها الأشعة الذاتية المميزة المقابلة لها التالية:

$$\boldsymbol{e_1} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \boldsymbol{e_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

وحسب هذه النظرية نجد أن:

$$A = \lambda_1\boldsymbol{e_1}\boldsymbol{e'_1} + \lambda_2\boldsymbol{e_2}\boldsymbol{e'_2} = 4\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} * \left(\frac{1}{\sqrt{2}} \ , \frac{1}{\sqrt{2}}\right) - 2\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix} * \left(\frac{1}{\sqrt{2}} \ , \frac{-1}{\sqrt{2}}\right) =$$

$$= 4\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} - 2\begin{bmatrix} \frac{1}{2} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} - \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} = A$$

2- نظرية المصفوفة المتعامدة:

إذا كانت A مصفوفة مربعة ومتناظرة ومن المرتبة $n*n$ وكانت قيمها الذاتية مختلفة، فإنه يكون لها مجموعة من الأشعة المتعامدة المميزة $\boldsymbol{e_1} \ \boldsymbol{e_2} \ \boldsymbol{e_3} \dots \boldsymbol{e_n}$ مقابل القيم الذاتية $\lambda_1 \ \lambda_2 \dots \lambda_n$، وإذا شكلنا من هذه الأشعة المتعامدة مصفوفة E ، فإن E ستكون متعامدة وتحقق العلاقات التالية (حسب النظرية (15)) :

$$E * E' = E' * E = I \quad \Longrightarrow \quad E' = E^{-1} \qquad (34-2)$$
$$E^{-1} * A * E = E' * A * E = \Lambda = diog\ (\lambda_1\ \ \lambda_2\ \ ...\lambda_n) \qquad (35-2)$$

حيث أن I هي المصفوفة الواحدية من المرتبة $n * n$، وإن Λ هي المصفوفة القطرية التي تتألف عناصر قطرها الرئيسي من القيم الذاتية $\lambda_1\ \ \lambda_2\ \ \lambda_3\ \ ...\ \lambda_n$ المختلفة .

ومن جهة أخرى نجد أنه لو ضربنا العلاقة (2–35) من اليمين بـ E' ومن اليسار بـ E فنجد أن:

$$E * E' * A * E * E' = E * \Lambda * E'$$

ومنها نستنتج أن :

$$A = E * \ \Lambda * E' \qquad (36-2)$$

وهذا يؤكد صحة العلاقة (2–33)

وبدمج هذه النظرية مع نظرية النشر الطيفي نستنتج أنه يمكننا كتابة أي مصفوفة مربعة ومتناظرة على الشكل التالي:

$$A = E * \ \Lambda * E' = \sum_{i=1}^{n} \lambda_i * \boldsymbol{e_i} * \boldsymbol{e'_i} \qquad (37-2)$$

حيث أن: e_i هي الأشعة الذاتية المميزة المقابلة للقيم الذاتية λ_i

وأن E هي المصفوفة المتعامدة من الأشعة المميزة $\boldsymbol{e_i}$

وأن Λ هي المصفوفة القطرية المؤلفة من القيم الذاتية λ_i

3- تعريف المصفوفة المحددة إيجابياً :

نقول عن أية مصفوفة مربعة ومتناظرة A ومن المرتبة $n * n$، إنها محددة إيجابياً إذا كانت نتيجة جداء الصيغة التربيعية $(X'A\,X)$ تساوي قيمة عددية موجبة، حيث X أي شعاع له نفس المرتبة n ولا يساوي الصفر . ونكتب ذلك كما يلي:

نقول عن المصفوفة A المربعة والمتناظرة إنها محددة إيجابياً إذا كان :

$$X'A\,X > 0 \qquad : \left(\text{بشرط } X \neq 0\right) \qquad (38-2)$$

ونقول عن المصفوفة A المربعة والمتناظرة إنها شبه محددة إيجابياً إذا كان :

$$X'A\,X \geq 0 \qquad : \left(\text{بشرط } X \neq 0\right) \qquad (39-2)$$

4- **إذا كانت A مصفوفة مربعة ومحددة إيجابياً** $(\boldsymbol{X'A\,X} > 0)$، فإن جميع قيمها الذاتية تكون موجبة، وبالمقابل يمكننا أن نقول عن أية مصفوفة مربعة ومتناظرة A أنها محددة إيجابياً إذا كانت جميع قيمها الذاتية موجبة $(\lambda_i > 0)$، وعندها فإن عدد تلك القيم يكون مساوياً لرتبتها r حيث $(r \leq n)$ وحيث r مرتبة A .

أما إذا كانت إحدى القيم الذاتية معدومة فإن المصفوفة A تسمى مصفوفة شاذة (Singular) لأن محددها يكون معدوماً $(|A| = 0)$ حسب النظرية (2) .

5- **إذا كانت A مصفوفة مربعة وشبه محددة إيجابياً** $(\boldsymbol{X'A\,X \geq 0})$، وكانت رتبتها تساوي r (حيث $r \leq n$)، فإنه يكون لها r قيمة ذاتية غير معدومة. ويكون لها $(n-r)$ قيمة ذاتية معدومة. وتستخدم هذه النظرية كثيراً في التحليل متعدد المتغيرات وخاصة في التحليل التمييزي .

6- إذا كانت A مربعة ومحددة إيجابياً فإن مقلوبها $\boldsymbol{A^{-1}}$ يساوي :

$$A^{-1} = E * \Lambda^{-1} * E' = \sum_{i=1}^{n} \frac{1}{\lambda_i} \boldsymbol{e_i} * \boldsymbol{e'_i} \quad : \lambda_i \neq 0 \qquad (40-2)$$

حيث أن Λ هي المصفوفة القطرية $\Lambda = Diag\,(\lambda_1 \;\; \lambda_2 \; \ldots \lambda_n)$

7- **إذا كانت A مصفوفة متناظرة وحقيقية** من $n * \mathrm{n}$ وكانت B مصفوفة متناظرة وحقيقية ومحددة إيجابياً فإنه يكون للمعادلة الذاتية العامة التالية:

$$A * \boldsymbol{e} = \lambda * B * \boldsymbol{e} \qquad (41-2)$$

n شعاعاً ذاتياً متعامداً $(e_1 \;\; e_2 \;\; e_3 \; \ldots e_n)$، وتكون هذه الأشعة متعامدة مع المصفوفة B ، أي أنها تحقق العلاقة التالية:

$$\begin{aligned} \boldsymbol{e'_i} * B * \boldsymbol{e_j} &= 0 \qquad : \quad i \neq j \\ \boldsymbol{e'_i} * B * \boldsymbol{e_j} &= 1 \qquad : \quad i = j \end{aligned} \qquad (42-2)$$

وبالمقابل يكون بالنسبة لـ A ما يلي:

$$\begin{aligned} \boldsymbol{e'_i} * A * \boldsymbol{e_j} &= 0 \qquad : \quad i \neq j \\ \boldsymbol{e'_i} * A * \boldsymbol{e_j} &= 1 \qquad : \quad i = j \end{aligned} \qquad (43-2)$$

ويمكن كتابة هاتين العلاقتين بشكل مصفوفي موحد كما يلي:

$$\begin{aligned} \boldsymbol{e'} * B * \boldsymbol{e} &= I \\ \boldsymbol{e'} * A * \boldsymbol{e} &= \Lambda \end{aligned} \qquad (44-2)$$

8- **نتيجة هامة:** إذا كان X_i حل خاص للمعادلة المميزة (2-11) ويقابل القيمة الذاتية λ_i فإنه يحقق المعادلة (2-10) ويمكن كتابته ومعالجته كما يلي:

$$A * X_i = \lambda_i * X_i \qquad (45-2)$$

لنضرب الطرفين من اليسار بـ X'_i فنحصل على أن:

$$X'_i * A * X_i = \lambda_i * X'_i * X_i = \lambda_i \; \|X_i\|^2 \qquad (46-2)$$

نقسم الطرفين على $\|X_i\|^2$ فنستنتج العلاقة الهامة التالية:

$$\frac{X'_i * A * X_i}{\|X_i\|^2} = \lambda_i \qquad (47-2)$$

وهي علاقة هامة ولها استخدامات متعددة .

وكذلك نجد أنه إذا كان $\boldsymbol{e_i}$ هو الحل الممعير للمعادلة (2-11)، المقابل للقيمة الذاتية λ_i فإنه يحقق المعادلة (2-10) ونكتب ذلك كما يلي:

$$A\boldsymbol{e_i} = \lambda_i \boldsymbol{e_i} \qquad (48-2)$$

لنضرب الطرفين من اليسار بـ e_i' فنحصل على أن :

$$\boldsymbol{e_i'} * A * \boldsymbol{e_i} = \lambda_i * \boldsymbol{e_i'} * \boldsymbol{e_i} = \lambda_i * 1 = \lambda_i \qquad (49-2)$$

ومنها نستنتج العلاقة الهامة التالية:

$$\boldsymbol{e_i'} * A * \boldsymbol{e_i} = \lambda_i \qquad (50-2)$$

مثال (2-3): نأخذ نتائج المثال(2-1) السابق حيث كان له القيمتين الذاتيتين التاليتين $\lambda_1 = 4$, $\lambda_2 = -2$ ويقابلهما الحلين الخاصين والممعيرين التاليين :

$$X_1 = \begin{bmatrix}1\\1\end{bmatrix} \qquad X_2 = \begin{bmatrix}1\\-1\end{bmatrix} \qquad e_1 = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} \qquad e_2 = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}$$

وبذلك نجد أن :

$$\|X_1\|^2 = 2 \qquad \|X_2\|^2 = 2 \qquad \|e_1\|^2 = 1 \qquad \|e_2\|^2 = 1$$

وبتعويض X_i في (2-47) نجد أن :

$$\frac{X_1' * A * X_1}{\|X_1\|^2} = \frac{(1\,,1)\begin{bmatrix}1 & 3\\3 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}}{2} = \frac{(1\,,1)\begin{bmatrix}4\\4\end{bmatrix}}{2} = \frac{8}{2} = 4 = \lambda_1$$

$$\frac{X_2' * A * X_2}{\|X_2\|^2} = \frac{(1\,,1)\begin{bmatrix}1 & 3\\3 & 1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}}{2} = \frac{(1\,,1)\begin{bmatrix}-2\\2\end{bmatrix}}{2} = \frac{-4}{2} = -2 = \lambda_2$$

وكذلك نجد أن تعويض $\boldsymbol{e_i}$ في (2-50) يعطينا أن :

$$\boldsymbol{e_1'} * A * \boldsymbol{e_1} = \left(\frac{1}{\sqrt{2}}\,,\frac{1}{\sqrt{2}}\right)\begin{bmatrix}1 & 3\\3 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} = \left(\frac{1}{\sqrt{2}}\,,\frac{1}{\sqrt{2}}\right)\begin{bmatrix}\frac{4}{\sqrt{2}}\\ \frac{4}{\sqrt{2}}\end{bmatrix} = 2 + 2 = 4 = \lambda_1$$

$$\boldsymbol{e_2'} * A * \boldsymbol{e_2} = \left(\frac{1}{\sqrt{2}}\,,-\frac{1}{\sqrt{2}}\right)\begin{bmatrix}1 & 3\\3 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix} = \left(\frac{1}{\sqrt{2}}\,,-\frac{1}{\sqrt{2}}\right)\begin{bmatrix}\frac{-2}{\sqrt{2}}\\ \frac{2}{\sqrt{2}}\end{bmatrix} = -1 - 1 = -2 = \lambda_2$$

2-5 مصفوفة الجذر التربيعي:

لنفترض أن A مصفوفة مربعة من المرتبة $(n * n)$ متناظرة ومحددة إيجابياً، فحسب نظرية التحليل الطيفي يمكن كتابتها كما يلي:

$$A = \sum_{i=1}^{n} \lambda_i * \boldsymbol{e_i} * \boldsymbol{e_i'} \qquad (51-2)$$

وليكن لدينا E مصفوفة الأشعة الذاتية الممعيرة والمتعامدة لها وهي :

$$E = [\boldsymbol{e_1}\ \ \boldsymbol{e_2}\ \ \boldsymbol{e_3}\ \ \dots \boldsymbol{e_n}] \qquad (52-2)$$

حيث أن E متعامدة وتحقق العلاقات التالية : $E * E' = E' * E = I \ \ , \ E' = E^{-1}$

وعندها فإن A يمكن أن تكتب حسب العلاقة (2−37) كما يلي:

$$A_{k*k} = \sum_{i=1}^{k} \lambda_i * \boldsymbol{e_i} * \boldsymbol{e_i'} = E_{k*k} * \Lambda_{k*k} * E'_{k*k} \qquad (53-2)$$

حيث أن Λ هي المصفوفة القطرية للقيم المميزة الموجبة للمصفوفة A وتساوي :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} \quad : \ \lambda_i > 0 \qquad (54-2)$$

والآن لنضرب طرفي العلاقة (1−53) بـ $E * \Lambda^{-1} * E'$ فنحصل على أن:

$$E\Lambda^{-1}E'A = E\Lambda^{-1}E' * E * \Lambda * E' = E * E' = I \qquad (55-2)$$

ثم نضرب طرفي (1−55) من اليمين بـ A^{-1} فنحصل على أن (بعد تبديل موقعي الطرفين) :

$$A^{-1} = E * \Lambda^{-1} * E' * A * A^{-1}$$

ومنها نحصل عل أن :

$$A^{-1} = E * \Lambda^{-1} * E' = \sum_{i=1}^{n} \frac{1}{\lambda_i} * \boldsymbol{e_i} * \boldsymbol{e_i'} \qquad (56-2)$$

ومن جهة أخرى لنفترض أن المصفوفة $\Lambda^{\frac{1}{2}}$ ترمز إلى المصفوفة القطرية التي عناصرها القطرية $\sqrt{\lambda_i}$، أي ان:

$$\Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sqrt{\lambda_k} \end{bmatrix} \quad : \ \lambda_i > 0 \qquad (57-2)$$

قياساً على (1−53) يمكننا أن نعرف مصفوفة الجذر التربيعي $A^{\frac{1}{2}}$ للمصفوفة A بواسطة العلاقة التالية:

$$A^{\frac{1}{2}} = \sum_{i=1}^{k} \sqrt{\lambda_i} * \boldsymbol{e_i} * \boldsymbol{e_i'} = E * \Lambda^{\frac{1}{2}} * E' \qquad (58-2)$$

وكذلك نجد أن $A^{-\frac{1}{2}}$ تساوي: $A^{-\frac{1}{2}} = \sum \frac{1}{\sqrt{\lambda_i}} * \boldsymbol{e_i} * \boldsymbol{e_i'} = E * \Lambda^{-\frac{1}{2}} * E'$

وهي مصفوفة مربعة ومتناظرة ومن المرتبة $n * n$ ، وتتمتع هذه المصفوفة الجذرية بالخواص التالية:

$$1) \ \left(A^{\frac{1}{2}}\right)' = A^{\frac{1}{2}} \qquad \left(\text{لأن } A^{\frac{1}{2}} \text{ قطرية}\right) \qquad (59-2)$$

$$2) \ \left(A^{\frac{1}{2}}\right) * \left(A^{\frac{1}{2}}\right) = A \qquad (60-2)$$

$$3) \ \left(A^{\frac{1}{2}}\right)^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} * \boldsymbol{e_i} * \boldsymbol{e_i'} = E\Lambda^{-\frac{1}{2}}\mathrm{E}' = A^{-\frac{1}{2}} \qquad (61-2)$$

وذلك لأن $\Lambda^{-\frac{1}{2}}$ مصفوفة قطرية وعناصرها القطرية هي $\frac{1}{\sqrt{\lambda_i}}$:

$$4)\ A^{+\frac{1}{2}} * A^{-\frac{1}{2}} = A^{-\frac{1}{2}} * A^{+\frac{1}{2}} = I \qquad (62-2)$$

$$5)\ A^{-\frac{1}{2}} * A^{-\frac{1}{2}} = A^{-1} \qquad (63-2)$$

$$6)\ A^{-\frac{1}{2}} = \left(A^{\frac{1}{2}}\right)^{-1} \Leftrightarrow \left(A^{\frac{1}{2}}\right)^{-1} = A^{-\frac{1}{2}} \qquad (64-2)$$

6-2 مصفوفة المتراجحات والتعظيم :

إن قضية التعظيم تلعب دوراً هاماً في التحليل المتعدد، وفي التحليل التمييزي الخطي. فعلى سبيل المثال فهي تفيدنا في مسألة توزيع أو تصنيف المشاهدات على المجموعات التي يتألف منها المجتمع المدروس. وإن قاعدة التوزيع غالباً ما تكون على شكل تابع خطي من المتحولات التي تعظم الفروقات بين المجموعات بالنسبة لتبايناتها الداخلية. وفي طريقة المركبات الأساسية والعوامل الرئيسية نعتمد على تركيب خطي من المتحولات لتعظيم الجدوى الاقتصادية لها .

وتلعب المتراجحات المصفوفية دوراً كبيراً في البرهان على بعض النظريات وفي تعظيم النتائج المستخرجة منها. وأهم هذه المتراجحات هي:

- **متراجحة (كوشي– شوارز)** Cauchy- shwarz :

لنفترض أنه لدينا شعاعين عموديين b, d من المرتبة$p*1$، فعندها يكون بينهما المتراجحة التالية:

$$(b' * d)^2 \leq (b' * b) * (d' * d) \qquad (65-2)$$

وتحدث المساواة فقط عندما يكون $b = cd$ أو $d = cb$ حيث c ثابت ما لا يساوي الصفر.

- **متراجحة (كوشي – شوارز) الموسعة :**

لنفترض أنه لدينا شعاعين b , d من المرتبة $p*1$، وأنه لدينا مصفوفة B محددة إيجابياً ومتناظرة، فعندها يكون لدينا بينهم المتراجحة التالية:

$$(b' * d)^2 \leq (b' * B * b) * (d' * B^{-1} * d) \qquad (66-2)$$

وتحدث المساواة فقط عندما يكون: $b = cB^{-1}d$ أو $d = cBb$ ، حيث c ثابت ما لا يساوي الصفر.

- **مسألة التعظيم:**

لنفترض أنه لدينا مصفوفة محددة إيجابياً ومتناظرة B، وأنه لدينا شعاعاً معلوماً d، وإذا كان X شعاعاً كيفياً (مجهولاً) وغير معدوم، فإن تعظيم النسبة التالية يساوي :

$$max\left[\frac{(X' * d)^2}{X' * B * X}\right] = d' * B^{-1} * d \qquad (67-2)$$

وهذه المساواة تحدث عندما $X = cB^{-1}d$

ويمكن البرهان على هذه النتيجة اعتماداً على متراجحة (كوشي– شوارتز) الموسعة بالنسبة لـ X، فنجد أن العلاقة (2–66) تأخذ الشكل التالي:

$$(X' * d)^2 \leq (X' * B * X)(d' * B^{-1} * d) \qquad (68-2)$$

وبما أن $X \neq 0$ وأن B محددة إيجابياً، فإن جداء الصيغة التربيعية لهما يساوي قيمة موجبة أي أن:

$$X' * B * X > 0$$

لذلك نقسم الآن طرفي المتراجحة (2–68) على الكمية العددية الموجبة السابقة $(X' * B * X)$، فنحصل على الحد الأعلى للنسبة المذكورة . أي نحصل على أن :

$$\frac{(X' * d)^2}{X' * B * X} \leq d' * B^{-1} * d \qquad (69-2)$$

وهذا يعني أن:

$$max\,\frac{(X'*d)^2}{X'*B*X} = d' * B^{-1} * d$$

ثم نقوم بالبحث عن قيمة X التي تقابل تلك القيمة العظمى في (2–69)، فنجد أن ذلك يحدث عندما تبلغ النسبة (2–69) حدها الأعلى $(d' * B^{-1} * d)$، ولقد تم البرهان على إن هذا يحدث عندما تأخذ X القيمة: $X = cB^{-1}d$ حيث c ثابت كيفي $0 \neq$.

وللتحقق من ذلك نعوض ذلك الحل في (2–69)، وبما أن B^{-1} متناظرة فإن $(B^{-1})' = B^{-1}$ ، وبالتالي يكون لدينا :

$$\frac{(cd' * B^{-1} * d)^2}{cd' * B^{-1} * B * cB^{-1} * d} = \frac{c^2(d'B^{-1}d)^2}{c^2 d' * B^{-1} * d} = d' * B^{-1} * d \qquad (70-2)$$

وهذا يعني أن الحل الذي يعظم النسبة $\frac{(X'*d)^2}{X'*B*X}$ هو الحل X الذي يساوي $X = cB^{-1}d$ ، وإن أكبر قيمة لتلك النسبة تساوي :

$$max\left[\frac{(X' * d)^2}{X' * B * X}\right] = d' * B^{-1} * d \qquad (70a-1)$$

2–7 مسألة تعظيم الصيغة التربيعية من نقطة على سطح الكرة الواحدية :

لنفترض أن B هي مصفوفة محددة إيجابياً ومتناظرة ومن $n* n$، ولها القيم الذاتية الموجبة والمرتبة التالية: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n > 0$ ، ويقابلها الأشعة الذاتية المميزة والمتعامدة $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \ldots, \boldsymbol{e_n}$ ، وتحقق العلاقات :

$$B\boldsymbol{e}_i = \lambda_i \boldsymbol{e_i} \qquad i: 1\ 2\ 3 \ldots . n$$
$$\boldsymbol{e_i'} * B * \boldsymbol{e_i} = \lambda_i$$

فعندها يكون لدينا :

$$max\,\frac{X' * B * X}{X' * X} = \lambda_1 \quad = \text{أكبر قيمة ذاتية} \qquad (71-2)$$

$$min\,\frac{X' * B * X}{X' * X} = \lambda_n \quad = \text{أصغر قيمة ذاتية} \qquad (72-2)$$

وعدا عن ذلك فإذا كان X متعامداً مع الأشعة الأولى $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_k}$، فعندها يكون لدينا :

$$max\,\frac{X' * B * X}{X' * X} = \lambda_{k+1} \qquad (73-2)$$

وهذا يحدث عندما يكون: $X = e_{k+1}$ حيث $k = 1\ 2\ 3 \ldots (n-1)$

وللبرهان على ذلك نأخذ المصفوفة المتعامدة E والتي أعمدتها مؤلفة من الأشعة المميزة والمتعامدة للمصفوفة B وهي: $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \ldots, \boldsymbol{e_n}$، مع المصفوفة القطرية Λ التي عناصرها القطرية هي جملة القيم الذاتية للمصفوفة B وهي: $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$، واعتماداً على العلاقة (2−58) نجد أنه يمكننا أن نعرف المصفوفة الجذرية $B^{\frac{1}{2}}$ بدلالة E و Λ كما يلي :

$$B^{\frac{1}{2}} = E * \Lambda^{\frac{1}{2}} * E' \qquad (74-2)$$

ولهذا نعوض ذلك في النسبة $\frac{X'*B*X}{X'*X}$ بعد تحليل B إلى جداء مصفوفتين للجذر التربيعي كما يلي: $B = B^{\frac{1}{2}} * B^{\frac{1}{2}}$، ثم نفرض أن $Y = E' * X$ فإن $Y \neq 0$ لأن $X \neq 0$ وبالتالي نحصل من (2−71) على أن :

$$\frac{X' * B^{\frac{1}{2}} * B^{\frac{1}{2}} * X}{X' * X} = \frac{X' * E * \Lambda^{\frac{1}{2}} * E' * E * \Lambda^{\frac{1}{2}} * E' * X}{X' * P * P'X} = \frac{Y' * \Lambda * Y}{Y' * Y} =$$

$$= \frac{\sum^n \lambda_i y_i^2}{\sum^n y_i^2} \leq \lambda_1 \frac{\sum^n y_i^2}{\sum y_i^2} = \lambda_1 \qquad (75-2)$$

وذلك بعد استبدال كل القيم الذاتية λ_i بأكبرها λ_1 وإخراجها خارج المجموع الأخير وبذلك نكون حصلنا على أن:

$$max \frac{X' * B * X}{X' * X} \leq \lambda_1 \qquad (76-2)$$

فمثلاً إذا أخذنا أي حل ممعير مثل $X = \boldsymbol{e_1}$ فإننا نجد أنه يحقق $\boldsymbol{e_i'} * \boldsymbol{e_1} = 1$، وبالتعويض نحصل على أن تلك النسبة تساوي اعتماداً على (2−76) ما يلي:

$$\frac{\boldsymbol{e}' * B * \boldsymbol{e}}{\boldsymbol{e}' * \boldsymbol{e}} = \boldsymbol{e_i'} * B * \boldsymbol{e_1} = \lambda_1 \qquad (77-2)$$

وهكذا يتم البرهان على البند الثاني فنحصل على أن أصغر قيمة لتلك النسبة تساوي λ_n أصغر القيم الذاتية لـ B

$$min \frac{X' * B * X}{X' * X} = \lambda_n \qquad (78-2$$

وللبرهان على البند الثالث نأخذ العلاقة العامة التالية :

$$X = E' * Y = [e_1, e_2, \ldots, e_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = e_1 y_1 + e_2 y_2 + \cdots + e_n y_n$$

وبما أن X متعامدة مع الأشعة الأولى المتعامدة $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_k}$، فإنه إذا ضربنا X من اليسار بـ $\boldsymbol{e_1}$ فإن الجداء $e_1' * X$ يكون معدوماً. ومنه نجد أن:

$$\boldsymbol{e_1'} * X = \boldsymbol{e_1'} * E' * Y \Longrightarrow \boldsymbol{e_1'}\boldsymbol{e_1} y_1 + \boldsymbol{e_1'}\boldsymbol{e_2} y_2 + \cdots + \boldsymbol{e_1'}\boldsymbol{e_n} y_n = 0 \quad (79-2)$$

$$= 1 * y_1 + 0 + 0 + \cdots + 0 = 0 \Longrightarrow y_1 = 0$$

وبطريقة مشابهة نجد أن: $y_1 = y_2 = y_3 = \cdots y_k = 0$

وبذلك تأخذ النسبة السابقة الشكل التالي :

$$\frac{X' * B * X}{X' * X} = \frac{\sum_{i=k+1}^{r} \lambda_i * y_i^2}{\sum_{i=k+1} y_i^2} \leq \lambda_{k+1} \qquad (80-2)$$

لأنه إذا وضعنا λ_i مساوية لأكبرها λ_{k+1}، فإننا نحصل على القيمة العظمى لها والتي تساوي λ_{k+1} .

الخلاصة : إذا أخذنا أي شعاع ممعير مثل $\boldsymbol{e_0}$ فإن النسبة $\frac{X'*B*X}{X'*X}$ يكون لها نفس القيمة التي للصيغة التربيعية $(\boldsymbol{e_0'} * B * \boldsymbol{e_0})$ ، حيث أن $\boldsymbol{e_0'}$ هو منقول الشعاع الممعير المساوي لـ: $\boldsymbol{e_0'} = \frac{X'}{\sqrt{X'*X}}$ والذي طوله يساوي الواحد .

وبالمقابل نجد أن المعادلة (1−53) تشير إلى أن القيمة الذاتية λ_1 هي القيمة الكبرى للصيغة التربيعية $(X' * B * X)$. من أجل جميع قيم X التي تبعد عن مركزها الأصلي بمقدار الواحد (قيم X التي تقع على محيط كرة نصف قطرها يساوي الواحد) .

وكذلك نجد أن القيمة الذاتية λ_n هي القيمة الصغرى للصيغة التربيعية $(X' * B * X)$ ، من أجل جميع قيم X التي تبعد عن المركز بمقدار الواحد (قيم X التي تقع على محيط كرة نصف قطرها يساوي الواحد).

أي أن القيمة الذاتية الكبرى λ_1 والقيمة الذاتية الصغرى λ_n تقدم لنا عرضاً للقيم الخارجية للصيغة $(X' * B * X)$ من أجل جميع قيم X التي تقع على محيط الكرة الواحدية .

وإن القيم الذاتية الوسطى للمصفوفة المحددة إيجابياً B وذات المرتبة $(n * n)$ أيضاً يمكن أن تفسر على أنها قيم خارجية . عندما تأخذ X أوضاعاً متعامدة مع الخيارات الأخرى .

المراجع :

1−مجدي الطويل: المصفوفات (النظرية والتطبيق) − الباب الثالث− جامعة القاهرة− كلية الهندسة− عام (غير مذكور).

2- Lancaster, k. Mathematical Economics, Macmilan Comb. Landon, New York, - Rusian.ed. Moscow- Sof. Padio. 1972

3- Johnson R.A. Wichern, D. Applied Multivariate statistical Analysis 2d Ed. Prentice- Hall Intenational Editions. London . 1988

4- Webb, A. Statistical Pattern Recignition ,3nd.ed, QinetiQ Ltd,Malvern, UK . 2002.

أسس
التحليل الإحصائي متعدد المتغيرات

الجزء الثاني
الاستدلال الاحصائي

الفصل الأول

الاستدلال بواسطة اختبارات الفرضيات البسيطة

1-1 : تمهيد :

تعتمد اختبارات الفرضيات على البيانات المتوفرة عن الظواهر المدروسة في المجتمعات الإحصائية. والبيانات هي قيم عددية أو حالات وصفية تعبر عن المتحولات المعرفة على الظاهرة المدروسة . وبذلك ويمكننا تصنيف البيانات إلى نوعين أساسيين هما:

أ- بيانات كمية: وهي بيانات عددية عن متحولات قابلة للقياس بواحدات قياس محددة، وهذه البيانات يمكن أن تكون:

- منقطعة: كعدد أفراد الأسرة- وعدد الطلاب- وعدد السيارات ...الخ .
- مستمرة: كعمر الإنسان- درجة الحرارة- مقدار الدخل ....الخ .

ب- بيانات نوعية: وهي حالات وصفية لمتحولات غير قابلة للقياس، وهذه المعلومات يمكن أن تكون:

- أسمية: كحالات الجنس- حالات العمل- الحالة الاجتماعية ...الخ .
- مرتبة: كحالات التعليم- حالات الوظيفة- حالات الرضا ..الخ .

ويتم تجميع البيانات عن الظاهرة المدروسة أو عن المتحولات المطلوبة من عناصر المجتمع الاحصائي بواسطة أحد الأسلوبين :

- الحصر الشامل: وهو يشمل جميع عناصر المجتمع الاحصائي المؤلف من N عنصراً
- المسح بالعينة: وهو يشمل جزء من المجتمع ويكون على شكل عينة حجمها n عنصراً، تسحب عشوائياً من عناصر ذلك المجتمع بدون إعادة أو مع الاعادة .

وتستخدم بيانات هذه العينة لتقدير معالم المجتمع المجهولة مثل: المتوسط μ أو التباين σ^2 أو النسبة فيه R، وذلك من خلال استخدام المؤشرات المقابلة لها والمحسوبة من العينة، والتي سنسميها (مؤشرات العينة)، وهي متوسط العينة $\bar{x}$ وتباين العينة S^2 والنسبة في العينة r . ويبُرهن في نظرية العينات أن مؤشرات العينة المصححة هي تقديرات غير متحيزة وفعالة ومتماسكة لمعالم المجتمع المقابلة لها .

والآن لنفترض أننا نقوم بدراسة خصائص أحد المتحولات الكمية X من عناصر المجتمع المدروس (وليكن X وزن الطالب في الجامعة)، لذلك نسحب عينة عشوائية من طلاب ذلك المجتمع بحجم n طالباً ، فنحصل من كل طالب i فيها على وزن محدد x_i ، ويكون لدينا القياسات التالية :

$$X: x_1 \quad x_2 \quad x_3 \dots \dots . x_i \dots \dots x_n$$

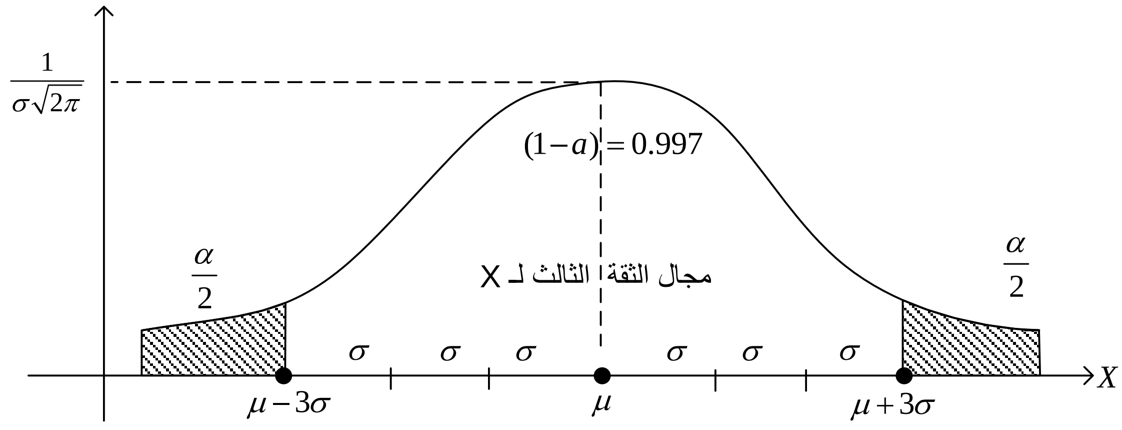
وبناء على نظرية العينات نقوم بتقدير معالم المجتمع المجهولة من مؤشرات العينة المعلومة وحساب مقدار الخطأ المرتكب في كل تقدير وفق الجدول التالي :

جدول $(1-1)$: معالم المجتمع وتقديراتها من مؤشرات العينة :

معالم المجتمع المجهولة للمتحول	تقديرات المعالم من مؤشرات العينة
1- قيمة المتوسط في المجتمع μ	تقدر من متوسط العينة: $\bar{x} = \frac{1}{n}\sum x_i$ ونكتب ذلك كما يلي: $\tilde{\mu} = \bar{x}$
2- قيمة تباين المجتمع σ^2	تقدر من تباين العينة المصحح والمعرف بالعلاقة: $S^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ ونكتب ذلك كما يلي: $\tilde{\sigma}^2 = S^2$
3- قيمة النسبة في المجتمع R للمتحول الثنائي (0 او 1)	تقدر من النسبة في العينة $r = \frac{m}{n}$ ونكتب ذلك كما يلي: $\tilde{R} = r$ حيث m عدد الظهور و n حجم العينة وتباينها يساوي $S^2 \approx r(1-r)$
4- الخطأ المعياري المرتكب في تقدير متوسط المجتمع في حالة السحب مع الاعادة $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$	يقدر من خلال ما يقابله في العينة : $\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$
5- الخطأ المعياري المرتكب في تقدير المتوسط في حالة السحب بدون إعادة : $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} * \frac{\sigma^2}{n}}$	يقدر من خلال العينة بالعلاقة: $\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N-1} * \frac{S^2}{n}} \approx \sqrt{\frac{N-n}{N} * \frac{S^2}{n}}$
6- الخطأ المعياري المرتكب في تقدير النسبة R في حالة السحب مع الاعادة: $\sigma_r = \sqrt{\frac{R(1-R)}{n}}$	يقدر من خلال خطأ النسبة r في العينة كمايلي: $\tilde{\sigma}_r = \sqrt{\frac{r(1-r)}{n}}$
7- الخطأ المعياري المرتكب في تقدير النسبة R في حالة السحب بدون إعادة : $\sigma_r = \sqrt{\frac{N-n}{N-1} * \frac{R(1-R)}{n}}$	يقدر من خلال خطأ النسبة r في العينة كما يلي: $\tilde{\sigma}_r = \sqrt{\frac{N-n}{N-1} * \frac{r(1-r)}{n}} = \sqrt{\frac{N-n}{N} * \frac{r(1-r)}{n}}$

ولكن عملية التقدير لا تنتهي عند ذلك ، بل يجب إنشاء مجال ثقة يحتوي المعلم الذي نقدره في المجتمع. وحتى نستطيع إنشاء مجال ثقة يجب أن يكون التوزيع الاحتمالي للمتحول X معلوماً. ومنه يجب أن يكون توزيع متوسط العينة $\bar{x}$ معلوماً أيضاً.

واختصاراً لهذه القضايا نفترض أن المتحول المدروس X يخضع في المجتمع للتوزيع الطبيعي العام $N(\mu\ ,\sigma^2)$ الذي متوسطه μ وتباينه σ^2 وهو يرسم الشكل التالي :



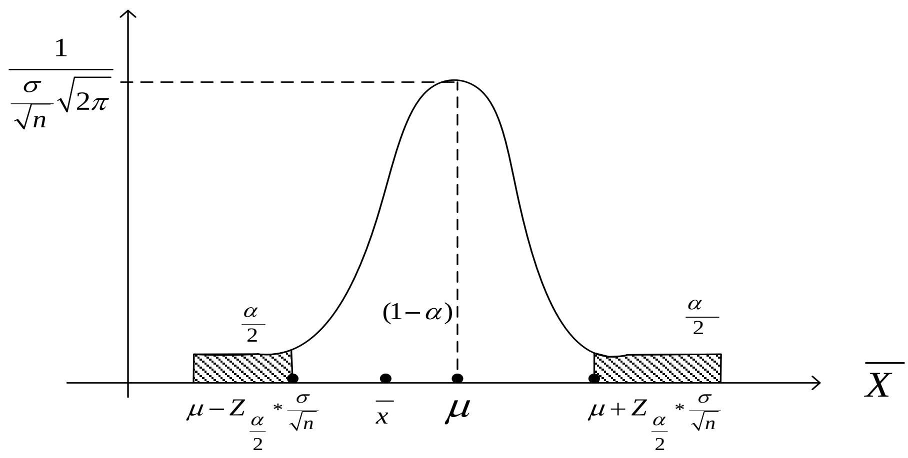
الشكل $(1-1)$ منحنى التوزيع الطبيعي للمتحول X ومجال الثقة الثالث

وبناءً على خواص التوزيع الطبيعي يمكننا إنشاء مجال الثقة لـ X المقابل لاحتمال الثقة $(1-\propto)$ أو لمستوى دلالة $(\propto)$ بحيث يكون:

$$P\left[\mu - Z_{\frac{\propto^*}{2}}\sigma \leq X \leq \mu + Z_{\frac{\propto^*}{2}}\sigma\right] = 1-\propto \qquad (1-1)$$

حيث أن $Z_{\frac{\propto}{2}}$ هي قيمة متحول التوزيع الطبيعي المعياري التي تترك نصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على يمينها و$\left(\frac{\propto}{2}\right)$ على يسار $-Z_{\frac{\propto}{2}}$، [حيث استبدلنا الرمز $Z_{1-\frac{\propto}{2}}$ بالرمز $Z_{\frac{\propto}{2}}$ للاختصار]. ولقد أنشأنا على الشكل $(1-1)$ مجال الثقة الثالث الذي يقابل احتمال ثقة $(0,997)$.

وعندما نسحب عينة من عناصر المجتمع بحجم n نحصل منها على متوسط هو $\bar{x}$ وعلى تباين هو S^2، ولكن هذه العينة ليست وحيدة بل يمكن أن يسحب غيرنا وغيرنا عينات أخرى، فيحصل على متوسطات أخرى $\bar{x}_k$ وتباينات أخرى S_k^2، وبما أن عدد العينات الممكنة يساوي C_N^n عينة (في حالة السحب بدون إعادة)، فإن هذا يعني أنه يمكننا أن نحصل على C_N^n متوسطاً $\bar{x}_k$، وكل منها يعتبر تقديراً لمتوسط المجتمع μ . وبناءً عليه يكون متوسط العينة $\bar{x}$ هو الآخر متحولاً عشوائياً جديداً متوسطه μ أيضاً، ولكن تباينه يساوي $\frac{\sigma^2}{n}$ (وانحرافه المعياري يساوي $\frac{\sigma}{\sqrt{n}}$) ويخضع للتوزيع الطبيعي العام $N\left(\mu\ ,\frac{\sigma^2}{n}\right)$، والذي يأخذ الشكل الضامر والمتطاول التالي:



الشكل $(2-1)$**: توزيع متوسط العينة** $\bar{x}$

وبناءً على شكل التوزيع $(2-1)$ وقياساً على العلاقة $(1-1)$ يمكننا أيضاً إنشاء مجال الثقة للمتوسط $\bar{x}$ المقابل لاحتمال الثقة $(1-\propto)$ أو لمستوى الدلالة $(\propto)$ بحيث يكون:

$$P\left[\mu - Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}} \leq \bar{\text{x}} \leq \mu + Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}}\right] = 1-\propto \qquad (2-1)$$

وحيث أن $Z_{\frac{\propto}{2}}$: هي القيمة الجدولية (الحرجة) لمتحول التوزيع الطبيعي المعياري المقابلة لنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على اليمين و $\left(\frac{\propto}{2}\right)$ على اليسار. وهذا المجال يضمن لنا أن تقع قيمة $\bar{x}$ ضمنه باحتمال يساوي $(1-\propto)$.

وحتى نستفيد من العلاقة (2-1) نطرح من أطرافها μ ثم نقسمها على $\frac{\sigma}{\sqrt{n}}$ فنحصل على ما يلي:

$$P\left[-Z_{\frac{\propto}{2}} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\propto}{2}}\right] = 1-\propto \qquad (3-1)$$

وهذا المجال يضمن لنا أن تقع قيمة المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ ضمن المجال $\left[-Z_{\frac{\propto}{2}}\ ,+Z_{\frac{\propto}{2}}\right]$ باحتمال يساوي $(1-\propto)$، وهنا نلاحظ أن المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ هو متحول عشوائي ثالث، وهنا نميز بين حالتين هما:

أ- إذا كانت قيمة σ^2 وبالتالي قيمة σ معلومة من المجتمع: فإن المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$ لأنه ناتج عن معايرة المتحول المتوسط $\bar{x}$، لذلك نرمز له بـ Z ونكتبه كما يلي:

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \qquad (4-1)$$

وهكذا نجد أنه يمكننا اعتبار هذا المقدار مؤشراً لاختبار أي فرضية حول متوسط المجتمع μ (مثل الفرضية $H_0 : \mu = \mu_0$)، وذلك عندما نعوض μ بـ μ_0 ونقوم بحساب قيمة Z من العلاقة $(4-1)$ وبشرط أن تكون قيمة σ معلومة .

ثم نقارن قيمة Z المحسوبة مع طرفي المجال المعرف في [3-1] وهو $\left[-Z_{\frac{\propto}{2}}\ ,+Z_{\frac{\propto}{2}}\right]$. ونتخذ القرار كمايلي : إذا كانت قيمة Z المحسوبة واقعة ضمن ذلك المجال نقبل تلك الفرضية $(H_0 : \mu = \mu_0)$ ،

أما إذا كانت قيمة Z المحسوبة واقعة خارجه من الطرفين، فإننا نرفض الفرضية $(H_0: \mu = \mu_0)$، وبعبارة أخرى إذا كانت $|Z| \leq Z_{\frac{\propto}{2}}$ نقبل الفرضية المذكورة ، وإذا كانت $|Z| > Z_{\frac{\propto}{2}}$ نرفض الفرضية السابقة. علماً بأن احتمال قبول الفرضية $(H_0: \mu = \mu_0)$ يساوي $(1-\propto)$ واحتمال رفضها من الطرفين يساوي $(\propto)$.

ب-أما إذا كانت قيمة σ^2 في المجتمع مجهولة، فإننا نلجأ إلى تقديرها من خلال تباين العينة المصحح S^2 ، ونقوم باستبدال قيمة σ في (1-4) بتقديرها S من العينة، فنحصل على متحول عشوائي جديد مركب من متحولين عشوائيين $\bar{x}$ و S ونرمز له بـ t ونكتبه كما يلي:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \qquad (5-1)$$

ومعلوم من نظرية الاحتمالات والاحصاء الرياضي أن المتحول t يخضع لتوزيع (ستودينت) ذي(n-1) درجة حرية. (وهو توزيع يتقارب مع التوزيع الطبيعي المعياري عندما تصبح $n \geq 30$) . وقياساً على العلاقة (1-3) يمكننا أن ننشأ مجال الثقة للمتحول t كما يلي:

$$P\left[-t_{\frac{\propto}{2}} \leq \frac{\bar{x} - \mu}{S/\sqrt{n}} \leq +t_{\frac{\propto}{2}}\right] = 1-\propto \qquad (6-1)$$

حيث أن $t_{\frac{\propto}{2}}$ هي قيمة متحول (ستودينت) الجدولية(أوالحرجة)، المقابلة لنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على اليمين ولـ $\left(\frac{\propto}{2}\right)$ على اليسار ولـ (n-1) درجة حرية، وهذا المجال يضمن لنا أن تقع قيمة t ضمن المجال $\left[-t_{\frac{\propto}{2}}\ , +t_{\frac{\propto}{2}}\right]$ باحتمال قدره $(1-\propto)$. وهكذا نجد أنه يمكننا الاستفادة من المتحول t في حالة العينات الصغيرة في اختبار أي فرضية حول متوسط المجتمع μ (مثل $H_0: \mu = \mu_0$) فنعوض μ بـ μ_0 ونحسب قيمة t من العلاقة (1-5) ، ثم نقارنها مع طرفي المجال المعرف في (1-6) وهو المجال $\left[-t_{\frac{\propto}{2}}\ ,\ +t_{\frac{\propto}{2}}\right]$، ونتخذ القرار كما يلي:

إذا كانت قيمة t المحسوبة واقعة ضمن ذلك المجال تقبل تلك الفرضية $(H_0: \mu = \mu_0)$، أما إذا كانت t المحسوبة واقعة خارجه فإننا نرفض الفرضية $(H_0: \mu = \mu_0)$.

وبعبارة أخرى: إذا كانت $|t| \leq t_{\left(\frac{\propto}{2}, n-1\right)}$ فإننا نقبل الفرضية $(H_0: \mu = \mu_0)$، أما إذا كانت $|t| > t_{\left(\frac{\propto}{2}, n-1\right)}$ فإننا نرفض الفرضية المذكورة. علماً بأن احتمال قبول الفرضية $(\mu = \mu_0)$ يساوي $(1-\propto)$ واحتمال رفضها من الطرفين يساوي $(\propto)$.

وهكذا نجد أنه يمكننا، وباتباع نفس الأسلوب، استنباط العديد من المؤشرات لاستخدامها في اختبارات الفرضيات المختلفة (كل حالة حسب طبيعتها وحسب توزيعها الاحتمالي) ، فنحصل على مؤشرات لتقدير النسبة R والتباين σ^2 وغيرهما.

ومن جهة أخرى يمكننا أن ننشأ مجال ثقة يحوي متوسط المجتمع μ وذلك بمعالجة العلاقة (1-2) وطرح المقدار $(\bar{x}-\mu)$ من أطرافها فنحصل على المجال التالي:

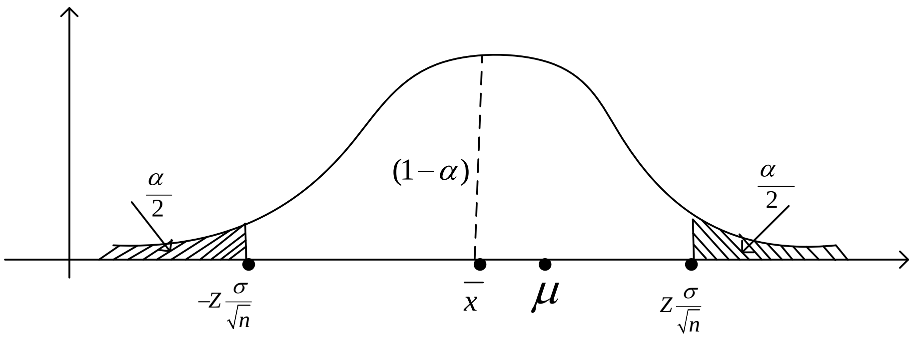
$$P\left[\bar{x}-Z_{\frac{\alpha}{2}}*\frac{\sigma}{\sqrt{n}}\leq\mu\leq\bar{x}+Z_{\frac{\alpha}{2}}*\frac{\sigma}{\sqrt{n}}\right]=1-\alpha \qquad (7-1)$$

وهو مجال مركزه متوسط العينة $\bar{x}$ (وليس μ) ونصف طوله يساوي $\left(Z_{\frac{\alpha}{2}}*\frac{\sigma}{\sqrt{n}}\right)$ ، ويضمن لنا أنه يحتوي على متوسط المجتمع المجهول μ باحتمال $(1-\alpha)$ والشكل (1-3) يوضح ذلك .

وإذا كان التباين σ^2 مجهولاً فإننا نستبدله بتقديره S^2، وعندها فإن مجال الثقة (1-7) يصبح معرفاً على توزيع (ستودينت) ذي $(n-1)$ درجة حرية ويأخذ الشكل التالي:

$$P\left[\bar{x}-t_{\frac{\alpha}{2}}*\frac{S}{\sqrt{n}}\leq\mu\leq\bar{x}+t_{\frac{\alpha}{2}}*\frac{S}{\sqrt{n}}\right]=1-\alpha \qquad (8-1)$$

وهو مجال مركزه $\bar{x}$ ونصف طوله $\left(t_{\frac{\alpha}{2}}*\frac{S}{\sqrt{n}}\right)$ ، ويضمن لنا أنه يحتوي على متوسط المجتمع المجهول μ باحتمال $(1-\alpha)$.



الشكل (1-3) مجال الثقة لـ μ

2-1 : أنواع وأسماء أهم الاختبارات :

توضع الفرضيات على معالم المجتمع (مثل μ أو R أو σ^2) أو على بعض خصائصه (مثل: الاستقلال، التوافق، التكرار، الالتواء ...الخ) وتختبر باستخدام معلومات العينة ومؤشرات الاختبار.

وتصنف الاختبارات إلى نوعين أساسيين :

أ- الاختبارات المعلمية: وتطبق على المتحولات الكمية .

ب- الاختبارات اللامعلمية: وتطبق على المتحولات النوعية والرتبية.

كما يمكن تصنيف الاختبارات حسب عدد المجتمعات والعينات كما يلي :

1-اختبارات لمجتمع واحد (عينة واحدة) .

2- اختبارات لمجتمعين (عينتين مستقلتين) .

3- اختبارات لعدة مجتمعات (لعدة عينات مستقلة) .

4-اختبارات لعينتين مترابطتين

5-اختبارات لعدة عينات مترابطة .

ويتضمن الجدول التالي أهم الاختبارات المعلمية واللامعلمية ومجالات تطبيقها

جدول (1−2): أهم الاختبارات الاحصائية المعلمية واللامعلمية:

أهم الاختبارات المعلمية	أهم الاختبارات اللامعلمية
1− اختبار Z الطبيعي لعينة واحدة وهو يطبق على متوسط المجتمع μ وعلى النسبة R فيه، مثل العلاقة (1−4)	**1− اختبار χ^2 : للاستقلال والارتباط بين متحولين نوعيين أو أحدهما نوعي . في عينة واحدة (غير مرتبة)**
2− اختبار (ستودينت) t لعينة واحدة صغيرة وهو يطبق على μ وعلى R مثل العلاقة (1−5)	**2− اختبار Gamma: للاستقلال والارتباط بين متحولين مرتبين من عينة واحدة**
3− اختبار χ^2 لعينة واحدة ويطبق على تباين المجتمع σ^2	**3− اختبارات الثبات والصدق ويستخدم في الاستبيانات (للمعلومات المرتبة)**
4− اختبار Z الطبيعي لعينتين مستقلتين ويطبق على الفرق بين متوسطي المجتمعين أو على الفرق بين النسبتين فيهما مثل العلاقة (1−24)	**4− اختبار مكنمارا: للحالات غير المرتبة (جدول رباعي)**
5− اختبار (ستودينت) t لعينيتين مستقلتين ويطبق على الفرق $\mu_1 - \mu_2$ وعلى الفرق $R_1 - R_2$، مثل العلاقة (1−25)	**5− اختبار: ويلكوكسن للمتحولات المرتبة**
6− اختبارF لعينتين مستقلتين ويطبق على تبايني مجتمعين $\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$، مثل العلاقة (1−35)	**6− اختبار كروسكال: للمتحولات المرتبة**
7− اختبار t للفرق بين الأزواج المتقابلة (عينتين مترابطتين) مثل العلاقة (1−54)	**7− اختبار مان ويتني: للمتحولات المرتبة**
8− اختبار تحليل التباين الأحادي لأكثر من عينتين مستقلتين، مثل العلاقة (1−40)	**8− اختبار t_B كيندال: للمتحولات المرتبة**
9− اختبار تحليل التباين الثنائي لمؤشرين على عدة مجتمعات .	**9− اختبار t_C كيندال: للمتحولات المرتبة**
10− اختبار χ^2 لتوافق التوزيعات الاحتمالية من عينة واحدة	**10− اختبار الاشارة للمتحولات الثنائية (1 , 0)**
11− اختبار كولموغوروف− سميرنوف لتوافق التوزيعات الاحتمالية	**11− اختبار كوكران− مينتال للبيانات المرتبة**
12− اختبار معنوية معامل الارتباط الخطي (معامل بيرسون)	**12− اختبار معنوية معامل الارتباط الرتبي (سبيرمان) للمتحولات المرتبة**

1-2-1: حساب موثوقية وقوة الاختبار

في الحقيقة إن عملية إجراء اختبار لأية فرضية عدم H_0، تتأثر بعدة عوامل أهمها حقيقة الفرضية H_0 في المجتمع ونوع القرار المتخذ بشأنها ، وبذلك نجد أنه لدينا الحالات التالية:

- إن فرضية العدم H_0 قد تكون بحقيقتها صحيحة أو خاطئة .
- إن القرار الذي سنأخذه حول H_0 يمكن أن يكون قبولاً أو رفضاً لها .

ويمكن وضع تقاطعات هذه الحالات الأربع في جدول كالتالي:

جدول(1-3) : حالات تقاطع حقيقة فرضية العدم مع نوع القرار المتخذ حولها

نوع القرار المتخذ / حقيقة الفرضية H_0	قبول	رفض
H_0 صحيحة	القرار صحيح واحتماله $=1-\alpha$	القرار غير صحيح واحتماله $=\alpha$
H_0 خاطئة	القرار غير صحيح واحتماله $=\beta$	القرار صحيح واحتماله $=1-\beta$

ومن الجدول السابق نلاحظ إنه عندما نتخذ القرار حول H_0 ، فيمكن أن يكون قرارنا غير صحيح في الحالتين التاليتين: رفض الفرضية الصحيحة، قبول الفرضية الخاطئة، وعندها سنرتكب أحد الخطأين التاليين:

- خطأ النوع الأول error type I: وهو قرار رفض الفرضية H_0 رغم إنها صحيحة، وإن احتمال وقوعنا في هذا الخطأ يسمى مستوى الدلالة α، ويسمى الاحتمال المتمم له $(1-\alpha)$ بدرجة الثقة أو بالموثوقية .
- خطأ النوع الثاني error type II : وهو قرار قبول الفرضية H_0 رغم إنها خاطئة أو غير صحيحة، وإن احتمال وقوعنا في هذا الخطأ يساوي عدداً آخراً β، ويسمى الاحتمال المتمم له $(1-\beta)$ بقوة الاختبار .

وبناءً على ذلك تم تعريف قوة الاختبار: بأنها احتمال رفض الفرضية H_0 عندما تكون خاطئة. وهو يتمم احتمال قبولها β، أي أن قوة الاختبار تعرف بالاحتمال المتمم لـ β وهو يساوي :

$$W = 1-\beta \qquad (23-1)$$

ويتم حساب قيم β من تكاملات شرطية معقدة لا مجال للخوض فيها في هذا الفصل.

3-1 : اختبارات معالم مجتمع طبيعي (من عينة واحدة):

1-3-1: اختبار متوسط المجتمع μ (أو النسبة R):

ويتألف من الخطوات التالية:

1- نحدد مستوى الدلالة (α) أو احتمال الثقة $(1-\alpha)$، وعادة يتم وضعه $(\alpha=0.05)$ أو $(\alpha=0.01)$ أو $(\alpha=0.10)$.

2- نضع على متوسط المجتمع μ (أو النسبة R فيه) فرضيتين متنافيتين ومتكاملتين كما يلي:

أ- فرضية العدم: نفترض أن متوسط المجتمع μ يساوي قيمة معلومة μ_0، وهذا يعني أنه لا يوجد فرق معنوي بينه وبين القيمة المفترضة μ_0 (أي عدم وجود فرق بينهما) ونكتب ذلك كما يلي:

$$H_0: \mu = \mu_0 \qquad (9-1)$$

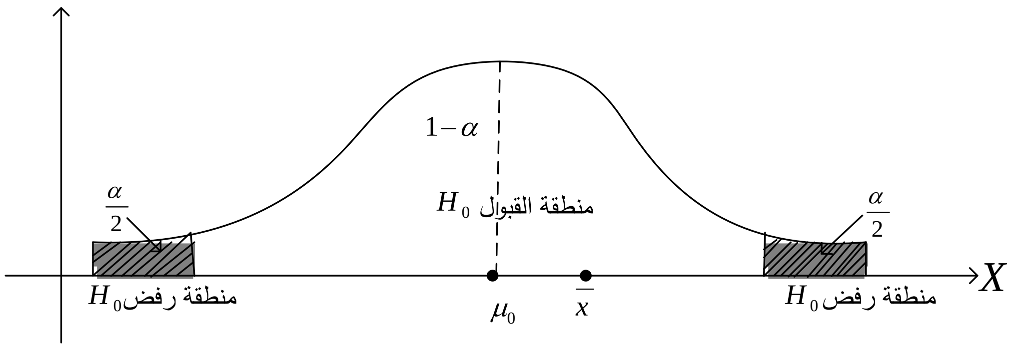
وتكون هذه الفرضية مقبولة إذا كان الفرق $(\mu - \mu_0)$ أو تقديره $(\bar{x} - \mu_0)$ واقعاً ضمن مجال الثقة المحدد للفرق $(\bar{x} - \mu)$ ، ونقرر ذلك من خلال مؤشر الاختبار المناسب.

ب- الفرضية البديلة: وهي الفرضية المعاكسة لفرضية العدم، ومن شكلها تتحدد منطقة الرفض، ويمكن أن تُكتب على أحد الأشكال الثلاثة التالية:

- الشكل الأول: الشكل الثنائي أو شكل عدم التساوي وتكتب الفرضية البديلة فيه كما يلي:

$$H_1: \mu \neq \mu_0 \qquad (10-1)$$

وفيه تكون منطقة الرفض واقعة على الجانبين، ولذلك يسمى هذا الشكل بالاختبار ثنائي الجانب، لأنه يخصص لكل جانب نصف مستوى الدلالة$\left(\frac{\propto}{2}\right)$ ، كما في الشكل التالي :



الشكل (1-4) منطقة القبول ومنطقتي الرفض على اليمين واليسار

- الشكل الثاني: الأحادي اليميني، وتكتب الفرضية البديلة H_1 كما يلي:

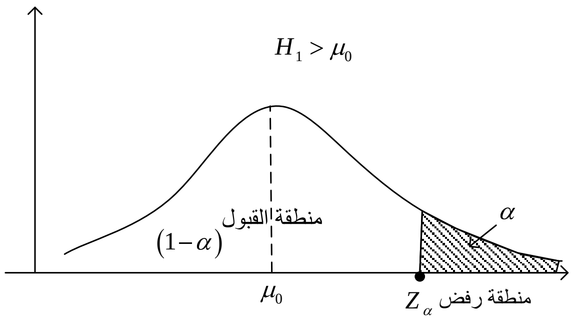
$$H_1: \mu > \mu_0 \qquad (11-1)$$

وفيه تكون منطقة الرفض على اليمين فقط، وتقابل كامل الاحتمال $\propto$ كما على الشكل (1-5).

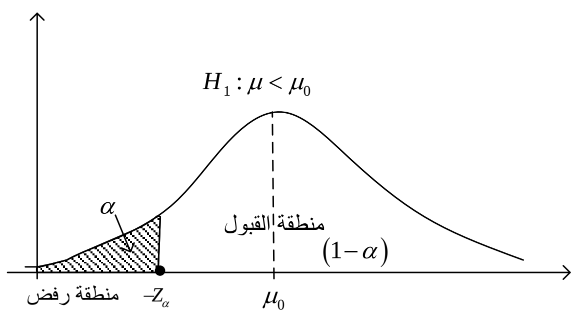
- الشكل الثالث: الأحادي اليساري، وتكتب الفرضية البديلة H_1 كما يلي:

$$H_1: \mu < \mu_0 \qquad (12-1)$$

- وفيه تكون منطقة الرفض على اليسار فقط، وتقابل كامل الاحتمال $\propto$ كما على الشكل (1-6).



الشكل (1-5) أحادي يميني



الشكل (1-6) أحادي يساري

وللتحقق من صحة أو عدم صحة فرضية العدم H_0، يجب علينا أن نسحب عينة عشوائية من المجتمع ونحسب متوسطها $\bar{x}$ ثم نقارنه مع متوسط المجتمع المفترض في الفرضية H_0 وهو μ_0. فإذا كان $\bar{x}$ يساويه أو قريباً منه نقبل فرضية العدم H_0، ونعترف بأن متوسط المجتمع يساوي μ_0، أما إذا كان $\bar{x}$ بعيداً عن μ_0 (يوجد فرق جوهري بينهما)، فإننا نرفض فرضية العدم H_0 ونقبل الفرضية H_1، ونعترف بأن متوسط المجتمع μ لايساوي μ_0، بل يساوي قيمة أخرى أكبر أو أصغر منها. وحتى لا تكون الأمور مزاجية فإن عملية مقارنة $\bar{x}$ مع μ_0، تحتاج إلى أداة إحصائية ورياضية تحدد لنا مقدار الفرق المقبول ومقدار الفرق المعنوي أو الجوهري، وهذه الأداة تسمى مؤشر الاختبار. وهكذا نجد أنفسنا بحاجة قبل كل شيء إلى سحب عينة عشوائية من المجتمع المدروس وحساب مؤشراتها المختلفة .

3- نقوم بتحديد طريقة سحب العينة (مع الإعادة أم بدون إعادة) ، وحسب طريقة السحب المختارة نقوم بحساب حجم العينة من إحدى العلاقتين الخاصتين بتقدير المتوسط (أو النسبة R ضمن القوسين) وهما: [انظر الفقرة (2-4) في الفصل الثاني] .

$$n = \frac{Z^2 S^2}{d^2} = \left(\frac{Z^2 * r(1-r)}{d^2}\right) \quad \text{للسحب مع الإعادة} \qquad (13-1)$$

$$n = \frac{NZ^2 S^2}{Nd^2 + Z^2 S^2} = \left(\frac{NZ^2 r(1-r)}{Nd^2 + Z^2 r(1-r)}\right) \quad \text{للسحب بدون اعادة} \qquad (14-1)$$

حيث أن : S^2 هو تباين العينة أو تقديره من أي عينة سابقة .

وأن: d هو مقدار الدقة المطلوبة وتحدد من قبل الجهات المعينة أومن قبل الباحث .

وأن: Z هي قيمة المتحول الطبيعي المعياري المقابل لنصف مستوى الدلالة $\frac{\propto}{2}$ على الطرفين.

وأن: r هو مقدار النسبة في المجتمع أو اي تقدير لها من خلال أي عينة اختبارية.

وفي حالة اختبار النسب المتوازنة في المجتمع نضع ($r = 0.50$) حتى نحصل على أكبر حجم ممكن للعينة، أما عندما يكون حجم المجتمع N كبيراً أو غير معروف، يفضل استخدام العلاقة (1-13) للسحب مع الإعادة ، ثم نقوم بسحب العينة المذكورة من المجتمع ونحسب متوسطها $\bar{x}$ وتباينها S^2 من العلاقتين:

$$\bar{x} = \frac{1}{n}\sum x_i \qquad (15-1)$$

$$S^2 = \frac{1}{n-1} * \sum (x_i - \bar{x})^2 \qquad (16-1)$$

4- نقوم بحساب مؤشر الاختبار للمتوسط (أو للنسبة ضمن القوسين) من العلاقة المعيارية التالية:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \left(\frac{|r - R_0|}{\sqrt{\frac{R_0(1-R_0)}{n}}}\right) \quad : \quad \left(\sigma \text{ معلوم}\right) \qquad (17-1)$$

وهو متحول عشوائي يخضع للتوزيع الطبيعي المعياري $N(0,1)$، ولكن بما أن تباين المجتمع σ^2 وبالتالي انحرافه المعياري σ يكون غالباً مجهولاً، فإن حساب قيمة Z السابقة يكون أمراً مستحيلاً . ولكي نتخلص من هذه المشكلة نستبدل σ^2 بتباين العينة S^2 كتقدير جيد له، ونعرف مؤشر جديد لاختبار متوسط المجتمع μ (دون تعديل المقام في مؤشر النسبة لأن R_0 تكون معلومة من فرضية العدم H_0) . بالعلاقة التالية :

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \left(\frac{|r - R_0| - \frac{1}{2n}}{\sqrt{\frac{R_0(1 - R_0)}{n}}} \right) \qquad (18 - 1)$$

وهو متحول جديد يخضع لتوزيع (ستودينت) بـ $(n - 1)$ درجة حرية عند اختبار المتوسط، وللتوزيع الطبيعي عند اختبار النسبة .

ملاحظة: إذا كان حجم العينة n كبيراً $(n > 30)$، فإن توزيع (ستودينت) يقترب من التوزيع الطبيعي المعياري، وعندها نعتبر t في العلاقة (18−1) خاضعاً للتوزيع الطبيعي المعياري .

5−نقوم بتحديد منطقتي الرفض والقبول واتخاذ القرار:

لتحديد منطقتي الرفض والقبول ولاتخاذ القرار اللازم حول صحة الفرضية H_0، يوجد طريقتان لاتخاذ القرار المناسب هما: طريقة القيمة الحرجة . وطريقة احتمال الدلالة P . وسنشرحهما كما يلي :

أ− **طريقة القيمة الحرجة** لـ Z أو t : لنفترض أن الاختبار ثنائي الجانب (أي أن $H_1 : \mu \neq \mu_0$)، فعندها يكون مستوى الدلالة ∝ موزعاً على الجانبين، وهنا يكون لدينا حالتان لـ σ هما: إما أن تكون قيمة σ معلومة، أو أن تكون σ مقدرة من العينة بـ S، ولذلك نعالجها كما يلي :

- إذا كانت قيمة σ معلومة: فإننا نحسب قيمة مؤشر الاختبار من العلاقة (17−1) ثم نقارنها مع القيمة الحرجة لمتحول التوزيع الطبيعي $Z_{\frac{\alpha}{2}}$، لذلك نبحث في جدول التوزيع الطبيعي المعياري عن القيمة الحرجة $Z_{\frac{\alpha}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ على الطرفين، ثم نقارن القيمة المحسوبة Z معها، ونتخذ القرار حول H_0 كما يلي:

إذا كانت $|Z| \leq Z_{\frac{\alpha}{2}}$، تكون Z واقعة في منطقة القبول، وعندها نقبل فرضية العدم ونقول بأن $\mu = \mu_0$ ، باحتمال ثقة يساوي ∝−1 .

إذا كانت $|Z| > Z_{\frac{\alpha}{2}}$، تكون Z واقعة في منطقة الرفض وعندها نرفض فرضية العدم ونقول بأن $\mu \neq \mu_0$ ، بمستوى دلالة يساوي ∝ .

- أما عندما يكون σ^2 مجهولاً، فإننا سنحسب قيمة المؤشر t من العلاقة (18−1) ، ثم نبحث في جدول توزيع (ستودينت) عن القيمة الحرجة $t_{\frac{\alpha}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ على الطرفين ولدرجة حرية $(n - 1)$ ونتخذ القرار بالمقارنة كما يلي:

إذا كانت $|t| \leq t_{\frac{\propto}{2}}$ نقبل فرضية العدم ونقول بأن $\mu = \mu_0$ باحتمال ثقة $1-\propto$

إذا كانت $|t| > t_{\frac{\propto}{2}}$ نرفض فرضية العدم ونقول بأن $\mu \neq \mu_0$ بمستوى دلالة $\propto$

ملاحظة: إذا كان الاختبار أحادي الجانب (يميني ويساري) ، فإن القيمة الحرجة لـ Z هي القيمة المقابلة لكامل مستوى الدلالة $\propto$ ونرمز لها بـ $Z_\propto$، وعندها نتخذ القرار عند المقارنة حول H_0 كما يلي:

- إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$ فإننا نقارن قيمة Z المحسوبة مع $Z_\propto$ ، فإذا كانت $Z \leq Z_\propto$ نقبل فرضية العدم H_0، ونقبل بأن $\mu = \mu_0$، أما إذا كانت $Z > Z_\propto$ فإننا نرفض H_0 ونعترف بأن $\mu > \mu_0$ كما في الشكل $(1-5)$ السابق .
- أما إذا كان الاختبار أحادي يساري $(H_1: \mu < \mu_0)$ فإننا نقارن قيمة Z المحسوبة مع $(-Z_\propto)$ السالبة . فإذا كانت $(Z \geq -Z_\propto)$ نقبل فرضية العدم H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كانت $(Z < -Z_\propto)$ فإننا نرفض H_0 ونقبل H_1 ونعترف بأن $\mu < \mu_0$ كما في الشكل $(1-6)$ السابق.

وكذلك الأمر عند استخدامنا لمؤشر (ستودينت) t فإن القيمة الحرجة لمتحوله t هي القيمة المقابلة لكامل مستوى الدلالة $\propto$ ولدرجة الحرية $(n-1)$ ونرمز لها بـ $t_\propto$. ونتخذ القرار كما يلي:

- إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$ وكان $t \leq t_\propto$ فإننا نقبل فرضية العدم H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كان $t > t_\propto$ فإننا نرفض H_0 ونقبل H_1 ونعترف بأن $\mu > \mu_0$.
- أما إذا كان الاختبار أحادي يساري $(H_1: \mu < \mu_0)$ فإننا نقارن قيمة t المحسوبة مع $(-t_\propto)$ السالبة، فإذا كانت $(t \geq -t_\propto)$ فإننا نقبل الفرضية H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كانت $(t < -t_\propto)$ فإننا نرفض H_0 ونقبل H_1 التي نقول بأن: $\mu < \mu_0$.

مثال (1−1): لنفترض إننا نريد اختبار أن يكون توقع X في المجتمع $\mu_0 = 50$، فسحبنا عينة عشوائية منه بحجم $n = 25$ عنصراً . فكان متوسطها: $\bar{x} = 53$ وتباينها: $S^2 = 400$ ، ثم حددنا مستوى الدلالة بـ $\propto = 0.05$ ووضعنا الفرضيتين كما يلي:

$$H_0: \mu = 50 \qquad H_1: \mu \neq 50 \qquad \text{(الاختبار ثنائي الجانب)}$$

ولإجراء هذا الاختبار نلاحظ أن تباين المجتمع σ^2 مجهول. لذلك نستخدم العلاقة $(1-18)$ والتي تأخذ الشكل التالي :

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{53 - 50}{20/5} = \frac{3}{4} = 0.75$$

وبما أن t يخضع لتوزيع (ستودينت) بـ $(n-1)$ درجة حرية ، فإننا نقوم بحساب القيمة الحرجة له $t_{n-1}\left(\frac{\propto}{2}\right)$ ، فنجد من جداول (ستودينت) أن:

$$t_{n-1}\left(\frac{\propto}{2}\right) = t_{24}\left(\frac{0.05}{2}\right) = t_{24}(0.025) = 2.064$$

وبمقارنة القيمة المحسوبة $t=0.75$ مع القيمة الحرجة $t_{24}\left(\frac{\propto}{2}\right)=2.064$ نجد أن $|t|<t_{24}\left(\frac{\propto}{2}\right)$ لذلك نقبل فرضية العدم H_0 والتي تقول أن $\mu=50$ ، وذلك باحتمال ثقة $1-\propto=0.95$.

ملاحظة: إذا كان تباين المجتمع σ^2 معلوماً نطبق العلاقة (1−7) ونستخدم المتحول المعياري Z .

ب− **طريقة اتخاذ القرار حول H_0 باستخدام احتمال الدلالة P** (P- Value) أو (Sig)

لتوضيح كيفية تطبيق هذه الطريقة لابد من توضيح معنى احتمال الدلالة (Signification probability) والذي يرمز له في البرامج الحاسوبية بالرمز P أو (P-value) أو بالرمز (Sig).

إن احتمال الدلالة P حسب التعريف هو: الاحتمال الذي تتركه القيمة المحسوبة لمؤشر الاختبار Z أو t أو غيرهما على طرفي التوزيع الاحتمالي، أو على أحد طرفيه، وهو يتأثر بنوع الاختبار وبحالاته المختلفة التالية:

فإذا كان الاختبار ثنائي الجانب $(H_1\,,\mu\neq\mu_0)$، فإن قيمة الاحتمال P، يتم توزيعها بالتساوي على طرفي التوزيع، بحيث يكون لكل طرف $\frac{P}{2}$، والطرفان هنا يقابلان المجالين المفتوحين $]+Z\,,+\infty[$ و $]-\infty\,,-Z[$.

أما إذا كان الاختبار أحادي يميني $(H_1\,,\mu>\mu_0)$، فإن كامل قيمة P تكون متوضعة على اليمين وتقابل المجال المفتوح $]+Z\,,+\infty[$.

وإذا كان الاختبار أحادي يساري $(H_1\,,\mu<\mu_0)$، فإن كان كامل قيمة P تكون متوضعة على اليسار وتقابل المجال المفتوح $]-\infty\,,-Z[$.

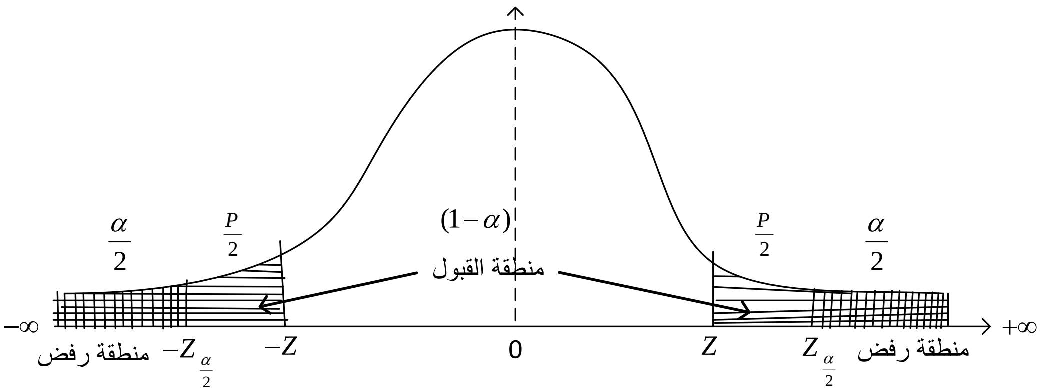
أي أن الاحتمال P يساوي المساحة التي تقع تحت منحني التوزيع وتقابل المجالات المذكورة (حسب كل حالة)، ويتم تحديد القيمة العددية لـ P بعد إعداد الحسابات اللازمة والقيام بإجراء الاختبار المفروض والحصول على القيمة المحسوبة Z أو t أو غيرهما . ثم حساب قيمة تكامل التوزيع الاحتمالي على المجالات المذكورة (حسب كل حالة)، كما سنرى لاحقاً .

واخيراً نشير إلى أن احتمال الدلالة P يختلف جذرياً عن مستوى الدلالة ∝ (Signification Level)، الذي يحدده الباحث أو المشرفون على البحث قبل إجراء البحث وقبل إجراء الاختبار نفسه . ويستفاد من P في اتخاذ القرارات حول الفرضية H_0 وذلك بمقارنتها مع ∝ .

ولتوضيح ذلك نأخذ حالة التوزيع الطبيعي المعياري، ونحسب منه قيمة احتمال الدلالة p، حسب حالات الاختبار التالية:

1) حالة الاختبار ثنائي الجانب: أي أن منطقة الرفض حسب القواعد السابقة تقع على الجانبين.

− فإذا كانت $|Z|\leq Z_{\frac{\propto}{2}}$ فإننا نقبل فرضية العدم H_0 ويكون لدينا الشكل التالي:



شكل (1−7) تحديد المساحة P

من هذا الشكل نلاحظ أن $\frac{\propto}{2}$ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال$\left]Z_{\frac{\propto}{2}}, +\infty\right[$.

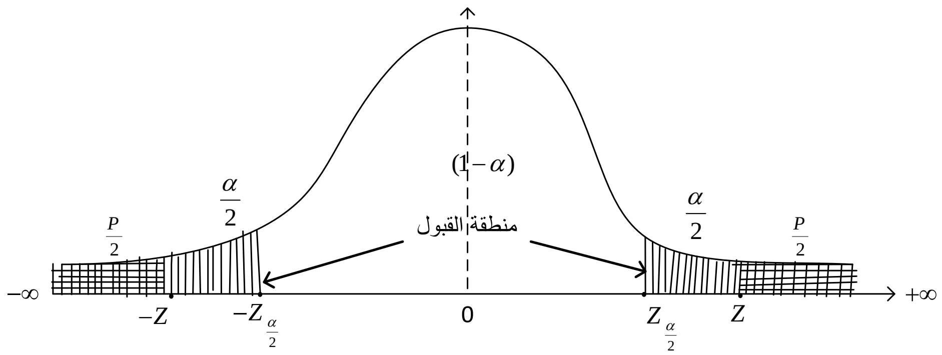
أما $\frac{P}{2}$ فهي المساحة المظللة بخطوط أفقية وهي التي تقابل المجال $]Z, +\infty[$، وعندما تكون $Z < Z_{\frac{\propto}{2}}$ كما في الشكل (1−7) يكون لدينا $\frac{P}{2} > \frac{\propto}{2}$ ، ويكون لدينا $P > \propto$، أي أنه علينا أن نقبل الفرضية H_0 إذا كانت $P > \propto$ (لأنه يكون لدينا $|Z| < Z_{\frac{\propto}{2}}$) .

ويتم حساب قيمة الاحتمال P في هذه الحالة من تكامل التوزيع الطبيعي المعياري على المجالين المتناظرين $]Z, +\infty[$ و$]-\infty, -Z[$ كما يلي:

$$P = \int_{-\infty}^{-Z} f(x)dx + \int_{Z}^{+\infty} f(x)dx = 2 * \int_{Z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = 2[1-\phi(Z)] \quad (19-1)$$

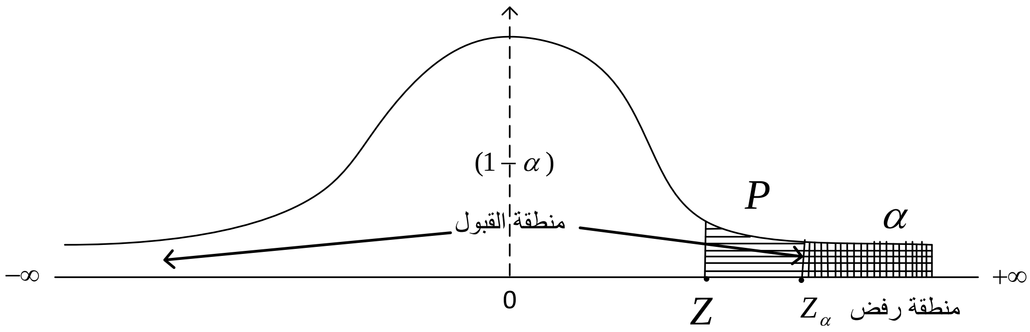
ويمكن الحصول على قيمة هذا التكامل من الجداول الإحصائية الجاهزة أو من الحواسيب المبرمجة .

- أما إذا كانت $|Z| > Z_{\frac{\propto}{2}}$ فإننا نرفض فرضية العدم H_0، وعندها يكون لدينا $\frac{P}{2} < \frac{\propto}{2}$ ، أي يكون لدينا $P < \propto$، أي أنه يجب علينا أن نرفض فرضية العدم H_0 إذا كان $P < \propto$ (لأن $|Z| > Z_{\frac{\propto}{2}}$) كما هو موضح في الشكل (1−8) التالي، حيث رمزنا للمساحة المظللة بخطوط أفقية من الطرفين للاحتمال P وللمساحة المظللة بخطوط عمودية لمستوى الدلالة $\propto$.



شكل (1−8) تحديد المساحة P

2) حالة الاختبار الأحادي اليميني: فإذا كانت $Z \leq Z_{\propto}$ فإننا نقبل فرضية العدم H_0، ويكون لدينا الشكل التالي:



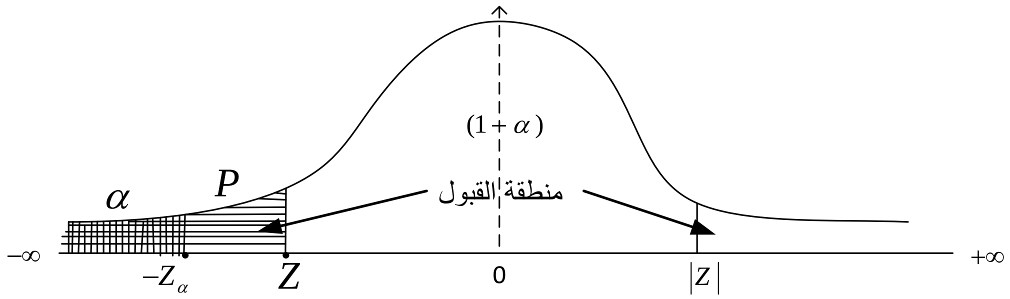
الشكل (1−9) **تحديد المساحة** P

ومن الشكل (الشكل (1−9) نلاحظ أن ∝ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال]$Z_\propto , +\infty$[، أما P فهي المساحة المظللة بخطوط أفقية وهي التي تقابل المجال]$Z , +\infty$[، وهنا يكون لدينا $P >\propto$، أي أنه علينا أن نقبل فرضية العدم H_0 إذا كان $P >\propto$ (لأن يكون $Z \leq Z_\propto$)

– أما إذا كانت $Z > Z_\propto$ (تقع على يمينها) فإننا نرفض الفرضية H_0، وعندها يكون لدينا $P <\propto$، أي أنه علينا أن نرفض الفرضية H_0 إذا كان $P <\propto$ (لأن $Z > Z_\propto$)، ويتم حساب قيمة P في هذه الحالة من تكامل التوزيع الطبيعي المعياري على المجال]$Z , +\infty$[كما يلي :

$$P = \int_{Z}^{+\infty} f(x)d_x = \int_{Z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = 1 - \phi(Z) \qquad (20-1)$$

3) حالة الاختبار الأحادي اليساري: إذا كانت $Z > -Z_\propto$ فإننا نقبل فرضية العدم H_0 ويكون لدينا الشكل التالي: (مع ملاحظة أن قيمة Z هي قيمة جبرية فقد تكون سالبة أو موجبة)



الشكل (1− 10) **تحديد المساحة** P

ومن الشكل (1− 10) نلاحظ أن ∝ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال]$-\infty, -Z_\propto$[، أما الاحتمال P فهو المساحة المظللة بخطوط أفقية وهي المساحة التي تقابل المجال]$-\infty , Z$[، وهذا يعني أن $P >\propto$، لذلك يجب علينا أن نقبل الفرضية H_0 إذا كان $P >\propto$ (لأن $Z > -Z_\propto$) . [انتبه إلى ذلك الاختلاف] .

– أما إذا كانت $Z < -Z_\propto$ (تقع على يسارها) فإننا نرفض الفرضية H_0، وعندها يكون لدينا$P <\propto$، أي أنه علينا أن نرفض الفرضية H_0 إذا كانت $P <\propto$ (لأن $Z < -Z_\propto$)، ويتم حساب P في هذه الحالة من العلاقة التالية:

$$P = \int_{-\infty}^{Z} f(x)d_x = \int_{|Z|}^{+\infty} f(x)d_x = \int_{|Z|}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = \phi(Z) \quad (21-1)$$

ملاحظة: إذا كان مؤشر الاختبار يخضع لتوزيع (ستودينت) t أو لأي توزيع آخر مثل X^2 أو F، فإن قيمة P تحسب حاسوبياً من تكاملات مشابهة للتكاملات السابقة على تلك التوزيعات وعلى المجالات المناسبة والمشابهة لتلك المجالات المذكورة. وهي أمور كثيرة وطويلة لا مجال للخوض فيها في هذا االفصل .

مثال (1−2) : لنفترض إننا نريد التأكد من نتيجة الاختبار في المثال (1−1)، لذلك قمنا بسحب عينة أخرى كبيرة بحجم $n = 100$ عنصراً ثم حسبنا متوسطها وتباينها فكانا كما يلي $\bar{x} = 54$ و $S^2 = 225$ ، ثم حددنا مستوى الدلالة بـ ($\propto = 0.05$) ووضعنا الفرضيتين كما يلي :

$H_0: \mu = 50 \qquad H_1: \mu \neq 50$ (الاختبار ثنائي الجانب)

وهنا نلاحظ أيضاً أن تباين المجتمع σ^2 مجهول . لذلك يجب علينا أن نطبق العلاقة (1−18)، ولكن بما أن حجم العينة n كبيراً ($n > 30$) فإن تلك العلاقة تقترب من العلاقة (1−17) ويصبح t متقارباً مع Z ونكتبها كما يلي:

$$t \approx Z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{54 - 50}{15/5} = \frac{4}{1.5} = 2.667$$

وبما أن Z يخضع للتوزيع الطبيعي المعياري، فإننا نقوم بإيجاد القيمة الحرجة $Z\left(\frac{\propto}{2}\right)$ من جداول التوزيع الطبيعي المعياري فنجد أن:

$$Z\left(\frac{\propto}{2}\right) = Z\left(\frac{0.05}{2}\right) = Z(0.025) = 1.96$$

وبمقارنة قيمة $Z = 2.667$ المحسوبة مع $Z\left(\frac{\propto}{2}\right) = 1.96$ الحرجة نجد أن:

$|Z| > Z\left(\frac{\propto}{2}\right)$. لذلك نرفض فرضية العدم H_0 التي تقول أن $\mu = 50$ ونقبل الفرضية H_1 التي تقول أن توقع X في ذلك المجتمع يختلف عن 50 . وذلك باحتمال ثقة 0.95 .

ملاحظة: يمكننا أن نستخدم طريقة احتمال الدلالة P لاتخاذ القرار حول H_0 ، لذلك نقوم بحساب P من العلاقة (1−19) فنجد من جداول التوزيع الطبيعي المعياري أن:

$$P = 2[1 - \phi(Z)] = 2[1 - \phi(2.667)] = 2[1 - 0.99615] = 0.0077$$

وبمقارنة P المحسوبة مع $\propto$ المفروضة نجد أن $P < \propto$. لذلك نرفض فرضية العدم H_0 بمستوى دلالة $\propto$.

وهنا نلاحظ أن نتيجة الاختبار في هذا المثال تختلف عن المثال (1−1) وذلك لأن العينة مختلفة عن العينة الأولى بالبيانات والحجم .

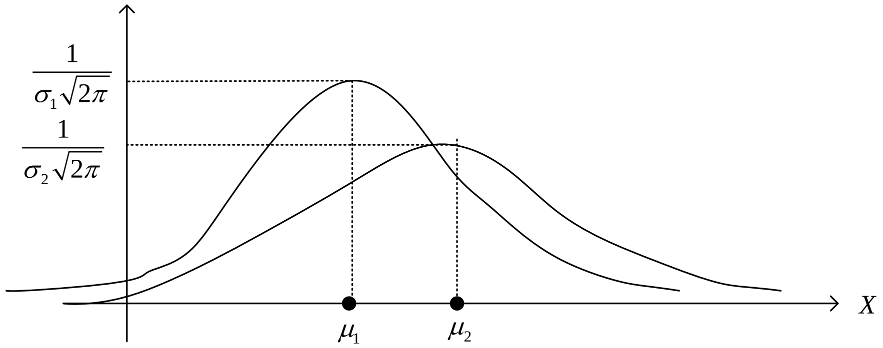
1-4 : اختبارات معالم مجتمعين طبيعيين (من عينتين مستقلتين):

1-4-1: اختبار الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد دراسة تغيرات متحول طبيعي X في مجتمعين منفصلين:

ولنفترض أن توقع X في المجتمع الأول هو μ_1 وتباينه فيه σ_1^2 وإن توزيعه الطبيعي هو $N(\mu_1 , \sigma_1^2)$ ، كما نفترض أن توقع X نفسه في المجتمع الثاني هو μ_2 وتباينه فيه σ_2^2، وإن توزيعه الطبيعي هو $N(\mu_2 , \sigma_2^2)$.

وبرسم هذين التوزيعين على شكل واحد نحصل على الشكل التالي :



الشكل (1-11) **شكلان طبيعيان فيهما** $\sigma_1^2 < \sigma_2^2$

ومن هذا الشكل نلاحظ أن عملية المقارنة بين μ_1 و μ_2 لا تتعلق بالفرق بينهما $(\mu_1 - \mu_2)$ فقط . بل تتعلق بشكل التوزيع الطبيعي وبتبايني X في هذين التوزيعين .

فإذا كان σ_1^2 و σ_2^2 مختلفان كثيراً فإن عملية المقارنة لا تكون متوازنة، لأنه إذا كان $\sigma_1^2 < \sigma_2^2$ فإن قيم X تكون متمركزة حول μ_1 أكثر من تمركز قيم X حول μ_2 .

وإن عملية المقارنة بين التوقعين μ_1 و μ_2 تكون أكثر فعالية عندما يكون $\sigma_1^2 = \sigma_2^2$ (أي عندما يكون الشكلان متشابهين) لذلك فإننا سنميز بين الحالتين التاليتين :

الحالة الأولى: الحالة التي يكون فيها $\sigma_1^2 \neq \sigma_2^2$

الحالة الثانية: الحالة التي يكون فيها $\sigma_1^2 = \sigma_2^2$

كما أننا سندرس الحالة التي يكون فيها σ_1^2 و σ_2^2 معلومين والحالة التي يكون فيها σ_1^2 و σ_2^2 مجهولين .

ولإجراء هذا الاختبار حول $(\mu_1 - \mu_2)$ علينا أن نسحب من هذين المجتمعين عينيتين عشوائيتين ومستقلتين بحجمين n_1 و n_2 ونحسب منهما مايلي:

أ- متوسط العينة الأولى $\bar{x}_1 = \frac{\sum x_i}{n_1}$ ويعتبر تقديراً غير متحيز لمتوسط المجتمع الأول μ_1 .

ب- متوسط العينة الثانية $\bar{x}_2 = \frac{\sum x_i}{n_2}$ ويعتبر تقديراً غير متحيز لمتوسط المجتمع الثاني μ_2 .

ج- نحسب التباين المصحح للعينة الأولى من العلاقة: $S_1^2 = \frac{1}{n_1-1}\sum(x_{i1} - \bar{x}_1)^2$ ،

ويعتبر هذا التباين تقديراً غير متحيز لتباين المجتمع الأول σ_1^2 .

د- نحسب التباين المصحح للعينة الثانية من العلاقة : $S_2^2 = \frac{1}{n_2-1}\sum(x_{i2}-\bar{x}_2)^2$، ويعتبر هذا التباين تقديراً غير متحيز لتباين المجتمع الثاني σ_2^2 .

هـ-نحسب الفرق بين متوسطين هاتين العينتين $(\bar{x}_1-\bar{x}_2)$، وهو يعتبر تقديراً غير متحيز للفرق بين متوسطي المجتمعين $(\mu_1-\mu_2)$ ، كما يعتبر الفرق $(\bar{x}_1-\bar{x}_2)$ متحولاً عشوائياً جديداً يخضع للتوزيع الطبيعي الذي توقعه $(\mu_1-\mu_2)$ وتباينه $Var(\bar{x}_1-\bar{x}_2)$.

ولاختبار الفرق بين متوسطي المجتمعين $(\mu_1-\mu_2)$ نضع الفرضيتين كما يلي :

فرضية العدم: $H_0:(\mu_1-\mu_2)=0$.

وتقابلها الفرضية البديلة والتي يمكن أن تكون على أحد الأشكال التالية:

على الشكل الثنائي الجانب : $H_1:(\mu_1-\mu_2)\neq 0$.

أو على الشكل الأحادي اليميني: $H_1:(\mu_1-\mu_2)>0$.

أو على الشكل الأحادي اليساري: $H_1:(\mu_1-\mu_2)<0$.

ثم نقوم بتشكيل مؤشر الاختبار المعياري للفرق $(\bar{x}_1-\bar{x}_2)$ من العلاقة التالية :

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{\sqrt{Var(\bar{x}_1-\bar{x}_2)}} \qquad (19-1)$$

ثم نقوم بمعالجته حسب الحالات السابقة لـ σ_1^2 و σ_2^2 التالية:

الحالة الأولى: إذا كان $\sigma_1^2 \neq \sigma_2^2$ وكان σ_1^2 و σ_2^2 معلومين عددياً فإننا نجد أن تباين الفرق $(\bar{x}_1-\bar{x}_2)$ لهاتين العينتين المستقلتين يساوي : [لعدم وجود ارتباط بين العينتين] .

$$Var(\bar{x}_1-\bar{x}_2)=Var(\bar{x}_1)+Var(\bar{x}_2)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\ \left(\text{معلوم}\right)$$

وبالتعويض في (1-19) نحصل على مؤشر الاختبار الطبيعي المعياري التالي :

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \quad : \left(\sigma_1^2\neq\sigma_2^2\ \text{ومعلومين}\right) \qquad (20-1)$$

وهو متحول يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$.

أما عندما يكون التباينان σ_1^2 و σ_2^2 مجهولين وغير متساويين، فإننا نستبدلهما بتقديريهما غير المتحيزين S_1^2 و S_2^2 والمحسوبين من العينتين فنحصل على مؤشر اختبار آخر t يخضع تقاربياً لتوزيع (ستودينت) وله درجة حرية معقدة (انظر المثال (1-3)) وهو يأخذ الشكل التالي :

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \quad : \left(\sigma_1^2\neq\sigma_2^2\ \text{ومجهولين}\right) \qquad (21-1)$$

ملاحظة: يمكن استخدام مؤشر الاختبار الأخير t في اختبارات الفرق بين المتوسطين $(\mu_1 - \mu_2)$ إذا كان حجما العينتين n_1 و n_2 كبيرين، وعندها نعتبر درجة الحرية مساوية لأصغر العددين $(n_1 - 1)$ أو $(n_2 - 1)$ لأن قيمتها الحقيقية تكون قريبة منها .

الحالة الثانية: وهي الحالة التي يكون فيها تباينا المجتمعين متساويين ومعلومين، أي عندما يكون لدينا $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ، حيث σ^2 هي القيمة المشتركة المعلومة لهما، وعندها تأخذ العلاقة (1− 20) الشكل التالي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sigma * \sqrt{\left(\frac{1}{n_1} + \frac{1}{\mathrm{n}_2}\right)}} \quad : \left(\sigma^2 \text{معلوم}\right) \qquad (\mathbf{22-1})$$

وهو متحول يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$.

أما عندما يكون التباين المشترك σ^2 مجهولاً ، فإننا نقدره من خلال المتوسط الحسابي للتباينين S_1^2 و S_2^2 المصححين والمحسوبين من العينتين والمثقلين بـ $(n_1 - 1)$ و $(n_2 - 1)$ على الترتيب، فنحصل من العلاقة المركبة لهما على ما يسمى بالتباين المدمج pooled ونرمز له بالرمز S_P^2 ونكتبه كما يلي :

$$S_P^2 = \widetilde{\sigma^2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad (23 - 1)$$

ويبرهن في الاحصاء الرياضي على أن التقدير S_P^2 هو تقدير غير متحيز للتباين المشترك σ^2، وبذلك تأخذ العلاقة (1−22) الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (24 - 1)$$

أو الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (25 - 1)$$

حيث استبدلنا في (1−22) التباين المشترك σ^2 بالتباين المدمج S_P^2 ، وهو مؤشر يخضع تقاربياً لتوزيع (ستودينت) بدرجة حرية $(n_1 + n_2 - 2)$، ويمكن استخدامه في اختبارات الفروق بشرط أن يكون تباينا المجتمعين متساويين، لذلك يجب أن نتحقق أولاً من أن: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ قبل تطبيق (1−25).

1−4−2: اختبار الفرق بين نسبتين في مجتمعين طبيعيين:

لاختبار الفرق بين نسبتي خاصتين R_1 و R_2 في مجتمعين طبيعيين، نفترض أولاً أن $\sigma_1^2 \neq \sigma_2^2$ ونضع فرضية العدم كما يلي: $H_0: R_1 - R_2 = 0$ ، والفرضية البديلة من الشكل : $H_1: R_1 - R_2 \neq 0$ أو غيره. وعندها يأخذ مؤشر الاختبار الأول (1−20) الشكل التالي:

$$Z = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{\sqrt{\frac{R_1(1-R_1)}{n_1} + \frac{R_2(1-R_2)}{n_2}}} \qquad (26-1)$$

حيث أن: R_1 و R_2 معلومتان وأن r_1 و r_2 هما النسبتان في العينتين المسحوبتين.

أما عندما يكون $\sigma_1^2 = \sigma_2^2 = \sigma^2$، فإننا نقدر σ^2 من النسبة المتوسطة $\bar{r}$ ونحسب تقديره من العلاقة التالية :

$$\widetilde{\sigma^2} = \bar{r}(1-\bar{r}) \qquad (27-1)$$

حيث أن النسبة المتوسطة $\bar{r}$ تحسب من المتوسط المثقل للنسبتين r_1 و r_2 كما يلي:

$$\bar{r} = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2} \qquad (28-1)$$

وعندها تأخذ العلاقة $(1- 26)$ شكلاً آخر هو التالي:

$$t = \frac{(r_1 - r_2) - (R_1 - \mathrm{R}_2)_0}{\sqrt{\frac{\bar{r}(1-\bar{r})}{n_1} + \frac{\bar{r}(1-\bar{r})}{n_2}}} = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{\sqrt{\bar{r}(1-\bar{r})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (29-1)$$

ومنها نحصل على المؤشر t الخاضع تقاربياً لتوزيع (ستودينت) بـ $(n_1 + n_2 - 2)$ درجة حرية، ولكنه عندما يكون $(n_1 + n_2 - 2) > 30$، فإنه يخضع تقاربياً للتوزيع الطبيعي المعياري. ولإجراء الاختبار نحسب قيمة المؤشر t من العلاقة $(1-29)$، ثم نقارنها مع قيمة $t_{\frac{\alpha}{2}}$ الحرجة والمقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ من الطرفين ولدرجة الحرية $(n_1 + n_2 - 2)$، وبناء على هذه المقارنة نقبل أو نرفض فرضية العدم H_0 وفق القواعد المذكورة سابقاً .

مثال (1-3): لدراسة حالة الفروقات بين كميتي البروتين ACTB-1 عند مرضى السرطان ومرضى الربو مقارنة مع الأشخاص الطبيعيين، أجريت التجارب اللازمة على ثلاث عينات واستخلصت منها النتائج والبيانات التالية [من نتائج تجارب رسلان في ألمانيا عام 2018]:

المؤشر / العينة	حجم العينة n_i	متوسط كمية البروتين في العينة $\bar{x}_i$	الانحراف المعياري SD_i
مرضى الربو	25	1063.126	669.1437
الأشخاص الطبيعيين	42	1535.488	479.3964
مرضى السرطان	14	2350.761	1116.602

والمطلوب: اختبار الفرق بين متوسط البروتين عند مرضى السرطان ومتوسطه عند الأشخاص الطبيعيين. ثم اختبار الفرق بين متوسطه عند مرضى الربو ومتوسطه عند الأشخاص الطبيعيين، وذلك بمستوى دلالة $\alpha = 0.05$.

الحل: لاختبار الفرق بين متوسطي البروتين ACTB-1 عند مرضى السرطان μ_1 وعند الأشخاص الطبيعيين μ_2، نضع الفرضيتين (العدم والبديلة) كما يلي:

فرضية العدم بين متوسطي المجتمعين وتشير إلى أنـه: $H_0: \mu_1 - \mu_2 = 0$

لا يوجد فرق بين المتوسطين μ_1 و μ_2

الفرضية البديلة: وتشير إلى أنه يوجد فروق بينـهما $H_1: \mu_1 \neq \mu_2$

الحالة الأولى: وهي الحالة التي يكون فيها $\sigma_1^2 \neq \sigma_2^2$، وهي الحالة التي تشير إليها بيانات الجدول .

وفي هذه الحالة نقوم بحساب قيمة مؤشر الاختبار t من العلاقة العامة:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(2350.761 - 1535.488) - 0}{\sqrt{\frac{(1116.602)^2}{14} + \frac{(479.3964)^2}{42}}}$$

$$t = \frac{815.273}{\sqrt{89057.1447 + 5471.926388}} = \frac{815.273}{307.4558} = 2.65167$$

ولاتخاذ القرار حول H_0 نستخدم كلا الطريقتين التاليتين:

- طريقة القيمة الحرجة $t_{\frac{\alpha}{2}}$

ولاتخاذ القرار المناسب بطريقة القيمة الحرجة $t_{\frac{\alpha}{2}}$ حول الفرضية H_0 نبحث في الجداول الإحصائية لتوزيع (ستودينت) عن القيمة الحرجة $t_{\frac{\alpha}{2}}$ المقابلة لدرجة حرية مساوية لأصغر العددين: $(n_1 - 1)$ أو $(n_2 - 1)$ فنجد أن $df = (n_1 - 1) = (14 - 1) = 13$، وبما أن الاختبار ثنائي الجانب، لأن $H_1: \mu_1 \neq \mu_2$ ، نجد أن القيمة الحرجة لـ t تساوي: $t_{\frac{\alpha}{2},13} = t_{0.25\,,13} = 2.1604$

وبمقارنة t المحسوبة مع $t_{\frac{\alpha}{2}}$ الحرجة نجد أن : $2.65167 > 2.1604$ أي أن $t > t_{\frac{\alpha}{2}}$.

لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تقول أن $\mu_1 \neq \mu_2$. أي أن متوسط البروتين عند مرضى السرطان لا يساوي متوسطه عند الأشخاص الطبيعيين باحتمال ثقة(0.95) على الأقل .

ملاحظة: كان يمكننا الاستفادة من بيانات العينة ووضع الفرضية البديلة H_1 على الشكل الأحادي اليميني $H_1: \mu_1 > \mu_2$ ، وفي هذه الحالة يجب علينا أن نقارن t المحسوبة مع القيمة الحرجة $t_{13}(\alpha)$ المقابلة لجانب واحد . لذلك نبحث في جداول (ستودينت) عن القيمة الحرجة $t_{13}(\alpha)$ فنجد أن :

$$t_{13}(\alpha) = t_{13}(0.05) = 1.7709$$

وبالمقارنة نجد أن: $t = 2.65167 > 1.7709$ لذلك نرفض فرضية العدم أيضاً . ونقبل بأن $\mu_1 > \mu_2$ ، أي نقبل بأن متوسط البروتين عند مرضى السرطان أكبر من متوسطه عند الأشخاص الطبيعيين باحتمال ثقة(0.95) على الأقل.

- <u>طريقة الاحتمال P:</u>

لاتخاذ القرار المناسب بطريقة احتمال الدلالة P، علينا أن نقوم بحساب قيمة P المقابلة للقيمة المحسوبة $t = 2.65167$ ، وبما أن الاختبار ثنائي فهي تساوي ضعف المساحة المحسوبة من تكامل توزيع (ستودينت) على المجال $]t, +\infty[$. وهذا يقتضي تحديد درجة الحرية الدقيقة المعرفة في توزيع (ستودينت) المستخدم في هذا الاختبار، وهناك عدة طرق لحساب الدرجة $d\!f$ وأهمها الطريقتين التاليتين:

<u>الطريقة الدقيقة</u> لحساب P: وهي طريقة معقدة وتطبق في البرامج الحاسوبية، ولحساب درجة الحرية اللازمة تستخدم العلاقة الآتية [Triola, P.390] :

$$d\!f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} = 14.62907983 \qquad (30-1)$$

وبما أن قيمة $d\!f$ التي حصلنا عليها كانت على شكل عدد كسري (غير صحيح)، لذلك نقوم بحساب قيمتي P المقابلتين لدرجتي الحرية الصحيحتين المجاورتين للقيمة $d\!f$ وهما 14 و15، فنجد من تكامل توزيع (ستودينت) على المجال $]2.65167, +\infty[$ أن :

$$P_{14} = 2 * (0.0094832) = 0.0189664 \qquad \text{(للجانبين)}$$

$$P_{15} = 2 * (0.0090654) = 0.0181308 \qquad \text{(للجانبين)}$$

ولحساب القيمة الحقيقية لـP المقابلة لدرجة الحرية الكسرية (14.629) نستخدم العلاقة التناسبية التالية:

$$P = P_{14} + (P_{15} - P_{14})(d\!f - 14)$$
$$P = 0.0189664 + (-0.0008356)(0.6290783)$$
$$P = 0.01844074$$

وهي قيمة قريبة جداً من قيمة P التي نحصل عليها من الحاسوب، وهذا يعني أن الحاسوب يتبع الطريقة الدقيقة والمعقدة في حساب P، وبما أن $P < \propto$ فإننا نرفض فرضية العدم H_0 التي تقول أن $\mu_1 = \mu_2$ ونقبل الفرضية البديلة التي تقوم أن $\mu_1 \neq \mu_2$، أي أن متوسط البروتين عند مرضى السرطان لا يساوي متوسطه الطبيعي باحتمال ثقة أكبر) بكثير من (0.95 ، وهو يساوي:

$$1 - P = 1 - 0.01844 = 0.98156$$

- <u>الطريقة التقريبية لحساب P:</u>

لحساب قيمة P التقريبية المقابلة لـ ($t = 2.65167$) في توزيع (ستودينت) نقوم بتحديد درجة الحرية $d\!f$ من أصغر العددين $(n_1 - 1)$ و$(n_2 - 1)$، فنجد أنها $(d\!f = 14 - 1 = 13)$، ;كما يمكن تقديرها من نتيجة الطريقة الدقيقة السابقة كما يلي: $d\!f = 14.62907983 - 2 = 13$.

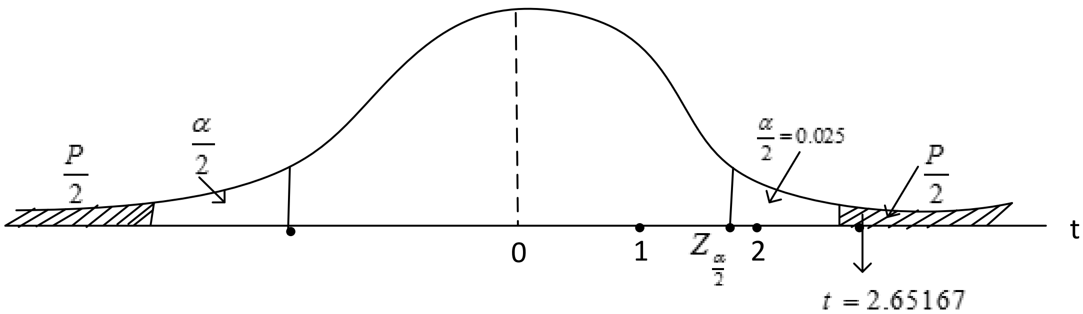
ومن الجداول الإحصائية لقيم متحول (ستودينت)، نجد أن قيمة ($t = 2.65167$) المحسوبة والموافقة لدرجة حرية $d\!f = 13$ تجعل الاحتمال P (من الطرفين) يساوي :

$$P = 0.0099741 * 2 = 0.0199482$$

وبما أن قيمة P أصغر من مستوى الدلالة $\propto = 0.05$ ، نرفض فرضية العدم H_0 التي تقول بعدم وجود فرق معنوي بين متوسطي المجتمعين، ونقبل الفرضية البديلة H_1 التي تقول أن متوسط المجتمع الأول (مرضى السرطان) لا يساوي المتوسط الطبيعي. وإن ذلك موثوق باحتمال ثقة أكبر بكثير من (0.95) ، وهو يقترب من الواحد لأنه يساوي :

$$1 - P = 1 - 0.0199482 = 0.9800518 \approx 98\%$$

والشكل التالي يوضح معنى P بالمنطقة المظللة تحت منحنى التوزيع الطبيعي المعياري (المقارب لتوزيع (ستودينت)) :



الشكل (1−12) **تحديد المنطقة** P

الحالة الثانية: وهي الحالة التي يكون فيها: $\sigma_1^2 = \sigma_2^2 = \sigma^2$، وعندها نقوم بحساب مؤشر الاختبار t من العلاقة (1− 24) التالية :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{(2350.761 - 1535.488) - 0}{\sqrt{\frac{13(1116.602)^2 + 41(479.3664)^2}{14 + 42 - 2}\left(\frac{1}{14} + \frac{1}{42}\right)}}$$

$$t = \frac{815.273}{212.60} = 3.8345$$

علماً بأن t يخضع تقاربياً لتوزيع (ستودينت) بدرجة حرية $n_1 + n_2 - 2$، ولهذا فإننا نقوم بحساب قيمة P المقابلة لـ $(t = 3.8345)$ ولدرجة حرية $(df = 14 + 42 - 2 = 54)$ ، فنجد أن الحاسوب يعطينا أن:

$$P = 2 * (0.00016548) = 0.00033096$$

وهي قيمة أصغر بكثير من مستوى الدلالة $\propto = 0.05$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية H_1 التي تقول بوجود فرق معنوي بين المتوسطين، ولكن هذه النتيجة محفوفة بالخطأ لأن الشرط

المستخدم فيها $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$ غير محقق في بيانات المثال المذكور، ولا يجوز الاعتماد عليها قبل إجراء اختبار لتساوي التباينين σ_2^2 و σ_1^2، ولقد قمنا بتطبيقها هنا للتدريب فقط .

ولاختبار الفرق بين متوسطي البروتين عند مرضى الربو والأشخاص الطبيعيين نتبع نفس الخطوات ونستخدم نفس العلاقات ونترك ذلك للقارئ على سبيل التدريب .

1-4-3: اختبار F لتساوي تبايني مجتمعين طبيعيين σ_2^2 و σ_1^2 :

لإجراء هذا الاختبار نضع فرضية العدم H_0 على الشكل التالي:

$$H_0 = \frac{\sigma_1^2}{\sigma^2} = 1 \qquad \Rightarrow (\sigma_1^2 = \sigma_2^2) \qquad (31-1)$$

ونضع الفرضية البديلة كما يلي:

$$H_1 = \frac{\sigma_1^2}{\sigma_2^2} > 1 \qquad : \ (\sigma_1^2 > \sigma_2^2) \qquad (32-1)$$

ثم نسحب من المجتمعين عينتين عشوائيتين بحجمين n_2 و n_1، ونحسب تباينيهما المصححين S_2^2 و S_1^2

ملاحظة: لتسهيل الحسابات تم تصميم جداول التوزيع F بحيث يكون رقم المجتمع الأول لصاحب التباين الأكبر (لذلك نرقم المجتمعين بحيث يكون $S_1^2 > S_2^2$ ، ونعدل الرموز في H_0 و H_1 حسب ذلك الترقيم) ثم نقوم بحساب مؤشر الاختبار F المعرف بالعلاقة التالية:

$$F = \frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2(n_1-1)}}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} = \frac{\dfrac{S_1^2}{\sigma_1}}{\dfrac{S_2^2}{\sigma_2^2}} = \frac{\sigma_2^2}{\sigma_1^2} * \frac{S_1^2}{S_2^2} \sim F_{v_1 v_2} \qquad (33-1)$$

ولكن بما أن فرضية العدم تنص على أن: $\sigma_1^2 = \sigma_2^2$ فإن المؤشر F يختصر ويأخذ الشكل التالي:

$$F = \frac{S_1^2}{S_2^2} \qquad : (S_1^2 > S_2^2) \qquad (34-1)$$

وهو متحول عشوائي يخضع لتوزيع فيشير $F_{(x)}$ بدرجتي حرية $v_1 = (n_1 - 1)$ للبسط و $v_2 = (n_2 - 1)$ للمقام، وبعد حساب قيمة F نقارنها مع القيمة الحرجة $F_{(\propto)}$ (لاتجاه واحد) والمقابلة لكامل $\propto$ ولدرجتي الحرجة $v_1 = (n_1 - 1)$ و $v_2 = (n_2 - 1)$ ونتخذ القرار كما يلي :

إذا كانت $F \leq F_{(\propto)}$ (أو كانت $P > \propto$) فإننا نقبل الفرضية H_0 التي تقول بتساوي التباينين $\sigma_1^2 = \sigma_2^2$.

أما إذا كانت $F > F_{(\propto)}$ نرفض H_0 ونقبل H_1 ونقول بعدم تساوي التباينين المذكورين .

1-5 : اختبارات معالم عدة مجتمعات طبيعية (من عدة عينات مستقلة):

1-5-1: اختبار تساوي متوسطات عدة مجتمعات طبيعية :

يطبق هذا الاختبار لمقارنة المتوسطات في أكثر من مجتمعين طبيعيين (3 فأكثر)، ولنفترض إننا سحبنا منهم عشوائياً عينات مستقلة بحجوم n_g ... n_3 و n_2 و n_1 (حيث أن g عدد المجتمعات و $g > 2$) .

وكانت متوسطاتها $\bar{x}_g$ ... $\bar{x}_3$ و $\bar{x}_2$ و $\bar{x}_1$ وكان متوسط المتوسطات هو $\bar{\bar{x}}$.

وكانت تبايناتها المصححة S_1^2 و S_2^2 و S_3^2 ... S_g^2

فإننا نضع فرضيتي العدم والبديلة حول متوسطات هذه المجتمعات كما يلي:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_g$$

$$H_1: \mu_k \neq \mu \quad \text{من أجل } k \text{ واحد على الأقل :} \qquad (35-1)$$

كما نفترض أن تباينات هذه المجتمعات متساوية وتساوي σ^2 .

ثم نحسب مجاميع مربعات الانحرافات المختلفة وهي :

مجموع مربعات الانحرافات داخل العينات، أي مربعات (الخطأ) :

$$SSE = \sum_{k=1}^{g} (n_k - 1) S_k^2 \qquad (36-1)$$

حيث g: عدد المجتمعات .

مجموع مربعات الانحرافات بين العينات :

$$SSB = \sum_{k=1}^{g} n_i (\bar{x}_k - \bar{\bar{x}})^2 \qquad (37-1)$$

مجموع مربعات الانحرافات الكلية لجميع عناصر العينات :

$$SST = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{\bar{x}})^2 = \sum_{k=1}^{g} \sum_{i=1}^{n_k} x_{ki}^2 - \frac{T^2}{n} \qquad (38-1)$$

حيث أن: $n = \sum n_k$ وأن: $T = \sum_{k=1}^{g} \sum_{i=1} x_{ki}$

ويكون لدينا:

$$SST = SSB + SSE \qquad (39-1)$$

علماً بأن درجات الحرية لكل منهم، هي و$(n-g)$ و$(g-1)$ و$(n-1)$ على الترتيب، ثم نضع النتائج في جدول كالتالي :

جدول(3-1) : نتائج تحليل التباين الأحادي ANOVA

مصدر التباين	مجموع المربعات ورمزه	درجة الحرية	متوسط المربعات	قيمة F المحسوبة	قيمة F الحرجة	قيمة P
التباين بين العينات	SSB	$g-1$	$MSSB = \frac{SSB}{g-1}$	$F = \frac{MSSB}{MSSE}$	$F_\propto$	P
التباين داخل العينات	SSE	$n-g$	$MSSE = \frac{SSE}{n-g}$	ـــــــ	ـــــــ	ـــــــ
التباين الكلي	SST	$n-1$	ـــــــ	ـــــــ	ـــــــ	ـــــــ

ثم نحسب قيمة مؤشر الاختبار F المعرف بالعلاقة التالية :

$$F = \frac{MSSB}{MSSE} \qquad (40-1)$$

وهو يخضع للتوزيع F بدرجتي حرية $(g-1\ ,n-g)$ ونتعامل معه كما تعاملنا مع F السابقة عند اتخاذ القرار حول H_0 في اختبار تساوي التباينين . فإذا كان $F \leq F(\propto)$ نقبل H_0 والعكس بالعكس .

ملاحظة: يسمى هذا الاختبار تحليل التباين ANOVA باتجاه واحد (Analysis Variance- one way) وسنقوم بدراسته بالتفصيل في الفصل السادس .

مثال (1-4): [مأخوذ من Copal P56 **بتصرف]**

لنفترض أنه لدينا (3) مجتمعات طبيعية، ونريد اختبار تساوي متوسطات متحول واحدX فيها، وبمستوى دلالة 0.05 =∝، لذلك سحبنا (3) عينات عشوائية منها بحجوم: $n_1 = 3 \quad n_2 = 5 \quad n_3 = 4$,

ولنفترض أن البيانات الأصلية (غير الموجودة) أعطتنا أن مجاميع قياسات X فيها كانت تساوي ما يلي:

$$\sum_{i=1}^{3} x_{1i} = 53.5 \quad , \quad \sum_{i=1}^{5} x_{2i} = 102.5 \quad , \quad \sum_{i=1}^{4} x_{3i} = 64.4$$

وإن مجموعها الكلي يساوي:

$$T = \sum_{k=1}^{3}\sum_{i=1}^{n_k} x_{ki} = 53.5 + 102.5 + 64.4 = 220.4$$

وإن متوسطات X في العينات المسحوبة تساوي:

$$\bar{x}_1 = \frac{53.5}{3} = 17.83 \quad , \quad \bar{x}_2 = \frac{102.5}{5} = 20.50 \quad , \quad \bar{x}_3 = \frac{64.4}{4} = 16.10$$

وإن المتوسط العام لـX فيها (أو المتوسط المثقل للمتوسطات) يساوي:

$$\bar{x} = \frac{T}{\sum n_i} = \frac{220.4}{12} = 18.37$$

ثم نقوم بحساب الكسر $\frac{T^2}{n}$ فنجد أن:

$$\frac{T^2}{n} = \frac{(220.4)^2}{12} = 4048.01$$

ثم نقوم بحساب SST من العلاقة (1-38) فنجد من البيانات الأصلية أن:

$$SST = \sum_{k=1}^{3}\sum_{i=1}^{n_k} x_{ki}^2 - \frac{T^2}{n} = [(954.43 + 2105.13 + 1037.98) - 4048.01]$$

$$SST = 4097.54 - 4048.1 = 49.53$$

ثم نقوم بحساب SSB من العلاقة (1-37) فنجد أن:

$$SSB = \sum_{k=1}^{3} x_k(\bar{x}_k - \bar{x})^2 = 3(17.83 - 18.37)^2 + 5(20.50 - 18.37)^2 + 416.10 - 18.37^2$$

$$SSB = 44.17$$

ثم نقوم بحساب SSE من العلاقة:

$$SSE = SST - SSB = 49.53 - 44.17 = 5.36$$

ثم نقوم بوضع نتائج هذه الحسابات في جدول كالتالي :

جدول (4-1): **جدول** ANOVA

مصدر التباين	مجموع المربعات	درجة الحرية	متوسطات المربعات	
بين العينات SSB	$SSB = 44.17$	$g-1=2$	$MSSB = 22.09$	$F = \frac{22.09}{0.556} = 37$
داخل العينات SSE	$SSE = 5.36$	$n-g=9$	$MSSE = 0.556$	———
التباين الاجمالي SST	$SST = 49.53$	$n-1=11$	———	———

ثم نقوم بحساب قيمة المؤشر F من العلاقة:

$$F = \frac{\frac{SSB}{g-1}}{\frac{SSE}{n-g}} = \frac{\frac{44.17}{2}}{\frac{5.36}{9}} = \frac{22.09}{0.556} = 37$$

ولمقارنة قيمة F المحسوبة مع قيمتها الحرجة $F_{v_1,v_2}(\propto)$ المقابلة لمستوى الدلالة ($\propto = 0.05$) ولدرجتي الحرية $v_1 = g-1 = 2$، $v_2 = n-g = 9$، علينا أن نبحث عن قيمة $F_{v_1,v_2}(\propto)$ في جداول F فنجد أنها تساوي :

$$F_{v_1,v_2}(\propto) = F_{2,9}(0.05) = 4.24$$

ثم نقوم بمقارنة F المحسوبة مع $F_{2,9}(0.05)$ الحرجة، فنجد أن $F > F_{2,9}(0.05)$، لذلك نرفض فرضية العدم H_0، التي تقول أن $H_0: \mu_1 = \mu_2 = \mu_s$، ونقبل الفرضية H_1 التي تقول أن أحد متوسطات هذه المجتمعات (على الأقل) يختلف عن الأخرى. (ولعله المجتمع الثاني لأن $\bar{x}_2 = 20.50$).

1-5-2: اختبار تساوي تباينات عدة مجتمعات طبيعية (من عدة عينات مستقلة):

1-2-5-1: اختبار (بارتليت Bartlitt) لتساوي تباينات عدة مجتمعات:

لنفترض إننا نريد دراسة تباينات متحول X في g مجتمعاً طبيعياً أو شبه طبيعي ($g > 2$). لذلك سحبنا منها g عينة عشوائية بحجوم مختلفة أو متساوية نرمز لها بـ $n_1, n_2, \dots n_k \dots n_g$، ولمجموع حجومها بـ $n = \sum_{k=1}^{n} n_k$. ثم نقوم بحساب متوسطاتها ورمزنا لها بـ $\bar{x}_1, \bar{x}_2, \dots \bar{x}_k \dots \bar{x}_g$، وحساب تبايناتها ورمزنا لها بـ $S_1^2, S_2^2, \dots S_k^2, \dots S_g^2$، والآن لنفترض أن تبايناتها X في هذه المجتمعات هي :

$$\sigma_1^2, \sigma_2^2, \dots \sigma_k^2, \dots \sigma_g^2$$

ثم نضع الفرضيتين حول تساوي هذه التباينات كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \dots = \sigma_g^2 \qquad (41-1)$$

$$H_1: \sigma_k^2 \neq \sigma_\ell^2$$: من أجل زوج واحد (k, ℓ) على الأقل

أي أننا نفترض في H_0 أن تباينات X في هذه المجتمعات متساوية، مقابل الفرضية البديلة H_1 التي تعني أنها غير متساوية من أجل مجتمعين على الأقل .

ولاختبار هذه الفرضية قام Bartlitt باستخراج مؤشر خاص وعرفه بالعلاقة التالية:

$$BT = \frac{(n-g)\ln S_p^2 - \sum_{k=1}^{g}(n_k - 1)\ln S_k^2}{1 + \left[\frac{1}{3(g-1)}\right]\left[\sum_{k=1}^{g}\frac{1}{n_k - 1} - \frac{1}{n-g}\right]} \sim \chi_{g-1}^2 \qquad (42-1)$$

حيث أن: n هو حجم العينة الكلية $n = \sum_{k=1}^{n} n_k$، و g عدد المجتمعات .

حيث أن: S_k^2 هو تباين X في العينة k .

وأن S_p^2 هو التباين المدمج المحسوب من العلاقة :

$$S_p^2 = \frac{\sum(n_k - 1)S_k^2}{n-g} \qquad (43-1)$$

وبرهن على هذا المؤشر يخضع تقاربياً لتوزيع χ_{g-1}^2 بدرجة حرية $(g-1)$.

لذلك فإننا عند اتخاذ القرار حول الفرضية H_0 نقارن القيمة المحسوبة BT مع القيمة الحرجة $\chi_{g-1}^2(\propto)$

ونتخذ القرار عند مستوى دلالة $\propto$ كما يلي:

(44−1) إذا كانت $BT \le \chi_{g-1}^2(\propto)$ نقبل الفرضية H_0 والتي تنص على أن التباينات متساوية،

أما إذا كان $BT > \chi_{g-1}^2(\propto)$ فإننا نرفض H_0 ونقبل الفرضية البديلة H_1، التي تنص على أن التباينات غير متساوية في مجتمعين على الأقل .

مثال (1−5): لنفترض أنه لدينا (5) خطوط لعصر الزيتون (معاصر) ونريد دراسة فيما إذا كانت تباينات الانتاج اليومي فيها متساوية أم مختلفة . لذلك وضعنا الفرضيتين كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$ من أجل خطين على الأقل :

ثم قمنا بسحب (5) عينات طبقية من إنتاج هذه الخطوط خلال أربعة أيام $(n_k = 4)$ فحصلنا منها على البيانات التالية:

جدول (1−5): كميات الإنتاج اليومية (بالكغ) حسب الخطوط والأيام

المجموع	الخط E	الخط D	الخط C	الخط B	الخط A	الخطوط / رقم اليوم	
	250	340	250	310	250	1	
	240	270	230	330	260	2	
	270	300	220	280	230	3	
	290	320	260	360	270	4	
$n = 20$	4	4	4	4	4	الأعداد	n_k
1382.5	262.5	307.5	240.0	320.0	252.5	المتوسطات	$\bar{x}_k$
3141.662	491.667	891.667	333.33	1133.33	291.667	التباينات	S_k^2
31.5094	6.1978	6.7931	5.8091	7.0329	5.6756	لوغاريتمات التباينات	$\ln S_k^2$

ولمتابعة الحل قمنا أولاً بحساب بعض الخصائص الاحصائية لتلك البيانات ووضعنا في أسفل الجدول السابق .

والآن نقوم بحساب الكميات التي تدخل في تعريف الاختبار BT فنجد أن التباين المدمج S_p^2 يساوي (انظر الجدول السابق) :

$$S_p^2 = \frac{\sum^g (n_k - 1)S_k^2}{n - g} = \frac{3(\sum S_k^2)}{20 - 5} = \frac{3 * (3141.667)}{15}$$

$$S_p^2 = \frac{9425}{15} = 628.333$$

كما نجد أن الحد الثاني في البسط يساوي :

$$\sum_{k=1}^{g} (n_k - 1)\, ln\, S_k^2 = 3\left(\sum_{k=1}^{g} ln\, S_k^2\right) = 3(31.5094) = 94.5292$$

ثم نقوم بحساب المقام فنجد أنه يساوي:

$$C = 1 + \left(\frac{1}{3(5-1)}\right)\left[\left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}\right) - \left(\frac{1}{20-5}\right)\right]$$

$$C = 1 + \frac{1}{12}\left[\frac{5}{3} - \frac{1}{15}\right] = 1.1333$$

نعوض نتائج هذه الحسابات في معادلة المؤشر BT فنجد أن:

$$BT = \frac{(20-5)\, ln(628.333) - 945292}{1.1333} = \frac{2.1167}{1.1333} = 1.8678$$

ثم نقوم بإيجاد القيمة الحرجة لـ $\chi_{g-1}^2(\propto)$ عند مستوى الدلالة $\propto = 0.05$ ودرجة الحرية $g - 1 = 5 - 1 = 4$ فنجد أن:

$$\chi_{g-1}^2(\propto) = \chi_4^2(0.05) = 9,488$$

وبمقارنة القيمة المحسوبة للمؤشر BT مع القيمة الحرجة $\chi_4^2(\propto)$ نجد أن $1.8678 < 9.488$ ، لذلك نقبل فرضية العدم H_0 التي نقول أن تباينات الانتاج على تلك الخطوط متساوية وباحتمال ثقة 0.95 .

ملاحظة: يمكننا أيضاً دراسة تساوي متوسطات الانتاج على هذه الخطوط وعندها يجب أن نستخدم الاختبار F وإجراء تحليل التباين ANOVA كما فعلنا في المثال (1- 4) السابق، ونترك ذلك للقارئ على سبيل التمرين .

2-2-5-1: اختبار (ليفيني Levene) لتساوي التباينات في عدة مجتمعات (من عدة عينات مستقلة) .

يستخدم هذا الاختبار لدراسة تساوي أو تجانس تباينات متحول X في عدة مجتمعات طبيعية، وهو يقدم لنا خدمة جليلة عند تطبيق الكثير من الاختبارات الإحصائية، التي تفترض أن تباينات X في المجتمعات المدروسة متساوية، لأنه يساعدنا على التحقق من صحة تلك الافتراضات، ويعتبر هذا الاختبار بديلاً لاختبار Bartlitt، ولكنه أقل حساسية منه في الاعتماد على التوزيع الطبيعي .

فإذا كان لدينا شك قوي بأن البيانات المستخدمة ليست مسحوبة من مجتمع طبيعي (أو شبه طبيعي) فإنه يفضل استخدام اختبار Bartlitt ، لأنه يعطينا نتائج أفضل منه .

ولإجراء هذا الاختبار نفترض أننا نريد اختبار تساوي تباينات متحول طبيعي X في عدة مجتمعات (أو مجموعات)، ولنفترض أن عدد تلك المجتمعات $g(g > 2)$ وسحبنا منها g عينة عشوائية بحجوم مختلفة أو متساوية: $n_1 , n_2 , \dots n_k \dots n_g$ حيث $(n = \sum n_k)$، وحصلنا منها على متوسطاتها: $\bar{x}_1 , \bar{x}_2 , \dots \bar{x}_k \dots \bar{x}_g$ ، وعلى تبايناتها التالية: $S_1^2 , S_2^2 , \dots S_k^2 , \dots S_g^2$. والآن لنفترض أن تباينات X في تلك المجتمعات هي: $\sigma_1^2 , \sigma_2^2, \dots , \sigma_k^2, \dots \sigma_g^2$.

وبناء على ذلك نصيغ الفرضيتين الاحصائيتين كما يلي:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \cdots = \sigma_g^2 \qquad (45-1)$$

$H_1 : \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد (k , ℓ) على الأقل

أما مؤشر الاختبار فيعرف حسب Levene بالعلاقة التالية:

$$W = \frac{(n-g)}{(g-1)} \frac{\sum_{k=1}^{g} (\bar{Z}_k - \bar{Z})^2}{\sum \sum (Z_{ki} - \bar{Z}_k)^2} \quad : \left(n = \sum n_k\right) \qquad (46-1)$$

حيث أن المتحول Z هو تحويل من المتحول X وفق إحدى العلاقات الثلاثة التالية:

1 − $Z_{ki} = |x_{ki} - \bar{X}_k|$: حيث أن : $\bar{X}_k$ متوسط X في العينة K (47 − 1)

2 − $Z_{ki} = |x_{ki} - X'_k|$: حيث أن : X'_k وسيط X في العينة K (48 − 1)

3 −: $Z_{ki} = |x_{ki} - X''_k|$: حيث أن : X''_k هو المتوسط المرتب لـ 10% الأولى من قيم X (49 − 1)

أما متوسطات Z فتحسب كما يلي:

$$\bar{Z}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} Z_{ki} \qquad (50-1)$$

: متوسط القيم Z_{ki} في العينة K

$$\bar{Z} = \frac{1}{g} \sum_{i=1}^{n_k} \bar{Z}_k = \frac{1}{n} \sum_{k=1}^{g} \sum_{i=1}^{n_k} Z_{ki} \qquad (47-1)$$

: المتوسط العام لـ Z_{ki}

وهنا نشير إلى أن التعاريف الثلاثة لـ Z_{ki} تساعدنا في تحديد حصانة وقوة اختبار (ليفيني)، ويقصد بمصطلح الحصانة قدرة الاختبار على عدم إعطاء إشارة مزيفة عن عدم تساوي التباينات عندما تكون البيانات غير خاضعة للتوزيع الطبيعي وتكون التباينات فعلياً متساوية، ويقصد بالقوة قدره الاختبار على اكتشاف التباينات غير المتساوية عندما تكون التباينات فعلياً غير متساوية .

وأخيراً نشير إلى أن مؤشر الاختبار W يخضع لتوزيع F بدرجتي حرية $v_1 = g - 1$ و $v_2 = \text{n} - g$ (حيث أن: $n = \sum_{k=1}^{g} n_k$) .

ولاتخاذ القرار حول نتيجة الاختبار نقارن قيمة W المحسوبة من العلاقة (1−46) بالقيمة الحرجة $F_{v_1,v_2}(\propto)$ ونتخذ القرار كما يلي:

إذا كانت $W \le F_{v_1,v_2}(\propto)$ نقبل فرضية العدم H_0

أما إذا كانت $W > F_{v_1,v_2}(\propto)$ نرفض H_0 ونقبل الفرضية H_1 $\quad (51-1)$

مثال (1−6): لنفترض أنه لدينا (10) مجموعات من الطلاب، ونريد اختبار تساوي تباينات أعمارهم في تلك المجموعات، فسحبنا من كل مجموعة عينة عشوائية بحجوم متساوية: $(n_k = 5)$ طلاب، فكان حجم العينة الكلية $n = 50$ طالباً . وبعد أخذ بيانات الأعمار في كل مجموعة وحساب متوسطاتها $\bar{x}_k$ وضعنا فرضيتي الاختبار كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \dots = \sigma_{10}^2$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد على الأقل

ثم نقوم بإجراء التحويلات من X إلى Z حسب إحدى العلاقات السابقة، ولتكن العلاقة (1−47) المستندة إلى المتوسطات $\bar{x}_k$، فنحصل على القيم Z_{ki}، ثم نقوم بحساب المتوسطات $\bar{Z}_k$ وأخيراً نقوم بحساب قيمة مؤشر الاختبار W ولنفترض أنها كانت تساوي:

$$W = \frac{(50-10)}{(10-1)} \frac{\sum 5(\bar{Z}_k - \bar{Z})^2}{\sum\sum(Z_{ki} - \bar{Z}_k)^2} = 1.75$$

وبما أن درجتي الحرية تساويان $v_1 = 10 - 1 = 9$ و $v_2 = 50 - 10 = 40$ ، فإننا نقارن القيمة المحسوبة W مع القيمة الحرجة $F_{v_1,v_2}(\propto)$. وباعتبار أن $\propto = 0.05$ نجد من جداول F أن:

$$F_{v_1,v_2}(\propto) = F_{9,40}(0.05) = 2.124$$

وبمقارنة القيمة المحسوبةW مع $F_{9,40}(0.05)$ نجد أن: $1.75 \le 2.124$ لذلك نقبل فرضية العدم H_0 التي تقول أن تباينات العمر X في هذه المجموعات متساوية وذلك باحتمال ثقة 0.95 على الأقل .

وبما أن التباينات متساوية فإننا نقوم بمقارنة الفروقات بين المتوسطات باستخدام العلاقة (1−40) أو (1−25) .

1−6 : اختبار الأزواج المتقابلة (من عينتين مرتبطتين):

يطبق هذا الاختبار لمقارنة نتائج إجابات أو علامات عينية من نفس الأشخاص قبل التجربة وبعدها، لذلك تسمى الدرجات الأولى بالدرجات القبلية وتسمى الدرجات الثانية بالدرجات البعدية، ويمكن وضع النتائج في جدول خاص على شكل أزواج متقابلة (كل زوج لشخص واحد) فنحصل على عينتين مرتبطتين من الدرجات كما يلي:

جدول (1−6): البيانات المتقابلة

رقم الشخص	1	2	3	4	.....	i	.....	n	المتوسط	
الدرجات القبلية	x_1	x_2	x_3	x_4	.....	x_i	.....	x_n	$\bar{x}$	
الدرجات البعدية	y_1	y_2	y_3	y_4	.....	y_i	.....	y_n	$\bar{y}$	
الفروقات d_i	d_1	d_2	d_3	d_4	.....	d_i	.....	d_n	$\bar{d}$	S_d

ثم نقوم بحساب الفروقات بين قيمتي كل زوج من العلاقة : $d_i = x_i - y_\mathrm{i}$

ثم نقوم بحساب متوسط وتباين هذه الفروقات من العلاقتين :

$$\bar{d} = \frac{1}{n}\sum d_i \qquad (52-1)$$

$$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right] \qquad (53-1)$$

ولإجراء هذا الاختبار حول الفرق بين العينتين نضع الفرضيتين كما يلي: $H_1: \bar{d} > 0 \quad H_0: \bar{d} = 0$

حيث $\bar{D}$: هو متوسط الفروقات في المجتمع .

ثم نحسب الانحراف المعياري لـ d_i ونرمز له بـ S_d ، ثم نحسب مؤشر الاختبار المعرف بالعلاقة:

$$t = \frac{\bar{d} - \bar{d}_0}{S_d/\sqrt{n}} \qquad (54-1)$$

حيث$\bar{D}_0$ هي قيمة متوسط الفروقات المفترضة في المجتمع، وتؤخذ قيمتها الصفرية من فرضية العدمH_0،

ثم نقارن قيمة t المحسوبة مع قيمة t الحرجة $t_{\propto,n-1}$ فإذا كانت $t < t_{\propto,n-1}$ تقبل فرضية العدم H_0،

ونقول بأنه لا يوجد فرق بين الدرجات القبلية والبعدية، والعكس بالعكس .

مثال (1−7): لدراسة تأثير أحد الأدوية على مستوى ضغط الدم عند المرضى المصابين به. قرر أحد الباحثين إجراء تجربة هذا الدواء على (8) مرضى. ولذلك قام أولاً بقياس مستويات الضغط عند هؤلاء المرضى قبل إعطائهم الدواء . ثم قام بإعطائهم الدواء وبعد مرور ساعة على ذلك أخذ قياسات مستويات الضغط لهم، فحصل على البيانات القبلية والبعدية ثم قام بحساب الفروقات الزوجية D_i وقام بتربيعها فحصل على الجدول التالي :

جدول (1−8): بيانات المثال:

رقم المريض	1	2	3	4	5	6	7	8	المجموع
مستوى الضغط قبل التجربة X	170	175	180	175	160	170	175	180	——
مستوى الضغط بعد التجربة Y	150	160	170	160	170	160	170	185	——
الفروقات الزوجية d_i	20	15	10	15	−10	10	5	−5	+60
d_i^2	400	225	100	225	100	100	25	25	1200

ثم قام بحساب متوسط تلك الفروقات فوجد أن:

$$\bar{d} = \frac{\sum_{i=1}^{8} d_i}{n} = \frac{+60}{8} = 7.5$$

ثم قام بحساب تباين تلك الفروقات فحصل على أن :

$$S_d^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right] = \frac{1}{7}[1200 - 8(7.5)^2] = 107.14$$

$$S_d = \sqrt{S_d^2} = \sqrt{107.14} = 10.35$$

ثم قام بوضع فرضيتي الاختبار كما يلي:

$$H_0: \bar{d} = 0 \qquad H_1: \bar{d} > 0$$

(الاختبار أحادي يميني)

ثم قام بحساب قيمة مؤشر الاختبار من العلاقة :

$$t = \frac{\bar{d} - \bar{d}_0}{S/\sqrt{n}} = \frac{7.5 - 0}{10.35/\sqrt{8}} = 2.05$$

وباعتماد $\propto = 0.05$ قام بمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $t_{n-1}(\propto)$ والتي تساوي: $t_{n-1}(\propto) = t_7(0.05) = 1.895$، فوجد أن: $t > t_7(0.05)$، لذلك رفض فرضية العدم H_0 التي تقول أن $(\bar{D} = 0)$ وتم قبول الفرضية البديلة H_1 التي تقول أن $(\bar{D} > 0)$. وهذا يعني أن متوسط الفروقات كان موجباً، وهو ما يؤكد أن الدواء المستخدم له تأثير إيجابي على مستويات الضغط عند المرضى .

الفصل الثاني
الاستدلال بواسطة الاستبيان

2-1 : تمهيد :

إن الاستبيان هو أداة لجمع البيانات عن أحوال عناصر موضوع الدراسة من خلال عينة مسحوبة عشوائياً من مجتمع البحث، ويكون الاستبيان مؤلفاً من عدة محاور، كل منها يتضمن أسئلة أو عبارات محددة عن جوانب الموضوع . ويقوم أفراد العينة بالإجابة عليها، بالطريقة المباشرة (المقابلة) أو عبر البريد العادي أو الالكتروني. وهناك شروط معينة لتصميم الاستبيان ولعدد الخيارات الممكنة للإجابة على الأسئلة الواردة فيه. وعادة ما يتم تقديم الاستبيان بتعليمات حول كيفية التعامل مع الأسئلة والإجابة عليها.
ويبدأ بطرح بعض الأسئلة عن أحوال الشخص المبحوث (جنسه + عمره+ تعليمه+ عمله ...الخ).
أما جسم الاستبيان فيتألف من عدة محاور تعبر عن المتحولات المعتمدة في البحث (لكل متحول محور)، ويتضمن كل محور عدداً محدداً من الأسئلة المباشرة أو العبارات الواضحة، التي تعبر عن ذلك المحور أو تشكل جزءاً منه، ويجب أن تكون الأسئلة أو العبارات ضمن المحاور قصيرة وواضحة وذات اتجاه واحد، وتتناسب مع مستوى المبحوثين، ولا تتضمن عبارات محرجة أو جارحة أو مسيئة أو سخيفة، ولا توحي للمبحوث باختيار إجابة معينة .
أما خيارات الأجوبة فيمكن أن تكون مغلقة أو مفتوحة .
وأهم الخيارات المغلقة هي خيارات (ليكرت) التالية:

1- الخيار الثنائي : ويكون أمام المبحوث خياران فقط للجواب على السؤال، مثل:

نعم	لا
1	0

2- الخيار الثلاثي: ويكون للسؤال ثلاثة خيارات للجواب، مثل:

موافق	محايد	معارض
3	2	1

3- الخيار الخماسي: ويكون للسؤال خمسة خيارات للجواب، مثل:

5	4	3	2	1

4- الخيار السباعي: ويكون للسؤال سبعة خيارات للجواب، مثل:

7	6	5	4	3	2	1

ويمكن أن تكون هذه الخيارات لغوية على الشكل التالي:

موافق جداً	**موافق**	**محايد**	**معارض**	**معارض جداً**
5	**4**	**3**	**2**	**1**

ثم يتم استبدالها بالأرقام المقابلة لها لتحليلها واستخلاص النتائج الممكنة .
وهنا لابد أن نشير إلى أن شكل السؤال أو العبارات يجب أن يتوافق مع شكل الخيارات وبالعكس. فإذا كان شكل السؤال مباشراً وكمياً فإن الخيارات يجب أن تكون رقمية مثل :

ماهي درجة التزامك بالرياضة ؟	5	4	3	2	1

أما إذا كان السؤال على شكل عبارة استفهامية فإن شكل الخيارات يكون لغوياً مثل:

انت شخص ملتزم بالرياضة؟	موافق جداً	موافق	محايد	معارض	معارض جداً

ثم يتم تحويل هذه الإجابة إلى أرقام مرتبة حسب ما يراه الباحث مناسباً .

كما نشير إلى أن الخيارات الرقمية أفضل وأدق من الخيارات اللغوية، لأن المسافات بين الخيارات الرقمية محددة ومتساوية (وتساوي الواحد)، أما المسافات بين الخيارات اللغوية فهي غير متساوية وغير محددة، فالمسافة بين (موافق جداً) و(موافق) غير معروفة ولا تساوي المسافة بين (موافق) و(محايد). عدا عن أن اتجاه ترتيبها اللغوي يجعلها عرضة للتحيز أثناء الإجابة .

ويجب أن يكون عدد الخيارات في الاستبيان الواحد موحداً. (ثلاثية أو خماسية أو سباعية لجميع الأسئلة). ولا يجوز اعتماد الاستبيان قبل عرضه على عدد من المختصين لتحكيمه وتصويبه، ثم القيام بتجربته وتمريره على عينة استطلاعية (لا تقل عن 30 فرداً) للتأكد من حسن صياغة الأسئلة ومن حسن فهمها ومن صواب الإجابة عليها، ويتم ذلك بحساب معامل الثبات (ألفا كرونباخ) من بيانات العينة الاستطلاعية فإذا كانت قيمته أكبر من (0.70) يمكننا اعتماد الاستبيان وتمريره على أفراد العينة الكلية ذات الحجم المحدد بـ n فرداً . وعند تحليل الثبات لا يجوز دمج الأسئلة ذات الخيارات المختلفة، بل يتم تحليل كل نوع على حده .

مثال (2-1): لنفترض أننا نريد دراسة أثر الهوايات المختلفة على صحة ونفسية كبار السن (أكبر من 70 عاماً) فصممنا استبياناً خاصاً مؤلفاً من (6) محاور (أو أسئلة) عن ممارسة الهوايات الممكنة (يومياً) لهؤلاء الأشخاص وعن حالتهم الصحية، النفسية، وكان على الشكل التالي:

جدول (2-1): الأسئلة والخيارات:

	نص السؤال أو العبارة (جميعها باتجاه واحد)	الخيارات أو الدرجات الممكنة للجواب				
X_1	ماهي درجة ممارستك للرياضة اليومية ؟	5	4 ✓	3	2	1
X_2	ماهي درجة تذوقك واستماعك للموسيقى ؟	5	4	3 ✓	2	1
X_3	ماهي درجة اهتمامك بالأخبار والسياسة ؟	5 ✓	4	3	2	1
X_4	ماهي درجة مواظبتك على المطالعة ؟	5	4 ✓	3	2	1
X_5	ماهي درجة تعاملك مع شبكات التواصل الاجتماعي ؟	5	4	3 ✓	2	1
X_6	ماهي درجة تقييمك لحالتك الصحية والنفسية ؟	5	4 ✓	3	2	1

وهنا نلاحظ أن كل سؤال من هذه الأسئلة يمكن أن يشكل محوراً خاصاً. ويمكننا أن نضع ضمنه عدة أسئلة أو عبارات حسب هدف البحث . وتتم الإجابة على هذه الأسئلة بسرعة وذلك بوضع إشارة معينة (مثل √ أو ×) على الدرجة المناسبة كما في الجدول السابق.

وأخيراً لنفترض إننا مررنا هذا الاستبيان السابق على عينة صغيرة مؤلفة من (10) أفراد (كمثال) ثم قمنا بوضع أجوبتهم حسب الأفراد (في الأسطر) وحسب السؤال (في الأعمدة) فكانت كما في الجدول التالي:

جدول (2-2): بيانات العينة الاستطلاعية (10 أفراد) :

الأسئلة / الأفراد i	X_1 / Y_1	X_2	X_3	X_4	X_5	X_6	T_{1i}	T_{2i}	T_i	المحك / $\bar{x}_i$
1	3 /3	2	3	2	2	2	8	6	14	2.33
2	4 /3	3	2	3	4	3	9	10	19	3.17
3	2 /3	2	4	2	3	2	8	7	15	2.50
4	3 /3	4	3	4	4	3	10	11	21	3.50
5	4 /3	3	4	3	3	4	11	10	21	3.50
6	3 /3	2	2	5	4	2	7	11	18	3.00
7	4 /4	3	3	3	2	3	10	8	18	3.00
8	3 /3	4	4	4	4	4	11	12	23	3.83
9	2 /3	3	2	2	2	2	7	6	13	2.17
10	3 /2	3	4	3	2	4	10	9	19	3.17
لدراسة الصدق										
متوسط السؤال $\bar{x}_j$	3.1 /3	2.9	3.1	3.1	3	2.9	9.1	9	18.1	3.017
الانحراف المعياري	0.7379 / 0.4714	0.7379	0.8756	0.9944	0.9429	0.8756	1.524	2.160	3.247	0.541
	$y_i = 0.3194$							$r = 0.540$		

المصدر: افتراضي من قبل المؤلف (الأرقام التي في زوايا خلايا العمود الثاني هي إجابات التجربة الثانية).

وهنا نطرح السؤالين التاليين:

هل هذه الإجابات ثابتة أم إنها تختلف من تجربة لأخرى أو من عينة لأخرى ؟

هل هذه الإجابات صادقة وتقترب من القيم الحقيقية أو المتوقعة لها ؟

للإجابة على هذين السؤالين نحتاج إلى استخدام أساليب مناسبة لقياس كل من الثبات والصدق. وهو ما سنعرضه فيما يلي :

2-2 أساليب قياس الثبات(Reliability) ويسمى أحياناً بـ (الاتساق الداخلي):

يعرف الثبات: بأنه استقرار الإجابات حول قيم معينة وعدم اختلافها كثيراً من تجربة لأخرى أو من عينة لأخرى . ويقصد بالاتساق الداخلي درجة انسجام الإجابات ضمن كل سؤال أو ضمن كل محور أو ضمن الاستبيان ككل . وهناك عدة أساليب لقياس هذا الثبات، أهمها ما يلي:

1-2-2 أسلوب إعادة التجربة (Parallel التوازي) :

وبحسب هذا الأسلوب يقوم الباحث بإعادة التجربة وتوزيع الاستبيان على نفس أفراد العينة، مع ضمان تحقيق نفس الشروط والظروف السابقة (بدون إعلامهم بهدف الإعادة) .

ثم يقوم بتسجيل الإجابات الجديدة مقابل الإجابات السابقة لكل فرد على كل سؤال. ولقياس ثبات النتائج يقوم بمقارنة الاجابات على كل سؤال في التجربتين، ثم يقوم بحساب معامل الارتباط (البيرسوني) بينهما من العلاقة المعروفة التالية :

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1) * \sigma_x * \sigma_y} \qquad (1-2)$$

حيث أن x_i هي قيمة الإجابة السابقة للفرد i على ذلك السؤال، وأن $\bar{x}$ متوسطها وأن σ_x هو انحرافها المعياري، وأن y_i هي قيمة الإجابة اللاحقة للفرد i على ذلك السؤال، وأن $\bar{y}$ هو متوسطها، و σ_y هو انحرافها المعياري . فإذا كانت قيمة هذا المعامل أكبر من (0.70) نعتبر أن ثبات هذه الإجابات مقبولاً أو جيداً ...الخ .

وكتطبيق على ذلك لنأخذ السؤال الأول فقط، ولنفترض أن إجابات الأفراد اللاحقة عليه هي الأرقام المسجلة في زوايا خلايا العمود X_1، والتي رمزنا لها بـ Y_1 ثم نقوم بحساب معامل الارتباط بين نتائج هاتين التجربتين من (2-1) فنجد أن قيمته تساوي:

$$r_1 = 0.3194 \qquad \text{(للسؤال الأول فقط)}$$

وهي قيمة صغيرة تدل على درجة ثبات ضعيفة للإجابات على السؤال الأول .

وهكذا نفعل مع بقية الأسئلة ونسجل الإجابات اللاحقة مقابل السابقة ونحسب معامل الارتباط لكل سؤال على حدة ونستخلص درجة الثبات لكل منها .

ولكن إذا كان حجم العينة كبيراً (وزوجياً) وكان تسلسل الأفراد في القائمة عشوائياً فيمكننا اتباع أسلوب آخر هو أسلوب التوازي النصفي ، والذي يتلخص بتجزئة إجابات كل سؤال إلى قسمين متساويين ، ثم وضع هذين القسمين مقابل بعضهما في عمودين جديدين وحساب معامل الارتباط بينهما، فنحصل على تقدير درجة الثبات لإجابات ذلك السؤال ، فمثلاً لو أخذنا إجابات السؤال الأول وقسمناها إلى قسمين (بفرض أن التسلسل عشوائي) كما يلي :

القسم الأول: يتألف من إجابات الأفراد الخمسة الأولى، والقسم الثاني: يتألف من إجابات الخمسة الأخرى ووضعناهما مقابل بعضهما ثم حسبنا معامل الارتباط بينهما لوجدنا أن: $r_1 = 0.4226$ ، وهو معامل صغير أيضاً ويدل على درجة ثبات ضعيفة لإجابات السؤال الأول .

ملاحظة: في الحقيقة أن معامل الارتباط لا يعبر بشكل جيد عن درجة ثبات الإجابات ، لأنه إذا طرحنا (واحد) من الإجابات السابقة فإنها ستصبح غير ثابتة، وإن قيمة معامل الارتباط بين الإجابات السابقة واللاحقة ستكون مساوية للواحد $r = 1$. رغم أن الإجابات أصبحت متحيزة، ولهذا فإننا سنحاول تطبيق

أساليب أخرى لقياس درجة ثبات الإجابات، ويفضل في هذه الحالة (حالة إعادة التجربة) تطبيق معامل التوافق X^2 بين نتائج التجربتين. أو استخدام اختبار الأزواج المتقابلة على الإجابات السابقة واللاحقة وحساب قيمة مؤشر ستودينت t من العلاقة:

$$t = \frac{\bar{d} - \bar{d}_0}{S_d/\sqrt{n}} \qquad (2-2)$$

حيث أن: $\bar{d}$ هو متوسط الفروقات بين نتائج التجربتين $d_i = (x_i - y_i)$.

وأن S_d هو الانحراف المعياري للفروقات d_i، و n حجم العينة .

أما $\bar{d}_0$ فهي قيمة متوسط الفروقات في المجتمع. ونأخذه من فرضية العدمH_0 (انظر العلاقة (1−54)).

2−2−2 أسلوب تجزئة الأسئلة بالمناصفة (Split- Half) أو أسلوب (سبيرمان− براون):

ويعتمد هذا الأسلوب على تجزئة الأسئلة (وليس الإجابات) إلى جزأين متساويين (بفرض أن عدد الأسئلة K زوجي) حسب تسلسلها أو حسب أي معيار آخر (فردي، زوجي) وتشكيل مجموعتين متقابلتين من الأسئلة. ثم نقوم بحساب متوسطات (أو مجاميع) إجابات الأفراد في أسئلة كل مجموعة، ونضعهما مقابل بعضهما، ثم نحسب معامل الارتباط بينها، وليكن مساوياً لـ r . ثم نقوم بحساب معامل الثبات الذي اقترحه (براون) لزيادة قيمة معامل الثبات وسُمي بمعامل (سبيرمان− براون) وهو يعرف بالعلاقة :

$$BC = \frac{2r}{1+r} \qquad (3-2)$$

ملاحظة: إذا كان عدد الأسئلة K فردياً يتم تجزئتها إلى $\frac{k+1}{2}$ سؤالاً ثم إلى $\frac{k-1}{2}$ سؤالاً (بفارق سؤال واحد)

ولقد قمنا بتجزئة أسئلة الاستبيان الوارد في الجدول (2−2) إلى جزأين متساويين كمايلي:

الجزء الأول: ويتألف من الأسئلة: الأول والثاني والثالث ، ورمزنا لمجموع إجابات الفرد i فيه (بدلاً من متوسطها) بالرمز T_{1i} .

الجزء الثاني: ويتألف من الأسئلة: الرابع والخامس والسادس، ورمزنا لمجموع إجابات الفرد i فيه (بدلا من متوسطها) بالرمز T_{2i} .

ثم حسبنا معامل الارتباط بين عمودي المجاميع T_{1i} و T_{2i} (أو المتوسطات) من العلاقة (2−1) فوجدنا أنه يساوي: $r = 0.540$.

وعندما حسبنا معامل (سبيرمان− براون) وجدنا أن قيمته تساوي : $BC = \frac{2(0.540)}{1.540} = 0.701$

وهي قيمة مقبولة وتدل على درجة ثبات مقبولة لإجابات ذلك الاستبيان .

2-2-3 أسلوب جوثمان (Guthman) :

ويعتمد هذا الأسلوب على أسلوب التجزئة بالمناصفة السابقة لأسئلة الاستبيان إلى نصفين متساويين، ثم القيام بحساب مجموع الإجابات في كل منهما فنحصل على T_{1i} و T_{2i}، ثم نقوم بحساب تباين مجاميع الإجابات T_{1i} و T_{2i} في كل نصف على حدة فنحصل على التباينين S_1^2 و S_2^2 . ثم نقوم بحساب المجموع الكلي للإجابات مقابل كل فرد فنحصل على العمود T_i ، ثم نقوم بحساب التباين الكلي S^2 لمجموع إجابات الاستبيان، وهو يساوي تباين المجاميع الكلية الواردة في العمود T_i في الجدول (2-2)، ثم نقوم بحساب قيمة معامل الثبات للاستبيان ككل من العلاقة التي عرفها (جوثمان) التالية :

$$\propto = 2\left[1-\frac{S_1^2+S_2^2}{S^2}\right] \qquad (4-2)$$

ومن بيانات الجدول (2-2) واعتماداً على التجزئة السابقة نجد أن:

$$\propto = 2\left[1-\frac{(1.524)^2+(2.160)^2}{(3.247)^2}\right] = 0.677$$

وهي قيمة شبه مقبولة، وتدل على أن درجة ثبات الإجابات في ذلك الاستبيان يمكن أن تكون مقبولة إذا ازداد حجم العينة n، كما يمكن تحسينها بإعادة التجزئة بأسلوب آخر أو بحذف سؤال واحد أو أكثر من أسئلة الاستبيان، وسنتعرض لتلك العمليات لاحقاً .

ملاحظة: نلاحظ أن قيمة معامل (جوثمان) تختلف عن قيمة معامل (سبيرمان- براون). وذلك لأمور تتعلق بأسلوب الحساب وهنا يبرز أمامنا السؤال الثاني: أي القيمتين نعتمدها في التحليل؟

إن الجواب على هذا السؤال يعتمد على قيمتي تبايني المجموعتين S_1^2 و S_2^2 .

فإذا كان هذان التباينان متجانسين فإننا نعتمد على قيمة معامل (سبيرمان- براون) .

أما إذا كان هذان التباينان مختلفين فإننا نعتمد على قيمة معامل (جوثمان) .

وفي مثالنا نجد أن التباينين مختلفان لذلك نعتمد على قيمة معامل (جوثمان) ونعمل على تحسينه .

2-2-4 أسلوب ألفا كرونباخ (Alpha,s Cronbache) :

وهو تعميم لطريقة (جوثمان) من أجل تطبيقها على جميع الأسئلة في الاستبيان المؤلف من K سؤالاً، لذلك قام (كرونباخ) باعتبار كل سؤال في الاستبيان وكأنه جزء خاص من أصل K جزءاً ، واستفاد من علاقة (جوثمان) ، وقام بإجراء التعميم على K جزءاً أو سؤالاً فتوصل إلى تعريف معامل جديد يسمى معامل (ألفا كرونباخ) ويحسب من العلاقة التالية:

$$\propto = \frac{k}{k-1}\left[1-\frac{S_1^2+S_2^2+S_3^2+\dots.\ S_k^2}{S^2}\right] \qquad (5-2)$$

وإذا قمنا بتطبيق هذه العلاقة على جميع أسئلة الاستبيان نجد من الجدول (2–2) أن:

$$\propto=\frac{6}{6-1}\left[1-\frac{(0.7379)^2+(0.7379)^2+(0.8756)^2+(0.9944)^2+(0.9429)^2+(0.8756)^2}{(3.247)^2}\right]$$

$$\propto=\frac{6}{5}\left[1-\frac{4.50023}{10.543000}\right]=0.6877$$

وهو معامل الثبات لإجمالي الاستبيان، وقيمته شبه مقبولة ويمكن تحسينها بحذف سؤال أو أكثر من الاستبيان كما سنرى لاحقاً. أو بزيادة حجم العينة n .

2-2-5: كيفية استخراج الصيغ المختلفة لمعامل (ألفا كرونباخ):

يوجد لمعامل (ألفا كرونباخ) عدة صيغ رياضية هي:

أ– **صيغة التباينات المشتركة** : وتستخرج من مصفوفة التباينات المشتركة لأسئلة الاستبيان. ولنفترض أن هذه المصفوفة المتناظرة تأخذ الشكل التالي :

$$\boldsymbol{cov(X)}=\begin{array}{c c} & \begin{array}{ccccc} X_1 & X_2 & X_3 & \dots & X_k \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k \end{array} & \begin{bmatrix} C_{11} & C_{12} & C_{13} & \dots & C_{1k} \\ C_{21} & C_{22} & C_{23} & \dots & C_{2k} \\ C_{31} & C_{32} & C_{33} & \dots & C_{3k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ C_{k1} & C_{k2} & C_{k3} & \dots & C_{kk} \end{bmatrix} \end{array} \qquad \boldsymbol{(6-2)}$$

حيث رمزنا بـ C_{ij} للتباين المشترك للسؤال i مع السؤال $j(i \neq j)$ ، والذي يعرف بالرمز $cov(X_i,X_j)$ ، أما إذا كان $(i=j)$ فإن C_{ii} هو تباين السؤال X_i نفسه، والذي يعرف بالرمز σ_i^2 .

واعتماداً على هذه المصفوفة يمكننا أن نعرف المجموع الكلي لعناصرها، ونرمز له بـ S ونحسبه من العلاقة :

$$S=\sum_{i=1}^{k}\sum_{j=1}^{k}C_{ij} \qquad \boldsymbol{(7-2)}$$

ثم نعرف متوسط عناصرها $\bar{S}$ ونحسبه من بالعلاقة :

$$\bar{S}=\frac{S}{k^2}=\frac{\sum\sum C_{ij}}{k^2} \qquad (8-2)$$

وحتى نبرز التأثيرات المختلفة للأسئلة على بعضها البعض، نأخذ مجموع التباينات المشتركة لها (غير القطرية) ونرمز له بـ C وهو يساوي :

$$C=\sum_{i\neq j}^{k}\sum_{j=1}^{k}C_{ij} \qquad (i\neq j) \quad \left(\text{غير القطرية}\right) \qquad \boldsymbol{(9-2)}$$

وهكذا نجد أن:

$$S=C+\sum_{i=1}^{k}\sigma_i^2 \qquad : \left(\text{غير القطرية + القطرية}\right) \qquad \boldsymbol{(10-2)}$$

حيث أن: $\sum_{i=1}^{k}\sigma_i^2$ هو مجموع العناصر القطرية في المصفوفة (2−6) لأن $C_{ii}=\sigma_i^2$.

إن المجموع المزدوج في (2−9) هو عبارة عن مجموع العناصر غير القطرية للمصفوفة cov والتي عددها يساوي $k(k-1)$ ، لذلك فإن متوسط عناصرها يساوي :

$$\bar{C}=\frac{1}{k(k-1)}\sum_{i\neq j}^{k}\sum_{j=1}^{k}C_{ij} \qquad (11-2)$$

وبناء على ذلك تم تعريف معامل (ألفا كرونباخ) الأساسي وسنرمز له بـ alpha من خلال العلاقة التالية:

$$alpha=\frac{\bar{C}}{\bar{S}}=\frac{\frac{1}{k(k-1)}\sum_{i\neq j}^{k}\sum_{j=1}^{k}C_{ij}}{\frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}C_{ij}}=\frac{k}{k-1}*\frac{C}{S} \qquad (12-2)$$

وبما أنه لدينا من (2−10) أن C تساوي المجموع الكلي S مطروحاً منه مجموع العناصر القطرية $\left(\sum_{i=1}^{k}\sigma_i^2\right)$ ، أي أن:

$$C=S-\sum_{i=1}^{k}\sigma_i^2 \qquad (a\ 12-2)$$

وبذلك نجد أن العلاقة (2−12) تأخذ الشكل التالي:

$$alpha=\frac{k}{k-1}*\left[\frac{S-\sum_{i=1}^{k}\sigma_i^2}{S}\right]$$

وبعد الإصلاح نجد أن :

$$alpha=\frac{k}{k-1}\left[1-\frac{\sum_{i=1}^{k}\sigma_i^2}{S}\right]=\frac{k}{k-1}\left[1-\frac{\sigma_1^2+\sigma_2^2+\cdots+\sigma_k^2}{S}\right] \qquad (\mathbf{13-2})$$

حيث أن: K هو عدد الأسئلة و σ_i^2 هو تباين السؤال i ، وأن S هو مجموع عناصر مصفوفة التباينات المشتركة لجميع الأسئلة . وهذه الصيغة هي الأكثر انتشاراً في التطبيقات العملية ، وهي تتطابق مع العلاقة السابقة (2−5) ، رغم اختلاف الرموز بينهما.

ب− **<u>صيغة المتوسطات</u>** : يمكن استخراج هذه الصيغة من العلاقة (2−12)، مع ملاحظة أنه لدينا من (2−10) أن: $S=\sum\sigma_i^2+C$ ، فإننا نعالج العلاقة (2−12)على الشكل التالي (نضرب ونقسم البسط بـ K ثم بـ$(k-1)$) :

$$alpha=\frac{k}{k-1}*\frac{C}{S}=\frac{\frac{C}{k-1}}{\frac{S}{k}}=\frac{\frac{k*C}{k(k-1)}}{\frac{\sum\sigma_i^2}{k}+\frac{C}{k}}$$

$$alpha=\frac{k*\bar{C}}{\bar{\sigma}^2+\frac{(k-1)C}{k(k-1)}}=\frac{k\bar{C}}{\bar{\sigma}^2+(k-1)\bar{C}} \qquad (14-2)$$

حيث أن $\bar{\sigma}^2$ هو متوسط التباينات القطرية للأسئلة المنفردة في المصفوفة (2−6)

وأن $\bar{C}$ هو متوسط التباينات المشتركة، أي متوسط العناصر غير القطرية في المصفوفة (2−6) .

ج- **الصيغة الارتباطية أو المعيارية:** وتعتمد هذه الصيغة على عناصر المصفوفة الارتباطية بين جميع الأسئلة X_i والتي يمكن كتابتها كما يلي:

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1k} \\ r_{21} & 1 & r_{23} & \dots & r_{2k} \\ r_{31} & r_{32} & 1 & \dots & r_{3k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \dots & 1 \end{bmatrix} \qquad (15-2)$$

ومنها يمكننا استنباط تعريف آخر للمعامل (ألفا كرونباخ) مشابه تماماً للتعريف الذي في (2-14)، وذلك باستخدام المصفوفة الارتباطية السابقة وباستبدال الرموز في (2-14) برموز مناسبة لمعاملات الارتباط ، فنحصل على أن :

$$alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}} \qquad (16-2)$$

حيث أن $\bar{r}$ هو متوسط معاملات الارتباط غير القطرية، وأما متوسط المعاملات القطرية فيساوي (1) لأن مجموعها K، وعددها K واحداً .

ومن جهة أخرى وبناء على العلاقة (2-13) يمكننا أيضاً أن نستنتج علاقة أخرى لحساب (ألفا كرونباخ) من عناصر المصفوفة الارتباطية R مباشرة كما يلي:

$$alpha = \frac{k}{k-1}\left[1 - \frac{k}{\sum_{i=1}^{k}\sum_{j=1}^{k} r_{ij}}\right] \qquad (17-2)$$

وذلك لأن مجموع العناصر القطرية يساوي K.

وعند حساب المصفوفة R وحساب قيمة $alpha$ للمثال (2-1) نجد أن: $alpha = 0.69400$، وهي قيمة مقبولة وتدل على ثبات مقبول للإجابات في الاستبيان السابق

د- **صيغة تحليل التباين باتجاهين:** من المعلوم أن تحليل التباين باتجاهين (tow way) يعتمد على عاملين مستقلين (الأسئلة والأشخاص) ليختبر تأثيرهما على عامل ثالث (الدرجات)، وهو يعطينا جدولاً لمربعات الانحرافات حسب الأسطر وحسب الأعمدة بالإضافة لتباين الأخطاء، كما في الجدول التالي: **(انظر الفصل السادس)**

جدول (2-3) تحليل التباين الثنائي (لـ K سؤالاً من عينة حجمها n فرداً)

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط المربعات	قيمة F	قيمة P Sig
حسب الأعمدة: column	SSC	$k-1$	$MSSC = \frac{SSC}{k-1}$	$F_C = \frac{MSC}{MSE}$	—
حسب الأسطر: raw	SSR	$n-1$	$MSSR = \frac{SSR}{n-1}$	$F_r = \frac{MSR}{MSE}$	—
الأخطاء: Error	SSE	$(n-1)(k-1)$	$MSSE = \frac{SSE}{(n-1)(k-1)}$	—	—
المجموع: Total	SST	$nk-1$	—	—	—

ومنه يمكننا حساب قيمة معامل (ألفا كرونباخ) من العلاقة :

$$alpha = \frac{MSE}{MSR} = \frac{\text{متوسط مربعات الخطأ}}{\text{متوسط مربعات الأسطر}} = \frac{1}{F_r} \qquad (18-2)$$

ملاحظة: عند حساب قيم (ألفا كرونباخ) للمثال نفسه من هذه الصيغ المختلفة ، قد نجد بعض الاختلاف بينها وذلك يعود لعوامل التقدير وطرق الحساب والتقريب.

2-2-6: قواعد تصنيف قيم المعامل (ألفا كرونباخ) :

لقد استقرت الآراء في أغلب المراجع على تصنيف قيم (ألفا كرونباخ) التي تقع في المجال [1 ، 0] إلى عدة مستويات كما في الجدول التالي:

جدول (2-4) مستويات تصنيف قيم ألفا كرونباخ

فئات التصنيف لـ $\propto$	تقدير الثبات أو الاتساق الداخلي
إذا كانت: $alpha \geq 0.90$	ممتاز
$0.80 \leq alpha < 0.90$	جيد
$0.70 \leq alpha < 0.80$	مقبول
$0.60 \leq alpha < 0.70$	هناك تساؤل
$0.50 \leq alpha < 0.60$	ضعيف
وعندما: $alpha \leq 0.50$	غير مقبول

ملاحظة: إذا كانت قيمة (ألفا كرونباخ) سالبة لأحد المحاور فهذا يدل عدم ثبات الإجابات فيه ويجب العمل على التخلص من هذه الحالة بحذف واحد أو أكثر من الأسئلة منه كما سنرى لاحقاً .

2-2-7: اختبار معنوية قيمة (ألفا كرونباخ) :

لاختبار معنوية $alpha$ في المجتمع من خلال القيمة المحسوبة لـ $alpha$ ضمن مستوى دلالة $\propto = 0.05$ ، نضع الفرضيتين بناء على الجدول (2-4) كما يلي:

وهو الحل الفاصل بين القبول والرفض:

$H_0: alpha \leq 0.70$

$H_1: alpha > 0.70$

ثم نقوم بحساب مؤشر الاختبار F المعرف كما يلي:

$$F = \frac{1-\propto_0}{1-alpha} = \frac{1-0.70}{1-alpha} \qquad (19-2)$$

وهو متحول يخضع للتوزيع $F(v_1, v_2)$، حيث أن درجتي الحرية تساويان: $v_1 = n-1$، و $v_2 = (n-1)(k-1)$. وكتطبيق على ذلك نحسب قيمة F من (2- 19) لقيمة (ألفا) المحسوبة من العلاقة (2-5) للمثال (2-1) فنجد أن $F = 0.9606$، ثم نقارن هذه القيمة مع القيمة الحرجة $F_{\propto(9.45)}$ المساوية لـ $F_{\propto} = 2.096$ ونتخذ القرار كمايلي:

بما أن $F \leq F_{\propto}$ نقبل فرضية العدم التي تقول أن قيمة $alpha$ أصغر أو تساوي 0.70 .
أما إذا كانت $F > F_{\propto}$ نرفض H_0 ونقبل H_1 التي تقول أن القيمة المحسوبة لـ $alpha$ معنوية وقيمتها أكبر من 0.70 .

2-2-8 حذف الأسئلة السيئة :

بعد أن نقوم بحساب قيمة $alpha$ واختبارها، فقد نجد أن قيمتها ضعيفة أو غير مقبولة أو مقبولة فقط، وفي مثل هذه الحالات علينا العمل على رفع قيمة $alpha$ المحسوبة، وذلك عن طريق حذف بعض الأسئلة السيئة من الاستبيان، ولإجراء ذلك الحذف علينا أن نحدد الأسئلة التي يجب حذفها. لذلك نقوم بدراسة قيم متوسطات وتباينات الأسئلة. ونركز على الأسئلة ذات المتوسطات الكبيرة أو الصغيرة ، أو ذات التباينات الكبيرة ، ونعمل على حذفها واحداً بعد الآخر أو دفعة واحدة، ثم نقوم بحساب قيمة $alpha$ بعد حذف كل سؤال سيئ، فإذا ازدادت قيمتها بشكل ملموس فإن ذلك يستوجب حذف ذلك السؤال من الاستبيان .

وهناك عملية خاصة في البرنامج الحاسوبي SPSS تسمى (Scale if item Deleted) (الدرجة إذا حُذف السؤال) لتنفيذ هذه العمليات وحساب تأثير حذف الأسئلة واحداً بعد الآخر على قيمة $alpha$، حيث يقوم بحساب معامل الارتباط لها مع عمود المجموع الإجمالي T، ويتم حذف كل سؤال يقابله قيمة صغيرة أو سالبة لذلك المعامل، مع الأخذ بعين الاعتبار قيمة $alpha$ بعد الحذف في كل حالة .
وكمثال على ذلك نأخذ المثال التالي :

مثال (2-2): لحساب قيمة (ألفا كرونباخ) لكامل الاستبيان السابق من العلاقة (2-13) علينا أن نحسب مصفوفة التباينات المشتركة cov، لذلك نلجأ إلى الحاسوب وإلى برنامج SPSS فنجد أنها تساوي:

$$\boldsymbol{cov(X)} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \\ 0.544 & 0.122 & -0.011 & 0.211 & 0.111 & 0.344 \\ 0.122 & 0.544 & 0.122 & 0.233 & 0.222 & 0.433 \\ -0.011 & 0.122 & 0.767 & -0.122 & -0.111 & 0.456 \\ 0.211 & 0.233 & -0.122 & 0.989 & 0.667 & 0.233 \\ 0.111 & 0.222 & -0.111 & 0.667 & 0.889 & 0.111 \\ 0.344 & 0.433 & 0.456 & 0.233 & 0.111 & 0.767 \end{bmatrix}$$

ومنها نجد أن مجموع عناصرها يساوي :

$$S = \sum_{i=1}^{k}\sum_{j=1}^{k} C_{ij} = 10.542$$

وأن مجموع العناصر القطرية فيها يساوي:

$$\sum_{i=1}^{k} \sigma_i^2 = \sum C_{ii} = 4.5$$

ومن العلاقة (2−13) نجد مباشرة أن:

$$alpha = \frac{6}{5}\left[1 - \frac{4.5}{10.542}\right] = 0.6878$$

ويمكن حساب هذا المعامل من المصفوفة الارتباطية (2−15) والعلاقة (2−17) لذلك نحسب المصفوفة الارتباطية لجميع أسئلة الاستبيان فنجد أنها تساوي :

$$R = \begin{bmatrix} 1 & 0.224 & -0.017 & 0.288 & 0.160 & 0.533 \\ 0.224 & 1 & 0.189 & 0.318 & 0.319 & 0.671 \\ -0.017 & 0.189 & 1 & -0.140 & -0.135 & 0.594 \\ 0.288 & 0.318 & -0.140 & 1 & 0.711 & 0.268 \\ 0.160 & 0.319 & -0.135 & 0.711 & 1 & 0.135 \\ 0.533 & 0.671 & 0.594 & 0.268 & 0.135 & 1 \end{bmatrix}$$

ولتطبيق العلاقة المعيارية (2−17) نقوم بحساب مجموع عناصر المصفوفة R فنجد أنه يساوي :

$$\sum_{i=1}^{6}\sum_{j=1}^{6} r_{ij} = 14.236$$

وهكذا نجد أن:

$$alpha = \frac{6}{5}\left[1 - \frac{6}{14.236}\right] = 0.694 \approx 0.70$$

وهي القيمة المعيارية لـ(ألفا كرونباخ) وهي تدل على درجة شبه مقبولة لثبات إجابات الاستبيان المدروس. ولتحسين هذه القيمة نستعين ببرنامج SPSS وندخل من Scale على (Reliability Analysis) ثم على Statistics فنجد أمامنا الخيار (Scale if item deleted)، وهو يعطينا قيمة (alpha) فيما إذا تم حذف سؤال معين من الاستبيان، ويقدم لنا بعد حسابات معقدة جدولاً مفصلاً يشير إلى قيمة (alpha) إذا تم حذف كل سؤال من أسئلة الاستبيان. وعلينا اتخاذ القرار المناسب لحذف سؤال أو أكثر لزيادة قيمة معامل الثبات . وفي مثالنا هذا نجد أنه يعطينا الجدول التالي:

جدول (2−5) مؤشرات حذف المتحولات

قيمة alpha بعد الحذف	مربع الارتباط المتعدد	معامل الارتباط المصحح للمجموع	تباين المجموع بعد الحذف	متوسط المجموع بعد الحذف	المتحول المرشح للحذف
0.664	0.583	0.363	8.444	15.00	X_1
0.611	0.657	0.552	7.733	15.20	X_2
0.738*	0.681	0.128	9.111	15.00	X_3
0.633	0.584	0.461	7.111	15.00	X_4
0.660	0.550	0.383	7.656	15.10	X_5
0.545	0.868	0.700	6.622	15.20	X_6

وأكثر ما يهمنا في هذا الجدول هو قيمة (alpha) بعد الحذف. فنجد أن هذه القيمة تزداد لتبلغ (0.738) إذا تم حذف المتحول X_3 . وهي تتناقص إذا تم حذف أي سؤال آخر. لذلك نقوم بحذف X_3 من الاستبيان ونعيد الحسابات فنحصل على أن:

عدد المتحولات	القيمة المعيارية لـ ألفا كرونباخ	قيمة ألفا كرونباخ
5	0.740	0.738

وهما تحسبان من العلاقتين السابقتين (2-13) و(2-17)، ولكن على المتحولات الخمسة المتبقية: $(X_1 , X_2 \ , X_4 \ , X_5 \ , X_6)$.

ويمكننا إعادة الحسابات للكشف عما إذا كان يمكن حذف متحول آخر لزيادة قيمة (alpha) . ولكن تلك الحسابات تشير إلى أنه إذا تم حذف أي متحول آخر فإن قيمة (alpha) ستنقص عما هي عليه. لذلك نتوقف عن الحذف ونكتفي بحذف السؤال الثالث X_3 ، لكي نحصل على أن alpha= 0.738، على أمل أنها ستتحسن عندما يزداد حجم العينة n .

وللتأكد من معنوية هذه القيمة يمكننا إجراء اختبار F عليها فنضع الفرضيتين كما يلي:

$$H_0 \leq 0.70 \qquad , \qquad H_1 > 0.70$$

ثم نحسب قيمة مؤشر الاختبار F فنجد أن:

$$F = \frac{1 - 0.70}{1 - 0.738} = 1.145$$

ثم نبحث عن $F_{0.05}$ الحرجة والمقابلة لدرجتي الحرية:

$$v_1 = n - 1 = 10 - 1 = 9$$
$$v_2 = (n-1)(k-1) = 9 * 4 = 36 \qquad \left(\text{لأن أصبحت } k = 5\right)$$

فنجد أنها تساوي: $F_{(0.05),9.36} = 2.153$

وبما أن $F < F_{0.05}$ فإننا نقبل فرضية العدم التي تقول أن قيمة (alpha) أصغر أو تساوي من (0.70) ، ونقول إن قيمة معامل الثبات مازالت ضعيفة ، لذلك يجب أن نعمل على زيادة قيمتها بزيادة حجم العينة n .

ملاحظة: يمكن حساب قيمة (alpha) لأي محور من محاور الاستبيان واعتباره كأنه استبيان خاص بحد ذاته. فمثلاً يمكننا أن نشكل من المتحولات الثلاثة الأولى $(X_1 , X_2 \ , X_3)$ محوراً خاصاً ونعتبره استبياناً كاملاً، ونحسب قيمة (alpha) له من العلاقة (2-5) فنجد أنها تأخذ الشكل التالي:

$$alpha = \frac{K}{K-1}\left[1 - \frac{S_1^2 + S_2^2 + S_3^2}{S_{T_1}^2}\right]$$

$$alpha = \frac{3}{2}\left[1 - \frac{(0.7379)^2 + (0.7379)^2 + (0.8756)^2}{(1.524)^2}\right]$$

$$alpha = \frac{3}{2}\left[1 - \frac{1.855668}{2.322576}\right] = 0.3015$$

وهي قيمة ضعيفة لثبات الإجابات في أسئلة المحور الأول (ربما بسبب X_3) . ثم نشكل محوراً آخر من المتحولات (X_6 ,X_5 ,X_4) ونحسب قيمة (alpha) لها من العلاقة (2–5) التي تأخذ الشكل التالي:

$$alpha = \frac{K}{K-1}\left[1 - \frac{S_4^2 + S_5^2 + S_6^2}{S_{T_2}^2}\right]$$

$$alpha = \frac{3}{2}\left[1 - \frac{(0.9944)^2 + (0.9429)^2 + (0.8756)^2}{(2.160)^2}\right]$$

$$alpha = \frac{3}{2}\left[1 - \frac{2.644567}{4.6656}\right] = 0.6479$$

وهي قيمة شبه مقبولة لثبات الإجابات في أسئلة المحور الثاني.

2–3 أساليب قياس الصدق :

إن مؤشرات الصدق في الاستبيان أمر هام لأنها تغني عن مؤشرات الثبات. فالإجابات الصادقة في الاستبيان هي إجابات ثابتة، ولكن العكس غير صحيح، لأنه ليست بالضرورة أن تكون الإجابات الثابتة صادقة (فقد تكون متحيزة وبالتالي تكون غير صادقة) .

والصدق: هو التطابق أو التوافق بين الإجابات التجريبية في الاستبيان مع القيم المعيارية أو الفعلية لها، إذاً لقياس الصدق لابد من وجود قيم معيارية معلومة لنقيس عليها الإجابات في كل سؤال. مثل: ماهو تقديرك لدرجة الحرارة الآن دون أن تعلم أنها على المقياس تساوي 25° ، وهنا تؤخذ القيمة (25) معياراً لصدق الإجابات .

وللصدق أنواع وأشكال هي:

- الصدق الظاهري أو الشكلي: وهو التطابق أو التوافق بين فقرات الاستبيان من حيث الشكل ومن حيث الغرض الذي نقيسه.
- صدق المحتوى أو المضمون: وهو أن يشمل الاستبيان جميع جوانب الموضوع المدروس ويعبر عن مضامينه .
- صدق المفهوم: وضوح العبارات والأسئلة ومطابقتها مع المفهوم العام للفكرة المطروحة (إن زيادة الدراسة يزيد من معدل النجاح) .
- الصدق العاملي: وهو كيفية تحليل الصفة المدروسة إلى عناصرها الأولية لتسهيل عملية قياسها .
- الصدق التنبؤي: يقصد به التطابق أو الارتباط لفكرة معينة حالية مع فكرة أخرى في المستقبل (تفوق الطالب في الثانوية يؤدي إلى تفوقه في الجامعة) ويتم معالجة هذه الأنواع السابقة من قبل المختصين والمحكمين .
- الصدق التلازمي: ويقصد به تطابق الإجابات أو متوسطاتها الأفقية مع متوسطات أو مع إجابات قياسية تسمى المعيار أو (المحك)، والمحك هو مقياس خارجي لا يتعلق بالاختبار، وهو مقياس

موضوعي، سبق إن تم التأكد من صدقه وثباته (مثل: متوسط ضغط الدم الطبيعي= 120 بار، أو متوسط درجة حرارة جسم الإنسان= 36.5 درجة). ويتم التأكد من الصدق التلازمي من خلال دراسة الارتباط بين إجابات الاستبيان لكل سؤال مع درجات المحك المقابلة لها . وكلما كان الارتباط شديداً بينهما كان الصدق التلازمي في الاختبار محققاً .

- الصدق التمييزي: وهو يعبر عن قدرة الاستبيان على تحديد التطابق بين طرفي الإجابات في المجموع العام أو في المتوسط العام لذلك الاستبيان .

2-3-1 أهم مؤشرات قياس الصدق هي:

هناك عدة مؤشرات لقياس الصدق ونذكر منها التالي:

1- **معامل الارتباط مع المحك:** وهو يستخدم لقياس الصدق التلازمي لإجابات كل سؤال مع المحك المعتمد، ويحسب من العلاقة (2-1) .

ونظراً لعدم وجود محكات جاهزة لأسئلة الاستبيان، لذلك نأخذ المتوسطات العامة لقيم الإجابات حسب الأفراد $\bar{x}_i$ ونشكل منها عموداً خاصاً. ونعتبره المحك المعياري لصدق الإجابات في ذلك الاستبيان.

ولإنشاء محك لمثالنا السابق (2-1) (قبل حذف المتحول X_3)، قمنا بحساب متوسطات إجابات الأفراد على جميع الأسئلة ووضعناها في عمود خاص ورمزنا له بـ $\bar{x}_i$ (لكل فرد i) . ثم قمنا بحساب معاملات الارتباط بين عمود الإجابات في كل سؤال مع ذلك المحك $\bar{x}_i$ فحصلنا على الجدول التالي:

المتحولات	X_1	X_2	X_3	X_4	X_5	X_6
معامل الارتباط مع المحك $\bar{x}_i$	0.552	0.700*	0.387	0.685*	0.617	0.825**
P=Sig (2-tailed)	0.098	0.024	0.269	0.029	0.057	0.003

ومن هذا الجدول نلاحظ أن قيم معامل الارتباط معنوية فقط مع ثلاثة متحولات هي: (X_6 ,X_4 ,X_2)، وإن القيم العددية لمعاملاتها تشير إلى درجة مقبولة لصدق الإجابات لتلك الأسئلة وإن قيمه غير معنوية مع المتحولات الأخرى (X_5 ,X_3 ,X_1)، وهي تشير إلى أن درجة الصدق في إجابات هذه المتحولات ضعيفة .

وهنا نلاحظ إن قيمته مع X_3 صغيرة جداً. لذلك نشكك في الإجابات على X_3 وننصح بحذفه من الاستبيان، وهذا ما أشرنا إليه سابقاً .

وأخيراً نشير إلى أن جميع هذه المعاملات ستتحسن وتستقر عند قيم معينة لكل منها عندما نزيد حجم العينة n، ويجب أن لا نتسرع بحذف المتحولات ذات القيم الصغيرة قبل أن نستكمل تمرير الاستبيان على كامل أفراد العينة، حيث يمكن حذف أي سؤال من الاستبيان بعد ذلك .

2- **معامل الصدق العام:** ويستخدم لقياس درجة الصدق العام، للاستبيان ككل أو لكل محور من محاوره، ويحسب من الجذر التربيعي لمعامل الثبات (alpha) كما يلي:

$$V = \sqrt{alpha} \qquad (20-2)$$

وفي مثالنا (2−1) نجد أن معامل الصدق لذلك الاستبيان تساوي :

$$V = \sqrt{0.6878} = 0.8293$$

وهي تعكس درجة جيدة لمصداقية الإجابات في ذلك الاستبيان.

ولحساب معامل الصدق للمحور الأول الذي يشمل (X_1, X_2 , X_3) نجد أن:

$$V_1 = \sqrt{0.3015} = 0.5490$$

وهي تعكس درجة ضعيفة لمصداقية الإجابات في ذلك المحور، وربما بسبب الإجابات في X_3 المرفوض .

ولحساب معامل الصدق للمحور الثاني الذي يشمل (X_4, X_5 , X_6) نجد أن:

$$V_2 = \sqrt{0.6497} = 0.8060$$

وهي تعكس درجة جيدة لمصداقية الإجابات في ذلك المحور .

3− **اختبار Z أو t** : وهو يستخدم لاختبار صدق إجابات كل سؤال j ، بمقارنة متوسطه العمودي $\bar{x}_j$ المسجل في السطر الأخير مع المتوسط المتوقع له. ولتحديد المتوسط المتوقع له نلجأ إلى اعتماد القيمة الوسطى لخيارات الأجوبة (مثل العدد 3 في المقياس الخماسي)، كما يمكن اعتبار المتوسط العام $\bar{\bar{x}}$ كقيمة متوقعة لجميع الإجابات في الاستبيان. ثم نضع الفرضيتين كما يلي:

$$H_0: \bar{x}_j = \bar{\bar{x}} = 3 \quad ، \qquad H_1: \bar{x}_j \neq \bar{\bar{x}}$$

وبعد ذلك يتم حساب قيمة مؤشر الاختبار t لكل سؤال j ، من العلاقة:

$$t_j = \frac{\bar{x}_j - \bar{\bar{x}}}{S_j/\sqrt{n}} \approx \frac{\bar{x}_j - 3}{S_j/\sqrt{n}} \qquad (21-2)$$

ثم نقارن قيمة t مع القيمة الحرجة $t_{\frac{\alpha}{2}}$ عند $(n-1)$ درجة حرية ، فإذا كانت $|t| < t_{\frac{\alpha}{2}}$ نقبل فرضية العدم التي تقول بعدم وجود فرق معنوي لمتوسط إجابات السؤال j مع القيمة المتوقعة لها .

وعند تطبيق ذلك على السؤال الأول في مثالنا السابق نجد من بيانات الجدول (2−2) أن: (باعتماد المتوسط العام) .

$$t_1 = \frac{\bar{x}_1 - \bar{\bar{x}}}{S_1/\sqrt{n}} = \frac{3.1 - 3.017}{0.7379/\sqrt{10}} = 0.3557$$

ومن جداول توزيع (ستودينت) نجد أن القيمة الحرجة المقابلة لمستوى دلالة $\alpha = 0.05$ ولدرجة حرية $(n-1=9)$ تساوي $t_{\frac{\alpha}{2}*9} = 1.833$ ، وعند المقارنة نجد أن $t_1 < t_{\frac{\alpha}{2}}$ ، لذلك نقبل فرضية العدم H_0 ونعتبر أن إجابات السؤال الأول صادقة باحتمال 0.95 . وهكذا يمكننا اختبار صدق الإجابات في الأسئلة الأخرى .

ملاحظة: إذا كان الاستبيان مؤلفاً من عدة محاور. فيمكن دراسة واختبار صدق متوسطات كل محور بمقارنتها مع المتوسط العام للاستبيان، وتطبيق نفس العلاقة (2−21) على متوسطات كل محور كما فعلنا أعلاه .

4- **اختبار الصدق التمييزي:** ويستخدم لاختبار تطابق طرفي الإجابات، ويشترط هذا الاختبار أن يكون حجم العينة n كبيراً، وهو يعتمد على مقارنة طرفي الإجابات المرتبة لكل سؤال أو للاستبيان ككل. وهذا يقتضي تشكيل مجموعتين طرفيتين من إجابات السؤال أو من المتوسطات العامة للإجابات في الاستبيان ككل. وسنقوم باختصار خطوات العمل لاختبار الصدق التمييزي للاستبيان ككل بما يلي:

أ- نرتب قيم المتوسطات العامة للاستبيان $\bar{x}_i$ تصاعدياً (أو تنازلياً) ونضعها في عمود جديد .

ب- نأخذ حوالي 25% أو 30% من المتوسطات العامة المرتبة الأولى، ونشكل منها المجموعة الطرفية الأولى ونضعها في عمود جديد آخر، ثم نحسب متوسطها $\bar{x}_1$ وتباينها S_1^2 .

ج- نأخذ حوالي 25% أو 30% من المتوسطات العامة المرتبة الأخيرة، ونشكل منها المجموعة الطرفية الثانية ونضعها في عمود جديد ثالث مقابل عمود المجموعة الأولى . ثم نحسب متوسطها $\bar{x}_2$ وتباينها S_2^2 .ويفضل أن يكون عدد العناصر في المجموعتين متساوياً .

د- نضع الفرضيتين حول الفرق بين هذين المتوسطين في المجتمع كما يلي:

$$H_0: \bar{y}_1 - \bar{y}_2 = 0 \qquad\qquad H_1: \bar{y}_1 - y_2 \neq 0$$

هـ- نقوم بحساب مؤشر الاختبار t للفرق بين عينتين مستقلتين من العلاقة:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad (22-2)$$

و- نقوم بمقارنة قيمة t المحسوبة مع قيمة $t_{\frac{\propto}{2}}$ الحرجة والمقابلة لدرجة حرية المساوية لأصغر العددين $(n_1 - 1)$ أو$(n_2 - 1)$ حسبما ورد في العلاقة (1−21)، فإذا كانت $|t| < t_{\frac{\propto}{2}}$ نقبل فرضية العدم ونعتبر أن الإجابات على ذلك الاستبيان صادقة باحتمال 0.95 . والعكس بالعكس .

2-4 كيفية حساب حجم العينة n من مجتمع طبيعي فيه N عنصراً :

2-4-1: تمهيد:

إن حجم العينة العشوائية n ، التي يجب سحبها من مجتمع الدراسة يرتبط بعدة أمور وهي:

1- نوع المعاينة: **المعاينة البسيطة- المعاينة الطبقية- المعاينة العنقودية ...الخ . ونظراً لتعقيدات الموضوع وتشعباته فإننا هنا سنقتصر على حساب حجم العينة n في المعاينة العشوائية البسيطة فقط .**

2-طريقة السحب: وهناك طريقتان للسحب مع الإعادة- السحب بدون إعادة .

3- نوع المؤشر المدروس: ونقصد به المؤشر ، الذي نريد تقديره من معلومات العينة ، وهو غالباً ما يكون أحد المؤشرين التاليين:

- متوسط مؤشر المجتمع μ والذي سنقدره من متوسط العينة $\bar{x}$ المعرف بالعلاقة: $\bar{x} = \frac{\sum x_i}{x}$.
- نسبة خاصة معينة في المجتمع R والتي سنقدرها من النسبة في العينة r والمعرفة بالعلاقة: $r = \frac{m}{n}$ ، حيث m :هو عدد المتصفين بالخاصة المدروسة في العينة .

4- درجة تجانس المجتمع: ويعبر عنه من خلال تباين المجتمع σ^2 ، فإذا كان التجانس ضعيفاً فإن التباين σ^2 يكون كبيراً، وعندها فإن حجم العينة n يجب أن يكون كبيراً ، أي أن حجم العينة n يتناسب طرداً مع التباين σ^2 . أما إذا كان تباين المجتمع σ^2 مجهولاً فيتم تقديره من خلال تباين العينة S^2 المعرف بالعلاقة : $S^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$

5-مقدار الدقة المطلوبة في التقدير: ويرمز له بـ d وهو الحد الأعلى للخطأ المسموح به عند تقدير المؤشرات: ويتم تحديده مسبقاً من قبل المسؤولين عن البحث وحسب نوع المؤشر المدروس.

فإذا كان المؤشر المطلوب تقديره هو متوسط المجتمع μ، فإننا نضع مقدار الدقة d مساوياً لأي عدد مطلق معقول، (فمثلاً إذا كنا نريد تقدير متوسط الدخل فيمكن أن نضع ل.س d = 200 أو 500 أو أي عدد آخر معقول) .

وهذا يعني أنه عند تقدير متوسط الدخل يجب أن يكون الخطأ المرتكب فيه (الخطأ المعياري) أقل أو يساوي المقدار المحدد للدقة d .

أما إذا كان المطلوب عند تقدير النسبة R في المجتمع، فإن الدقة تحدد على شكل نسبة مئوية مثل d=6% أو d=3% أو d=2% . وهذا يعني أنه عند تقدير النسبة R يجب أن يكون الخطأ المرتكب فيه (الخطأ المعياري) أقل أو يساوي النسبة المحددة للدقة d .

6- الخطأ المعياري للمؤشر المدروس: وهو يحسب لكل مؤشر على حدة ويبرهن في نظرية العينات والاحتمالات على ما يلي:

- إن الخطأ المعياري المرتكب في تقدير متوسط المجتمع μ من خلال متوسط العينة $\bar{x}$ والذي سنرمز له بـ $\sigma_{\bar{x}}$ يتأثر بطريقة السحب، وهو يساوي حسب حالتي السحب ما يلي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{S}{\sqrt{n}} = \tilde{\sigma}_{\bar{x}} \qquad \text{(في حالة السحب مع الإعادة)} \qquad (23-2)$$

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N}} * \frac{\sigma}{\sqrt{n}} \approx \sqrt{\frac{N-n}{N}} * \frac{s}{\sqrt{n}} = \tilde{\sigma}_{\bar{x}} \qquad \text{(في حالة السحب بدون الإعادة)} \qquad (24-2)$$

حيث S: هو الانحراف المعياري للعينة، ويعتبر تقديراً غير متحيز للانحراف المعياري في المجتمع .

- إن الخطأ المعياري المرتكب في تقدير النسبة R في المجتمع من خلال النسبة في العينة r والذي سنرمز له بـ σ_r يساوي حسب حالتي السحب ما يلي:

$$\sigma_r = \sqrt{\frac{R*(1-R)}{n}} \approx \sqrt{\frac{r*q}{n}} = \tilde{\sigma}_r \qquad \text{(في حالة السحب مع الإعادة)} \qquad (25-2)$$

$$\sigma_r = \sqrt{\frac{N-n}{N}} * \sqrt{\frac{R*(1-R)}{n}} \approx \sqrt{\frac{N-n}{N} * \frac{r*q}{n}} = \tilde{\sigma}_r \qquad \text{(في حالة السحب بدون الإعادة)} \qquad (26-2)$$

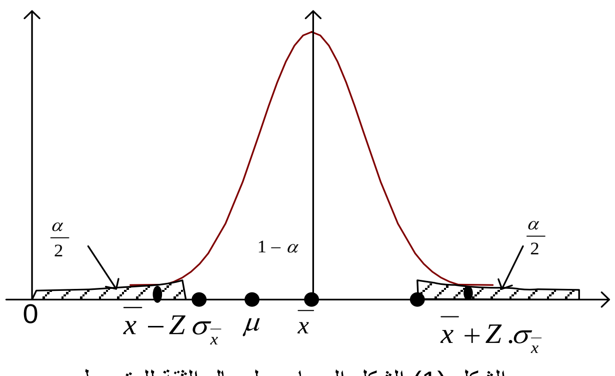
حيث r هي النسبة في العينة أو في أية عينة تجريبية و $q = 1 - r$

7-مستوى الدلالة المطلوب $\propto$: ويتم تحديده مسبقاً من قبل المسؤولين عن البحث وهو يعبر عن احتمال أن يكون التقدير المأخوذ من العينة مرفوضاً، أو أن يكون مقبولاً باحتمال $(1-\propto)$، ويسمى الاحتمال $(1-\propto)$ باحتمال الثقة في التقدير، لأنه يضمن لنا أن يكون المؤشر المطلوب تقديره واقعاً ضمن مجال محدد يسمى مجال الثقة باحتمال $(1-\propto)$. وعندما نحدد مستوى الدلالة $\propto$ نوزعه على طرفي المجال ونضع لكل طرف $\frac{\propto}{2}$. وعندها تتحدد معنا القيمتان العدديتان لمتحول التوزيع الطبيعي المعياري $Z_{\frac{\propto}{2}}$ و $Z_{1-\frac{\propto}{2}}$ الفاصلتان بين منطقتي القبول والرفض وبسبب التناظر يكون لدينا $Z_{1-\frac{\propto}{2}} = -Z_{\frac{\propto}{2}}$ لذلك نأخذ القيمة $Z_{1-\frac{\propto}{2}}$ ونقوم بإنشاء مجال الثقة المقابل لذلك . وتكون منطقة الرفض خارج مجال الثقة المذكور .

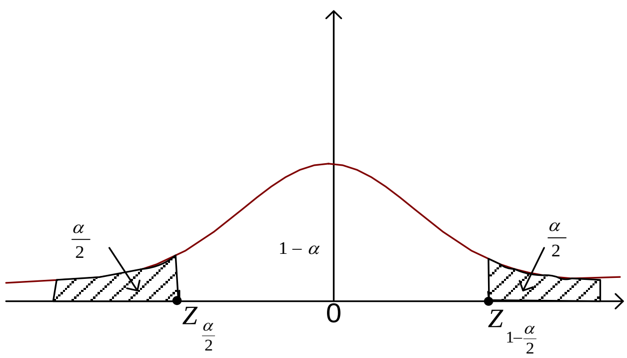
ولإنشاء مجال الثقة لمتوسط المجتمع μ نجعل مركزه متوسط العينة $\bar{x}$ ونصف طوله يساوي $(Z * \sigma_{\bar{x}})$ فنحصل على مجال الثقة الذي يحقق احتمال الثقة المطلوب وهو المجال التالي: [انظر (7-1) الفصل الأول]

$$P\left[\bar{x} - Z_{1-\frac{\propto}{2}} * \sigma_{\bar{x}} \le \mu \le \bar{x} + Z_{1-\frac{\propto}{2}} * \sigma_{\bar{x}}\right] = 1-\propto \qquad (27-2)$$

وهو يأخذ الشكل البياني التالي :



الشكل (1) الشكل المعياري لمجال الثقة للمتوسط



الشكل (2) الشكل العام لمجال الثقة

الشكل (2-1): مجال الثقة للمتوسط

ولإنشاء مجال الثقة للنسبة R في المجتمع نجعل مركزه النسبة في العينة r ونصف طوله $(Z * \sigma_r)$ فنحصل على مجال الثقة الذي يحقق احتمال الثقة $(1-\propto)$ التالي:

$$P = \left[r - Z_{1-\frac{\propto}{2}} * \sigma_r \le R \le r + Z_{1-\frac{\propto}{2}} * \sigma_r\right] = 1-\propto \qquad (28-2)$$

2-4-2: كيفية حساب حجم العينة (في حالة السحب مع الإعادة): لمتوسط المجتمع μ وللنسبة فيه R :

لحساب حجم العينة n في المعاينة العشوائية البسيطة علينا أن نراعي جميع الشروط المذكورة أعلاه ، وخاصة شرط الدقة d (الحد الأعلى للخطأ المسموح به عند التقدير)، وهنا علينا أن نحسب حجم العينة n لكل مؤشر على حدة (للمتوسط المجتمع μ وللنسبة فيه R).

1- إذا كان المطلوب تقدير متوسط المجتمع الطبيعي μ، وكان السحب مع الإعادة، وكان مقدار الدقة المطلوبة d، فإننا نجعل نصف طول مجال الثقة للمتوسط μ أصغر أو يساوي مقدار الدقة المحدد d ، فنحصل على ما يلي:

$$Z_{1-\frac{\alpha}{2}} * \sigma_r \le d$$
$$Z^2 * \sigma_{\bar{x}}^2 \le d^2 \qquad (29-2)$$

وبما أن $\left(\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\right)$ نجد أن :

$$Z^2 * \frac{\sigma^2}{n} \le d^2$$

ومنها نجد أن :

$$n \ge \frac{Z^2 * \sigma^2}{d^2} \approx \frac{Z^2 S^2}{d^2} \qquad \text{(وهو الحد الأدنى لحجم العينة للمتوسط مع الاعادة)} \qquad \mathbf{(2-30)}$$

حيث أن: s^2 هو تباين العينة أو أي عينة تجريبية عشوائية ، و يؤخذ كتقدير لتباين المجتمع σ^2 .

مثال (2-3): لتقدير متوسط الوزن للأفراد البالغين في مجتمع ما ، نريد سحب عينة منه مع الإعادة لتحقيق دقة $d = 5\ kg$ وباحتمال 95% . إذا علمت أنه تم تقدير تباين المجتمع σ^2 من خلال تباين عينة تجريبية فكان $S^2 = 1500$.

ولحساب حجم هذه العينة تطبق العلاقة (2-30) ونلاحظ أن مستوى الدلالة $\alpha = 0.05$ وأن $Z_{1-\frac{\alpha}{2}} = 1.96$ ومنها نجد أن :

$$n \ge \frac{(1.96)^2(1500)}{(5)^2} = 230.496 \approx 231$$

وهو حجم العينة الكافي لتحقيق الشروط المفروضة على المسألة من حيث الدقة واحتمال الثقة . علماً بأنه يجب تقريب حجم العينة إلى الأعلى دائماً .

ملاحظة : وبعد سحب هذه العينة وإجراء الحسابات المطلوبة نقوم بحساب تباينها الحقيقي S^2 ، ثم نعود ونحسب حجم العينة من جديد للتأكد من أن حجم العينة المسحوبة يحقق الشروط المفروضة على المسألة المدروسة . فإذا كان حجم العينة المسحوبة أقل من حجمها الجديد، فهذا يعني أنها لا تحقق الشروط المفروضة على المسألة لأسباب تتعلق بعشوائية السحب أو بصحة البيانات أو بغير ذلك، ولتصحيح ذلك نقوم بزيادة حجم العينة إلى الحجم المطلوب ونعيد الحسابات من جديد .

2- إذا كان المطلوب تقدير النسبة R في المجتمع، وكان السحب مع الإعادة، فإننا نجعل نصف طول مجال الثقة أصغر أو يساوي نسبة الدقة d والمحددة لنسبة مئوية صغيرة مثل (2% أو 3%...) كما يلي:

$$Z_{1-\frac{\alpha}{2}} * \sigma_r \leq d$$
$$Z^2 * \sigma_r^2 \leq d^2 \qquad (31-2)$$

وبما أن $\left(\sigma_r^2 = \frac{R(1-R)}{n}\right)$ نجد أن:

$$d^2 \geq Z^2 \frac{R(1-R)}{n} \approx Z^2 * \frac{r * q}{n}$$

$$n \geq \frac{Z^2 * r * q}{d^2} \quad \text{(وهذا هو الحد الأدنى لحجم العينة للنسبة مع الإعادة)} \qquad (32-2)$$

حيث أن: r هي النسبة في العينة أو في أي عينة تجريبية أو سابقة و أن $q = 1 - r$

مثال (2-4): لنفرض إننا نريد حساب حجم العينة n المطلوب سحبها مع الإعادة لتقدير نسبة المدخنين في مجتمع ما بدقة قدرها d= 0.4 وباحتمال ثقة 95% أي أن: $\left(Z_{1-\frac{\alpha}{2}} = 1.96\right)$

علماً بأن المعلومات السابقة تشير إلى أن نسبتهم r = 0.25 نطبق العلاقة (2-32) فنجد أن:

$$n \geq \frac{(1.96)^2(0.25)(0.75)}{(0.04)^2} = \frac{0.7203}{0.0016} = 450.19 \approx 451$$

وبعد سحب هذه العينة وحساب النسبة r فيها نعود ونحسب حجم العينة من جديد للتأكد من أن حجم العينة المسحوبة يحقق الشروط المفروضة على المسألة المدروسة .

2-4-3: كيفية حساب حجم العينة (في حالة السحب بدون إعادة): لمتوسط المجتمع μ ثم للنسبة فيه R :

1- إذا كان المطلوب تقدير متوسط المجتمع الطبيعي μ وكان السحب بدون إعادة، وكان مقدار الدقة المطلوبة d . فإننا نجعل نصف طول مجال الثقة أصغر أو يساوي مقدار الدقة d ، فنحصل على ما يلي:

$$Z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}} \leq d$$
$$Z^2 * \sigma_{\bar{x}}^2 \leq d^2 \qquad (33-2)$$

وبما السحب بدون إعادة فإن: $\sigma_{\bar{x}}^2 = \frac{N-n}{N} * \frac{\sigma^2}{n}$ وبذلك نجد أن:

$$Z^2 * \frac{N-n}{N} * \frac{\sigma^2}{n} \leq d^2$$

ومنها نجد أن :

$$nNd^2 \geq Z^2 * N\sigma^2 - Z^2 n\sigma^2$$
$$nNd^2 + Z^2 * n\sigma^2 \geq Z^2 * N\sigma^2$$
$$n(Nd^2 + Z^2\sigma^2) \geq Z^2 N\sigma^2$$

$$n \geq \frac{NZ^2\sigma^2}{Nd^2 + Z^2\sigma^2} \qquad (34-2)$$

وباستبدال σ^2 بتباين العينة s^2 أو لأي عينة تجريبية نحصل على ما يلي:

$$n \geq \frac{NZ^2S^2}{Nd^2 + Z^2S^2} \qquad (35-2)$$

وهو الحد الأدنى لحجم العينة للمتوسط(في حالة السحب بدون إعادة)

مثال (2-5): نريد تقدير متوسط دخل الأسرة في الشهر في مجتمع مؤلف من N= 1000 أسرة بدقة لا تتجاوز ل.س 500 = d وباحتمال ثقة 95% . والمطلوب حساب حجم العينة n اللازم لذلك، إذا علمت أن تباين الدخل في ذلك المجتمع ($\sigma^2 = 7562000$)، وإن السحب سيجري بدون إعادة ، لحساب حجم العينة n نلاحظ أن مستوى الدلالة 0.05 =∝ وأن $Z_{1-\frac{\propto}{2}} = 1.96$ وبتطبيق العلاقة (2-35) ، نحصل على أن:

$$n \geq \frac{1000(1.96)^2(7562000)}{1000(500)^\wedge 2 + (1.96)^2(7562000)} = \frac{29050179200}{279050179.2} = 104.10 \approx 105$$

وهو حجم العينة n الكافي لتحقيق شروط المسألة في الدقة واحتمال الثقة، مع ملاحظة أنه يجب تقريب حجم العينة إلى الأعلى دائماً .

ملاحظة: إذا كان تباين المجتمع σ^2 مجهولاً فإننا نعمل على إيجاد تقدير له من خلال سحب عينة تجريبية بحجم معقول ثم حساب تباينها S^2 ثم تعويضه في العلاقة (2-35)، فنحصل على الحجم الأدنى المطلوب. وأخيراً نشير إلى أنه بعد سحب العينة وإجراء الحسابات اللازمة من بياناتها، يجب أن نتأكد من إنها تحقق الشروط المفروضة على المسألة .

ملاحظة: من مقارنة العلاقتين (2-30) و (2-35) نجد أن حجم العينة n المحسوبة بنفس الشروط في حالة السحب بدون إعادة أصغر من حجمها n_0 في حالة السحب مع الإعادة، وهما يرتبطان بالعلاقة:

$$n = \frac{N * n_0}{N + n_0} \qquad (36-2)$$

تعليق: إن العلاقة (2-35) هي العلاقة الأساسية لحساب حجم العينة في حالة السحب بدون إعادة ولكن بعض المؤلفين والباحثين قام بتحويرها إلى أشكال أخرى . فقام أحدهم بتقسيم البسط والمقام على S^2 ، فحصل على أن:

$$n \geq \frac{NZ^2}{N\left(\frac{d^2}{S^2}\right) + Z^2} \approx \frac{NZ^2}{Ne^2 + Z^2} \quad : \left(e^2 = \frac{d^2}{S^2}\right) : \text{حيث أن} \qquad (37-2)$$

وقام بعضهم بتقسيم البسط والمقام على ($Z^2\sigma^2$) ، فحصل على أن :

$$n \geq \frac{N}{N\left(\frac{d^2}{Z^2S^2}\right) + 1} \approx \frac{N}{NE^2 + 1} \quad : \left(E^2 = \frac{d^2}{Z^2 * S^2}\right) : \text{حيث أن} \qquad (38-2)$$

وقام بعضهم بتقسيم البسط والمقام على $\bar{x}^2$ وادخال معامل الاختلاف CV في المعادلة (2−35) ، فحصل على أن:

$$n \geq \frac{NZ^2 \frac{S^2}{\bar{x}^2}}{N\left(\frac{d^2}{\bar{x}^2}\right) + Z^2 * \frac{S^2}{\bar{x}^2}} = \frac{NZ^2 * CV^2}{N\delta^2 + Z^2 * CV^2} \;;: \left(\text{الدقة النسبية } \delta = \frac{d}{\bar{x}} \text{ وأن } CV = \frac{S}{\bar{x}}\right) : \text{حيث أن} \quad \mathbf{(39-2)}$$

وقام بعضهم بتقسيم البسط والمقام على Nd^2 ، فحصل على العلاقة التالية:

$$n \geq \frac{Z^2\left(\frac{\sigma^2}{d^2}\right)}{1 + Z^2\left(\frac{\sigma^2}{d^2}\right)} \approx \frac{Z^2\left(\frac{S^2}{d^2}\right)}{1 + Z^2\left(\frac{S^2}{d^2}\right)} = \frac{Z^2 * C^2}{1 + Z^2C^2} \quad : \left(C = \frac{S}{d}\right) \qquad \mathbf{(40-2)}$$

وقام بعضهم بتقسيم البسط والمقام على حجم المجتمع N ، فحصل على أن:

$$n \geq \frac{Z^2\sigma^2}{d^2 + \frac{Z^2\sigma^2}{N}} \approx \frac{Z^2s^2}{d^2 + \frac{Z^2S^2}{N}} \qquad (41-2)$$

ومنها استنتج أنه عندما يكون حجم المجتمع N كبيراً فإن المقدار $\frac{Z^2S^2}{N}$ يصبح مهملاً وإن العلاقة السابقة تأخذ الشكل التالي :

$$n \geq \frac{Z^2S^2}{d^2} \qquad (42-2)$$

وهي نفس العلاقة (2−30) لحالة السحب مع الإعادة . ومنها يمكننا أن نستنتج مايلي:

نتيجة هامة: عندما يكون حجم المجتمع N كبيراً ، فإن حجم العينة n في حالة السحب بدون إعادة يصبح متقارباً مع حجمها في حالة السحب مع الإعادة، وعندها يمكننا حساب حجم العينة في كلتا حالتي السحب من العلاقة :

$$n \geq \frac{Z^2S^2}{d^2} \quad : (n << N) \qquad (43-2)$$

حيث s^2 هو تباين العينة أو أي عينة تجريبية .

علماً بأنه كان يمكننا الحصول على هذه النتيجة من قيمة الخطأ المعياري للمتوسط المعرف في العلاقة (2−24) . حيث نجد أنه عندما يكون N كبيراً فيكون $(n << N)$ ، وعندها فإن كسر معامل التصحيح يصبح قريباً من الواحد أي أن :

$$\frac{N-n}{N} \approx 1 \quad : (n << N) \qquad (45-2)$$

وعندها نجد أن الخطأ المعياري لحالة السحب بدون إعادة يصبح متقارباً مع الخطأ مع الإعادة أي أن :

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N}} * \frac{\sigma}{\sqrt{n}} \approx \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} \quad : (n << N) \qquad (46-2)$$

2-إذا كان المطلوب تقدير النسبة R في المجتمع وكان السحب بدون إعادة ، فإننا نجعل نصف طول مجال الثقة أصغر أو يساوي نسبة الدقة المحددة d فنجد أن :

بما أن $\sigma_r = \sqrt{\frac{N-n}{n}}\sqrt{\frac{R(1-R)}{n}}$ نجد أن: $Z_{1-\frac{\alpha}{2}} * \sigma_r \le d$

$$Z^2 * \frac{N-n}{N} * \frac{R(1-R)}{n} \le d^2 \qquad (47-2)$$

$$nNd^2 \ge NZ^2R(1-R) - nZ^2R(1-R)$$

$$n\big(Nd^2 + Z^2R(1-R)\big) \ge Nz^2R(1-R)$$

ومنها نجد أن:

$$n \ge \frac{NZ^2R(1-R)}{Nd^2 + Z^2R(1-R)} \qquad (48-2)$$

وباستبدال النسبة R بالنسبة r في العينة أو في أي عينة تجريبية نحصل على أن :

$$\text{n} \ge \frac{N * Z^2 * r * q}{Nd^2 + Z^2 * r * q} \qquad \text{حيث أن } q = 1 - r : \qquad (49-2)$$

وهو الحد الأدنى لحجم العينة للنسبة في حالة السحب بدون إعادة

وهي العلاقة الأساسية لحساب حجم العينة n عند تقدير النسبة R ، ويمكن تحويرها إلى أشكال أخرى كما رأينا في حالة المتوسط من خلال علاقات مشابهة للعلاقات السابقة. فمثلاً لو قسمنا البسط والمقام على Nd^2 نحصل على العلاقة الآتية :

$$n \ge \frac{(Z/d)^2 * r * q}{1 + (1/N)(Z/d)^2 * r * q} \qquad (50-2)$$

ملاحظة: عند حساب حجم العينة n غالباً ما يتم حساب النسبة r من أي عينة تجريبية ، وقد تكون قيمتها الحقيقية 20% أو 30% أو 40% أو غير ذلك ، وبعد تعويضها في العلاقة (2-49) نحصل على حجم العينة المطلوبة. ولكن عندما تكون النسبة متوازنة في المجتمع المدروس (كنسبة الإناث في المجتمع العام التي تأخذ حوالي R=0.50) فإن حجم العينة n المحسوبة من تلك العلاقة يبلغ أكبر قيمة له . وذلك لأن الجداء($R^*(1-R)$) يأخذ أكبر قيمة له عندما تكون (R = 0.50) ويساوي:

$$R^*(1-R^*) = (0.50)(0.50) = (0.25)$$

وبناءً على هذه الخاصة يقوم بعض الباحثين عند حساب حجم n من العلاقة (2-49) بافتراض أن النسبة متوازنة في المجتمع ويضعون فيها (r = 0.50) و (q= 0.50) ويعوضون قيم باقي المؤشرات d, N, Z, بقيمها ، فيحصلون على أكبر حجوم للعينات المقابلة لإحجام مختلفة لمجتمعات متوازنة (وذلك في حالة r= 0.50) . وبناءً على ذلك قام بعضهم بتطبيق العلاقة (2-49) بحساب أحجام العينات المقابلة لمستويات معينة من الدقة d وعند مستوى دلالة محدد ∝= 0.05 والمقابل لـ Z= 1.96 ، فحصل على الجدول التالي :

الجدول (1–2) أحجام العينات في حالة السحب بدون إعادة لتقدير النسبة R المقابلة لأحجام مختلفة من مجتمعات متوازنة وحسب قيم مختلفة للدقة ومقابل لـ $\propto = 0.05$ (أي أن قيمة $Z_{1-\frac{\propto}{2}} = 1.96$) .

حجم المجتمع N	نسبة الدقة =d هامش الخطأ في تقدير النسبة R			
	%5	%3	%2	%1
50	44	48	49	50
100	79	91	96	99
150	108	132	141	148
200	132	168	185	196
250	151	203	226	244
300	168	234	267	291
400	196	291	343	384
500	217	340	414	475
750	254	440	571	696
1000	278	516	706	906
2000	322	696	1091	1655
5000	357	879	1622	3288
10000	370	964	1936	4899
100000	383	1056	2345	8762
1000000	384	1066	2395	9513
10000000	384	1067	2400	9595

المصدر : sounders lewis & thornhill, 2009

مثال (5-2): لتوضيح كيفية حساب أحجام هذه العينات ، نفترض أننا نريد أن نسحب بدون إعادة عينة بحجم n من مجتمع طبيعي (كلية الاقتصاد) حجمه معلوم N = 1000 لتقدير نسبة الإناث فيه وبدقة محددة d=0.03 وباحتمال ثقة قدره 95% . (أي بمستوى دلالة $\propto = 0.05$ وهو يقابل قيمة حرجة تساوي $Z_{1-\frac{\propto}{2}} = 1.96$)، وللحصول على أكبر حجم للعينة المطلوبة نفترض أن هذا المجتمع متوازن جنسياً ونضع نسبة الإناث R=0.50 ونعوض في العلاقة الأخيرة (2–49) فنحصل على أن:

$$n \geq \frac{1000(1.96)^2(0.50)(0.50)}{1000(0.03)^2 + (1.96)^2(0.50)(0.50)} = \frac{960.4}{1.8604} = 516.23 \approx 517$$ (انظر الجدول) :

ولكن إذا علمنا من شؤون الطلاب أن نسبة الإناث في ذلك المجتمع R = 0.40 فإننا نكون في مجتمع غير متوازن من حيث الجنس . وعندها نقوم بتعويض هذه القيمة في العلاقة (2–49) فنحصل على أن حجم العينة المطلوب يساوي :

$$n \geq \frac{1000(1.96)^2(0.40)(0.60)}{1000(0.03)^2 + (1.96)^2(0.40)(0.60)} = \frac{921.984}{1.821984} = 506$$

وهو حجم العينة الكافي لتحقيق الشروط المفروضة على هذه المسألة .

ولكن إذا كنا نريد تقدير نسبة المدخنين في المجتمع المذكور وبنفس الشروط السابقة وكانت هذه النسبة مجهولة ، فإننا نكون في مجتمع غير متوازن من حيث خاصة التدخين، لذلك يجب أن نبحث عن تقدير أولي لنسبة المدخنين فيه، ولهذا نسحب عينة تجريبية بحجم معقول (وليكن n=100) ونستخلص منها تقدير لنسبة المدخنين الأولية في المجتمع ، ولنفترض إنها كانت تساوي r= 0.35 . وعندها نقوم بحساب حجم العينة من العلاقة السابقة (2–49) فنجد أن:

$$n \geq \frac{1000(1.96)^2(0.35)(0.65)}{1000(0.03)^2+(1.96)^2(0.35)(0.65)} = \frac{873.964}{1.77396} = 492.48 \approx 493$$

وهو حجم العينة الكافي لتحقيق شروط المسألة وهي أن يكون احتمال الثقة أكبر من 95% وأن تكون نسبة الخطأ المعياري (المرتكب) في تقدير النسبة لا تزيد عن مقدار الدقة المحددة بـ 3%. وكل ذلك بفرض أن جميع عناصر هذه العينة ستستجيب للمعاينة وتعطينا البيانات الصحيحة عن أحوالها .

ملاحظة هامة: أما إذا كانت الاستجابة على الاستبيان غير كاملة فيجب أن نزيد الحجم، وذلك بتقسيم حجم العينة المحسوبة n على نسبة الاستجابة R المقدرة، فنحصل على حجم العينة $\grave{n}$ اللازم للمعاينة الميدانية من العلاقة $\left(\grave{n} = \frac{n}{R}\right)$.

وفي جميع الأحوال يجب على الباحث أن يقوم بإعادة حساب حجم العينة $\grave{n}$ اللازم لتحقيق شروط المعاينة اعتماداً على بيانات العينة المستخدمة في البحث نفسها. وذلك للتأكد من حجمها ومن صلاحيتها ومن أنها تحقق الشروط المفروضة على البحث.

ملاحظة: لحساب حجم العينة الكلية في المعاينة الطبقية يمكننا استخدام نفس العلاقات وتطبيقها على المجتمع الكلي، فنحصل على حجم العينة الكلية n ، ثم نقوم بتوزيعها على الطبقات حسب أحجامها النسبية من المجتمع . كما يمكن العمل بالعكس واعتبار كل طبقة مجتمعاً مستقلاً وحساب أحجام العينات الطبقية، وذلك بتطبيق نفس العلاقات السابقة على كل طبقة ، ثم دمج هذه العينات الطبقية فنحصل على العينة الكلية . كما يمكن أخذ التكلفة بعين الاعتبار ...الخ . وهناك أمور كثيرة أخرى لا مجال لبحثها في هذا المنشور .

الفصل الثالث

الاستدلال حول طبيعية المتحولات العشوائية

3-1: مقدمة عن التوزيعات الاحتمالية:

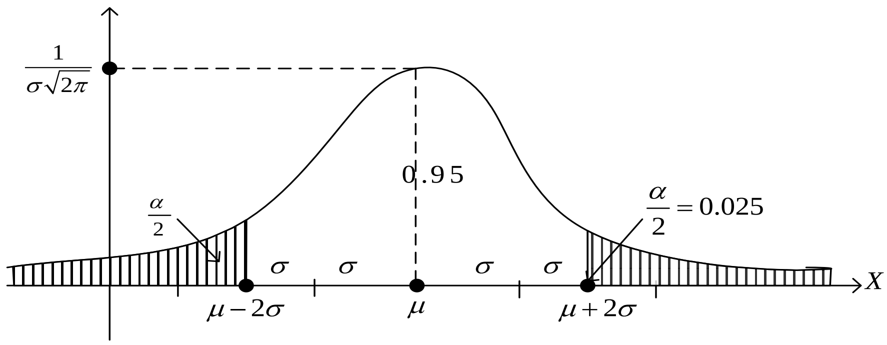
يتناول هذا الفصل قضايا الاستدلال الاحصائي لعدة متحولات . ويدرس كيفية إجراء الاختبارات بين متوسطاتها . ولكن قبل الدخول في الموضوع نستعرض أهم التوزيعات الاحتمالية المستخدمة في هذه الاختبارات وهي:

3-1-1: التوزيع الطبيعي العام : (General Normal distribution)

وهو أهم التوزيعات الاحتمالية المستخدمة في الاحصاء والاحتمالات، ويعبّر هذا التوزيع عن التوزيع الاحتمالي لمتحول طبيعي X توقعه μ وتباينه σ^2 ، ويرمز له بـ $N(\mu\,,\sigma^2)$ ويعطى بالعلاقة التالية:

$$\ell(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad : -\infty < x < +\infty \qquad (1-3)$$

وهو يرسم منحني متناظر حول التوقع μ ويأخذ الشكل التالي:



الشكل (3-1): منحني التوزيع الطبيعي العام

ويتمتع هذا التوزيع بعدة خواص تميزه عن غيره من التوزيعات وهي:

1- إنه يعرف على متحول مستمر X يأخذ قيمه من $-\infty$ حتى $+\infty$ ، وإن توقعه μ وتباينه σ^2.

2- إنه يرسم منحني متناظر بالنسبة للتوقع μ فوق المحور OX .

3- إنه يأخذ أكبر قيمة له عندما $X = \mu$ وتساوي $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$.

فإذا ازدادت قيمة σ فإن قيمته العظمى تنخفض، وإذا نقصت قيمة σ فإن قيمته العظمى ترتفع .

4- إن طرفاه الأيمن والأيسر ينتهيان إلى الصفر، أي أنه عندما $x \to \pm\infty$ فإن $f(x) \to 0$.

5- إن قيمة المنوال Mod تساوي قيمة الوسيط Me . وتساويان معاً قيمة التوقع μ أي أن:

$$\mu = Mod = Me \qquad (2-3)$$

6- إن قيمة معامل (الالتواء) تساوي الصفر $K = 0$

كما إن قيمة معامل التطاول المعدل (التفرطح) تساوي $\ell = L - 3 = 0$

7-إن قيمة معامل الاختلاف v تساوي:

$$v = \frac{\sigma}{\mu} \leq 0.33 \qquad (3-3)$$

8-إن الاحتمالات المقابلة لمجالات الثقة المختلفة تساوي المساحات المقابلة لها والواقعة تحت المنحني $f(x)$ وهي كما يلي :

- إن حوالي 66% من الاحتمالات تقع مقابل المجال $\mu \pm \sigma$.
- إن حوالي 90% من الاحتمالات تقع مقابل المجال $\mu \pm 1.64\,\sigma$.
- إن حوالي 95% من الاحتمالات تقع المجال $\mu \pm 2\,\sigma$.
- إن حوالي 99.72% من الاحتمالات تقع مقابل المجال $\mu \pm 3\,\sigma$.

ويستفاد من هذه الخواص عند اختبار طبيعية أي متحول عشوائي (Normaity) وتحديد التوزيع الاحتمالي التجريبي له قبل استخدامه في التحليل الاحصائي .

3-1-2: التوزيع الطبيعي المعياري: (Standart Normal Distribution) :

وهو حالة خاصة من التوزيع الطبيعي العام، يكون فيها $\mu = 0$ و $\sigma^2 = 1$، لذلك يرمز له بالرمز $N(0,1)$ ، ويخضع لهذا التوزيع جميع المتحولات المعيارية المحولة من المتحول الطبيعي X إلى المتحول Z وفق العلاقة :

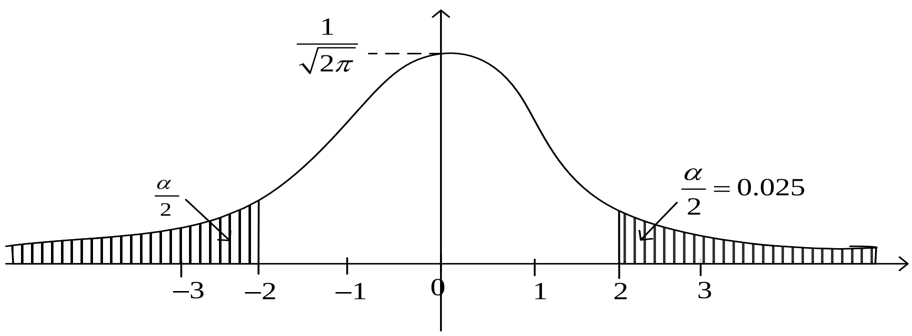
$$Z = \frac{X - \mu}{\sigma} \qquad (4-3)$$

وهذه العلاقة تعطينا أن: $E(Z) = 0$ ، وأن: $Var(Z) = 1$

وهو يعطى بالعلاقة التالية:

$$\varphi(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \quad : -\infty < Z < +\infty \qquad (5-3)$$

ويرسم المنحني المتناظر حول $(Z = 0)$ الشكل التالي:



الشكل (3-2) التوزيع الطبيعي المعياري

ويستخدم هذا التوزيع لحساب جميع الاحتمالات الطبيعية بعد تحويلها من التوزيع الطبيعي العام إلى التوزيع المعياري وفق العلاقة (3-4) السابقة .

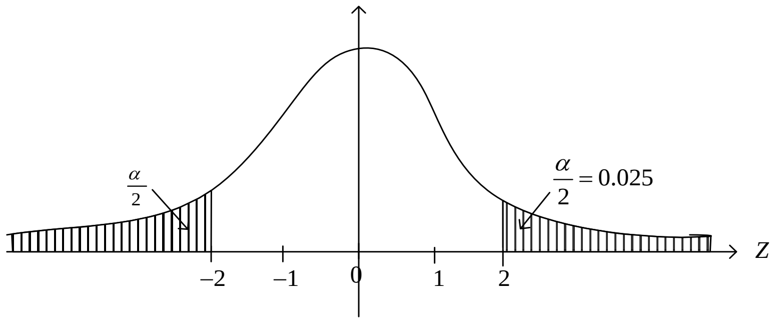
3-1-3 : توزيع (ستودينت) T:

وهو توزيع شبيه بالتوزيع الطبيعي المعياري، ولكنه يستخدم للعينات الصغيرة $(n < 30)$. ولبعض المتحولات الخاصة الأخرى . ويرمز لمتحوله بالرمز t وإن توقعه $\mu = 0$ وتباينه $\sigma^2(t) = \frac{v}{v-2}$ ، وهو يعطى بالعلاقة التالية:

$$T(t) = C * \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} \qquad : -\infty < t < +\infty \qquad (6-3)$$

حيث أن: v هو عدد صحيح موجب يسمى بدرجة الحرية و $(v > 2)$ وهي تساوي عدد المشاهدات المستقلة .

أما C فهو عدد موجب مرتبط بدرجة الحرية v، ويتم تحديده بحيث تكون المساحة تحت المنحني مساوية للواحد . والتوزيع يرسم الشكل التالي:



الشكل (3-3): منحني توزيع (ستودينت)

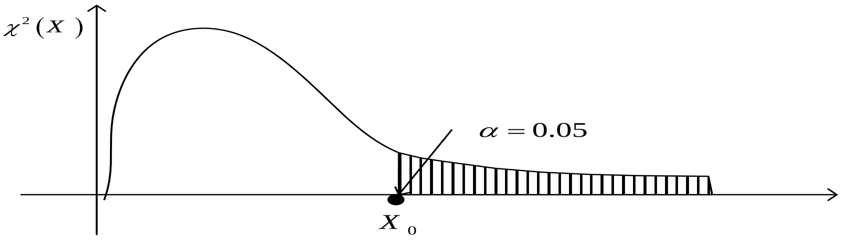
4-1-3 : توزيع كاي مربع χ^2 :

وهو توزيع غير متناظر ويعرف للمتحولات الموجبة X، وإن توقعه يساوي $\mu = v$ وتباينه $\sigma^2(X) = 2v$ ويعطى بالعلاقة التالية:

$$\chi^2(X) = C * X^{\frac{v}{2}-1} * e^{-\frac{x}{2}} \qquad : \quad x \geq 0 \qquad (7-3)$$

حيث أن: C عدد موجب مرتبط بدرجة الحرية v ليجعل المساحة تحت المنحني مساوية للواحد .

وهو يرسم المنحني التالي:



الشكل (3-4): منحني توزيع χ^2

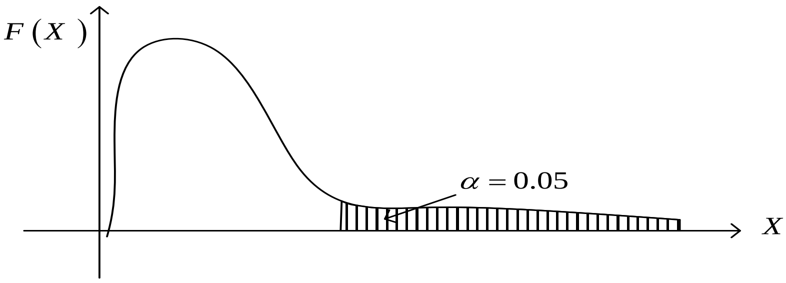
3-1-5 : توزيع (فيشر) F :

وهو توزيع للنسبة بين تبايني عينتين $\left(F = \frac{S_1^2}{S_2^2}\right)$ ، وهو توزيع غير متناظر ويعرف بالعلاقة التالية:

$$F(X) = C * \frac{X^{\frac{v}{2}-1}}{\left(1 + \frac{v}{w}X\right)^{\frac{v+w}{z}}} \qquad : \qquad X \geq 0 \qquad\qquad (8-3)$$

حيث أن: C هو عدد موجب مرتبط بدرجتي الحرية v و w، ليجعل المساحة تحت المنحني مساوية للواحد . علماً بأن توقعه الرياضي $\mu = \frac{w}{w-2}$ وتباينه $\sigma^2(X) = \frac{2w^2(v+w-2)}{v(w-2)^2(w-4)}$

وهو يرسم المنحني التالي:



الشكل (3-5): منحني توزيع فيشر F

3-1-6 : التوزيع الطبيعي المتعدد (لعدة متحولات طبيعية) :

لاستخراج العلاقة الرياضية لهذا التوزيع الطبيعي نعود إلى العلاقة (3-1) . ومنها نلاحظ أنه يمكننا كتابة الأس فيها على الشكل التالي :

$$\left(\frac{X-\mu}{\sigma}\right)^2 = (X-\mu) * \sigma^{-1} * (X-\mu) \qquad\qquad (9-3)$$

وعندما يكون لدينا عدة متحولات طبيعية $X_1\ X_2\ X_3 \ldots X_P$ فإننا نضعها على شكل شعاع عمود ثم نحسب شعاع توقعاتها كما يلي :

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_P \end{bmatrix} \quad \boldsymbol{\mu} = E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \\ \vdots \\ E(X_P) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_P \end{bmatrix} \qquad\qquad (10-3)$$

ثم نقوم بحساب مصفوفة التباين المشترك لها V من العلاقة :

$$\boldsymbol{V} = COV(\boldsymbol{X}) = (\boldsymbol{X}-\boldsymbol{\mu})'(\boldsymbol{X}-\boldsymbol{\mu}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \ldots & \sigma_{1P} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \ldots & \sigma_{2P} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \ldots & \sigma_{3P} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \sigma_{P1} & \sigma_{P2} & \sigma_{P3} & \ldots & \sigma_{PP} \end{bmatrix} \qquad (11-3)$$

ثم نحسب محددها $|V|$ ومقلوبها V^{-1} (بفرض أنها نظامية)، وبذلك يمكننا أن نصيغ قانون التوزيع الطبيعي المشترك لهذه المتحولات كما يلي:

$$f(X_1\ X_2\ X_3\ \dots\ X_P) = \frac{1}{(2\pi)^{\frac{P}{2}}|V|^{\frac{1}{2}}}\,\bar{e}^{\frac{1}{2}[(\boldsymbol{X}-\boldsymbol{\mu})'*\boldsymbol{V}^{-1}*(\boldsymbol{X}-\boldsymbol{\mu})]} \qquad (12-3)$$

حيث أن الأس فيها يساوي الجداء الشعاعي المصفوفي التالي:

$$[(\boldsymbol{X}-\boldsymbol{\mu})'*\boldsymbol{V}^{-1}*(\boldsymbol{X}-\boldsymbol{\mu})] = [(\boldsymbol{X_1}-\boldsymbol{\mu_1}),(\boldsymbol{X_2}-\boldsymbol{\mu_2})\dots(\boldsymbol{X_P}-\boldsymbol{\mu_P})]\begin{bmatrix}\boldsymbol{\sigma_{11}} & \boldsymbol{\sigma_{12}} & \dots & \boldsymbol{\sigma_{1P}}\\ \boldsymbol{\sigma_{21}} & \boldsymbol{\sigma_{22}} & \dots & \boldsymbol{\sigma_{2P}}\\ \vdots & \vdots & \dots & \vdots\\ \boldsymbol{\sigma_{P1}} & \boldsymbol{\sigma_{P2}} & \dots & \boldsymbol{\sigma_{PP}}\end{bmatrix}^{-1} * \begin{bmatrix}(\boldsymbol{X_1}-\boldsymbol{\mu_1})\\(\boldsymbol{X_2}-\boldsymbol{\mu_2})\\ \vdots\\(\boldsymbol{X_P}-\boldsymbol{\mu_P})\end{bmatrix} \qquad (\boldsymbol{13-3})$$

ملاحظة :عندما يكون $\boldsymbol{\mu}$ و $\mathbf{V}$ مجهولين ، فإننا نقوم بتقدير شعاع التوقع μ بواسطة شعاع المتوسطات $\bar{X}$ المحسوب من بيانات العينة ذات الحجم n كما يلي:

$$\widetilde{\boldsymbol{\mu}} = \bar{\boldsymbol{X}} = \begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \bar{x}_3\\ \vdots\\ \bar{x}_P\end{bmatrix} \qquad : \ \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \qquad (14-3)$$

ونقوم بتقدير مصفوفة التباين المشترك V بواسطة مصفوفة التباين المشترك S المحسوبة من بيانات العينة ذات الحجم n ، كما يلي:

$$\widetilde{\boldsymbol{V}} = \boldsymbol{S} = \begin{bmatrix}S_{11} & S_{12} & S_{13} & \dots & S_{1P}\\ S_{21} & S_{22} & S_{23} & \dots & S_{2P}\\ S_{31} & S_{32} & S_{33} & \dots & S_{3P}\\ \vdots & \vdots & \vdots & \dots & \vdots\\ S_{P1} & S_{P2} & S_{P3} & \dots & S_{PP}\end{bmatrix} : \quad S_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k) \qquad (15-3)$$

3-1-7 : التوزيع المشترك لـ $\bar{X}$ و S :

بما أن $\bar{X}$ و S متحولان عشوائيان (لأنهما يتغيران من عينة لأخرى). فإن توزيعهما المشترك يتعلق بتوزيع كل منهما . لذلك نورد النظرية التالية:

- **نظرية:** إذا كانت $X_1\ X_2\ X_3\ \dots\ X_P$ متحولات عشوائية طبيعية، وكان توقعها هو الشعاع $\boldsymbol{\mu}$ وتباينها المشترك هو المصفوفة **V**. وإذا سحبنا عينة عشوائية من ذلك المجتمع بحجم n(حيث $n > P$). فإنه يكون لدينا ما يلي:

1- إن المتوسط $\bar{X}$ يخضع للتوزيع الطبيعي المتعدد الذي توقعه $\boldsymbol{\mu}$ وتباينه المشترك ($\frac{\boldsymbol{1}}{\boldsymbol{n}}\boldsymbol{V}$) ونكتب ذلك كما يلي: $\bar{X}\sim N(\boldsymbol{\mu}\ ,\ \frac{\boldsymbol{1}}{\boldsymbol{n}}\boldsymbol{V})$.

2- إن المتحولين العشوائيين، الشعاع $\bar{\boldsymbol{X}}$ والمصفوفة **S** مستقلان .

3- إن المصفوفة: $(n-1)\boldsymbol{S}$ تخضع لتوزيع جديد يسمى بتوزيع (ويشارت Wishart) للمصفوفات العشوائية، وبدرجة حرية (n-1) .

وإن أهم خواص توزيع (ويشارت) هي: (انظر Johnston, Wichern p. 143)

1- إذا كانت A_1 مصفوفة خاضعة لتوزيع (ويشارت)$W_{m_1}(\boldsymbol{A_1/V})$ بدرجة حرية m_1، وكانت مستقلة عن المصفوفة A_2 الخاضعة لتوزيع (ويشارت) $W_{m_2}(\boldsymbol{A_2/V})$، فإن مجموعهما $(\boldsymbol{A_1+A_2})$ يكون خاضعاً لتوزيع (ويشارت) $W_{m_1+m_2}(\boldsymbol{A_1+A_2/V})$.

2- إذا كانت $\mathbf{A}$ مصفوفة خاضعة لتوزيع (ويشارت) $W_m(\boldsymbol{A/V})$، فإن التركيب $\boldsymbol{CAC'}$ يكون خاضعاً لتوزيع (ويشارت) التالي: $W_m[\boldsymbol{CAC'/CVC'}]$.

- **نظرية (قانون الأعداد الكبيرة):** لنفترض أنه لدينا متحولاً عشوائياً Y، توقعه الرياضي μ وتباينه σ^2. وإذا أخذنا n مشاهدة مستقلة عنه هي: $Y: Y_1\ Y_2\ Y_3\ \dots\ Y_{\boldsymbol{n}}$, ومسحوبة عشوائياً من مجتمع ما .

فإن متوسط هذه المشاهدات $\bar{Y}$ المحسوب من العلاقة:

$$\bar{Y}=\frac{Y_1+Y_2+Y_3+\cdots+Y_{\boldsymbol{n}}}{n} \qquad (16-3)$$

يقترب احتمالياً من التوقع الحقيقي $\boldsymbol{\mu}$ كلما ازدادت n (بدون حدود) .

وإن تباين هذه المشاهدات S^2 المحسوب من العلاقة:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 \qquad (17-3)$$

يقترب احتمالياً من التباين الحقيقي σ^2 ، كلما ازدادت n (بدون حدود) .

- **نظرية النهاية المركزية:** لنفترض أنه لدينا مصفوفة المشاهدات المستقلة لجملة من المتحولات $X_1\ X_2\ \dots\ X_P$ والمسحوبة من مجتمع ما، توقعها الشعاع $\boldsymbol{\mu}$ وتباينها المصفوفة المحددة $\mathbf{V}$. فعندها يكون المتحول المركب $\sqrt{n}(\boldsymbol{\bar{X}}-\boldsymbol{\mu})$ خاضعاً للتوزيع الطبيعي المتعدد $N(0,\boldsymbol{V})$. وذلك من أجل القيم الكبيرة لحجم العينة n (ويشترط أن يكون $(n-P)$ كبيراً) . ويمكن أن نستخلص منها أن المتحول المركب $\sqrt{n}(\boldsymbol{\bar{X}}-\boldsymbol{\mu})$ يخضع تقاربياً للتوزيع الطبيعي المتعدد $N(0,S)$ ، حيث أن $\mathbf{S}$ هي مصفوفة التباين المشترك لتلك المشاهدات، لأن $\boldsymbol{\tilde{V}}=\boldsymbol{S}$. كما يمكننا أن نستخلص أن الجداء التالي:

$$n(\boldsymbol{\bar{X}}-\boldsymbol{\mu})'*\boldsymbol{S^{-1}}*(\boldsymbol{\bar{X}}-\boldsymbol{\mu}) \qquad (18-3)$$

يخضع تقاربياً لتوزيع χ_P^2 بدرجة حرية P . وذلك عندما يصبح $(n-P)$ كبيراً .

- **نظرية التراكيب الخطية:** إذا كان الشعاع X خاضعاً للتوزيع الطبيعي $N(\boldsymbol{\mu},\boldsymbol{V})$ ، فإن أي تركيب خطي لمتحولاته مثل: $\boldsymbol{a'X}=a_1X_1+a_2X_2+a_3X_3\dots a_P\ X_P$ يخضع للتوزيع الطبيعي $N(\boldsymbol{a'\mu},\boldsymbol{a'Va})$. وبالعكس فإنه إذا كان التركيب $\boldsymbol{a'X}$ يخضع للتوزيع الطبيعي $N(\boldsymbol{a'\mu},\boldsymbol{a'Va})$ من أجل كل الأعداد a ، فإن الشعاع X يكون خاضعاً للتوزيع الطبيعي $N(\boldsymbol{\mu},\boldsymbol{V})$.

3-2: مراحل معالجة البيانات التجريبية:

قبل إجراء أي استدلال إحصائي من البيانات التجريبية المتوفرة يجب معالجتها وتشذيبها وتبويبها، وذلك لاستخلاص أهم خصائصها . وتهدف عملية المعالجة إلى تحقيق الأهداف التالية:

- التعرف على خصائص البيانات وتحديد قيمها المميزة: كالمتوسط والمنوال والوسيط، والتباين والانحراف المعياري، ومدى الانتشار وكذلك معاملات الالتواء والتطاول (التفرطح) والاختلاف . وكذلك حساب الاحتمالات الأولية المقابلة لمجالات الثقة الأساسية ...الخ .
- تحديد القيم الشاذة (المتطرفة يميناً أو يساراً) والقيام بحذفها (إن وجدت) من تلك البيانات . وذلك من أجل زيادة فعالية البيانات عند استخدامها في قضايا الاستدلال الاحصائي، علماً بأن هذه القيم الشاذة قد تحدث نتيجة خطأ في القياس أو التسجيل أو النقل أو الادخال، أو قد تكون بسبب عدم تجانس المجتمع المدروس. وإن وجود هذه القيم ضمن بيانات العينة سيؤثر سلباً على تقديرات معالم المجتمع، وبالتالي على نتائج الاختبارات الاحصائية . ويكفي أن نذكر هنا أن وجود (10%) من القيم الشاذة في بيانات العينة يؤدي إلى تضخم تقدير تباين المجتمع σ^2 بمقدار الضعفين (أي 200%) .
- القيام بتبويب البيانات المتبقية (بعد حذف القيم الشاذة) تبويباً تصاعدياً ضمن مجالات متواصلة وغير متقاطعة وحساب تكراراتها النسبية، ثم العمل على اختبارها فيما إذا كانت متوافقة مع التوزيع الطبيعي أم لا، فإذا كانت تتوافق مع التوزيع الطبيعي نتابع التحليل الاحصائي المطلوب . أما إذا كانت لا تتوافق مع التوزيع الطبيعي فإننا نحاول إجراء تحويل عليها لنجعلها خاضعة تقاربياً للتوزيع الطبيعي ، كما سنرى لاحقاً .

وهكذا نجد أن معالجة البيانات التجريبية تمر بعدة مراحل نلخصها بما يلي :

3-2-1 : مرحلة حساب الخصائص الوصفية :

لنفترض إننا سحبنا عينة عشوائية من المجتمع المدروس بحجم n عنصراً لدراسة خصائص متحول X، ولنفترض أيضاً أن هذه العينة أعطتنا البيانات االخام عن X ، التالية :

$$X: x_1\ x_2\ x_3\ \dots\dots\ x_i\ \dots\dots\ x_n \qquad (19-3)$$

لحساب المؤشرات الوصفية لـ X من بيانات هذه العينة االخام (أي من البيانات قبل ترتيبها أو تبويبها)، نقوم بحساب المؤشرات التالية:

- المتوسط الحسابي: $$\bar{x} = \frac{1}{n}\sum^{n} x_i$$
- التباين العادي = العزم المركزي الثانى): $$S^2 = \frac{1}{n}\sum^{n}(x_i - \bar{x})^2 = m_2$$ (20-3)
- التباين المصحح : $$\bar{S}^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

- الانحراف المعياري العادي : $S = \sqrt{S^2}$
- الانحراف المعياري المصحح : $\bar{S} = \sqrt{\bar{S}^2}$
- العزم المركزي الثالث : $M_3 = \frac{1}{n}\sum^{n}(x_i - \bar{x})^3$
- العزم المركزي الرابع : $M_4 = \frac{1}{n}\sum^{n}(x_i - \bar{x})^4$
- معامل الالتواء : $K = \frac{M_3}{\bar{S}^3}$
- معامل التطاول (التفرطح) العام : $L = \frac{M_4}{\bar{S}^4}$
- معامل التطاول المعدل : $\ell = L - 3$
- معامل الاختلاف : $v = \frac{\bar{S}}{\bar{x}}100$
- أكبر قيمة (Max (x ، وأصغر قيمة (Min (x
- تراكم النسب المئوية لتوزيع قيم X ضمن مجال تحول X
- حساب أي مؤشرات أخرى يمكن أن يطلبها الباحث

وبعد دراسة هذه المؤشرات الوصفية والتأكد من سلامتها . نركز اهتمامنا على معاملي الالتواء K والتطاول المعدل ℓ ، وعلى معامل الاختلاف v .

فإذا كانت قيمتا معاملي الالتواء K والتطاول ℓ قريبتان من الصفر، فهذا يعني أن توزيع البيانات التجريبية يأخذ شكلاً متناظراً حول المتوسط $\bar{x}$. وهذا يشير (ولا يبرهن) إلى إمكانية أن يكون توزيع بيانات هذه العينة متوافقاً مع التوزيع الطبيعي العام . أما إذا كانت قيمتاهما تختلفان عن الصفر، فإن ذلك يدل على أن توزيع بيانات X غير طبيعي .

ثم ننتقل إلى دراسة قيمة معامل الاختلاف v فإذا كانت قيمته $v \leq 0{,}33$ ، فإننا نستدل على أن توزيع بيانات هذه العينة يمكن أن يكون طبيعياً . أما إذا كانت قيمة $v > 0{,}33$ فإننا نستخلص مباشرة أن توزيع بيانات هذه العينة غير طبيعي . وفي هذه الحالة $(v > 0{,}33)$ علينا أن نقوم بإجراء تحويل تلك البيانات إلى بيانات جديدة شبه طبيعية . وذلك باستخدام إحدى العلاقات التالية وحسب الحالات المقابلة لها :

$$(21-3) \quad \text{إذا كان الالتواء إلى اليمين} \quad Y = X^{1,5} \quad \text{أو} \quad Y = x^2$$

$$(22-3) \quad \text{إذا كان الالتواء إلى اليسار} \quad Y = \sqrt{X} \quad \text{أو} \quad Y = ln\, X$$

$$(23-3) \quad \text{إذا كان الالتواء إلى اليسار} \quad Y = \frac{1}{\sqrt{X}} \quad \text{أو} \quad Y = \frac{1}{X}$$

وبعد حساب القيم الجديدة لـ Y من إحدى هذه العلاقات، نقوم بحساب المؤشرات الوصفية للمتحول الجديد Y . وثم نقوم بدراستها من جديد للاستدلال على شكل التوزيع الاحتمالي لـ Y . وإذا لم ننجح في عملية التحويل الأولى، نقوم بإجراء التحويل باستخدام علاقة أخرى من العلاقات السابقة وحسب الحالة المرافقة .

لقد قدم (Box – Cox) علاقة عامة لإجراء التحويلات الأسية وهي المعرفة في [Johnson Wichern p. 157]:كما يلي:

$$Y(\lambda)=\begin{cases}\dfrac{x^{\lambda}-1}{\lambda} & : \quad \lambda\neq 0\\ \ln x & : \quad \lambda=0\end{cases} \qquad : \; x>0 \qquad (24-3)$$

حيث أن λ : هو عبارة عن عدد حقيقي (موجب أو سالب)، وتحدد قيمته من بيانات العينة للمتحول X، وبحيث يتم تعظيم قيمة المقدار التالي:

$$\ell(\lambda)=-\frac{n}{2}\ln\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i(\lambda)-\bar{Y}(\lambda)\right)^2\right]+(\lambda-1)\sum_{i=1}^{n}\ln x_i \qquad (25-3)$$

وحيث أن $\bar{Y}(\lambda)$ هو المتوسط الحسابي للبيانات المحولة $Y(\lambda)$ عند القيمة λ .

ويحسب من العلاقة:

$$\bar{Y}(\lambda)=\frac{1}{n}\sum_{i=1}^{n}Y_i(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i^{\lambda}-1}{\lambda}\right) \qquad (26-3)$$

لذلك يتم حساب قيم المقدار $\ell(\lambda)$ المقابلة لعدة قيم لـ λ بواسطة الحاسوب ثم رسم المخطط البياني للنقاط $[\lambda,\ell(\lambda)]$ ، ومنه نحدد أكبر قيمة للمقدار $\ell(\lambda)$ ، ثم نستخلص قيمة λ المقابلة لتلك القيمة العظمى (ولو بشكل تقريبي) .

فإذا كانت قريبة من $\lambda=0$ فإننا نستخدم التحويل $Y=\ln X$

وإذا كانت قريبة من $\lambda=\frac{1}{2}$ فإننا نستخدم الجذر $Y=\frac{\sqrt{X}-1}{\frac{1}{2}}\approx\sqrt{X}$

وإذا كانت قريبة من $\lambda=-1$ فإننا نستخدم المقلوب $Y=\frac{X^{-1}-1}{-1}=-\frac{1}{X}+1\approx\frac{1}{X}$

وهكذا دواليك ...

وعندما يكون لدينا عدة متحولات $X_1\,X_2\,X_3\,...\,X_P$ ، فإننا سنحصل عند تحويلها بواسطة العلاقة (24 –3) على P قيمة لـ λ (مقابل كل منها) ، ونرمز لتلك القيم بـ $\lambda_1\,\lambda_2\,\lambda_3\,...\,\lambda_P$ ، والتي كل منها تعظم المقدار $\ell_K(\lambda)$ المقابل لـ X_K

كما يمكن تحديد جملة القيم $(\lambda_1\,\lambda_2\,\lambda_3\,...\,\lambda_P)$ دفعة واحدة، بحيث يتم تعظيم التابع المتعدد التالي:

$$\ell(\lambda_1\,\lambda_2\,\lambda_3\,...\,\lambda_P)=-\frac{n}{2}\ln|S(A)|+(\lambda_1-1)\sum_{i=1}^{n}\ln X_{1i}+(\lambda_2-1)\sum\ln X_{2i}+\cdots.(\lambda_P-1)\sum_{i=1}^{n}\ln X_{pi} \qquad (27-3)$$

حيث ان $S(A)$ هي مصفوفة التباين المشترك للمتحولات الجديدة $Y_1\,Y_2\,...\,Y_P$ و $|S(A)|$ قيمة محددها.

3-2-2 : مرحلة استبعاد القيم الشاذة من البيانات :

يوجد في المراجع العلمية عدة طرائق لاستبعاد القيم الشاذة، ولكننا هنا سنقتصر على أبسطها وهي :

- **الطريقة البيانية:**

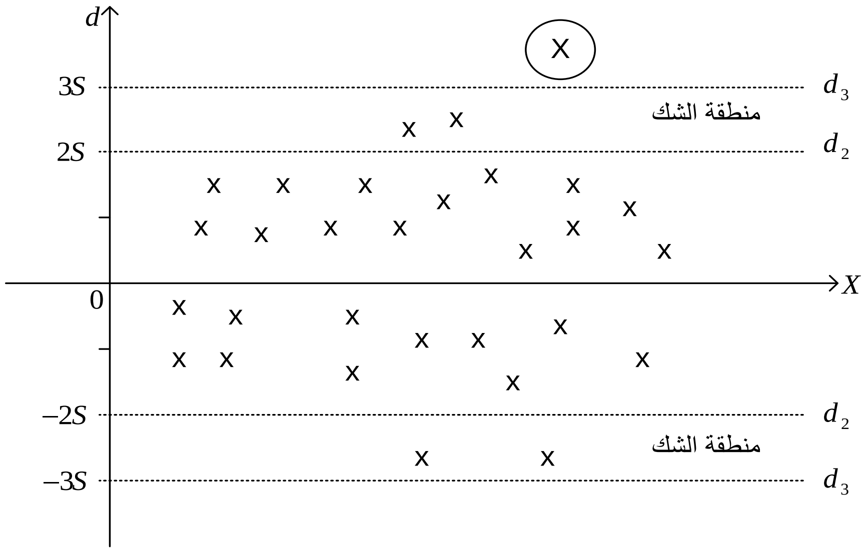
نقوم بحساب متوسط القيم $\bar{x}$ وانحرافها S، ثم نقوم بحساب الفروقات عن المتوسط $\bar{x}$ من العلاقة التالية :

$$d_i = x_i - \bar{x} \qquad (28-3)$$

ونضعها في عمود خاص إلى جانب قيم X المدخلة في الحاسوب . ثم نقوم برسم هذه الفروقات على شكل نقاط بيانية، إحداثياتها (x_i , d_i) في المستوى Xod (علماً بأن $\bar{d} = \frac{1}{n}\sum d_i = 0$) .

ونحدد على ذلك المستوى الخطوط المستقيمة $d_2 = \pm 2S$ و $d_3 = \pm 3S$

(التي تمثل حدود مجالي الثقة الثاني والثالث) فنحصل على شكل بياني كالتالي:



الشكل (3-6): شكل انتشار الفروقات d_i

وبعدها ندرس كيفية انتشار الفروقات d_i على ذلك المستوى، وقبل كل شيء نحدد النقاط التي تقع خارج المجال الثالث، أي النقاط التي يكون فيها:

$$d_i > 3S \qquad (29-3)$$

وهي التي تقابل القيم الشاذة لـ X . فإذا كان عددها قليلاً (أو نسبتها صغيرة 1%) ،فإننا نقوم باستبعادها دفعة واحدة من جملة تلك البيانات دون أي تحفظ . أما إذا كان عددها كبيراً (أو نسبتها كبيرة) فإن ذلك يدل على عدم طبيعية البيانات المدروسة . ويجب معالجتها بطريقة أخرى .

ثم ننتقل إلى دراسة النقاط التي تقع بين مجالي الثقة الثاني والثالث . أي النقاط التي يكون فيها :

$$2S < |d_i| < 3S \qquad (29a-3)$$

فإن هذا يعني أن النقطة (x_i , d_i) تقع ضمن منطقة تسمى منطقة الشك . وعندها يترك أمر استبعادها أو الإبقاء عليها للباحث، ولكن إذا أراد الباحث استبعاد بعض هذه النقاط ، عليه أن يأخذها نقطة بعد نقطة، وأن يبدأ من النقطة الأبعد عن OX . وأن يقوم بإعادة حساب المؤشرات الوصفية بعد الحذف

للبيانات المتبقية من جديد . وأن يكرر ذلك بعد حذف كل نقطة شاذة، حتى يتم استبعاد جميع النقاط الشاذة، والحصول على بيانات طبيعية أو شبه طبيعية .

أما بالنسبة للنقاط التي تقع في الشريط الأوسط ، والتي تحقق العلاقة :

$$|d_i| < 2S \qquad (29\,b-3)$$

فإنها تقابل القيم الطبيعية لـ X، ولا يلزم استبعاد أي منها (إلا إذا كانت خطأً واضحاً) .

ملاحظة: لتسهيل عمليات الاستبعاد في منطقة الشك (عندما $(2S \le |d_i| \le 3S)$) يفضل أن نبدأ بدراسة القيمتين الكبرى والصغرى لـ X، وحساب فرقيهما عن $\bar{x}$. فإذا كانت قيمتا الفرقين (بالقيمة المطلقة) تحققان العلاقتين التالتين معاً :

$$\left.\begin{aligned} |max\,d_i| &= |max\,X_i - \bar{x}| < 2S \\ |min\,d_i| &= |min\,X_i - \bar{x}| < 2S \end{aligned}\right\} \qquad (30-3)$$

فإننا نعتبر جميع قيم X قيماً عادية وليست شاذة ونتابع التحليل .

أما إذا كانت قيمة أحدهما أو كلاهما (بالقيمة المطلقة) أكبر من $3S$ فإننا نستبعدها من البيانات . ثم نعيد حساب جميع المؤشرات السابقة من جديد للبحث عن قيم شاذة جديدة . ونبدأ أيضاً من القيمتين الكبرى والصغرى الجديدتين ... وهكذا دواليك .

- **طريقة الانحراف النسبي τ** :

وهي تعالج مسألة استبعاد القيم الشاذة في العينات الصغيرة $(n \le 25)$ ونعرضها كما يلي : $[Lvovosky.P.24]$

لنفترض أن x_i نقطة شاذة أو متفرقة في بيانات إحدى العينات الصغيرة . فلدراسة إمكانية استبعادها من هذه البيانات نقوم بحساب الفرق بينها وبين المتوسط $\bar{x}$ من العلاقة :

$$d_i = x_i - \bar{x}$$

ثم نضع فرضتي الاختبار كما يلي :

H_0 فرضية العدم $\quad d_i = 0 \quad$ (لا يوجد فرق معنوي بين x_i و $\bar{x}$)

H_1 فرضية البديلة $\quad d_i \ne 0 \quad$ (يوجد فرق معنوي بين x_i و $\bar{x}$)

ثم نقوم بحساب قيمة مؤشر الاختبار المعرف بالعلاقة :

$$\tau == \frac{|x_i - \bar{x}\,|}{\bar{S}} \qquad (31-3)$$

ثم نقوم بمقارنة قيمة τ (تاو)المحسوبة من العلاقة (3-31) مع قيمة $\tau_{\propto,n}$ الحرجة والمقابلة لمستوى الدلالة $\propto$ ولحجم العينة n ، وهي تؤخذ من جدول خاص لـ $\tau_{\propto}$ كالجدول المختصر التالي:

جدول (3-1): القيم الحرجة لـ $\tau_{\propto}$ حسب $\propto, n$:

حجم العينة n / مستوى الدلالة ∝	3	4	5	6	7	8	9	10	15	20	25
0.10	1.41	1.64	1.79	1.89	1.97	2.04	2.10	2.15	2.33	2.45	2.54
0.05	1.41	1.69	1.87	2.00	2.09	2.17	2.24	2.29	2.49	2.62	2.72
0.01	1.41	1.72	1.96	2.13	2.26	2.37	2.46	2.54	2.80	2.96	3.07

المصدر : $[Lvovosky.P.24]$

فإذا كانت $\tau \le \tau_{\propto,n}$ نقبل الفرضية H_0 ونعتبر x_i نقطة عادية .

أما إذا كانت $\tau > \tau_{\propto,n}$ نرفض الفرضية H_0، ونعتبر القيمة x_i قيمة شاذة لـ X ونقوم باستبعادها من جملة البيانات المعروضة .

ولمتابعة دراسة بقية النقاط الشاذة (المتوقعة) نقوم بإعادة حساب جميع مؤشرات العينة من جديد، ثم نقوم بإعادة الاختبار لكل نقطة شاذة جديدة .

ملاحظة: يمكن استخدام العلاقة (3-31) حتى في حالة العينات الكبيرة $(n > 25)$ ولكن حساب القيمة الحرجة لـ $(\tau_{\propto,n})$ يتم من خلال علاقتها بمتحول (ستودينت) t المعرفة بالعلاقة التالية:

$$\tau_{(\propto,n)} = \frac{t_{(\propto,n-2)} * \sqrt{n-1}}{\sqrt{(n-2) + \left[t_{(\propto,n-2)}\right]^2}} \qquad (31a-3)$$

ولتطبيق هذه العلاقة على العينات الكبيرة يفضل اتباع الخطوات التالية :

أ- نقوم بحساب أكبر الفروقات لـ d_i من :

$$d = max\, d_i = \max x_i - \bar{x} \qquad (32-3)$$

ب- نقوم بحساب قيمة τ من العلاقة المعدلة التالية :

$$\tau = \frac{|x_i - \bar{x}\,|}{S} = \frac{|x_i - \bar{x}\,|}{\bar{\mathrm{S}}\sqrt{\frac{n-1}{n}}} \qquad (33-3)$$

ج- نقوم بإيجاد القيمتين الحرجتين لمتحول (ستودينت) t عند $(n-2)$ درجة حرية والمقابلتين لمستويين من الدلالة هما :

مستوى الدلالة $\propto_1 = 0.05$ الموافق لمجال الثقة الثاني

مستوى الدلالة $\propto_2 = 0.001$ الموافق لمجال الثقة الثالث

وذلك من جداول توزيع (ستودينت) (أو من جداول التوزيع الطبيعي المعياري) فنحصل على القيمتين الحرجتين التاليتين :

$$t_{(0.05\,,n-2)} \quad \text{و} \quad t_{(0.001\,\,,n-2)} \qquad (34-3)$$

د- نقوم بحساب قيمتي $\tau_{(\propto,n)}$ المقابلتين لهاتين القيمتين من العلاقة $(31a - 3)$ السابقة، فنحصل على القيمتين الحرجتين لـ $\tau_{(\propto,n)}$ وهما :

$$\tau_{(0.05\ ,\ n)} \quad \text{و} \quad \tau_{(0.001\ ,\ n)} \qquad (35-3)$$

هـ-نقوم بمقارنة قيمة τ المحسوبة من العلاقة $(31-3)$ مع هاتين القيمتين الحرجتين، فيكون لدينا الحالات التالية :

- إذا كانت $\tau \leq \tau_{(0.05\ ,\ n)}$ تكون النقطة المقابلة لها نقطة عادية ولا يجوز حذفها
- إذا كانت $\tau_{(0.05\ ,\ n)} < \tau \leq \tau_{(0.001\ ,\ n)}$ تعتبر النقطة المقابلة لها موضع شك، وإن مسألة استبعادها أو إبقائها تعود لتصورات الباحث .
- أما إذا كانت $\tau > \tau_{(0.001\ ,\ n)}$ فإن تلك النقطة تستبعد وتحذف من بيانات العينة المسحوبة .

3-2-3 : مرحلة تبويب البيانات وتهيئتها للاختبار :

لتبويب البيانات المتبقية (بعد استبعاد القيم الشاذة) :

أ- نقوم أولاً بتحديد طول المجال الكلي لها من العلاقة :

$$L = X_{max} - X_{min} \qquad (36-3)$$

ب- نقوم بتحديد عدد المجالات الجزئية m ، التي سنعتمدها لتبويب تلك البيانات . ويمكن للباحث الخبير أن يقدر العدد m حسب أهداف البحث أو حسب طبيعة المتحول X . كما يمكنه أن يستأنس بعلاقة (ستورجيس) لحساب العدد m المناسب لحجم العينة n . وذلك من العلاقة :

$$m = 1 + 3.32 * log_{10}\, n \qquad \text{(غير ملزمة)} \qquad (37-3)$$

ج-نقسم المجال الكلي L إلى مجالات جزئية متساوية عددها m مجالاً . ونحسب طول كل مجال جزئي من العلاقة :

$$\ell = \frac{L}{m} \qquad (38-3)$$

د- نصمم جدولاً خاصاً يتضمن هذه المجالات الجزئية بشكل متصل ومتصاعد . وبحيث تكون تلك المجالات الجزئية متواصلة وغير متقاطعة . ونخصص في ذلك الجدول سطراً لحساب مراكز تلك المجالات وسطراً للتكرارات المطلقة n_j المقابلة للمجال j ، وسطراً للتكرارات النسبية $\frac{n_j}{n}$ المقابلة للمجال j .

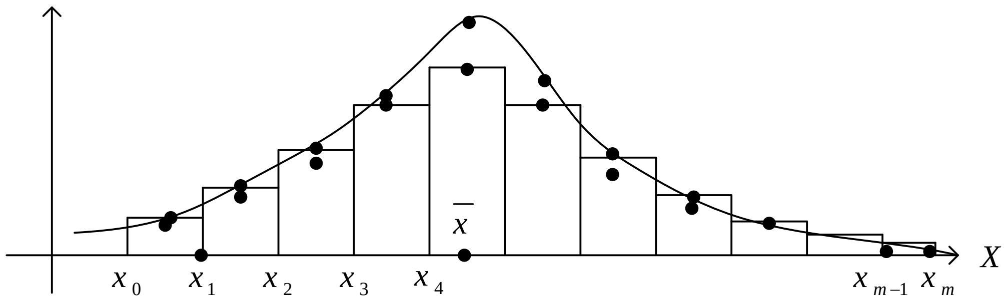
وهكذا نجد أن جدول التبويب يأخذ الشكل التالي:

ملاحظة: بما أن متحول التوزيع الطبيعي يتحول من $-\infty$ إلى $+\infty$ ، فإنه يمكننا عند تبويب بيانات X أن نضيف إلى المجال الكلي L مجالاً احتياطياً على اليسار مثل $]x_0\ ,\ x_1[$ ، ومجالاً احتياطياً آخر على اليمين مثل $]x_m\ ,\ x_{m+1}[$ على أن نضع التكرارات المقابلة لهما مساوية للصفر .

جدول (3-2): بيانات العينة المبوبة تصاعدياً

رقم المجال j	1	2	3			$\sum$
المجال $[x_{j-1} - x_j[$	$[x_0 - x_1[$	$[x_1 - x_2[$	$[x_2 - x_3[$	……	$x_{m-1} - x_m$	–
x'_j مركز المجال	x'_1	x'_2	x'_3	……	x'_m	–
n_j التكرارات المطلقة	n_1	n_2	n_3	……	n_m	n
P_j التكرارات النسبية	$\frac{n_1}{n} = P_1$	$\frac{n_2}{n} = P_2$	$\frac{n_3}{n} = P_3$	……	$\frac{n_m}{n} = P_m$	1
$\not{f}_j$ الاحتمالات النظرية	$\not{f}_1$	$\not{f}_2$	$\not{f}_3$	……	$\not{f}_m$	1

هـ- نرسم مدرج التكرارات النسبية لهذه البيانات (وهو الذي يمثل الاحتمالات التجريبية) . وندرس فيما إذا كان يحقق بعض مواصفات التوزيع الطبيعي العام (كأن يكون شبه متناظر ويكون معامل الالتواء $K \approx 0$ ويكون معامل التطاول $\ell \approx 0$) . ولنفترض أنه يأخذ الشكل التالي:



الشكل (3-7): توزيع التكرارات النسبية التجريبية

و- نقوم بحساب الاحتمالات النظرية الطبيعية $\not{f}_j$ المقابلة لكل مجال جزئي $[x_{j-1} - x_j[$ بشكل تقريبي، وذلك بتعويض مركز ذلك المجال x'_j في العلاقة (3-1) وحساب قيمة $\not{f}(x_j)$ منها . وهنا يجب أن نميّز بين حالتين:

- إذا كان μ و σ معلومين (أو مفروضين) فإننا نحسب الاحتمالات النظرية $\not{f}_j$ المقابلة للمجال $[x_{j-1} - x_j[$ من العلاقة :

$$\not{f}_j = \not{f}(x'_j) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x'_j-\mu}{\sigma}\right)^2} \qquad (39-3)$$

- إذا كان μ و σ مجهولين، نستبدلهما بتقديريهما $\bar{x}$ و S . ثم نحسب الاحتمالات $\not{f}_j$ بشكل تقريبي من العلاقة :

$$\tilde{\not{f}}_j = \not{f}(x'_j) = \frac{1}{S\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x'_j-\bar{x}}{S}\right)^2} \qquad (40-3)$$

ملاحظة: يفضل حساب الاحتمالات النظرية $\not{f}_j$ بشكل دقيق من تكامل (3−1) على المجال j كما يلي:

$$\not{f}_j = P(x_{j-1} \le x < x_j) = \int_{x_{j-1}}^{x_j} \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} * dx \qquad (41-3)$$

$$\not{f}_j = \phi\left(\frac{x_j-\mu}{\sigma}\right) - \phi\left(\frac{x_{j-1}-\mu}{\sigma}\right) \qquad j: 1\ 2\ 3 \ldots . m$$

وإذا كان μ و σ مجهولتين نستبدلهما بتقديريهما $\bar{x}$ و S فنحصل على الاحتمالات التقريبية $\tilde{\not{f}}_j$.

ز− نضع الاحتمالات النظرية $\not{f}_j$ في سطر خاص ملحق بالجدول (3−2) ثم نحسب الفروقات بين الاحتمالات التجريبية والنظرية بالقيمة المطلقة في كل مجال j ، وذلك لاستخدام مجموعها $\sum_{j=1}^{m}|P_j - \not{f}_j|$ في دراسة جودة التوافق بين التوزيع التجريبي والتوزيع الطبيعي المفروض .

ح− نصيغ فرضيتي الاختبار بشكل عام وفق التالي :

H_0 – فرضية العدم: لا توجد فروقات معنوية بين التوزيع التجريبي والتوزيع الطبيعي المفروض، أي أن المتحول المدروس X يخضع للتوزيع الطبيعي المفروض $\not{f}(x)$.

H_1 – الفرضية البديلة: توجد فروقات معنوية بين التوزيع التجريبي والتوزيع الطبيعي المفروض، أي أن X لا يخضع للتوزيع الطبيعي المفروض .

ط− بناءً على معطيات الجدول (3−2) نقوم بحساب المؤشرات الوصفية السابقة من جديد، للحصول على المتوسط والمنوال والوسيط، التباين معاملا الالتواء والتطاول ومعامل الاختلاف ...الخ . ثم نقوم بدراسة الأمور التالية:

1−دراسة تساوي أو تقارب قيم المتوسط والمنوال والوسيط . لأن تقاربها يشير إلى طبيعية البيانات .

2−دراسة قيم معاملي الالتواء والتطاول . وهما معاملان يساويان أو يقتربان من الصفر إذا كانت البيانات طبيعية ، والعكس بالعكس .

3−دراسة قيمة معامل الاختلاف v ومقارنتها مع القيمة الحدية 0.33 ، ويكون لدينا حالتان :

إذا كانت $v \le 0.33$ فإن البيانات يمكن أن تكون طبيعية .

أما إذا كانت $v > 0.33$ فإن البيانات لا تعتبر طبيعية، إلا إذا تم تحويلها إلى شبه طبيعية . وفي هذه الحالة (حالة $v > 0.33$) نعود إلى دراسة معامل الالتواء ونجري التحويلات كما يلي:

إذا كان الالتواء إلى اليمين نجري أحد التحويلين: $Y = X^{1.5}$ أو $Y = X^2$.

أما إذا كان الالتواء إلى اليسار نجري أحد التحويلات التالية:

$$Y = \ln x \qquad Y = \frac{1}{X} \qquad Y = \sqrt{X} \qquad Y = \frac{1}{\sqrt{X}} \qquad (42-3)$$

ثم نطبق التحليل على بيانات المتحول الجديد Y .

ي–نقوم بدراسة متوسط الانحرافات المطلقة للبيانات عن متوسطها. ونحسبه من العلاقة التالية:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

وحتى تكون هذه البيانات لـ X طبيعية، يجب أن تتحقق العلاقة التالية [Lvovsky P.28] :

$$\left|\frac{\bar{d}}{\bar{\bar{S}}} - 0.7979\right| < 0.4\sqrt{n} \qquad (42-3)$$

3-3: اختبارات التوزيع الطبيعي (الطبيعية Normality) :

نشير – قبل كل شيء – إلى أن هذه الاختبارات تطبق على المتحولات المستمرة (أو المرتبة) X، ولكل متحول بمفرده، وذلك لأن التوزيع الطبيعي هو توزيع مستمر . وهي تعتمد على بيانات العينة (كبيرة الحجم) المسحوبة من المجتمع المدروس . وعلى حساب الاحتمالات التجريبية منها، ثم مقارنتها مع الاحتمالات النظرية المقابلة لها والمحسوبة من التوزيع الطبيعي المفروض [انظر العلاقات (3–39)، (3–40)، (3–41)] ،

أي أنه لتطبيق هذه الاختبارات لابد من إجراء ما يلي :

1–سحب عينة عشوائية بحجم n من المجتمع المدروس ، والحصول منها على البيانات اللازمة عن المتحول المدروس X وعن المتحولات الأخرى .

2–حساب المؤشرات الوصفية للمتحول المدروس X . ودراسة خصائص هذه المؤشرات ومقارنة تلك الخصائص مع خصائص متحول التوزيع الطبيعي .

3–القيام باستبعاد القيم الشاذة من تلك البيانات باستخدام أحد الأساليب المذكورة في الفقرة (3–2–2) .

4–تصنيف أو تبويب البيانات المتبقية وحساب التكرارات المطلقة ثم التكرارات النسبية التي تمثل الاحتمالات التجريبية P_j المقابلة للمجالات j .

5–حساب الاحتمالات النظرية الطبيعية من أحد العلاقات (3–39)، (3–40)، (3–41)، ثم تطبيق أحد الاختبارات التالية :

3-3-1: اختبار التوافق $\mathcal{X}^2$ (اختبار بيرسون Pearson) .

ويعرف هذا الاختبار حسب الرموز السابقة بالعلاقة التالية:

$$\mathcal{X}^2 = n\sum_{j=1}^{m}\left[\frac{\left(\not{f}_j - P_j\right)^2}{\not{f}_i}\right] \qquad (44-3)$$

حيث أن n : هو حجم العينة، و m : عدد المجالات الجزئية المتساوية، و $\not{f}_j$: هي قيم الاحتمالات النظرية، وP_j : هي قيم الاحتمالات التجريبية . أما $\mathcal{X}^2$: فهي قيمة مؤشر الاختبار، وهي متحول عشوائي يخضع لتوزيع $\mathcal{X}^2(x)$ بدرجة حرية $(m-1)$ (لأن الاحتمالات مرتبطة بالعلاقة: $\sum_{j=1}^{m}\not{f}_j = \sum P_j = 1$) .

وهناك صيغة أخرى لهذا الاختبار تعطى بدلالة التكرارات التجريبية n_j والنظرية (المتوقعة) m_j ، وتأخذ الشكل التالي :

$$\mathcal{X}^2 = \sum_{j=1}^{m}\left[\frac{\left(m_j - n_j\right)^2}{m_j}\right] \qquad (45-3)$$

حيث أن: $n_j = n * P_j$ ، وأن $m_j = n * f_j$ ، ويشترط هنا أن يكون: $m_j \geq 5$

ولإجراء هذا الاختبار نضع الفرضيتين كما يلي:

- فرضية العدم: X يخضع للتوزيع الطبيعي: H_0 .
- الفرضية البديلة: X لا يخضع للتوزيع الطبيعي: H_1 .

وبعد حساب قيمة $\mathcal{X}^2$ من إحدى العلاقتين (3−44) أو (3−45)، نقارنها مع القيمة الحرجة $\mathcal{X}^2_{\propto}$ المقابلة لـ $(m-1)$ درجة حرية ، ونتخذ القرار حول فرضية العدم H_0 كما يلي:

إذا كانت $\mathcal{X}^2 \leq \mathcal{X}^2_{\propto_{m-1}}$ نقبل فرضية العدم H_0 التي تقول أن X يخضع للتوزيع الطبيعي .

أما إذا كانت $\mathcal{X}^2 > \mathcal{X}^2_{\propto_{m-1}}$ نرفض فرضية العدم H_0، وعندها نكون مضطرين لقبول الفرضية البديلة H_1 . التي تقول أن X لا يخضع للتوزيع الطبيعي .

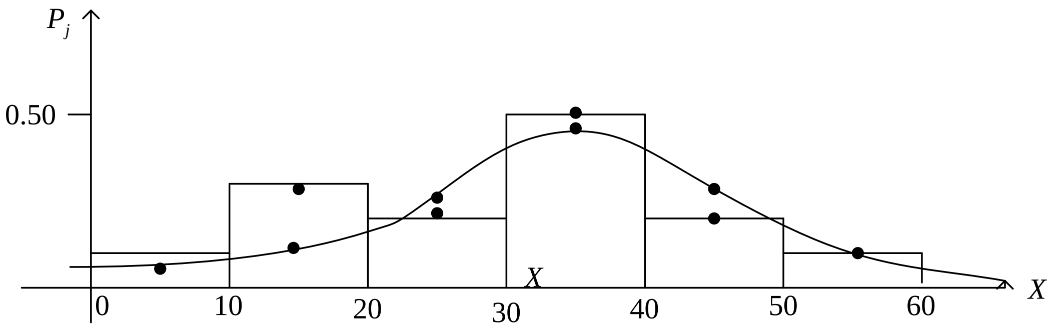
مثال (3−1): لنفترض أن دراسة لأعمار المسافرين عبر أحد المنافذ الحدودية، شملت عينة من المسافرين بحجم $(n = 100)$ مسافر، ثم تم تبويب هؤلاء المسافرين حسب العمر X إلى (6) مجالات متساوية فكانت كما يلي :

جدول (3−3): بيانات المثال المبوبة تصاعدياً

j	1	2	3	4	5	6	$\sum$
مجالات العمر X	$[0-10[$	$[10-20[$	$[20-30[$	$[30-40[$	$[40-50[$	$[50-60[$	ــــ
التكرارات المطلقة n_j	5	20	15	45	10	5	100
التكرارات النسبية $n_j/n = P_j$	0.05	0.20	0.15	0.45	0.10	0.05	1
مراكز المجالات X'_j	5	15	25	35	45	55	−
$n_j\, X'_j$	25	300	375	1575	450	275	3000
$n_j\left(X'_j - \bar{x}\right)^2$	3125	4500	375	1125	2250	3125	14500

والمطلوب: اختبار فيما إذا كان العمر X خاضعاً للتوزيع الطبيعي وذلك بمستوى دلالة $\propto = 0.05$.

الحل: نقوم برسم مدرج التكرارات النسبية $P_j = \frac{n_j}{n}$ فنجد أنه كما يلي :



الشكل (3–8): مدرج التكرارات النسبية للمثال

وهنا نلاحظ أن شكل هذا المدرج لا يتصف بالتناظر حول المتوسط . ولكننا حتى نختبر خضوع العمر X إلى التوزيع الطبيعي العام ، يجب علينا أن نحدد قيمة التوقع μ والتباين σ^2 من معطيات المجتمع المدروس ، وإذا كانا مجهولين نقوم بتقدير μ بواسطة متوسط العينة $\bar{x}$ والتباين σ^2 بواسطه التباين في العينة S^2 ومن معطيات الجدول (3–3) نجد أن:

$$\tilde{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{6} n_j X'_j = \frac{1}{100}(3000) = 30$$

$$\tilde{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{j=1}^{6} n_j\left(X'_j - \bar{x}\right)^2 = \frac{1}{99}(14500) = 146.546$$

وبذلك نجد أن الانحراف المعياري يقدر بـ $\tilde{\sigma} = \sqrt{146.546} = 12.1$

ثم نقوم بتعويض ذلك في معادلة التوزيع الطبيعي العام فنحصل على التوزيع المحدد التالي:

$$f(x) = \frac{1}{12.1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-30}{12.1}\right)^2} \qquad (46-3)$$

والآن نضع فرضيتي الاختبار كما يلي:

H_0: فرضية العدم: إن العمر X للمسافرين يخضع للتوزيع الطبيعي المحدد بالعلاقة (3–46)، أي لا توجد فروقات معنوية بين الاحتمالات التجريبية P_j والنظرية f_j .

H_1: الفرضية البديلة: إن العمر X لا يخضع للتوزيع الطبيعي المذكور، ثم نقوم بحساب الاحتمالات النظرية الطبيعية بشكل دقيق في كل مجال j من تكامل العلاقة (3–46) على ذلك المجال، فنجد من (3–41) ومن جداول تابع التوزيع الطبيعي $\phi(z)$ أن:

$$\begin{aligned} f_1 = P(0 \le X < 10) &= \phi\left(\frac{10-30}{12.1}\right) - \phi\left(\frac{0-30}{12.1}\right) = \phi(-1.65) - \phi(-2.48) \\ &= 1 - \phi(1.65) - [1 - \phi(2.48)] = \phi(2.48) - \phi(1.65) \\ &= 0.9934 - 0.9505 = 0.0429 \end{aligned}$$

$$\begin{aligned} f_2 = P(10 \le X < 20) &= \phi\left(\frac{20-30}{12.1}\right) - \phi\left(\frac{10-30}{12.1}\right) = \phi(-0.83) - \phi(-1.65) \\ &= \phi(1.65) - \phi(0.83) = 0.9505 - 0.7967 = 0.1538 \end{aligned}$$

$$f_3 = P(20 \le X < 30) = \phi\left(\frac{30-30}{12.1}\right) - \phi\left(\frac{20-30}{12.1}\right) = \phi(0) - \phi(-0.83)$$
$$= 0.50 - (1 - 0.7967) = 0.2967$$

$$f_4 = P(30 \le X < 40) = \phi\left(\frac{40-30}{12.1}\right) - \phi\left(\frac{30-30}{12.1}\right) = \phi(0.83) - \phi(0)$$
$$= 0.7967 - 0.50 = 0.2967$$

$$f_5 = P(40 \le X < 50) = \phi\left(\frac{50-30}{12.1}\right) - \phi\left(\frac{40-30}{12.1}\right) =$$
$$= 0.9505 - 0.7967 = 0.1538$$

$$f_6 = P(50 \le X < 60) = \phi\left(\frac{60-30}{12.1}\right) - \phi\left(\frac{50-30}{12.1}\right) =$$
$$= 0.9934 - 0.9505 = 0.0429$$

ولحساب قيمة مؤشر الاختبار χ^2 نضع هذه الاحتمالات النظرية f_j مع الاحتمالات التجريبية P_j في جدول مناسب ونجري عليه الحسابات اللازمة فنحصل على الجدول التالي:

جدول (3–4): الحسابات اللازمة لـ χ^2 :

المجال j	1	2	3	4	5	6	$\sum$
الاحتمالات التجريبية P_j	0.05	0.20	0.15	0.45	0.10	0.05	1
الاحتمالات النظرية f_j	0.0429	0.1538	0.2967	0.2967	0.1538	0.0429	1
$(f_j - P_j)^2/f_j$	0.001175	0.013878	0.072534	0.079208	0.0188195	0.0011750	0.1867896

ومن مجموع السطر الأخير في الجدول (3–4) السابق نحسب قيمة مؤشر الاختبار χ^2 فنجد أن :

$$\chi^2 = n\sum_{j=1}^{6}\left[\frac{(f_j - P_j)^2}{f_j}\right] = 100(0.1867896) = 18.697$$

ولاتخاذ القرار المناسب حول H_0 : نقارن قيمة χ^2 المحسوبة مع قيمة χ^2_{α} الحرجة المقابلة لـ $\alpha =$ 0.05 ولـ $(\nu = 6 - 2 - 1 = 3)$ درجة حرية والتي تساوي $\chi^2_{0.053} = 7.815$ (لقد نقصت درجة الحرية بمقدار (2) لأننا استخدمنا بيانات العينة في تقدير المعلمتين μ و σ^2 في التوزيع) . وبالمقارنة نجد أن $\chi^2 > \chi^2_{\alpha}$ لذلك نرفض فرضية العدم H_0 ونقبل H_1 . والتي تقول أن العمر X لا يخضع للتوزيع الطبيعي المفروض .

- **اختبار رومانوفسكي:** لقد لاحظ (Romanovsky) أن التوزيع الاحتمالي $\chi^2_{(x)}$، يتمتع بالخواص التالية: [Lvovsky P.30]

$$E(\chi^2) = \nu$$
$$\sigma^2(\chi^2) = 2\nu$$

حيث : ν عدد درجات الحرية

(47 – 3)

وبناءً على قاعدة (3σ) فإنه عندما يكون v كبيراً فإن التوزيع $\chi^2_{(x)}$ يقترب من الطبيعي ويكون لدينا :

$$\begin{aligned} &P\left[v-3\sqrt{2v} \le \chi^2 \le v+3\sqrt{2v}\right] = \\ &= P\left[-3\sqrt{2v} \le (\chi^2 - v) \le 3\sqrt{2v}\right] = 1-\propto \end{aligned} \quad (48-3)$$

وهذا يكافئ ما يلي:

$$P\left[\frac{|\chi^2-v|}{\sqrt{2v}} \le 3\right] = 1-\propto \quad (49-3)$$

وبناءً على ذلك اقترح (رومانوفسكي) استخدام المؤشر $\frac{|\chi^2-v|}{\sqrt{2v}}$ ، لاتخاذ القرار حول توافق التوزيعين [التجريبي $P(x)$ مع النظري $f(x)$] كما يلي:

فإذا كان: $\frac{|\chi^2-v|}{\sqrt{2v}} \le 3$ نقبل فرضية العدم H_0 ونعتبر أن المتحول X يخضع للتوزيع الاحتمالي المفروض $f(x)$.

أما إذا كان: $\frac{|\chi^2-v|}{\sqrt{2v}} > 3$ نرفض فرضية العدم H_0، ونقبل H_1: التي تقول أن المتحول X لا يخضع للتوزيع المفروض $f(x)$. لأن الفروقات بين التوزيعين جوهرية، ويتميز هذا الاختبار بأنه سهل التطبيق ولا يحتاج إلى الجداول الإحصائية ، ولتطبيقه على بيانات المثال (3−1) نلاحظ أن قيمة χ^2 المحسوبة $\chi^2 = 18.679$ ، وأن $v = 6-2-1 = 3$.

وبذلك نجد أن:

$$\frac{|\chi^2-v|}{\sqrt{2v}} = \frac{|18.679-3|}{\sqrt{2*3}} = \frac{15.679}{2.449} = 6.402 > 3$$

لذلك نرفض H_0 ونقبل H_1 التي تقول أن X لا يخضع للتوزيع الطبيعي المفروض .

- **اختبار (Jorque- Bera)**: وينسبه بعضهم إلى Fisher [Gopal P.51]: ولقد قام (Jorque- Bera) بوضع معادلة لاختبار طبيعة المتحول المدروس X باستخدام معامل الالتواء k ومعامل التطاول L وعرفاه بالعلاقة التالية[عناني صـ 723] :

$$JB = n\left[\frac{k^2}{6} + \frac{(L-3)^2}{24}\right] \sim \chi^2_2 \quad (49\,a-3)$$

وهو يخضع لتوزيع χ^2_2 بدرجة حرية $v = 2$. ولإجراء اختبار طبيعة أي متحول عشوائي X ، نقوم بحساب القيمة العددية لـ JB من المعادلة السابقة، ثم نقارنها مع القيمة الحرجة $\chi^2_2(\propto)$ ونتخذ القرار كما يلي:

إذا كانت $\chi^2 \le \chi^2_2(\propto)$ نقبل فرضية العدم H_0 ونعترف بأن X يخضع للتوزيع الطبيعي .

أما إذا كانت $\chi^2 > \chi^2_2(\propto)$ فإننا نرفض H_0 ونقبل H_1 التي تقول أن X لا يخضع للتوزيع الطبيعي .

وعندها نقوم بإجراء إحدى التحويلات (3−21) أو (3−23) ونعالج البيانات من جديد.

3-3-2: اختبار (كولموغوروف– سميرنوف Kolmogorov- Smirnov) .

يعتمد هذا الاختبار على مقارنة التكرارات التجميعية المتصاعدة مع تابع التوزيع الطبيعي التكاملي(أوأي توزيع آخر) . لذلك نفترض أننا سحبنا عينة عشوائية من المجتمع المدروس بحجم كبير n، وحصلنا منها على قيم X المرتبة والمبوبة مع تكراراتها المطلقة والمتراكمة، كما في الجدول التالي :

جدول (3-4): بيانات العينة للمتحول المستمر X:

j	1	2	3	…	K	…	m	$\sum$
مجالات X $[X_{j-1}, X_j[$	$[x_0, x_1[$	$[x_1, x_2[$	…	…	$[x_{j-1}, x_j[$	…	$[x_{m-1}, x_m[$	
التكرارات n_j	n_1	n_2	n_3	…	n_j	…	n_m	n
التكرارات التجميعية المتصاعدة K_j	n_1	n_1+n_2	$n_1+n_2+n_3$	…	$\sum_{j=1}^{K} N_j = N_j$	…	n	

فهل التكرارات التجميعية المتصاعدة لهذه البيانات تتوافق مع تابع التوزيع التكاملي $F(x)$ للتوزيع المفروض $f(x)$ ؟

لدراسة فيما إذا كان تابع التوزيع التجميعي لهذه البيانات يتوافق مع تابع التوزيع التكاملي للتوزيع المفروض $F(x)$، نضع الفرضيتين كما يلي:

H_0 : فرضية العدم : إن المتحول X حسب بيانات العينة يخضع للتوزيع المفروض $f(x)$.

H_1 : الفرضية البديلة: إن المتحول X لا يخضع للتوزيع المفروض $f(x)$.

وحتى نتحقق من ذلك نقوم بحساب قيم تابع التوزيع التجميعي التصاعدي حتى نهاية كل مجال $[x_{K-1}, x_K[$ فنجد أنها تساوي :

$$P(x_K) = \frac{N_K}{n} \qquad K: 1\ 2\ 3\ \ ....m \qquad (50-3)$$

حيث أن: N_K هو مجموع التكرارات المطلقة حتى نهاية المجال K، حيث $\left(N_K = \sum_{j=1}^{K} n_j\right)$ ،

ثم نعرف تابع التوزيع التكاملي للتوزيع المفروض $F(x)$ حتى نهاية المجال k من العلاقة التالية:

$$F(x_K) = \int_{-\infty}^{x_K} f(x)dx \qquad (51-3)$$

حيث x_K هو الحد الأعلى للمجال k .

وبعد حساب قيم $P(x_K)$ و $F(x_K)$ المتقابلة، نضعها في جدول خاص ونحسب الفروقات بينهما (بالقيمة المطلقة) ، فنحصل على الفروقات التالية:

$$|F(x_K) - P(x_K)| \qquad k: 1\ 2\ 3\ \ ...m \qquad (52-3)$$

ثم نقوم بالبحث عن أكبر قيمة لتلك الفروقات ونرمز لها بـ D ، ونكتب ذلك كما يلي:

$$D = \max_{j=1}^{m} \left| F(x_j) - P(x_j) \right| \qquad (53-3)$$

ولقد لاحظ (كولموغوروف) أن المقدار D هو عبارة عن متحول عشوائي، وبرهن على أنه إذا ضربنا المقدار D بجذر حجم العينة $\sqrt{n}$. فإننا نحصل على متحول عشوائي جديد هو:

$$\lambda = D * \sqrt{n} \qquad (54-3)$$

ثم برهن على أن المتحول الجديد λ يخضع تقاربياً (عندما تكبر n) إلى توزيع احتمالي خاص اطلق عليه اسم توزيع (كولموغوروف) ، وإن تابع توزيعه يعطى بالعلاقة التالية:

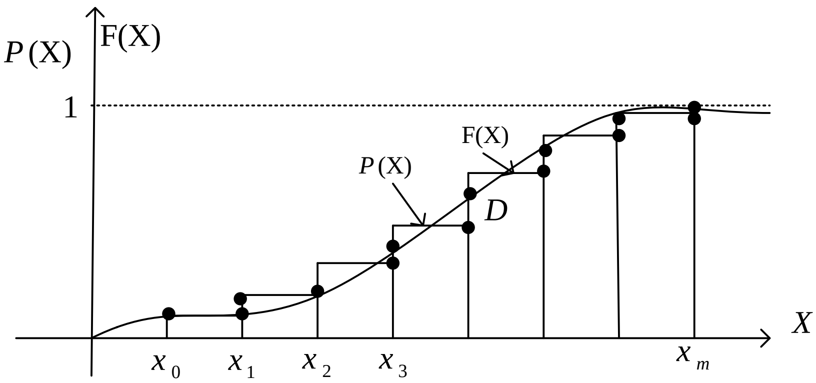
$$K(\lambda) = \lim_{n \to \infty} P\left(D * \sqrt{n} \le \lambda\right) = \sum_{K=-\infty}^{K=+\infty} (-1)^K e^{-\frac{2K^2}{\lambda^2}} \quad : \lambda \ge 0 \qquad (55-3)$$

وذلك من أجل أي تابع توزيع $F(x_K)$.

وبناءً على العلاقة (3–55) وعندما يكون حجم العينة n كبيراً ، قام (سميرنوف) بحساب جداول الاحتمالات المقابلة لمجالات الثقة للاحتمالات $P(x_j)$ من العلاقة التالية:

$$K(\lambda) \approx P\left[F(x_j) - \frac{\lambda}{\sqrt{n}} \le P(x_j) \le F(x_j) + \frac{\lambda}{\sqrt{n}}\right] \qquad (56-3)$$

وهو الاحتمال المقابل لمجال الثقة لانتشار قيم تابع التوزيع التجريبي $P(x_j)$ حول منحنى تابع التوزيع النظري $F(x_j)$ بنصف طول قدره $\frac{\lambda}{\sqrt{n}}$. وذلك كما هو موضح على الشكل التالي:



الشكل (3–9): تابعاً التوزيع $P(x)$ و $F(x)$

وبالمقابل يمكننا إيجاد الاحتمال P المتمم لاحتمال الثقة، وهو الذي يساوي احتمال الدلالة P أو Sig . وذلك من العلاقة التالية :

$$P = Sig = P\left(D * \sqrt{n} > \lambda\right) = 1 - K(\lambda) \qquad (57-3)$$

وهناك جداول خاصة لحساب هذا الاحتمال مقابل لكل قيمة λ ولكل مستوى دلالة $\propto$.

وباختصار فإن تطبيق اختبار (كولموغوروف – سميرنوف) يقتضي أن نقوم بتبويب البيانات ضمن مجالات متساوية (غير طويلة) وحساب التكرارات المقابلة لكل منها، ثم حساب تابع التوزيع التجريبي المتراكم $P(x_k)$ (حيث x_k وهو الحد الأعلى للمجال k) . ثم حساب قيم تابع التوزيع النظري $F(x_k)$

المكامل من $-\infty$ حتى نهاية المجال k، ثم حساب الفروقات بين قيم $P(x_k)$ و $F(x_k)$ في كل مجال k . ثم تحديد أكبر قيمة مطلقة لهذه الفروقات D وحساب قيمة المؤشر $\lambda = D * \sqrt{n}$، ثم إيجاد قيمة احتمال الدلالة المقابل لها $P = 1 - K(\lambda)$ من الجداول الخاصة بذلك .

ولاتخاذ القرار المناسب نقارن قيمة D مع القيمة الحرجة $D_\propto$، فإذا كانت $D \le D_\propto$ نقبل فرضية العدم H_0 ونعتبر أن بيانات العينة تخضع لذلك التوزيع $f(x)$. علماً بأن: $D = \frac{\lambda}{\sqrt{n}}$.

أما إذا كانت قيمة $D > D_\propto$ فإننا نرفض فرضية العدم H_0 ونقبل H_1 ونعتبر بيانات هذه العينة X لا تخضع للتوزيع المفروض $f(x)$ الذي تابع توزيعه $F(x)$.

ملاحظة 1: عندما تكون أطوال مجالات التبويب طويلة ، فإنه يمكننا تقليصها أو ضغطها وتحويلها إلى مجالات واحدية (طول كل منها يساوي الواحد) . وذلك باستخدام التحويلة التالية :

$$x_j' = \frac{x_j - x_0}{\ell} \qquad (58-3)$$

حيث أن ℓ : هو طول أحد المجالات المتساوية . وأن x_0 أي نقطة قريبة من المتوسط العادي .

ملاحظة 2:**اختبار (سميرنوف)** : قام (سميرنوف) بتطوير تطبيق اختبار (كولموغورف – سميرنوف) لاختبار التوافق بين توزيعين تجريبيين مأخوذين من عينتين مستقلتين مسحوبتين من مجتمع واحد أو من مجتمعين مختلفين ، وبعد إجراء المعالجات اللازمة على البيانات التجريبية وتبويبها بنفس الأسلوب .

أفترض أن $F_{n_1}(x_j)$ و $E_{n_2}(x_j)$ هما تابعا التوزيعين التجريبيين لـ X في العينتين المستقلتين وذات الحجمين الكبيرين n_1 و n_2 . ثم قام بحساب أكبر الفروقات المطلقة بينهما D من العلاقة :

$$D = \max_{j=1}^{m} \left| F_1(x_j) - E_2(x_j) \right| \qquad (59-3)$$

ولكنه قام بحساب قيمة λ في هذه الحالة(حالة عينتين) من العلاقة التالية:

$$\lambda = D * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}} \qquad \text{(اختبار سميرنوف)} \qquad (60-3)$$

وبعدها تابع عمليات الاختبار كما فعلنا سابقاً، وأعد جداول خاصة للقيم الحرجة لـ λ ، تسمى جداول (سميرنوف) ، ويتم التعامل معها كما في حالة اختبار (كولموغوروف– سميرنوف).

مثال (3–2): لنأخذ بيانات المثال (3–1) ولندرس فيما إذا كانت بيانات تلك العينة خاضعة للتوزيع الطبيعي العام باستخدام اختبار (كولموغورف – سميرنوف) . لذلك نعود إلى الجدول (3–3) ونحسب منه الاحتمالات التجميعية التجريبية $P(x_k)$ فنجد أنها تساوي :

$$P(x_k): 0.05 \quad 0.25 \quad 0.40 \quad 0.85 \quad 0.95 \quad 1$$

وبناءً على العلاقة (1–51) نقوم بحساب الاحتمالات التجميعية الطبيعية $F(x_k)$ من التكامل التالي :

$$F(x_k) = \int_{-\infty}^{x_k} f(x)d(x) = \phi\left(\frac{x_k - \bar{x}}{S}\right)$$

وبما اننا وجدنا سابقاً أن $\bar{x} = 30$ وأن $S = 12.1$ فإن تلك الاحتمالات التكاملية تحسب كما يلي :

$$F(x_{10}) = \phi\left(\frac{10-30}{12.1}\right) = \phi(-1.65289) = 0.049177$$
$$F(x_{20}) = \phi\left(\frac{20-30}{12.1}\right) = \phi(-0.82645) = 0.20427$$
$$F(x_{30}) = \phi\left(\frac{30-30}{12.1}\right) = \phi(0.00) = 0.50$$
$$F(x_{40}) = \phi\left(\frac{40-30}{12.1}\right) = \phi(0.82645) = 0.79573$$
$$F(x_{50}) = \phi\left(\frac{50-30}{12.1}\right) = \phi(1.65289) = 0.95082$$
$$F(x_{60}) = \phi\left(\frac{60-30}{12.1}\right) = \phi(2.47434) = 0.99342$$

ثم نضع هذه الاحتمالات في الجدول التالي :

جدول (3–6): الحسابات اللازمة لاختبار K-S

المواصفات	6	5	4	3	2	1	المجال j
$\bar{P} = 0.583333$ $S_P = 0.402078$	1	0.95	0.85	0.40	0.25	0.05	$P(X_j)$
$\bar{K} = 0.565566$ $S_F = 0.37916$	0.9934	0.9508	0.7957	0.50	0.2043	0.0492	$F(X_j)$
$a = 2.284536$ $b = 0.930379$	0.0066	0.0008	0.0543	0.10	0.0457	0.008	$D_j = \lvert F(X_j) - P(X_j) \rvert$
$r = 0.986617$							

ومن الجدول السابق نستخلص أن أكبر الفروقات هو الفرق المقابل للمجال الثالث، حيث نجد أن:

$$D = max\, D_j = |F(X_3) - P(X_3)| = 0.10$$

ثم نبحث في جداول D عن القيمة الحرجة $D_{0.05}$ المقابلة لحجم العينة $n = 100$ فنجد أنها تساوي :

$$D_{0.05} = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{100}} = 0.136$$

وبمقارنة D مع $D_{\propto}$ نجد أن : $D < D_{0.05}$ لذلك نقبل فرضية العدم H_0 التي تقول أن العمر X لهؤلاء المسافرين يخضع للتوزيع الطبيعي $N(30\,,12.1)$ (وهذا يخالف اختبار χ^2) .

ملاحظة : يمكننا من جداول الاختبار $(K-S)$ حساب احتمال الدلالة P (أو Sig) لنتيجة هذا الاختبار بعد حساب λ وتحديد مستوى الدلالة المعتمد $\propto$، ونتخذ القرار حول الفرضية H_0 كما يلي:

إذا كانت قيمة $P \geq \propto$ ، فإننا نقبل فرضية العدم H_0، التي تقول أن المتحول X يخضع للتوزيع الطبيعي المفروض .

أما إذا كانت قيمة $P < \propto$ فإننا نرفض فرضية العدم H_0 ، ونقبل الفرضية البديلة H_1 ، التي تقول أن المتحول X لا يخضع للتوزيع الطبيعي المفروض .

لذلك يجب علينا حساب قيمة λ من العلاقة : $\lambda = D\sqrt{n}$ ثم البحث في جداول λ عن الاحتمالP ، الذي يقابلها ، وذلك حسب مستوى الدلالة $\propto$ وحجم العينة n فنجد أن:

$$1 - K(\lambda) = P(\lambda , \propto) = P$$

ملاحظة وتعديل: نلاحظ من المثال السابق أن أكبر الفروقات D يحدث مقابل الطرف الأيمن لأحد المجالات الوسطى وليكن $]x_{K-1}, x_K]$ ، ولقد أظهرت بعض التجارب العملية على المتحولات المستمرة X، أن ذلك الفرق D قد لا يحدث في نقطة واحدة، بل يمكن أن يحدث في عدة نقاط أخرى متجاورة . لذلك يفضل عند دراسة طبيعية المتحولات المستمرة X أن نقوم بدراسة الفروقات D عند الطرف الأيمن للمجال ، الذي يليه وهو المجال $]x_K , x_{K+1}]$ ، ومقارنة الاحتمال النظري $F(X_{k+1})$ مع الاحتمال التجريبي $P(X_K)$. أوبعبارة أخرى مقارنة $F(X_k)$ مع $P(X_{K-1})$.

وهذا يقتضي منا أن نقوم بحساب فروقات إضافية هي الفروقات بين تابع التوزيع النظري $F(X_k)$ وتابع التوزيع التجريبي $P(X_{K-1})$ المزاح إلى اليمين بمقدار خطوة واحدة فقط . ثم حساب أكبر الفروقات الجديدة D_2 من العلاقة :

$$D_2 = max|F(X_K) - P(X_{K-1})| \qquad (61-3)$$

حيث يفترض أن نضع $P(X_0) = 0$. في المجال الأول (الصفري) .

وإذا رمزنا لأكبر الفروقات السابقة المعرفة بالعلاقة (3−53) بالرمز D_1 . فإنه يكون لدينا :

$$D_1 = max|F(X_K) - P(X_K)| \qquad (62-3)$$

فإننا نقوم بحساب أكبر هاتين القيمتين ، ونرمز لها بالرمز D ونحسبها من العلاقة :

$$D = max|D_1 , D_2| \qquad (63-3)$$

ولاتخاذ القرار حول H_0 نقارن D مع قيمة $D_\propto$ الحرجة ونتخذ القرار كما يلي:

إذا كانت $D \le D_\propto$ نقبل فرضية العدم H_0

أما إذا كانت $D > D_\propto$ نرفض فرضية العدم H_0 ونقبل H_1 .

مثال (3−3): لنتابع معالجة المثال (3−1) ونحسب كل من D_1 و D_2 ونكتب بيانات الجدول (3−6) كما يلي:

جدول (3−7): توابع التوزيعات $F(X_K)$ و $P(X_K)$ و $P(X_{K-1})$ المزاح

رقم المجال j	1	2	3	4	5	6
المجال $]X_K , X_K[$	0 −10	10 −30	20 −30	30 −40	40 −50	50 −60
$P(X_K)$	0.05	0.25	0.40	0.85	0.95	1
$F(X_K)$	0.0492	0.2043	0.50	0.7957	0.9508	0.9939
$\|F(X_K) - P(X_K)\|$	0.0008	0.0497	0.10	0.0543	0.0008	0.0061
$P(X_{K-1})$	0	0.05	0.25	0.40	0.85	0.95
$\|F(X_K) - P(X_{K-1})\|$	0.0492	0.1543	0.25	0.3957	0.1008	0.0439

وهكذا نجد من السطر الخامس أن D_1 تساوي :

$$D_1 = max|F(X_K) - P(X_K)| = 0.10$$

ومن السطر السابع نجد أن قيمة D_2 تساوي :

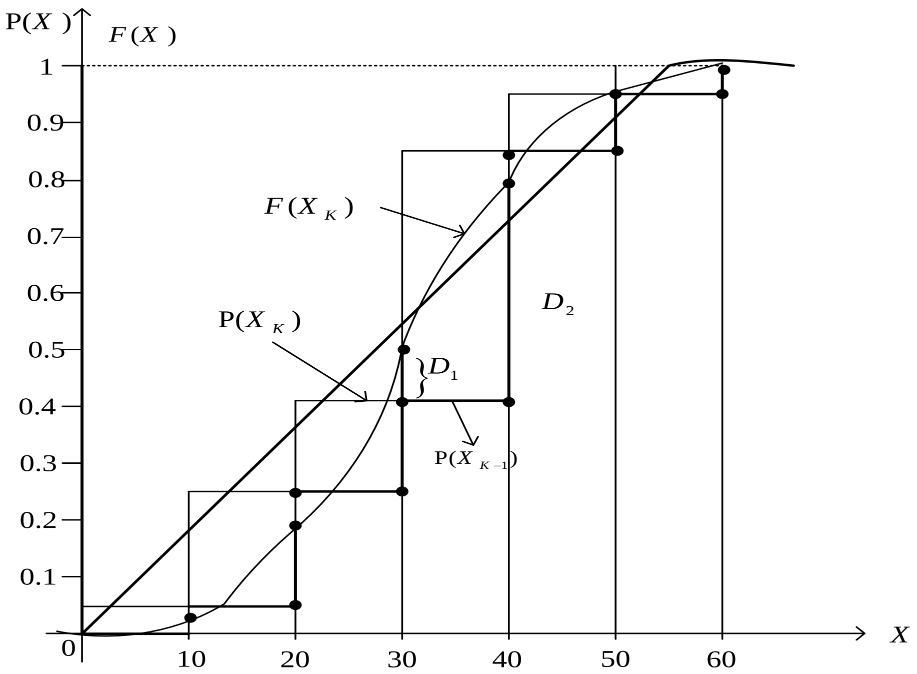
$$D_2 = max|F(X_K) - P(X_{K-1})| = 0.3957$$

ومن هاتين القيمتين نحسب القيمة الكبرى D من العلاقة :

$$D = max[D_1, D_2] = max[0.10\ \ , 0.3957] = 0.3957$$

ومن جداول $(K - S)$ نجد أن قيمة $D_\propto$ الحرجة تساوي $D_\propto = D_{0.05} = \frac{1.36}{\sqrt{100}} = 0.136$

وعند المقارنة نجد أن $D > D_\propto$ لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تقول أن X لا يخضع للتوزيع الطبيعي المفروض (وهذا يتفق مع اختبار χ^2) .



الشكل (3–10): تمثيل الفروقات D_1 و D_2 بعد انزياح المدرج التكراري خطوة واحدة إلى اليمين

ملاحظة: يفضل عند اختبار طبيعية المتحولات المستمرة استخدام التعديل السابق وتطبيق العلاقات (3–61) و(3–62) و(3–63). أما إذا كان المتحول المدروس X منقطعاً (ومرتباً) فلا ضرورة لذلك.

3–3–3: اختبار (ليليفورز Lilliefors) للتوزيع الطبيعي المعياري $N(0\ ,1)$.

وهو حالة خاصة من اختبار (كولموغوروف – سميرنوف) لأنه يدرس مدى توافق بيانات العينة مع التوزيع الطبيعي المعياري الذي متوسطه يساوي الصفر وتباينه يساوي الواحد .

ولتطبيق هذا الاختبار، علينا أولاً أن نقوم بمعالجة بيانات العينة لتصبح صالحة للمقارنة مع التوزيع الطبيعي المعياري $N(0\ ,1)$ ، لذلك نقوم بالإجراءات التالية:

1−نقوم بحساب متوسط العينة $\bar{x}$ وتباينها S^2 من العلاقتين التاليتين :

$$\bar{x} = \sum_{j=1}^{m} \frac{n_j}{n} x'_j \qquad \left[x'_j = \text{متوسط المجال } j\right] \qquad (64-3)$$

$$S^2 = \frac{1}{n-1} \sum \frac{n_j}{n} \left(x'_j - \bar{x}\right)^2 \qquad (65-3)$$

2− نقوم بحساب القيم المعيارية لبيانات العينة من العلاقة:

$$Z_i = \frac{x_i - \bar{x}}{S} \qquad (66-3)$$

وبعد تبويبها نحصل على مجالات جديدة ذات أطراف جديدة $[Z_{K-1}, Z_K[$ ، مقابلة لنفس تابع التوزيع التجريبي $P(X_K)$ ونرمز له بـ $P(Z_K)$.

3− نقوم بحساب قيم تابع التوزيع الطبيعي المعياري $\phi(Z_K)$ ، المتكاملة من $-\infty$ حتى نهاية المجال k، ومن العلاقة التكاملية التالية :

$$\phi(Z_K) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_K} e^{-\frac{Z^2}{2}} dz \qquad (67-3)$$

ثم نضعها مقابل الاحتمالات التجريبية $P(Z_K)$ في جدول مناسب كالتالي:

جدول (3−8): بيانات المثال

j	1	2	3	4	5	6
المجالات $[Z_K, Z_{K+1}[$	$[Z_0, Z_1[$	$[Z_1, Z_2[$	$[Z_2, Z_3[$	$[Z_3, Z_4[$	$[Z_4, Z_5[$	$[Z_5, Z_6[$
$P(Z_K)$	$P(Z_1)$	$P(Z_2)$	$P(Z_3)$	$P(Z_4)$	$P(Z_5)$	$P(Z_6)$
$\phi(Z_K)$	$\phi(Z_1)$	$Q(Z_2)$	$\phi(Z_3)$	$\phi(Z_4)$	$\phi(Z_5)$	$\phi(Z_6)$
$T_K = \lvert P(Z_K) - \phi(Z_K) \rvert$	T_1	T_2	T_3	T_4	T_5	T_6

4−ونضع الفرضيتين كما يلي : $H_0: P(Z_K) = \phi(Z_K) \quad , \quad H_1: P(Z_K) \neq \phi(Z_K)$

5−نقوم بحساب الفروقات بين تابعي التوزيع (النظري والتجريبي) ثم نحدد أكبر هذه الفروقات بالقيمة المطلقة ، فنحصل على أن:

$$T = max\lvert P(Z_K) - \phi(Z_K) \rvert \qquad (68-3)$$

6−لاتخاذ القرار حول H_0 . نحدد مستوى الدلالة $\propto$ ثم نبحث في جداول اختبار (ليليفورز) عن قيمة $T_\propto$ الحرجة ، ونتخذ القرار كما يلي :

إذا كانت قيمة : $T \leq T_\propto$ نقبل فرضية العدم H_0 .

أما إذا كانت قيمة $T > T_\propto$ نرفض فرضية العدم H_0 ونقبل H_1 .

مثال (3-4): لنتابع معالجة المثال (3-1) وندرس فيما إذا كانت بيانات تلك العينة خاضعة للتوزيع الطبيعي العام باستخدام اختبار (ليليفورز) .

نلاحظ أن المطلوب في هذا المثال هو مقارنة بيانات هذه العينة مع التوزيع الطبيعي المعياري لذلك علينا أن نجري ما يلي:

1-نقوم بحساب متوسط العينة وتباينها $\bar{x}$ و S^2 فنجد من الحسابات السابقة أن:

$$\bar{x} = \sum_{j=1}^{6} \frac{n_j}{n} x'_j = 30$$

$$S^2 = \sum_{j=1}^{6} \frac{n_j}{n} \left(x'_j - \bar{x}\right)^2 = 146.546$$

ومنها نجد أن الانحراف المعياري للعينة: $S = \sqrt{146.546} = 12.1$

2-نقوم بحساب القيم المعيارية لأطراف مجالات الجدول (3-8) فنحصل على مجالات جديدة $]Z_{K-1}, Z_K[$ تقابل نفس الاحتمالات التجميعية لتابع التوزيع الاحتمالي التجريبي $P(X_K)$ والذي سنرمز له بـ $P(Z_K)$.

3-نقوم بحساب قيم تابع التوزيع الطبيعي المعياري $\phi(Z_K)$. فنحصل على الجدول التالي:

جدول (3-9): بيانات المثال:

j	1	2	3	4	5	6
المجال المعياري (Z_{K-1}, Z_K)	−2.48 −1.65	−1.65 −0.83	−0.83 0	0 0.83	0.83 1.65	1.65 2.48
$P(Z_K)$	0.05	0.25	0.40	0.85	0.95	1
$\phi(Z_K)$	0.04947	0.2033	0.50	0.7967	0.9505	0.9934
$T = \lvert P(Z_K) - \phi(Z_K) \rvert$	0.00053	0.0467	0.10	0.0533	0.0005	0.0066
المزاح $P(Z_{K-1})$	0	0.05	0.25	0.40	0.85	0.95
$T_2 = \lvert \phi(Z_K) - P(Z_{K-1}) \rvert$	0.04947	0.1533	0.25	0.3967	0.1005	0.0434

ومن السطر الخامس في الجدول (3-9) نلاحظ أن أكبر قيمة لـ T تساوي :

$$T_1 = max\lvert P(Z_K) - \phi(Z_K) \rvert = 0.10$$

ومن جداول اختبار (ليليفورز) نجد أن القيمة الحرجة لـ T عند مستوى دلالة $\propto = 0.05$ تساوي تقريباً ما يلي: $T_{0.05} = \frac{0.886}{\sqrt{n}} = 0.0886$، وبالمقارنة نجد أن $T_1 > T_\propto$، لذلك نرفض فرضية العدم H_0 ونقبل H_1 التي تقول أن Z لا يخضع للتوزيع الطبيعي المعياري، وبالتالي فإن X لا يخضع للتوزيع الطبيعي العام .

ولتدارك أي خطأ ممكن نقوم بحساب الفروقات المزاحة T_2 ونضعها في السطر الأخير . ومنه نجد أن :

$$T = max[T_1 , T_2] = max[0.10 ,0.3967] = 0.3967$$

وبالمقارنة نجد أن : $T > T_{\propto}$ لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تقول أن المتحول Z لا يخضع للتوزيع الطبيعي المعياري وبالتالي فإن X لا يخضع للتوزيع الطبيعي العام .

3-3-4: اختبار (P- P) للاحتمالات:

إن الرمز $(P-P)$ مأخوذ من الكلمات (مخطط الاحتمالات مقابل الاحتمالات Probability-Probability plot)، وإن هذا الاختبار ينطلق من أفكار (كولموغوروف) ويعتمد على حساب تابعي الاحتمالات التجميعيين تصاعدياً وهما:

- **تابع الاحتمالات االتجميعي التجريبي ونرمز له بـ** $P(X_K)$ **وهو يحقق** $0 \le P(x) \le 1$.
- **تابع الاحتمالات التجميعي النظري ونرمز له بـ** $F(X_K)$ **وهو يحقق** $0 \le F(x) \le 1$.

ونكتب قيمهما المقابلة لقيم أو لمجالات X كما يلي (انظر الجدول (3-6))

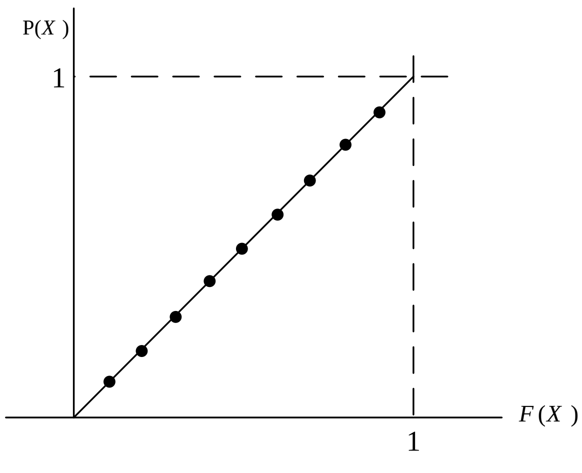
$$\begin{array}{llllllll} X: & X_1 & X_2 & X_3 & \dots & X_K & \dots & X_m \\ P(X_K): & P(X_1) & P(X_2) & P(X_3) & \dots & P(X_K) & \dots & P(X_m)=1 \\ F(X_K): & F(X_1) & F(X_2) & F(X_3) & \dots & F(X_K) & \dots & F(X_m)\approx 1 \end{array} \quad (69-3)$$

ثم نضع الفرضتين كما يلي:

فرضية العدم H_0 **: إن** $P(X_K) = F(X_K)$ **من أجل جميع القيم** X_K .

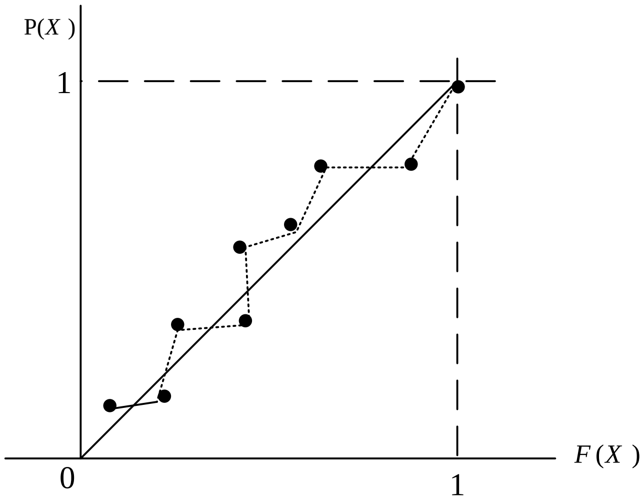
الفرضية البديلة H_1**: إن** $P(X_K) \neq F(X_K)$ **من أجل قيمة ما** X_K **(واحدة على الأقل)**

وهنا نلاحظ أنه إذا كانت فرضية العدم محققة فإن جميع قيم $P(X_K)$ تساوي قيم $F(X_K)$ المقابلة لها (بحدود خطأ معينة) ، وهذا يعني أنه لو رسمنا القيم المتقابلة لهما $\big(F(X_K), P(X_K)\big)$ على شكل نقاط على المستوى FOP ، فإننا سنحصل (في حالة التساوي) على نقاط متفرقة ولكنها تقع على الخط المستقيم المنصف للربع الأول ، وتقع ضمن المربع الأول الذي طول ضلعه يساوي الواحد (لأن أكبر قيمة لهما تساوي الواحد) وسيظهر الشكل كما يلي :



الشكل (3-11): مخطط $P(X_K)$ **مقابل** $F(X_K)$ **عند** H_0

ومن الشكل (3−11) نستنتج ببساطة أن المستقيم المنصف للربع الأول يمثل حالة تحقق فرضية العدم H_0
أما إذا كانت H_0 غير محققة فإن H_1 تكون هي المحققة، وعندها يكون لدينا $P(X_K) \neq F(X_K)$ من أجل بعض النقاط X_K، وعندما نرسم القيم المتقابلة لها على شكل نقاط هندسية $\big(F(x), P(x)\big)$ على المستوى FOP، فإننا سنحصل على نقاط لا تقع على المستقيم المنصف للربع الأول . بل تقع على جانبيه أو على جانب واحد منه . وعندها سيظهر لدينا الشكل التالي :



الشكل (3−12): مخطط $P(X_K)$ مقابل $F(X_K)$ عند H_1

فإذا كانت تلك النقاط قريبة جداً من المستقيم المنصف فإن ذلك يدل على أن X يخضع للتوزيع المفروض $\not{f}(x)$ (قد يكون التوزيع الطبيعي أو غيره) وإن الفرضية H_0 محققة .
أما إذا كانت مواقع تلك النقاط بعيدة نسبياً عن ذلك المستقيم فإن ذلك يؤشر على أن X لا يخضع للتوزيع المفروض $\not{f}(x)$، وللتأكد من ذلك نقوم بحساب معامل الارتباط الخطي بين التابعين التراكميين $P(X_K)$ و $F(X_K)$ من العلاقة المعروفة التالية :

$$r = \frac{\sum^{m}(F_K - \bar{F})(P_K - \bar{P})}{(m-1) * \sigma_P * \sigma_F} \qquad (70-3)$$

فإذا كانت قيمة r كبيرة $(r \geq 0.70)$ فإننا نعتبر أن X يخضع للتوزيع $\not{f}(x)$ ويمكن اختبار معنوية قيمة r من اختبار (ستودينت) التالي: (ضمن فرضية العدم: $H_0: \rho = 0$ مقابل $H_1: \rho > 0$)

$$t = \frac{r\sqrt{m-2}}{\sqrt{1-r^2}} \qquad (71-3)$$

فإذا كانت $|t| \leq t_{\propto}$ تقبل $(H_0: \rho = 0)$ التي تقول أن معامل الارتباط ρ معدوماً، ولذلك نرفض H_0 الأساسية حول X ونعتبره غير خاضع لـ $\not{f}(x)$ والعكس بالعكس .

مثال (3-5): نعود إلى المثال (3-1) وإلى بياناته المبوبة في الجدول (3-5) فنجد أنها تأخذ الشكل التالي:

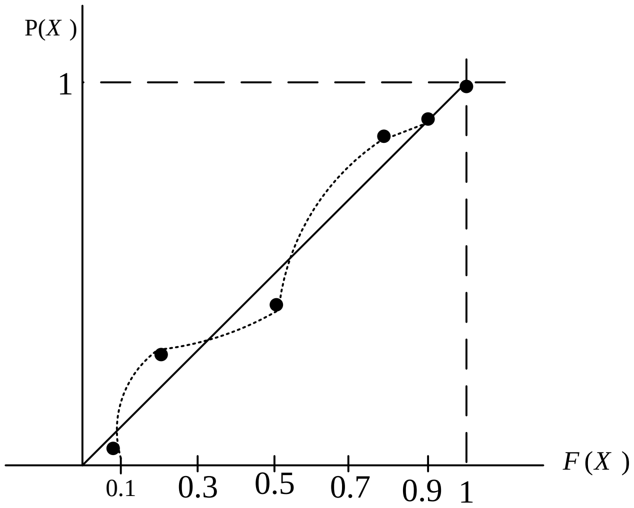
جدول (3-10):

المجال K	1	2	3	4	5	6	المتوسط$\bar{P}$	σ_P
$P(X_K)$	0.05	0.25	0.40	0.85	0.95	1	0.5893	0.40208
$\phi(X_K) = F(X_K)$	0.0492	0.2043	0.50	0.7957	0.9508	0.9934	0.5656	0.3792

حيث أن: $P(X_K)$ هو تابع التوزيع التجريبي التجميعي والمحسوب من بيانات العينة

وأن: $F(X_K)$ هو تابع التوزيع النظري التجميعي والمحسوب من التوزيع الطبيعي $N(30,\ 12.1)$ ، الذي نوقعه 30 وانحرافه المعياري 12.1 (مقدرين من بيانات العينة) .

ثم نقوم برسم نقاط الاحتمالات التجميعية المتقابلة $[F(x), P(x)]$ على المستوى FOP فنحصل على الشكل التالي:



الشكل (3-13): مخطط رسم $P(X)$ مقابل $F(X)$ للمثال (3-1)

ونلاحظ من الشكل أن النقاط الهندسية $[F(x), P(x)]$ تقع على جانبي المستقيم (ثلاثة منها قريبة جداً منه وثلاثة ليست بعيدة عنه) . لذلك يمكن اعتبار بيانات العمر X المأخوذة من بيانات العينة المسحوبة طبيعية أو شبه طبيعية، وللتأكد من ذلك نحسب معامل الارتباط بين $P(X_K)$ و $F(X_K)$ فنجد أن:

$$r = \frac{\sum(F_K - \bar{F})(P_K - \bar{P})}{(m-1) * \sigma_F * \sigma_P} = 0.9866$$

وهي قيمة ممتازة تدل على ارتباط متين بين $P(X_K)$ و $F(X_K)$. ولاختبار معنوية r نحسب المؤشر t من العلاقة :

$$t = \frac{r\sqrt{m-2}}{\sqrt{1-r^2}} = 12.0938$$

ومن جداول توزيع (ستودينت)، نجد أن القيمة الحرجة لمتحول (ستودينت) t المقابلة لدرجة حرية v تساوي لـ $(v = m - 2 - 1 = 3)$ ولمستوى دلالة $\propto = 0.05$ تساوي $t_{0.05\,,3} = 2.3534$

وبمقارنةt المحسوبة مع $t_{\propto}$ الحرجة نجد أن: $t > t_{\propto}$، لذلك نرفض فرضية العدم حول معامل الارتباط $(H_0 : \rho = 0)$ ونعتبر أن قيمة معامل الارتباط المحسوبة r هي قيمة معنوية . ومنها نستنتج أن العلاقة الخطية بين التابعين $P(X_K)$ و $F(X_K)$ هي علاقة معنوية، وبالتالي يمكننا أن نعتبر أن متحول العمر X يخضع للتوزيع الطبيعي المفروض $N(30\,,12.1)$.

3-3-5: اختبار (Q- Q) للكميات:

إن الرمز $(Q - Q)$مأخوذ من الكلمات (مخطط الكميات مقابل الكميات Quantile-Quantile plot) ويعتمد هذا الاختبار على مقارنة القيم الكمية للمتحول المدروس X مع القيم الكمية لمتحول التوزيع الطبيعي المعياري Z، وينطلق من الفكرة التالية:

إن قيم المتحول الطبيعي المعياري Z تتحول من $-\infty$ إلى $+\infty$، ولكنها تتجمع حول مركزها $(\bar{Z} = 0)$ وتزداد كثافتها كلما كانت قريبة من ذلك المركز، ولذلك فإنه إذا سحبنا عينة عشوائية من مجتمع طبيعي، فإن قيم Z المقابلة والمرتبة ستتوزع على الجانبين، ولكنها ستكون متقاربة أو متلاصقة أو متطابقة مع بعضها في المنطقة الوسطى (منطقة المركز).

وكذلك الأمر بالنسبة للمتحول المدروس X . فإذا كان X خاضعاً للتوزيع الطبيعي العام أو المعياري . فإنه لابد وأن تتوزع قيم المشاهدات المأخوذة من عينة عشوائية بحجم مقبول n على الجانبين، وأن تتجمع حول المركز $\bar{x}$. وتزداد كثافتها وتقاربها من بعضها كلما اقتربت من المركز $\bar{x}$.

وللاستفادة من هذه الخاصة نقوم بمقارنة القيم الكمية المرتبة لـ X مع القيم الكمية المرتبة لـ Z . فإذا كانت قيم X تتجمع بشكل مشابه لتجمع Z . فإننا نستنتج أن المتحول X يخضع للتوزيع الطبيعي العام .

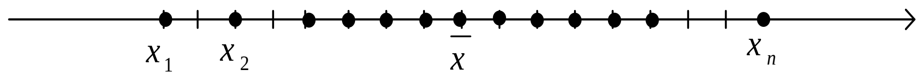
والسؤال الآن هو: كيف سنقوم بمقارنة القيم الكمية لهذين المتحولين X و Z ؟

للجواب على هذا السؤال يجب أولاً تحديد القيم الكمية لكل من X وZ . ثم ترتيبها ومقارنة كل قيمة من قيم X المرتبة مع القيمة المقابلة لها مع قيم Z المرتبة .

إن تحديد القيم الكمية لـ X يتم من خلال سحب عينة عشوائية بحجم n مفردة والحصول منها على n مشاهدة، ثم القيام بترتيبها تصاعدياً، فنحصل على القيم الكمية المرتبة التالية:

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_K \leq \cdots x_{n-1} \leq x_n \qquad (72-3)$$

ثم نقوم برسمها على المحور OX كما يلي:

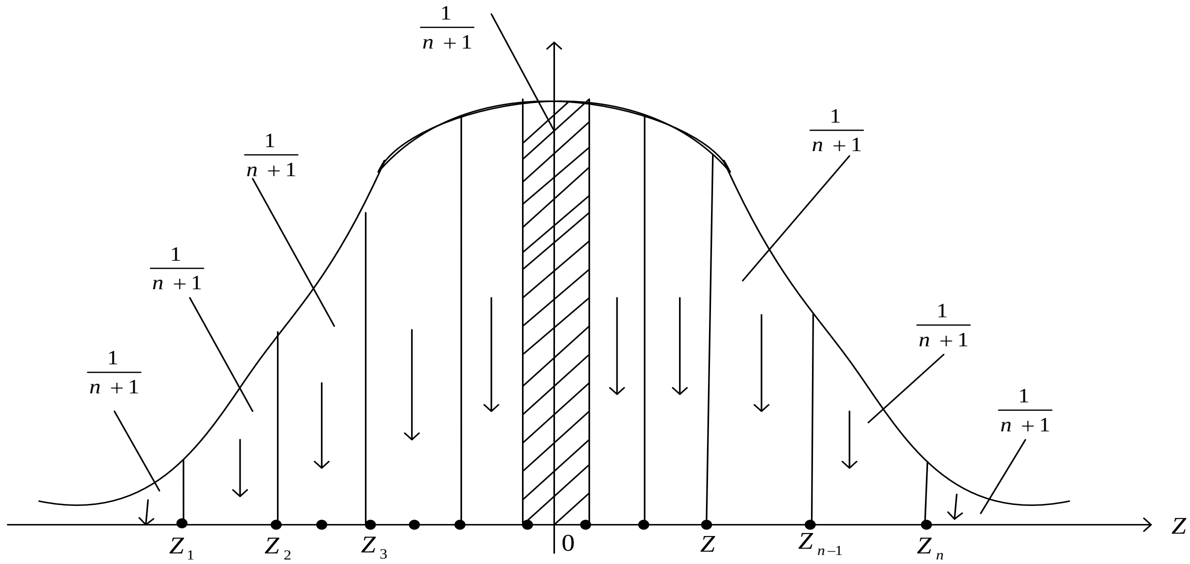


الشكل (3–14): توضع قيم X على المحور OX

ولتحديد القيم الكمية لـ Z المعياري، نلجأ إلى تقسيم المساحة تحت منحني التوزيع الطبيعي المعياري إلى مساحات متساوية، ولذلك نرسم عدداً n من النقاط على المحور OZ (بعدد نقاط X). وبذلك يتشكل لدينا $(n+1)$ مجالاً جزئياً، ويتم تحديد مواقع القيم الكمية لتلك النقاط على المحور OZ، بحيث تكون قيم الاحتمالات المقابلة لتلك المجالات الجزئية متساوية، أي نشترط أن تكون المساحات التي تحت المنحني والمقابلة لكل مجال جزئي متساوية كما في الشكل (3−15) .

وبما أن المساحة الكلية تحت منحني التوزيع تساوي الواحد، وأن عدد المجالات يساوي $(n+1)$ مجالاً جزئياً . فإن الاحتمال المقابل لكل مجال جزئي يجب أن يساوي $(\frac{1}{n+1})$.

ولإنجاز هذه المهمة نقوم بتقسيم المساحة التي تحت المنحني إلى $(n+1)$ قسماً متساوياً، فنحصل على مساحات متساوية، ولكنها تقابل مجالات جزئية غير متساوية . والشكل التالي يوضح ذلك :



الشكل (3−15): تقسيم المساحة تحت المنحني للحصول على قيم Z

وهنا نلاحظ أن الاحتمال الذي يقابل كل مجال جزئي أصبح يساوي $(\frac{1}{n+1})$، وبذلك نجد أن تابع التوزيع التجميعي للتوزيع الطبيعي المعياري $\phi(Z_K)$ يساوي مجموع أو تكامل الاحتمالات المقابلة لتلك المجالات حتى النقطة Z_k ويساوي :

$$\phi(Z_K) = P(Z \leq Z_K) = \frac{K}{n+1} = \int_{-\infty}^{Z_K} \frac{1}{\sqrt{2\pi}} e^{\frac{Z^2}{2}} d_z \qquad (73-3)$$

حيث أن: k هو عدد المجالات المحسوبة .

واعتماداً على هذه العلاقة يمكننا حساب قيم Z_K المقابلة للاحتمال $\frac{K}{n+1}$ من جداول $\phi(Z)$ فنجد أن:

$$Z_K = \phi^{-1}\left(\frac{K}{n+1}\right) \qquad (74-3)$$

علماً بأن Z_K هو الحد الأعلى للمجال k، ولكن بما أن القيم الكمية لـ X عندما يكون مستمراً قد تكون مقابلة لمراكز المجالات الجزئية للمعلومات المبوبة، فقد تم تعديل الصيغة (3−73) إلى صيغة ملائمة

لمنتصفات المجالات، وذلك بأخذ التكامل حتى منتصف المجال الأخير (Z_{k-1}, Z_k) ، وبذلك تأخذ العلاقة (3−73) الصيغة التالية:

$$\phi(Z_K) = P(Z \le Z_K) = \frac{K - \frac{1}{2}}{n} \qquad (75-3)$$

ومنها يتم حساب Z_K كما يلي:

$$Z_K = \phi^{-1}\left(\frac{K - \frac{1}{2}}{n}\right) \qquad (76-3)$$

ملاحظة 1: هناك تعديلات كثيرة على الصيغة (3−73) نذكر أهمها كما وردت في [Wikipedia] وهي:

$$\phi(Z_K) = \frac{K - 0.3}{n + 0.4} \quad , \quad \phi(Z_K) = \frac{K - 0.375}{n + 0.25}$$

$$\phi(Z_K) = \frac{K - 0.3175}{n + 0.365} \quad , \quad \phi(Z_K) = \frac{K - \frac{1}{3}}{n + \frac{1}{3}} \qquad (77-3)$$

$$\phi(Z_K) = \frac{K - 0.4}{n + 0.12} \quad , \quad \phi(Z_K) = \frac{K - 0.567}{n + 0.134}$$

$$\phi(Z_K) = \frac{K - 1}{n - 1}$$

وأخيراً نشير إلى أنه يمكن استنتاج جميع هذه العلاقات من العلاقة العامة التالية:

$$\phi(Z_K) = \frac{K - a}{n + 1 - 2a} \qquad (78-3)$$

حيث: $0 \le a \le 1$. وعند إعطاء a أية قيمة في ذلك المجال نحصل على إحدى العلاقات (3−77)، وعندما نضع $a = \frac{1}{2}$ نحصل على العلاقة (3−75) .

كما نشير إلى أن أحد الإحصائيين [Fillipin] قد اقترح الاعتماد على العلاقة التالية :

$$\phi(Z_K) = \begin{cases} 1 - 0.5^{\frac{1}{n}} & : \ K = 1 \\ \dfrac{\mathrm{k} - 0.3175}{n + 0.365} & : \ K = 2\ 3\ 4 \ldots n - 1 \\ 0.5^{\frac{1}{n}} & : \ K = n \end{cases} \qquad (79-3)$$

ملاحظة 2: نلاحظ من الشكل (3−15) أنه إذا وضعنا المجال الجزئي الأيمن (الأخير) مساوياً لنصف مستوى الدلالة $\frac{\propto}{2}$ (إذا كان الاختبار ثنائي الجانب) ، وبذلك يكون لدينا :

$$\frac{1}{n+1} = \frac{\propto}{2} \implies n = \frac{2}{\propto} - 1 \approx \frac{2}{\propto} \qquad (80-3)$$

ومنه يمكننا حساب الحد الأدنى لحجم العينة اللازم لتحقيق مستوى الدلالة $\propto$ من العلاقة: $n \approx \frac{2}{\propto}$

فإذا كان $\propto = 0.05$ فإننا نجد أن:

$$n = \frac{2}{0.05} - 1 = 40 - 1 = 39 \approx 40$$

وإذا كان $\propto = 0.10$ فإننا نجد أن:

$$n = \frac{2}{0.10} - 1 = 19 \approx 20$$

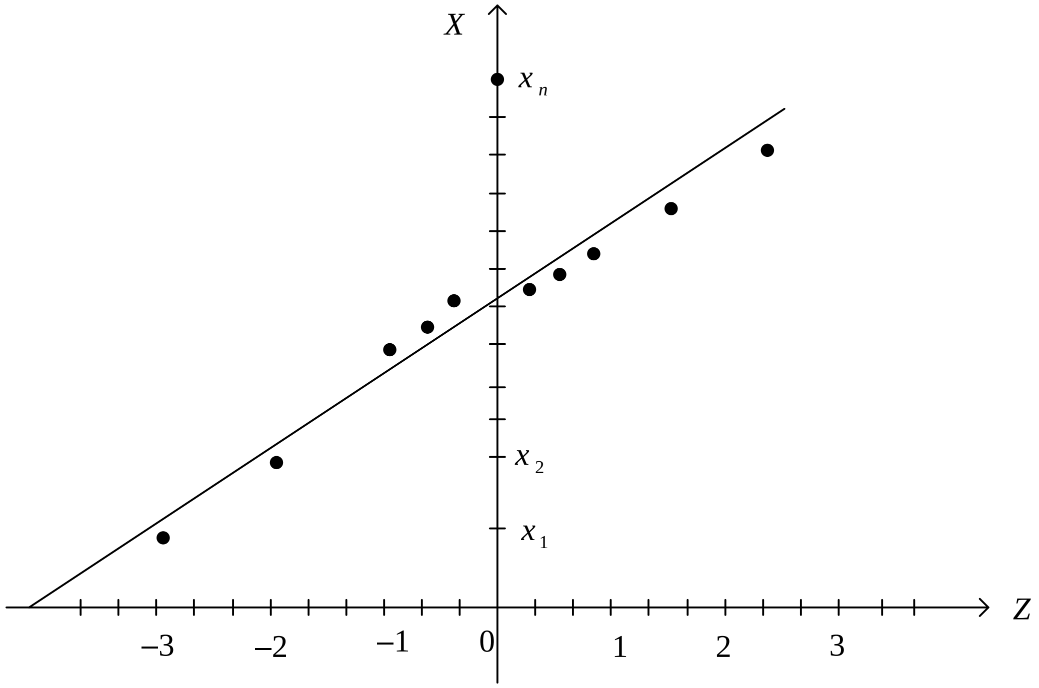
والآن نعود إلى العلاقة (3−76) فنحصل منها على n قيمة لـ Z_K تكون مرتبة تصاعدياً كما يلي:

$$Z_1 < Z_2 < Z_3 \ldots Z_K < \cdots Z_n \qquad (81-3)$$

وبعدها نقارن توضع قيم X المبينة في العلاقة (3−72) مع توضع قيم Z المبينة في (3−81) ونشكل منهما الأزواج المتقابلة التالية:

$$(Z_1, X_1)\ (Z_2\ , X_2)\ (Z_3\ , X_3) \ldots (Z_K\ , X_K) \ldots\ (Z_n\ , X_n) \qquad (82-3)$$

ثم نرسم هذه النقاط الهندسية على المستوى ZOX، فنحصل على الشكل التالي:



الشكل (3−16): مخطط Q-Q لـ X على Z

فإذا كان توضع القيم X_k متوافقاً مع توضع القيم Z_k، يكون توزيع X متوافقاً مع توزيع Z، وعندها فإن تلك النقاط تكون واقعة تقريباً على خط مستقيم، وإن ذلك يعني أن X يخضع للتوزيع الطبيعي العام، أما إذا كان توضع النقاط مشتتاً عنه، فإن X لا يخضع للتوزيع الطبيعي العام . وللتأكد من ذلك نقوم بحساب معامل الارتباط الخطي بين X و Z من العلاقة :

$$r = \frac{\sum(X_K - \bar{X})(Z_K - \bar{Z})}{(n-1)\sigma_X * \sigma_Z} = \frac{\sum(X_K - \bar{X})Z_K}{(n-1)\sigma_X} \qquad (83-3)$$

وذلك لأن: $\bar{Z} = 0$ و $\sigma_Z = 1$

ثم نقوم بمقارنة قيمة r مع القيمة الحرجة $r_{\propto,n}$ المقابلة لـ $\propto$ وn، فإذا كانت $r \geq r_{\propto}$ نعتبر الارتباط بين X و Z معنوياً ونقبل فرضية العدم التي تقول أن X يخضع للتوزيع الطبيعي العام .

ومن أجل تنفيذ هذا الاختبار بشكل صحيح نستعين بالجدول المختصر للقيم الحرجة لمعامل الارتباط r عند مستويي دلالة $\propto= 0.05$ و$\propto= 0.10$ وحسب حجوم مختلفة للعينة n التالي:

جدول (3–11): القيم الحرجة لمعامل الارتباط r حسب n و$\propto$:

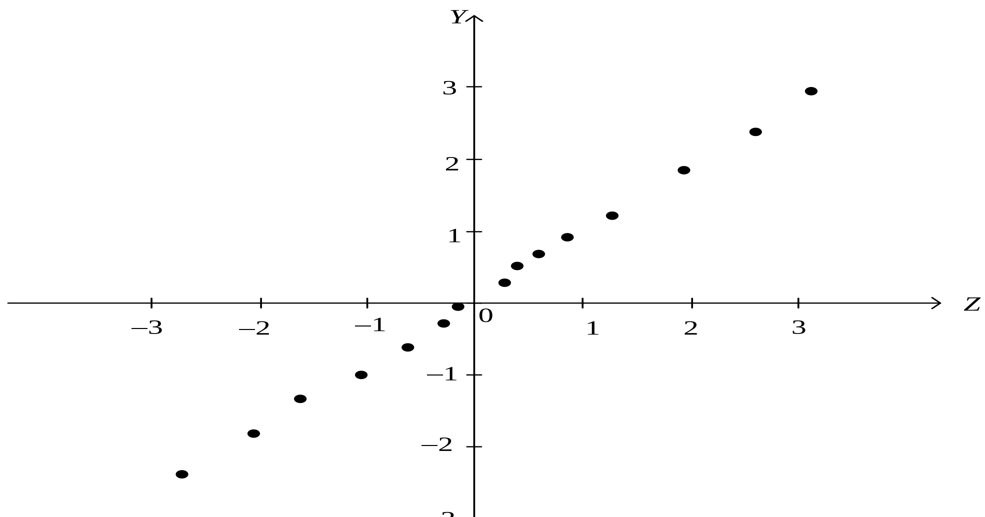
حجم العينة n	10	15	20	25	30	40	50	60	75	100	150	200
مستوى الدلالة $\propto= 0.05$	0.918	0.938	0.950	0.958	0.964	0.972	0.976	0.980	0.984	0.986	0.991	0.993
مستوى الدلالة $\propto= 0.10$	0.935	0.951	0.960	0.966	0.971	0.977	0.981	0.984	0.987	0.989	0.992	0.994

المصدر: [Johnston P. 151]

ملاحظة 3: يمكننا إجراء معيرة(معايرة) لقيم X قبل مقارنتها مع قيم Z (وهو الأفضل) وذلك بإجراء التحويل التالي عليها :

$$Y_K = \frac{X_K - \bar{X}}{\sigma_{X_k}} \quad : \quad K = 1\ 2\ 3 \ldots n \quad \text{حيث} \qquad (84-3)$$

وبذلك نحصل على متحول معياري Y_K متوسطه $\bar{Y} = 0$ وتباينه $\sigma_Y^2 = 1$، وبعد الحصول على القيم Y_K نشكل الأزواج المتقابلة $(Z_K\,, Y_K)$ ونرسمها على المستوى ZOY فنحصل على الشكل الممعير التالي :



الشكل (3–17): المخطط الممعير لـ Q – Q

وهنا نضع الفرضتين كما يلي:

- فرضية العدم H_0: وتصاغ كما يلي $H_0: Y_k = Z_k$ من أجل جميع قيم k، وهذا يعني أن قيم المتحول Y تتطابق مع قيم المحول المعياري Z، أي Y يخضع للتوزيع الطبيعي المعياري . ومنه نستنتج أن المتحول الأصلي X يخضع للتوزيع الطبيعي العام .
- الفرضية البديلة H_1 : وتصاغ كما يلي: $H_1: Y_k \neq Z_k$ من أجل بعض قيم k . وهذا يعني أن Y لا يخضع للتوزيع الطبيعي المعياري وبالتالي فإن X لا يخضع للتوزيع الطبيعي العام .

ولاختبار تحقق الفرضيةH_0، نقوم بحساب معامل الارتباط الخطي بين المتحولين Y وZ من العلاقة :

$$r = \frac{\sum(Y_K - \bar{Y})(Z_K - \bar{Z})}{(n-1)\sigma_Y * \sigma_Z} = \frac{\sum Y_K * Z_K}{(n-1)} \qquad (85-3)$$

وذلك لأن $\bar{Z} = 0$ و $\sigma_Z = 1$ ولأن $\bar{Y} = 0$ و $\sigma_Y = 1$

ثم نقارن قيمة r المحسوبة مع القيمة الحرجة $r_{\propto,n}$ حسب $\propto$ و n ، فإذا كانت $r \geq r_{\propto}$ فإننا نقبل فرضية العدم H_0 الأساسية ونعتبر المتحول Y خاضعاً للتوزيع المعياري . وبالتالي يكون المتحول X خاضعاً للتوزيع الطبيعي العام .

أما إذا كان $r < r_{\propto}$ فإننا نرفض H_0 ونقبل H_1 ...الخ .

ملاحظة 4: **يمكن اختبار معنوية** r **باستخدام مؤشر (ستودينت)** t **وحسابه من العلاقة** (ضمن $H_0: \rho = 0$ مقابل $H_1: \rho > 0$

$$t = \frac{r-0}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (86-3)$$

ثم نقارن قيمة t المحسوبة مع قيمة $t_{\propto}$ الجدولية المقابلة لمستوى الدلالة $\propto$ ولدرجة الحرية $n-2$ ، ونقرر كما يلي:

إذا كان $t \leq t_{\propto}$ فإننا نقبل فرضية $H_0: \rho = 0$، ونعتبر أن معامل الارتباط بين Y وZ غير معنوي، وإن Y لا يخضع لتوزيع Z .

أما إذا كانت $t > t_{\propto}$ نرفض فرضية $H_0: \rho = 0$ ونقبل فرضية العدم الأساسية $H_0: Y_k = Z_k$. ونعتبرY يخضع للتوزيع الطبيعي المعياري ونستخلص أن X يخضع للتوزيع الطبيعي العام . وإذا كان العكس فبالعكس .

ملاحظة 5: إن مخططات $Q-Q$ لا تكون مفيدة، إذا كان حجم العينة n صغيراً $(n < 20)$. ولقد لاحظنا مما سبق أن الحد الأدنى لحجم العينة n وعند مستوى دلالة $\propto= 0.10$ يساوي $n = 19$.

ولهذا يجب عدم الاعتماد على مخططات $Q-Q$ في حالة العينات الصغيرة، لتحديد مدى تقارب توزيع المحول X من التوزيع الطبيعي العام أو المعياري .

مثال (3–6): [مأخوذ من Johnston, Wichern P.148 بتصرف] لنفترض أننا نريد معرفة التوزيع الاحتمالي، الذي تخضع له درجة الحرارة عند الفجر في مكان ما وخلال شهر الاعتدال (شهر آذار)، فأخذنا عينة عشوائية مؤلفة من $n = 10$ أيام من آذار (مارس) فوجدنا بعد القياس أن درجات الحرارة عند الفجر في تلك الأيام كانت كما يلي:

$$T: 23.5, -10.0\,, 17.1\,, -1.0\,, 15.4\,, 1.6\,\,, 1.26\,\,, 4.1\,, 8.5\,\,, 6.2$$

ولمعالجة هذه البيانات واختبار خضوعها للتوزيع الطبيعي (أو غيره) نقوم بما يلي:

1- نرتب قيم هذه المشاهدات تصاعدياً . ونضعها في السطر الأول من الجدول المخصص لذلك ونرمز لها بـ X_K .

2- نرقم المشاهدات المرتبة من 1 حتى 10 . ونضعها أرقامها في السطر الثاني ونرمز للرقم العام لها بالرمز k .

3- نحسب قيم الاحتمالات التراكمية للتوزيع الطبيعي المعياري التي تقابل كل مشاهدة من هذه المشاهدات، من العلاقة المعدلة (3-75) التالية :

$$P_K = \frac{K - \frac{1}{2}}{n}$$

ونضع هذه الاحتمالات التراكمية في السطر الثالث ، فمثلاً نجد أن:

$$P_1 = \frac{1 - \frac{1}{2}}{10} = 0.05 \qquad P_2 = \frac{2 - \frac{1}{2}}{10} = 0.15 \quad \dots\dots\dots.$$

4- نقوم بحساب قيم المتحول الطبيعي المعياري Z_K من العلاقة (3-76) فنجد أن:

$$Z_K = \phi^{-1}\left(\frac{K - \frac{1}{2}}{n}\right)$$

فمثلاً نجد أن: $Z_1 = \phi^{-1}(0.05) = -1.645$

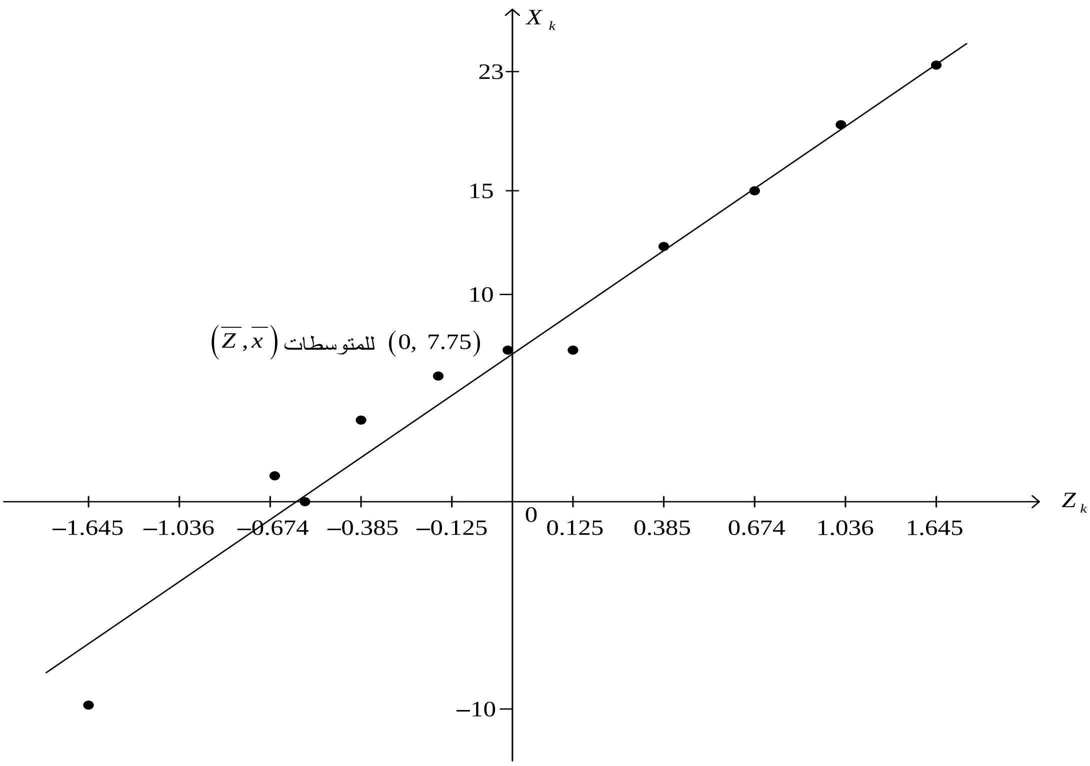
وهكذا $Z_2 = \phi^{-1}(0.15) = -1.036 \; \dots\dots\dots$

ثم نضعها في السطر الرابع، فنحصل على الجدول التالي:

جدول (3-12): بيانات المثال

قيم المشاهدات X_K المرتبة	−10.0	−1.0	1.6	4.1	6.2	8.0	12.6	15.4	17.1	23.5
أرقام المشاهدات المرتبة K	1	2	3	4	5	6	7	8	9	10
الاحتمالات التراكمية المقابلة للمشاهدات المرتبة P_K	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
قيم المتحول الطبيعي Z_K المقابلة للاحتمالات P_K	−1.645	−1.036	−0.674	−0.385	−0.125	0.125	0.385	0.674	1.036	1.645

5- نقوم برسم النقاط الهندسية (X_K, Z_K) على المستوى ZOX فنحصل على الشكل التالي:



الشكل (3−18): تمثيل بيانات المثال

ومن الشكل (3−18) السابق نلاحظ أن النقاط قريبة من الخط المستقيم المرسوم . أي أن درجة الحرارة T يمكن أن تكون خاضعة للتوزيع الطبيعي العام .

6−نقوم بحساب معامل الارتباط بين المتحولين X_K و Z_K من العلاقة (3−85) فنجد أنه يساوي:

$$r = \frac{\sum(Z_K - \bar{Z})(X_K - \bar{X})}{(n-1)\sigma_Z * \sigma_X} = 0.994$$

وبمقارنة هذه القيمة المحسوبة لـ r مع القيمة الحرجة $r_{\propto}$ المقابلة لمستوى دلالة $\propto = 0.05$ ولـ $n = 10$ والمساوية $r_{\propto} = 0.918$ ، نجد أن $r > r_{\propto}$. لذلك نرفض فرضية العدم حول معامل الارتباط $(H_0: \rho = 0)$. ونعتبر أن قيمة r التي حصلنا عليها هي قيمة معنوية . أي أن العلاقة الارتباطية بين المتحولين X و Z هي علاقة معنوية ومتينة، ومنها نستنتج أن X يخضع للتوزيع الطبيعي العام .

لأن هذا يجعلنا نقبل فرضية العدم حول X التي تقول أن X يخضع للتوزيع الطبيعي العام .

ملاحظة: كان يمكن معالجة هذه البيانات بعد تحويلها إلى بيانات طبيعية معيارية باستخدام علاقة التحويل التالية:

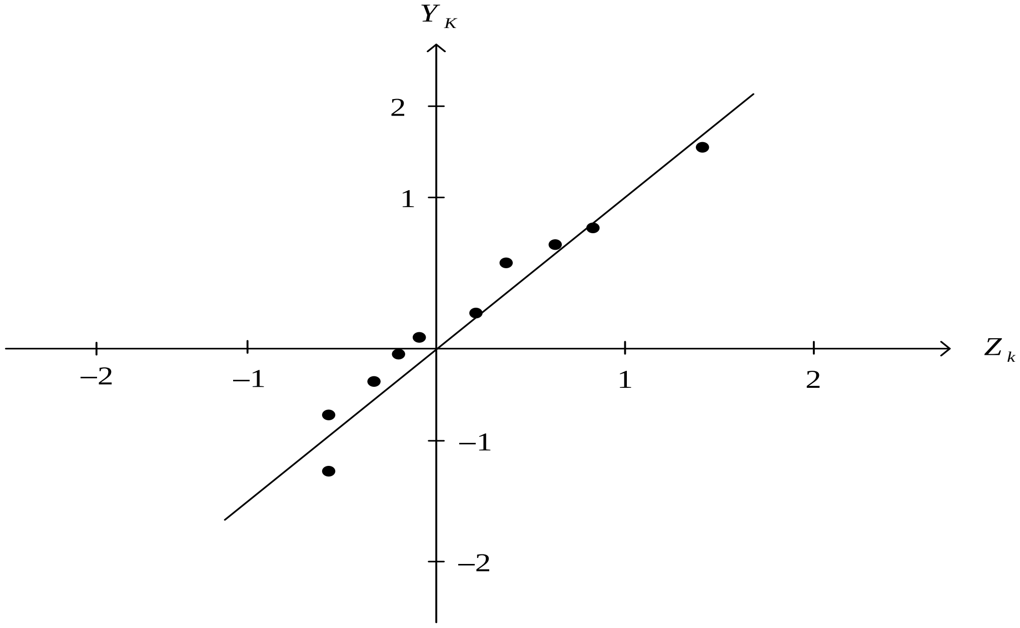
$$Y_K = \frac{(X_K - \bar{X})}{\sigma_{X_k}} \qquad (87-3)$$

من الجدول (3−12) نجد أن $\bar{X} = 7.75$ وأن $\sigma_X = 9.70$ ، وبذلك نجد أن قيم Y_K المعيارية المقابلة لقيم Z_K المعيارية تأخذ الشكل التالي:

جدول (3-13): البيانات المعيارية للمثال (3-6)

K	1	2	3	4	5	6	7	8	9	10
Y_K	−1.825	−0.897	−0.629	−0.371	−0.155	0.031	0.505	0.794	0.59	1.577
Z_K	−1.645	−1.036	−0.674	−0.385	−0.125	0.125	0.385	0.674	1.036	1.645

ثم نقوم برسم النقاط الهندسية (Z_K, Y_K) على المستوى ZOY فنحصل على الشكل التالي:



الشكل (3-19): تمثيل البيانات الممعيرة للمثال

ثم نقوم بحساب معامل الارتباط من العلاقة:

$$r = \frac{\sum(Z_K - \bar{Z})(Y_K - \bar{Y})}{(n-1)\sigma_Z * \sigma_X} = \frac{\sum Z_K * Y_K}{n-1} = 0.994$$

وبمقارنة هذه القيمة لـ r مع القيمة الحرجة $r_{0.05} = 0.918$ نجد أن $r > r_{\propto}$ لذلك نعتبر العلاقة معنوية ونقبل بأن X يخضع للتوزيع الطبيعي العام .

3-3-6: اختبار التوزيع الطبيعي لمتحولين (X_1, X_2) أو أكثر:

لنرمز لشعاع المتحولين بالرمز بـ $Z = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. ولشعاع التوقع الرياضي لهما بـ $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$.

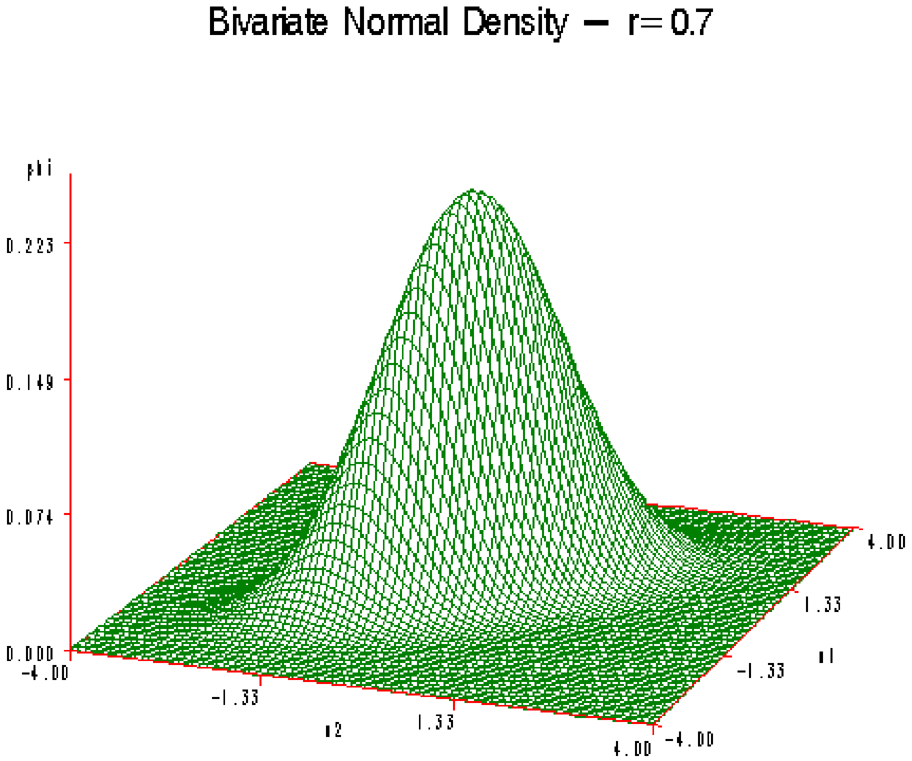
ولمصفوفة التباين المشترك لها بالرمز V :

$$V = cov(X) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad (89-3)$$

وعندها نجد أن التوزيع الاحتمالي المشترك للمتحولين (X_1, X_2) يعطى حسب (3-12) بالعلاقة :

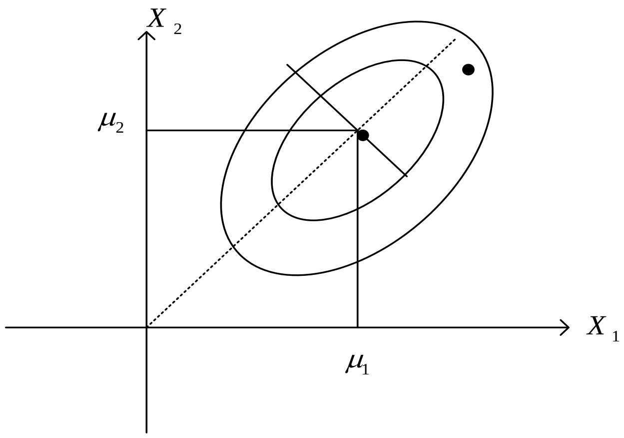
$$f(X_1, X_2) = \frac{1}{2\pi|V|^{\frac{1}{2}}} e^{-\frac{1}{2}[(X-\mu)' * V^{-1}(X-\mu)]} \qquad (90-3)$$

وهو يشكل سطحاً على شكل هضبة انسيابية فوق المستوى $X_1 O\ X_2$ كما في الشكل التالي:



الشكل (3-20): سطح التوزيع الطبيعي الثنائي

وإن أي مقطع له موازٍ للمستوى $X_1 O\ X_2$ ، يكون على شكل قطع ناقص . وإن مسقطه على $X_1 O\ X_2$ يكون أيضاً على شكل قطع ناقص كما في الشكل التالي :



الشكل (3-21): مساقط مقاطع سطح التوزيع الطبيعي الثنائي

ويتم تحديد معادلة القطع الناقص عند أي مقطع C (محدد) بوضع المعادلة (3-90) تساوي C، فنحصل على المعادلة التالية:

$$f(x_1\,, x) = \frac{1}{2\pi|V|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\boldsymbol{X}-\boldsymbol{\mu})'*\boldsymbol{V}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]} = C \qquad (90-3)'$$

وهذا يعني أن الأس يجب أن يساوي قيمة ثابتة C_1، وللحصول على ذلك القطع الناقص كان يكفي أن نضع الأس مساوياً لعدد ثابت C_1 فنحصل على المعادلة التالية:

$$(\boldsymbol{X}-\boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X}-\boldsymbol{\mu}) = C_1 \qquad (91-3)$$

وهي تمثل معادلة مساقط جميع القطوع الناقصة الناتجة عن مقاطع مجسم التوزيع الطبيعي المشترك $f(x_1, x_2)$.

علماً بأن الطرف الأيسر من العلاقة (3–91) هو عبارة عن مربع المسافة الإحصائية D^2 لأي نقطة من المستوى $X_1 O X_2$ عن المركز الحقيقي $\boldsymbol{\mu}(\mu_1, \mu_2)$ وهو يخضع لتوزيع $\chi^2(D^2)$ ذي $P = 2$ درجة حرية .

لذلك يمكننا أن نكتب المعادلة (3–91) على الشكل التالي:

$$D^2 = (\boldsymbol{X}-\boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi_2^2(D^2) \qquad (92-3)$$

وعندما يكون μ و V مجهولين فإننا نقدرهما من متوسط العينة $\bar{x}$ ومن مصفوفة التباين المشترك للعينة S فنحصل على تقدير لـ D^2 من العلاقة:

$$d^2 = (\boldsymbol{X}-\overline{\boldsymbol{X}})' * \boldsymbol{S}^{-1} * (\boldsymbol{X}-\overline{\boldsymbol{X}}) \sim \chi_2^2(d^2) \qquad (93-3)$$

ولقد تم تحديد وتمييز أهم ثلاثة قطوع ناقصة لهذه التوزيعات الاحتمالية هي:

1- القطع الناقص الأول: وهو القطع الذي يكون الاحتمال الذي فوقه مساوياً لـ 0.50، أي هو القطع الذي يكون الحجم الذي فوقه وتحت سطح التوزيع مساوياً لـ 50% . من أصل الحجم الكلي المساوي للواحد . ويتم تحديد هذا القطع من العلاقة :

$$P(d^2 \le C_1) = 0.50 \qquad (94-3)$$

وبما أن d^2 يخضع لـ $\chi_2^2(d^2)$ فإنه يمكننا كتابة هذه العلاقة كما يلي:

$$P\left(d^2 \le \chi_2^2(0.50)\right) = 0.50 \qquad (95-3)$$

ومنها ومن جداول $\chi_P^2(X)$ يمكننا حساب قيمة $\chi_2^2(0.50)$ الحرجة والمقابلة للاحتمال 0.50 . فنجد من تلك الجداول أن $\chi_2^2(0.50) = 1.39$. لذلك نكتب العلاقة (3–95) كما يلي:

$$P(d^2 \le 1.39) = 0.50 \qquad (96-3)$$

وهي معادلة القطع الناقص الأول .

2- القطع الناقص الثاني: وهو القطع الذي يكون الاحتمال الذي فوقه مساوياً لـ 0.95 . ويتم تحديد معادلته بنفس الأسلوب، فنحصل على المعادلة التالية:

$$P(d^2 \le 5.99) = 0.95 \qquad (97-3)$$

3- القطع الناقص الثالث: وهو القطع الذي يكون الاحتمال الذي فوقه مساوياً لـ 0.99 . ويتم تحديد معادلته بنفس الأسلوب ، فنحصل على المعادلة التالية :

$$P(d^2 \le 9.21) = 0.99 \qquad (98-3)$$

ويستفاد من هذه القطوع . وخاصة من القطع الأول في اختبار طبيعية التوزيع الاحتمالي لأي متحولين (X_1 , X_2). ولذلك نقوم بالإجراءات التالية :

1–نسحب عينة عشوائية من المجتمع المدروس بحجم n ونأخذ منها البيانات عن X_1 و X_2 .

2–نحسب المتوسطين $\bar{x}_1$ و $\bar{x}_2$ من بيانات العينة .

3–نحسب التباينين المصححين S_1^2 و S_2^2 من بيانات العينة .

4–نحسب مصفوفة التباين المشترك ونرمز لها بـ $\mathbf{S}$ ثم نقوم بحساب مقلوبها $\boldsymbol{S^{-1}}$.

5–نحسب معادلة مربع المسافة d^2 من العلاقة التالية:

$$d^2 = (\boldsymbol{X} - \boldsymbol{\bar{X}})' * \boldsymbol{S^{-1}} * (\boldsymbol{X} - \boldsymbol{\bar{X}}) \qquad (99-3)$$

6–نقوم باختبار كل نقطة $\left(X_{1j} , X_{2j}\right)$ في بيانات العينة فيما إذا كانت تحقق شرط الوقوع ضمن القطع الأول من العلاقة :

$$d_j^2 \leq 1.39 \qquad j: 1\ 2\ 3 \dots n \qquad (100-3)$$

فإذا كانت نسبة النقاط التي تقع ضمن ذلك القطع أكثر من 50% فإن ذلك يشير إلى أن بيانات العينة المدروسة تخضع للتوزيع الطبيعي الثنائي . ولكن ذلك ليس كافياً بل يجب تطبيق الاختبار $Q - Q$ على مربعات المسافات d_j^2 . لذلك نتابع الاختبار بإجراء ما يلي :

7–نقوم بترتيب المربعات d_j^2 تصاعدياً فنحصل على ما يلي:

$$d_K^2 :\ d_1^2 \leq d_2^2 \leq d_3^2 \leq \cdots d_K^2 \dots \leq d_n^2 \qquad (101-3)$$

ثم نقوم بحساب الاحتمالات التراكمية المقابلة لهذه المربعات فنحصل من العلاقة (3–75) على أن:

$$P_K :\ \frac{1-\frac{1}{2}}{n}, \frac{2-\frac{1}{2}}{n}, \frac{3-\frac{1}{2}}{n} \ \dots\ \frac{K-\frac{1}{2}}{n} \ \dots\ \frac{n-\frac{1}{2}}{n} \qquad (102-3)$$

وبما ان d^2 يخضع للتوزيع $\chi_2^2(X)$ فإننا نحسب القيم الكمية للتوزيع $\chi_2^2(X)$ من العلاقة:

$$\chi_K^2 = \chi_2^2\left(\frac{K-\frac{1}{2}}{n}\right) \qquad (103-3)$$

فنحصل على القيم الكمية التالية:

$$\chi_K^2 = \chi_2^2\left(\frac{1-\frac{1}{2}}{n}\right)\ \chi_2^2\left(\frac{2-\frac{1}{2}}{n}\right)\ \chi_2^2\left(\frac{3-\frac{1}{2}}{n}\right)\dots\ \chi_2^2\left(\frac{K-\frac{1}{2}}{n}\right)\dots\ \chi_2^2\left(\frac{n-\frac{1}{2}}{n}\right) \qquad (104-3)$$

8–نقوم بتشكيل ورسم النقاط الهندسية التالية:

$$\left[d_K^2 \quad , \quad \chi_2^2\left(\frac{K-\frac{1}{2}}{n}\right)\right] \qquad (105-3)$$

على المستوى الاحداثي d^2OX^2 فنحصل على شكل الانتشار للمسافات d_K^2 مقابل القيم $\chi_2^2(X)$ ، فإذا كان ذلك الشكل على هيئة مستقيم ، فإننا نعتبر المتحولين (X_1 , X_2) خاضعين للتوزيع الطبيعي الثنائي والعكس بالعكس .

وللتحقق من ذلك نقوم بحساب معامل الارتباط الخطي r بين القيم الكمية d_K^2 والقيم الكمية $\chi^2(X)$ ، ثم نقارنه مع القيمة الحرجة $r_\propto$. فإذا كان $r \geq r_\propto$ فإننا نعتبر أو نقبل بأن المتحولين (X_1 , X_2) يخضعان للتوزيع الطبيعي الثنائي (وذلك حسب بيانات العينة المسحوبة). كما يمكن اختبار معنوية r من العلاقة (3–86) ومقارنة قيمة t المحسوبة مع القيمة الحرجة $t_{\propto,\ n-1}$ ، فإذا كانت $t > t_\propto$ نقبل فرضية بأن المتحولين (X_1 , X_2) يخضعان للتوزيع الطبيعي الثنائي . والعكس بالعكس .

ملاحظة 7: يمكن تعميم أفكار وإجراءات هذه الفقرة على P متحولاً، مثل : $X_1\ X_2\ X_3 \ldots X_P$ وتطبيق نفس الخوارزمية لاختبار خضوع هذه المتحولات إلى التوزيع الطبيعي المتعدد، مع الانتباه إلى أن القيم الكمية $\chi_P^2(X)$ ستصبح بـ P درجة حرية .

مثال (3–7): [مأخوذ من Johnston ,Wichern P.152 بتصرف وإضافة] لدراسة أحوال للشركات الرئيسية في أمريكا تم اعتماد البيانات المالية لأكبر /10/ شركات صناعية في الولايات المتحدة الأمريكية وفي عام 1978 فكانت كما يلي:

جدول (3–14) بيانات الشركات:

اسم الشركة	G.M	Exxon	Ford	Mobil	Texaco	Std.oil	IBM	Gulf	G.E	Chresler
الموجودات X_1 (مليون دولار)	26.7	38.4	19.2	20.6	18.9	14.8	19.0	14.2	13.7	7.7
الدخل الصافي X_2 (مليون دولار)	3.3	2.4	1.7	1.0	0.9	1.0	2.7	0.8	1.1	2.5

حيث رمزنا للموجودات بـ X_1 وللدخل الصافي بـ X_2 والمطلوب اختبار فيما إذا كان المتحولان (X_1 , X_2) يخضعان للتوزيع الطبيعي الثنائي .

الحل: لدراسة هذه القضية وإجراء ذلك الاختبار، نقوم بحساب المتوسطين $\bar{x}_1$ و $\bar{x}_2$. ثم حساب مصفوفة التباين المشترك S فنجد أن:

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 19.32 \\ 1.51 \end{bmatrix} \qquad S = \begin{bmatrix} 70.41 & 5.87 \\ 5.87 & 0.97 \end{bmatrix}$$

ثم نحسب المقلوب S^{-1} فنجد أنه يساوي :

$$\boldsymbol{S^{-1}} = \frac{1}{(70.41)(0.97) - (5.87)^2} \begin{bmatrix} 0.97 & -5.87 \\ -5.87 & 70.41 \end{bmatrix} = \begin{bmatrix} 0.029 & -0.173 \\ -0.173 & 2.080 \end{bmatrix}$$

وهكذا نجد أن معادلة مربع المسافة من أية نقطة $(x_1\ x_2)_K$ إلى المركز $(\bar{x}_1, \bar{x}_2)$ تعطى بالعلاقة التالية:

$$d_K^2 = (X - \bar{x})' * S^{-1} * (X - \bar{x})_K$$

وإن معادلة النقاط التي داخل وعلى محيط القطع الأول تساوي:

$$d_K^2 = \begin{bmatrix} (x_1 - 19.32) \\ (x_2 - 1.51) \end{bmatrix}' * \begin{bmatrix} 0.029 & -0.173 \\ -0.173 & 2.080 \end{bmatrix} * \begin{bmatrix} (x_1 - 19.32) \\ (x_2 - 1.51) \end{bmatrix} \leq 1.39$$

ثم نقوم بتعويض إحداثيات كل نقطة $(x_1\ x_2)$ في معادلة القطع الأول . ولنبدأ من النقطة الأولى (26.7 , 3.3) لشركة جنرال موتورز G.M فنجد أن مربع المسافة d_1^2 يساوي :

$$d_1^2 = \begin{bmatrix}(26.7-19.32)\\(3.3-1.51)\end{bmatrix}' * \begin{bmatrix}0.029 & -0.173\\-0.173 & 2.080\end{bmatrix} * \begin{bmatrix}(26.7-19.32)\\(3.3-1.51)\end{bmatrix} = 3.67$$

وبمقارنة هذه القيمة مع العدد (1.39) الخاص بمحيط القطع الأول نجد أن:

$$\mathrm{d}_1^2 = 3.67 \geq 1.39$$

وهذا يعني أن هذه النقطة تقع خارج القطع الأول، ثم نقوم بحساب مربعات المسافات d_K^2 للشركات الأخرى فنحصل على أنها تساوي:

$$\mathrm{d}_1^2 = 3.67 \quad \mathrm{d}_2^2 = 6.32 \quad \mathrm{d}_3^2 = \underline{0.07} \quad \mathrm{d}_4^2 = \underline{0.81} \quad \mathrm{d}_5^2 = \underline{0.69}$$
$$\mathrm{d}_6^2 = \underline{0.34} \quad \mathrm{d}_7^2 = 3.08 \quad \mathrm{d}_8^2 = \underline{0.56} \quad \mathrm{d}_9^2 = \underline{0.47} \quad \mathrm{d}_{10}^2 = 2.21$$

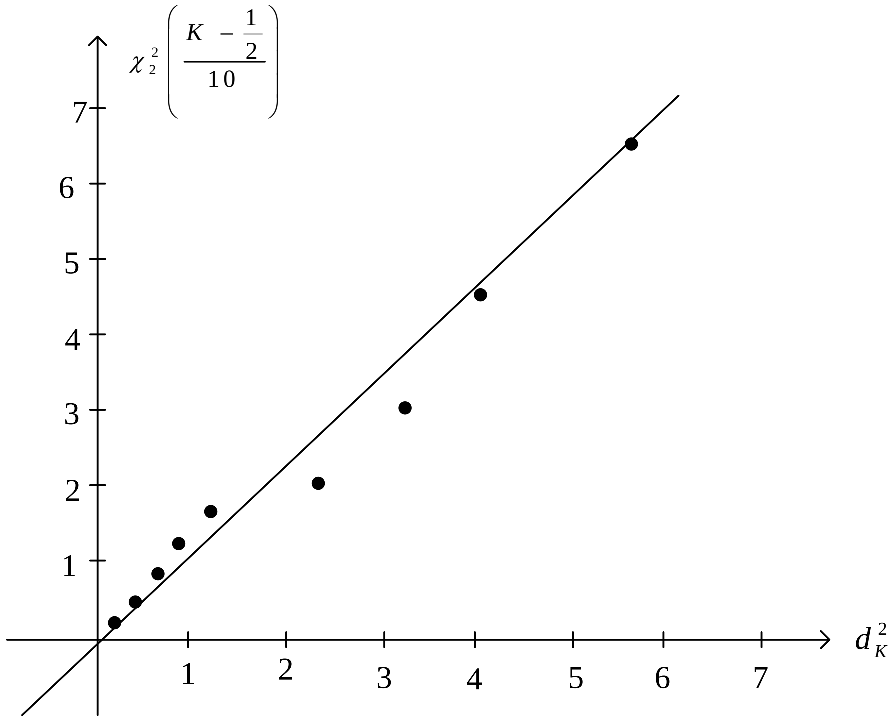
وبمقارنة هذه المربعات مع العدد (1.39) الخاص بالقطع الأول نجد أن /6/ شركات تحقق الشرط $d_j \leq 1.39$ ، أي أن نسبة الشركات التي تقع ضمن القطع الأول تساوي 60% . وهذا يشير إلى أن بيانات هذه الشركات (X_1, X_2) تخضع للتوزيع الطبيعي . أي أنه لا يمكننا أن نرفض فرضية خضوع بيانات هذه الشركات إلى التوزيع الطبيعي الثنائي .

ولمتابعة التحليل والاختبار نقوم بتطبيق الاختبار $Q-Q$ على d_K^2 . لذلك نقوم بترتيبها تصاعدياً وحساب الاحتمالات التراكمية المقابلة لكل منها $\frac{K-\frac{1}{2}}{n}$ ، ثم حساب القيم الكمية لـ $\chi_2^2(\frac{K-\frac{1}{2}}{n})$ فنحصل على الجدول المرتب تصاعدياً حسب d_K^2 التالي :

جدول (3–15): حساب بيانات المثال (3–7) المرتبة تصاعدياً حسب d_K^2

قيم d_K^2 المرتبة	0.07	0.34	0.47	0.55	0.69	0.81	2.21	3.08	3.69	6.32	1.821
الرقم k	1	2	3	4	5	6	7	8	9	10	
الاحتمالات $\frac{K-\frac{1}{2}}{10}$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	
قيمة الكميات $\chi_2^2(\frac{K-\frac{1}{2}}{10})$	0.10	0.33	0.58	0.86	1.20	1.60	2.10	2.77	3.79	5.99	1.932

ثم نقوم برسم النقاط الهندسية $\left[d_K^2,\ \chi^2(\frac{K-\frac{1}{2}}{10})\right]$ فنحصل على الشكل التالي :



الشكل (3−22): تمثيل العلاقة بين d_K^2 و $\chi^2(\frac{K-\frac{1}{2}}{10})$.

ومن الشكل (3−22) نلاحظ أن معظم النقاط متقاربة من المستقيم المرسوم . وللتأكد من ذلك نقوم **بحساب قيمة معامل الارتباط بين القيم الكمية d_K^2 والقيم الكمية** $\chi^2(\frac{K-\frac{1}{2}}{10})$ **فنجد أنها تساوي:** $r = 0.9877$ وبمقارنة هذه القيمة مع القيمة الحرجة $r_\propto$ عند $\propto= 0.05$ و$n = 10$ والمساوية 0.917 نجد أن: $r \geq r_\propto$ لذلك نقبل فرضية أن يكون الارتباط معنوياً، وبالتالي نقبل م يكون المتحولان $(x_1\ x_2)$ خاضعين للتوزيع الطبيعي الثنائي .

الفصل الرابع
الاستدلال حول شعاع المتوسطات (لعدة متحولات طبيعية)

4-1: تمهيد:

يتناول هذا الفصل مسألة الاستدلال حول شعاع المتوسطات (لعدة متحولات) في المجتمع من خلال تطبيق الاختبارات المناسبة على الفرضيات المطروحة . ولهذا فإننا نقدم بعض المفاهيم اللازمة لاستعراض خواص ومركبات المتوسطات، وذلك اعتماداً على مفاهيم مجالات ومناطق الثقة المعروفة . علماً بأن قواعد التحليل متعدد المتحولات تقتضي أن تتم معالجة المتحولات المترابطة بشكل مشترك ، وهذا ما سنعمل به عند استعراض الطرائق المستخدمة في هذا المنشور .

لنفترض- في البداية- إننا ندرس أحوال متحول واحد X في مجتمع طبيعي توقعه μ (مجهول)، وتباينه مجهول، وإن القيمة الظاهرية لمتوسط X في ذلك المجتمع تساوي μ_0، لذلك نضع حول μ_0 الفرضيتين الإحصائيتين كما يلي:

فرضية العدم H_0 : ونكتبها كما يلي : $H_0: \mu = \mu_0$

الفرضية البديلة H_1 : ونكتبها (ثنائية الجانب) كما يلي: $H_1: \mu \neq \mu_0$

وللتحقق من الفرضية H_0 نسحب عينة عشوائية بحجم n من عناصر ذلك المجتمع ونأخذ منها قيم المتحول المدروس X ، فنحصل على القيم التالية :

$$X: x_1 , x_2 , x_3 , \dots , x_n \qquad (1-4)$$

ثم نحسب متوسطها $\bar{x}$ وتباينها المصحح S^2 من العلاقتين :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (2-4)$$

ولإجراء اختبار صحة فرضية العدم H_0 حول μ_0 نستخدم مؤشر اختبار (ستودينت) المعرف بالعلاقة التالية:

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \qquad (3-4)$$

علماً بأن المؤشر t هو متحول عشوائي جديد يخضع لتوزيع (ستودينت) بـ $(n-1)$ درجة حرية .

وبعد حساب قيمة t نقارنها مع القيمة الحرجة لمتحول (ستودينت) المقابلة لنصف مستوى الدلالة $\propto$ ولـ $(n-1)$ درجة حرية . والتي يرمز لها بـ $t_{n-1}\left(\frac{\propto}{2}\right)$ (لأن الاختبار ثنائي الجانب، $\frac{\propto}{2}$ لكل جانب) .

فإذا كانت القيمة المطلقة لـ $|t|$ أصغر أو تساوي $t_{n-1}\left(\frac{\propto}{2}\right)$ ، فإننا نقبل فرضية العدم H_0 بمستوى دلالة $\propto$ ($\frac{\propto}{2}$ لكل جانب) ونكتب القاعدة كما يلي:

نقبل الفرضية H_0 إذا كانت :

$$|t| \le t_{n-1}\left(\frac{\propto}{2}\right) \qquad (4-4)$$

أما إذا كانت العلاقة (4−4) غير محققة، فإننا نضطر إلى رفض الفرضية H_0 . وعندها نستخلص أن القيمة الظاهرية μ_0 هي قيمة غير مقبولة لتمثيل التوقع المجهول μ في المجتمع الطبيعي المدروس .
وهنا نطرح السؤال التالي: هل هناك قيم أخرى (غير μ_0) يمكن اعتبارها قيماً مقبولة لـ μ ؟ . والجواب هو: نعم . لأنه في الحقيقة يوجد دائماً مجموعة من القيم المقبولة لتقدير توقع المجتمع μ . وهي كما هو معلوم، تشكل مجالاً (أو منطقة) تسمى منطقة قبول الفرضية H_0 مقابل H_1 . كما يطلق عليها اسم مجال (أو منطقة) الثقة لـ μ_0 المقابل لمستوى الدلالة $\propto$.
ولتحديد ذلك المجال نكتب العلاقة (4−4) كما يلي :

$$\left|\frac{\bar{x}-\mu_0}{S/\sqrt{n}}\right| \le t_{n-1}\left(\frac{\propto}{2}\right) \qquad (5-4)$$

ثم نعالجها كما يلي فنجد أن:

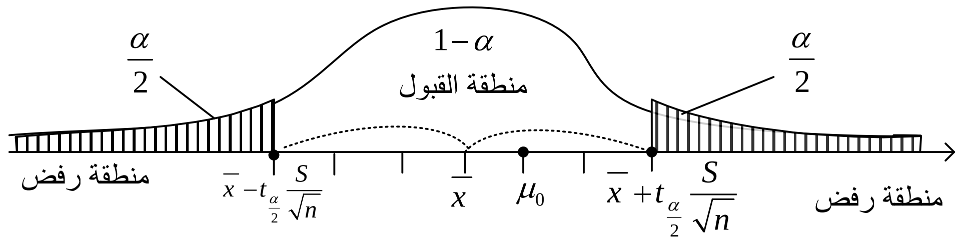
$$-t_{n-1}\left(\frac{\propto}{2}\right) \le \frac{\bar{x}-\mu_0}{S/\sqrt{n}} \le +t_{n-1}\left(\frac{\propto}{2}\right)$$

$$-t_{n-1}\left(\frac{\propto}{2}\right)\frac{S}{\sqrt{n}} \le \bar{x}-\mu_0 \le +t_{n-1}\left(\frac{\propto}{2}\right) * \frac{S}{\sqrt{\text{n}}}$$

ومنها نحصل على مجال الثقة التالي (نطرح $\bar{x}$ من الأطراف ثم نضربها بـ -1 ونغير الاتجاهات) .

$$\bar{x}-t_{n-1}\left(\frac{\propto}{2}\right)\frac{S}{\sqrt{n}} \le \mu_0 \le \bar{x}+t_{n-1}\left(\frac{\propto}{2}\right)\frac{S}{\sqrt{n}} \qquad (6-4)$$

وهو يرسم بيانياً كما يلي:



الشكل (4−1): مجالات القبول والرفض المقابلة لمستوى الدلالة $\propto$

وهنا نلاحظ أن هذا المجال يتضمن جميع القيم الممكنة لـ μ_0 التي تجعلنا نقبل فرضية العدم المذكورة $H_0: \mu = \mu_0$ ، وأن طرفي ذلك المجال هما مقداران عشوائيان، لأنهما مرتبطان بالمتحولين العشوائيين $\bar{x}$ و S، اللذين يختلفان من عينة لأخرى .
كما أن احتمال أن يتضمن ذلك المجال قيمة المتوسط μ يساوي $(1-\propto)$ ، ويسمى احتمال الثقة . أي أنه لو سحبنا عدداً كبيراً من العينات الممكنة، فإننا سنحصل على عدد كبير من مجالات الثقة المقابلة والمستقلة، وستكون نسبة المجالات منها، التي تتضمن قيمة التوقع μ مساوية $100(1-\propto)$ % .

والآن نعود إلى العلاقة (4–4) ونشير إلى أننا سنرفض الفرضية H_0، إذا كانت قيمة $|t|$ كبيرة نسبياً (عندما تكون $|t| > t_{\frac{\alpha}{2}}$). وهذا يكافئ القول: أننا سنرفض الفرضية H_0 . إذا كانت قيمة t^2 كبيرة أيضاً .

ولهذا نربع t ونكتبه على الشكل التالي :

$$t^2 = \left(\frac{\bar{x} - \mu_0}{S/\sqrt{n}}\right)^2 = n(\bar{x} - \mu_0) * (S^2)^{-1} * (\bar{x} - \mu_0) \qquad (7-4)$$

وهنا يجب أن نشير إلى أن المقدار t^2 هو عبارة عن متحول عشوائي جديد (موجب) . يعبر عن مربع المسافة من المتوسط $\bar{x}$ حتى القيمة المختبرة μ_0 ، وإن تلك المسافة مقاسة بواحدات الانحراف المعياري المقدر للمتوسط $\bar{x}$ (والتي تساوي $\frac{S}{\sqrt{n}}$) . وعندما يكون $\bar{x}$ و S معلومين، فإن الاختبار السابق للمتحول المنفرد X يصبح كما يلي:

نقبل فرضية العدم H_0 مقابل H_1 باحتمال ثقة $(1-\alpha)$ إذا كان :

$$t^2 \le t^2_{n-1}\left(\frac{\alpha}{2}\right) \qquad (8-4)$$

أي إذا كان مربع المسافة يحقق المتراجحة التالية :

$$n(\bar{x} - \mu_0)(S^2)^{-1}(\bar{x} - \mu_0) \le t^2_{n-1}\left(\frac{\alpha}{2}\right) \qquad (9-4)$$

حيث أن $t^2_{n-1}\left(\frac{\alpha}{2}\right)$: هو مربع القيمة الحرجة لمتحول (ستودينت) المقابل لنصف مستوى الدلالة $\left(\frac{\alpha}{2}\right)$ وعند $(n-1)$ درجة حرية .

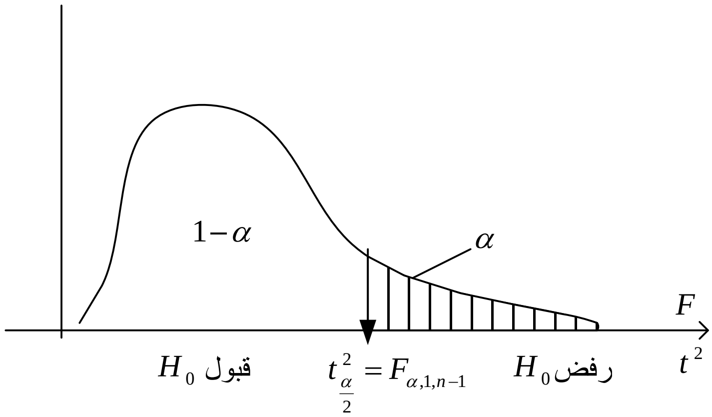
وإن حلول هذه المتراجحة تشكل مجالاً يسمى مجال القبول، وتترك على يمينها مجالاً يسمى مجال الرفض (المقابل لمستوى الدلالة α) . ولتسهيل ذلك نعتمد على علاقة متحول توزيع (ستودينت) t بمتحول توزيع (فيشر) F ، والتي تأخذ في حالة متحول واحد ، الشكل التالي:

$$t^2_{n-1}\left(\frac{\alpha}{2}\right) = F_{1,n-1}(\alpha) \qquad (10-4)$$

وهذا يعني أن مربعات قيم متحول (ستودينت) t^2 عند $(n-1)$ درجة حرية وبمستوى دلالة α ($\left(\frac{\alpha}{2}\right)$ لكل طرف) تساوي قيم متحول (فيشر) F عند درجتي الحرية $v_1 = 1$ و $v_2 = n-1$ وبمستوى دلالة α (لطرف واحد)، ويمكنك التحقق من ذلك من خلال مقارنة القيم في جداول التوزيعين المذكورين، أي أن العلاقة (4–9) يمكن أن تكتب بدلالة توزيع فيشر على الشكل التالي :

$$n(\bar{x} - \mu_0)(S^2)^{-1}(\bar{x} - \mu_0) \le F_{1,n-1}(\alpha) \qquad (11-4)$$

وإن حلول هذه المتراجحة تشكل على يسارها مجال القبول (مجال الثقة) باحتمال ثقة $(1-\alpha)$، وتترك على يمينها مجال الرفض المقابل لمستوى الدلالة α، ويمكن تمثيل ذلك بيانياً على الشكل التالي (قارنه مع الشكل (4–1)) .



الشكل (4-2): مجالا الرفض والثقة بمستوى دلالة ∝ لـ F

مثال (4-1): لاختبار القيمة الظاهرية السائدة 36 لتوقع درجة حرارة جسم الإنسان الصحيح، وضعنا الفرضيتين كما يلي: $H_0: \mu = 36$ و $H_1: \mu \neq 36$، ثم حددنا 0.05 =∝، وسحبنا عينة عشوائية من الأشخاص الأصحاء بحجم $n = 20$ شخصاً، فوجدنا أن متوسطها: $\bar{x} = 37$ وأن تباينها: $S^2 =$ 5.68 وأن $S = 2.38$ ، ثم قمنا بحساب قيمة مؤشر الاختبار t فوجدنا أن:

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{37 - 36}{2.38/\sqrt{20}} = 1.876$$

ومن جداول توزيع (ستودينت) نجد أن: $t_{n-1}\left(\frac{\propto}{2}\right) = t_{19}(0.025) = 2.093$

وعند المقارنة نجد أن: $|t| < t_{n-1}\left(\frac{\propto}{2}\right)$ لذلك نقبل فرضية العدم H_0 .

ويمكن إجراء هذا الاختبار بحساب t^2 من العلاقة:

$$t^2 = n(\bar{x} - \mu_0)(S^2)^{-1}(\bar{x} - \mu_0)$$

$$t^2 = 20(37 - 36)(5.68)^{-1}(37 - 36) = \frac{20}{5.68} = 3.52$$

ولاتخاذ القرار المناسب نقارن قيمة t^2 مع القيمة الحرجة لمتحول توزيع (فيشر) $F_{1,n-1}(\propto)$ ، فنجد من جداول F عند مستوى دلالة 0.05 =∝ ودرجتي حرية $v_1 = 1$ و $v_2 = 19$ أن:

$$F_{1,n-1}(\propto) = F_{1,19}(0.05) = 4.38$$

وعند المقارنة نجد أن:

$$(t^2 = 3.52) < \left(F_{1,19}(0.05) = 4.38\right)$$

لذلك نقبل فرضية العدم H_0 . ونعتبر أن القيمة الظاهرية المفترضة $\mu = 36$ هي قيمة مقبولة لدرجة حرارة الإنسان الصحيح .

4-2: اختبار شعاع المتوسطات لعدة متحولات طبيعية:

لنفترض الآن أننا نريد دراسة أحوال P متحولاً مستمراً وطبيعياً . لذلك نرمز لها بالحروف الكبيرة كما يلي:

$$X_1 \quad X_2 \quad X_3 \quad \dots \quad X_P$$

ثم نشكل منها شعاع العمود $\mathbf{X}$ ونكتبه مع شعاع توقعاته الفعلية كما يلي:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_P \end{bmatrix} \qquad \boldsymbol{\mu} = E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \\ \vdots \\ E(X_P) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_P \end{bmatrix} \qquad (12-4)$$

كما نرمز لمصفوفة التباينات المشتركة الفعلية لهذه المتحولات في المجتمع بالرمز V ونكتبها كما يلي:

$$\boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & \cdots & V_{1P} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2P} \\ V_{31} & V_{32} & V_{33} & \cdots & V_{3P} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ V_{P1} & V_{P2} & V_{P3} & \cdots & V_{PP} \end{bmatrix} \qquad (13-4)$$

ولنفترض أننا نريد اختبار فيما إذا كان الشعاع الافتراضي $\boldsymbol{\mu_0} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0P} \end{bmatrix}$، المؤلف من P قيمة ظاهرية

لتوقعات تلك المتحولات في المجتمع، يمثل شعاع التوقع الفعلي $\boldsymbol{\mu}$. وذلك بمستوى دلالة $\propto$.

لذلك نضع الفرضيتين H_0 و H_1 كما يلي:

$$H_0 = \boldsymbol{\mu} = \boldsymbol{\mu_0} \quad \Leftrightarrow \quad H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0P} \end{bmatrix} \qquad (14-4)$$

$$H_1 = \boldsymbol{\mu} \neq \boldsymbol{\mu_0} \quad \Leftrightarrow \quad H_1 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0P} \end{bmatrix}$$ **من أجل قيمة واحدة على الأقل** μ_j

وللتحقق من هذه الفرضية نقوم بسحب عينة عشوائية n من عناصر المجتمع المدروس، ونأخذ من كل عنصر منها قيم المتحولات المطلوبة $X_1 \;\; X_2 \;\; \ldots \;\; X_P$، فنحصل على مصفوفة من المرتبة $(p * n)$ من القيم العددية المجدولة . نرمز لها بـ X (كبيرة) . ونكتبها كما يلي:

عناصر العينة

$$\boldsymbol{X_{P*n}} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_P \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & & n \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{21} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ x_{P1} & \cdots & \cdots & \cdots & x_{Pn} \end{bmatrix} \end{array} \qquad (15-4)$$

ثم نقوم بحساب متوسطاتها من بيانات العينة (أفقياً)، فنحصل على شعاع متوسطات هذه المتحولات، ونرمز له بالرمز $\bar{X}$ (كبيرة) ونكتبه كما يلي:

$$\boldsymbol{\bar{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \\ \vdots \\ \bar{X}_P \end{bmatrix} \quad : \; \bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \qquad (16-4)$$

ثم نقوم بحساب مصفوفة التباينات المشتركة لها، مأخوذة من بيانات العينة ، ونرمز لها بـ **S** ونكتبها كما يلي:

$$\boldsymbol{S_{P*P}} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1P} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2P} \\ S_{31} & S_{32} & S_{33} & \cdots & S_{3P} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ S_{P1} & S_{P2} & S_{P3} & \cdots & S_{PP} \end{bmatrix} \; : \; S_{kj} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{iK}-\bar{X}_k)(X_{ij}-\bar{X}_j) \quad (17-4)$$

وللتحقق من صحة الفرضية H_0 مقابل H_1 نعتمد على مؤشر مربع المسافة T^2 بين شعاع المتوسطات $\boldsymbol{\bar{X}}$ وشعاع التوقع $\boldsymbol{\mu}$ ، والذي يعطى بواسطة علاقة مشابهة للعلاقة (4–7) ونكتبها بدلالة الأشعة والمصفوفات وحسب الرموز السابقة كما يلي :

$$T^2 = n(\boldsymbol{\bar{X}}-\boldsymbol{\mu})' * \boldsymbol{S^{-1}} * (\boldsymbol{\bar{X}}-\boldsymbol{\mu}) \qquad (18-4)$$

حيث أن **S** هي مصفوفة التباينات المشتركة للمتحولات المدروسة X_P ... X_2 X_1 وهي من المرتبة $(P*P)$ وأن الرمز $(\boldsymbol{X}-\boldsymbol{\mu})'$ هو منقول شعاع العمود $(\boldsymbol{\bar{X}}-\boldsymbol{\mu})$.

كما يمكننا كتابة ذلك المؤشر T^2 بطريقة أخرى كما يلي :

$$T^2 = (\boldsymbol{\bar{X}}-\boldsymbol{\mu})' * \left(\frac{\boldsymbol{S}}{\boldsymbol{n}}\right)^{\boldsymbol{-1}} * (\boldsymbol{\bar{X}}-\boldsymbol{\mu}) \qquad (19-4)$$

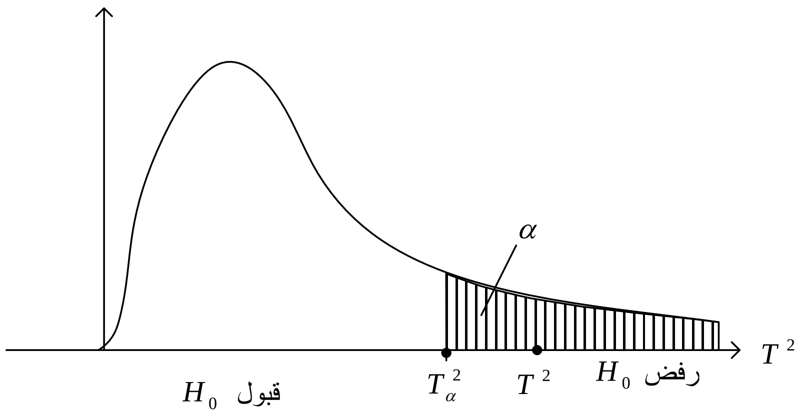
حيث أن $\left[\frac{S}{\boldsymbol{n}}\right]$ هي مصفوفة التباينات المشتركة لشعاع المتوسطات $\bar{X}$، ويمكن حسابها كما في (4–17).

ويسمى المقدار T^2 بمؤشر Hotelling لاختبار الفرضية H_0 حول شعاع المتوسطات μ . وهو يعبر عن مربع المسافة الإحصائية من رأس الشعاع $\boldsymbol{\bar{X}}$ حتى رأس الشعاع $\boldsymbol{\mu_0}$ في الفضاء R^P ذي الـ P بعداً [راجع الفقرة (1–12) في الفصل الأول)

وبما أن الشعاع $\boldsymbol{\bar{X}}$ هو متحول عشوائي، فإن المقدار T^2 هو متحول عشوائي جديد، يخضع إلى توزيع احتمالي محدد يسمى بتوزيع Hotelling (نسبته لرائد التحليل المتعدد Harold Hotelling الذي استخرجه عام 1931) ، وهو التوزيع الذي يعطى بالعلاقة التالية:

$$f_P(T^2) = C(T^2)^{\frac{P}{2}-1} * \left(1+\frac{T^2}{n}\right)^{-\frac{n+1}{2}} \qquad (20-4)$$

حيث أن: C هو عدد لجعل المساحة تحت المنحني تساوي الواحد، وأن درجتي الحرية $v_1 = p$ و $v_2 = n-1$ مأخوذتان من عدد المتحولات p وحجم العينة n .



الشكل (4-3): منحني توزيع Hotelling

ويستفاد من هذا التوزيع في اختبار الفرضية H_0 حول شعاع المتوسطات $\boldsymbol{\mu}$ (وهي $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$) كما يلي:

نقبل فرضية العدم $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ باحتمال ثقة $(1 - \propto)$ إذا كان :

$$T^2 \leq T^2_{P\,,n-1}(\propto) \qquad (21-4)$$

ونرفض فرضية العدم H_0 إذا كانت النتيجة بالعكس

حيث أن $T^2_{P\,,n-1}(\propto)$ هي القيمة الحرجة لمتحول Hotelling عند درجتي الحرية $v_1 = P$ و $v_2 = n-1$ والمقابلة لمستوى دلالة $\propto$ كما في الشكل (4-3) .

وإن تطبيق العلاقة (4-21) لاتخاذ القرار تقتضي إعداد جداول خاصة لاستخراج القيم الحرجة لتوزيع Hotelling $T_{P\,,n-1}(\propto)$. وذلك من أجل استخدامها في معالجة مختلف المسائل الإحصائية، ولكن البحوث اللاحقة برهنت على أن متحول Hotelling، وهو $T^2_{P\,,n-1}(\propto)$ مرتبط بمتحول (فيشر) $F_{P\,,n-\mathrm{p}}(\propto)$ ذي $v_1 = P$ و $v_2 = n - P$ درجتي حرية ، من خلال العلاقة التالية:

$$T^2_{P\,,n-1}(\propto) = \frac{P(n-1)}{n-P} F_{P\,,n-P}(\propto) \qquad (22-4)$$

لذلك تم الاستغناء عن جداول $T_{P\,,n-1}(\propto)$ واستبدالها بجداول التوزيع $F_{P\,,n-P}(\propto)$ المتداولة . ولهذا السبب فإنه عند إجراء الاختبار حول $H_0: \mu = \mu_0$ نقوم بمقارنة المقدار T^2 المحسوب من العلاقة (4-18) مع مقدار القيمة الحرجة المأخوذة من العلاقة (4-22) . فإذا كانت :

$$T^2 \leq \frac{P(n-1)}{n-P}\, F_{P\,,n-P}(\propto) \qquad (23-4)$$

نقبل فرضية العدم $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$، وذلك باحتمال ثقة $(1 - \propto)$.

وأخيراً يمكننا عند قبول الفرضية $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$، أن نكتب ذلك كما يلي:

$$P\left[n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})' * \boldsymbol{S^{-1}} * (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \frac{P(n-1)}{n-P} F_{P\,,n-P}(\propto)\right] = 1 - \propto \qquad (24-4)$$

أما إذا كانت:

$$T^2 > \frac{P(n-1)}{n-P} F_{P\,,n-P}(\propto) \qquad (25-4)$$

فإننا نرفض فرضية العدم $\boldsymbol{H_0}:\boldsymbol{\mu}=\boldsymbol{\mu_0}$ ونعترف بأن أحد (أو أكثر) مركبات شعاع المتوسطات المفترض $\boldsymbol{\mu_0}$ لا يصلح لأن يكون ممثلاً لما يقابله في شعاع التوقع $\boldsymbol{\mu}$ في المجتمع .

مثال (4-2): لنفترض أننا نريد دراسة أحوال متحولين طبيعيين X_1 و X_2 في مجتمع ما، واختبار فيما إذا كان الشعاع $\boldsymbol{\mu_0}=\begin{bmatrix}9\\5\end{bmatrix}$ ، يمثل شعاع التوقع $\boldsymbol{\mu}=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}$ لهذين المتحولين .

لإجراء ذلك الاختبار نحدد مستوى الدلالة بـ $\propto=0.05$، ثم نضع الفرضيتين كما يلي:

$H_0:\mu=\mu_0$ و $H_1:\mu\neq\mu_0$ من أجل أحد المتحولين على الأقل . ثم نسحب عينة من عناصر ذلك المجتمع بحجم $n=3$ ، ونأخذ قيم هذين المتحولين منها . ولنفترض إنها كانت كما يلي:

$$\boldsymbol{X}=\begin{bmatrix}X_1\\X_2\end{bmatrix}=\begin{array}{c}\begin{matrix}1 & \;\;2 & \;\;3\end{matrix}\\\begin{bmatrix}x_{11} & x_{12} & x_{13}\\x_{21} & x_{22} & x_{23}\end{bmatrix}\end{array}=\begin{bmatrix}6 & 10 & 8\\9 & 6 & 3\end{bmatrix}$$

ثم نقوم بحساب شعاع المتوسطات لهما، فنجد أن:

$$\bar{\boldsymbol{X}}=\begin{bmatrix}\bar{X}_1\\\bar{X}_2\end{bmatrix}=\begin{bmatrix}\dfrac{6+10+8}{3}\\\dfrac{9+6+3}{3}\end{bmatrix}=\begin{bmatrix}8\\6\end{bmatrix}$$

ثم نقوم بحساب عناصر مصفوفة التباين المشترك لهما فنجد أن:

$$S_{11}=\frac{1}{n-1}\sum_{i=1}^{3}(X_{1i}-\bar{X}_1)^2=\frac{(6-8)^2+(10-8)^2+(8-8)^2}{2}=4$$

$$S_{22}=\frac{1}{n-1}\sum_{i=1}^{3}(X_{2i}-\bar{X}_2)^2=\frac{(9-6)^2+(6-6)^2+(3-6)^2}{2}=9$$

$$S_{12}=S=\frac{1}{n-1}\sum_{i=1}^{3}(X_{1i}-\bar{X}_1)\,(X_{2i}-\bar{X}_2)=$$

$$=\frac{(6-8)(9-6)+(10-8)(6-6)+(8-8)(3-6)}{2}=-3$$

وبذلك نحصل على مصفوفة التباين المشترك **S** التالية :

$$\boldsymbol{S}=\begin{bmatrix}4 & -3\\-3 & 9\end{bmatrix}$$

ومنها نجد أن مقلوبها يساوي:

$$\boldsymbol{S^{-1}}=\frac{1}{4*9-(-3)(-3)}\begin{bmatrix}9 & 3\\3 & 4\end{bmatrix}=\begin{bmatrix}\dfrac{1}{3} & \dfrac{1}{9}\\\dfrac{1}{9} & \dfrac{4}{27}\end{bmatrix}$$

ثم نقوم بحساب قيمة T^2 من العلاقة (4-18) التالية:

$$T^2 = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0})' * \boldsymbol{S^{-1}} * (\overline{\boldsymbol{X}} - \boldsymbol{\mu_0}) =$$

$$= 3[(8-9),(6-5)] * \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix} * \begin{bmatrix} (8-9) \\ (6-5) \end{bmatrix} =$$

$$T^2 = 3(-1\,,1) \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 3(-1\,,1) \begin{bmatrix} -\frac{2}{9} \\ \frac{1}{27} \end{bmatrix} = \frac{7}{9}$$

وبناءً على العلاقة (4–22) فإن T^2 تخضع للتوزيع التالي :

$$T^2 \sim \frac{p(n-1)}{n-p} F_{P\,,n-p} = \frac{2(3-1)}{3-2} F_{2\,1} = 4F_{2\,,1}$$

ولاتخاذ القرار المناسب حول فرضية العدم $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ نقارن قيمة T^2 المحسوبة من بيانات العينة $\left(T^2 = \frac{7}{9}\right)$ مع القيمة الحرجة لتوزيعها وهي التي تساوي :

$$T^2_{P\ n-1}(0.05) = \frac{p(n-1)}{n-p} F_{P\,,n-p}(0.05) = 4F_{2\,,1}(0.05) = 4(199.5) = 798$$

وبمقارنة T^2 مع $T^2_{2\ 1}(\propto)$ نجد أن :

$$\left(T^2 = \frac{7}{9}\right) < (T_{2\,1}(0.05) = 798)$$

لذلك نقبل فرضية العدم $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ ونعتبر أن الشعاع $\boldsymbol{\mu_0} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ يمثل شعاع التوقع $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ للمتحولين المدروسين X_1 و X_2 باحتمال ثقة $1 - \propto = 0.95$.

ملاحظة: لقد جعلنا حجم العينة في هذا المثال صغيراً جداً لتسهيل وتوضيح الحسابات . ولا يجوز الاعتماد على مثل هذه العينات الصغيرة لإجراء الاختبارات المشابهة .

مثال (4–3): [مأخوذ من Johnson, Wichern , P, 173 بتصرف] لدراسة خواص التعرق عند النساء، تم اختيار (3) متحولات هي:

X_1 معدل التعرق، X_2 كمية الصوديوم، X_3 كمية البوتاسيوم .

والمطلوب اختبار فيما إذا كان شعاع القيم الظاهرية $\boldsymbol{\mu_0} = \begin{bmatrix} 4 \\ 50 \\ 10 \end{bmatrix}$ لتوقعات هذه المتحولات يمثل شعاع التوقعات الفعلية في المجتمع $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$ وذلك بمستوى دلالة $\propto = 0\,,10$.

لإجراء هذا الاختبار وضعنا الفرضيتين كما يلي: $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ مقابل $H_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$، ثم سحبنا عينة عشوائية من 20 امرأة وأخذنا قياسات تلك المتحولات ' ثم حسبنا متوسطاتها والتباين المشترك لها فكانت كما يلي:

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \end{bmatrix} = \begin{bmatrix} 4,640 \\ 45,400 \\ 9,965 \end{bmatrix} \quad , \quad \boldsymbol{S} = \begin{bmatrix} 2,879 & 10,002 & -1,810 \\ 10,002 & 199,798 & -5,627 \\ -1,810 & -5,627 & 3,628 \end{bmatrix}$$

ومنها حسبنا المقلوب $\boldsymbol{S^{-1}}$ فوجدنا أن:

$$\boldsymbol{S^{-1}} = \begin{bmatrix} 0.586 & 0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix}$$

ثم قمنا بحساب قيمة T^2 من العلاقة (4−18) فوجدنا أن:

$$T^2 = 20[\![(4,640-4),(45,400-50),(9,965-10) * \begin{bmatrix} 0.586 & 0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix} * \begin{bmatrix} (4,640-4) \\ (45,400-50) \\ (9,965-10) \end{bmatrix} =$$

$$T^2 = 20[0.640\,,-4.600\,,-0.035] \begin{bmatrix} 0.467 \\ -0.042 \\ 0.160 \end{bmatrix} = 9,74$$

ولمقارنة القيمة المحسوبة $T^2 = 9,74$ مع القيمة الحرجة $T^2_{(\propto)}$ نقوم بحسابها من العلاقة التالية:

$$T^2_{(\propto)} = \frac{p(n-1)}{n-p} F_{P\,,n-p}(\propto) = \frac{3*19}{17} * F_{3,17}(0,10) = 3.353(2.44) = 8.18$$

وبالمقارنة نجد أن:

$$(T^2 = 9,74) > (T^2_{\propto} = 8.18)$$

لذلك نرفض فرضية العدم $H_0{:}\,\boldsymbol{\mu} = \boldsymbol{\mu_0}$ بمستوى دلالة $\propto = 0\,,10$, ونقبل الفرضية البديلة، $H_1{:}\,\boldsymbol{\mu} \neq \boldsymbol{\mu_0}$ ، التي تقول أن واحدة أو أكثر أو بعض مركبات شعاع المتوسطات $\bar{X}$ تختلف كثيراً (قارن المركبة الثانية) عن القيمة المقابلة لها في الشعاع المفترض $\boldsymbol{\mu_0} = \begin{bmatrix} 4 \\ 50 \\ 10 \end{bmatrix}$. وفي هذه الحالة لا يكون لدينا أي مبرر للقول بأن بيانات العينة المسحوبة تدعم القيم المفترضة في $\boldsymbol{\mu_0}$.

بما أننا نفترض أن بيانات العينة تخضع للتوزيع الطبيعي المتعدد، فإن مخططات $Q-Q$ المستخرجة من التوزيعات الهامشية (المنفردة) للمتحولات X_1 و X_2 و X_3 ستشكل خطوط مستقيمة تقريبا . كما أن أشكال الانتشار لأزواج المشاهدات سيكون لها أشكال قطوع ناقصة . وهذا ما يجعلنا نستخلص أن الفرضيات الطبيعية تكون حقيقية في مثل هذه الحالات (كما سنرى في الأمثلة اللاحقة) .

4−3: تحديد مناطق الثقة لشعاع المتوسطات $\boldsymbol{\mu}$ ولمركباته :

حتى نحصل على استدلال كامل عن شعاع المتوسطات $\boldsymbol{\mu}$ ، نقوم بتعميم فكرة مجال الثقة للمتوسط الوحيد μ على الشعاع المتعدد $\boldsymbol{\mu}$ ونشكل له ما يسمى منطقة الثقة في الفضاء R^P . وإن هذه المنطقة تتحدد من خلال بيانات العينة في تلك اللحظة . لذلك نفترض أنه لدينا P متحولاً هي: $X_1\ X_2\ X_3\ \dots\ X_P$ ، وإن قيمها العددية المأخوذة من عينة حجمها n تشكل المصفوفة X التالية :

$$\begin{array}{cc} & \text{عناصر العينة} \\ & \begin{array}{ccccc} 1 & 2 & 3 & & n \end{array} \\ \boldsymbol{X} = \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_P \end{array} & \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{21} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_{P1} & x_{P2} & x_{P3} & \cdots & x_{Pn} \end{bmatrix} \end{array} \qquad (26-4)$$

أن متوسطاتها تشكل الشعاع $\overline{\boldsymbol{X}}$ وتبايناتها المشتركة تشكل المصفوفة **S** كما يلي :

$$\overline{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_P \end{bmatrix} \qquad \boldsymbol{S} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1P} \\ S_{21} & S_{21} & \cdots & S_{2P} \\ \vdots & \vdots & \dots & \vdots \\ S_{P1} & S_{P2} & \cdots & S_{PP} \end{bmatrix} \qquad (27-4)$$

فعندها نجد أن منطقة الثقة للشعاع $\boldsymbol{\mu}$ في الفضاء R^P تتحدد، وبدرجة ثقة $(1-\propto)$. من خلال العلاقة (24–4) والتي نكتبها على الشكل التالي:

$$P\left[n(\overline{\mathbf{X}}-\boldsymbol{\mu})' * \boldsymbol{S^{-1}} * (\overline{\boldsymbol{X}}-\boldsymbol{\mu}) \le \frac{p(n-1)}{n-p} F_{P\,,n-P}(\propto)\right] = 1-\propto \qquad (28-4)$$

وهذا يعني أنه هناك في الفضاءR^P منطقة نرمز لها بـ $R(1-\propto)$ ، لأنها تضمن لنا تحقق هذه العلاقة باحتمال قدره $(1-\propto)$ أو بنسبة $100(1-\propto)\%$. لذلك نسميها منطقة الثقة المقابلة للاحتمال $(1-\propto)$.

ولتحديد فيما إذا كان أي شعاع مفترض $\boldsymbol{\mu_0}$ يقع ضمن تلك المنطقة (أو يشكل قيمة محتملة لـ $\boldsymbol{\mu}$) ، نحسب مربع المسافة ما بين $\overline{\boldsymbol{X}}$ و $\boldsymbol{\mu_0}$ من العلاقة: $[n(\overline{\boldsymbol{X}}-\boldsymbol{\mu_0})' * \boldsymbol{S^{-1}} * (\overline{\boldsymbol{X}}-\boldsymbol{\mu_0}) = T^2]$ ، ثم نقوم بمقارنته مع المقدار $\left[\frac{P(n-1)}{n-P} F_{P\,,n-P}(\propto) = C^2(\propto)\right]$ ، فإذا كانت $T^2 \le C^2(\propto)$ فإن الشعاع $\boldsymbol{\mu_0}$ يكون واقعاً داخل منطقة الثقة، وعندها نقبل فرضية العدم التي تقول أن $H_0\!: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ ، أما إذا كانت $T^2 > C^2(\propto)$ فإننا نرفض فرضية العدم $H_0\!: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ ونقبل الفرضية البديلة التي تقول أن $H_1\!: \boldsymbol{\mu} \ne \boldsymbol{\mu_0}$ وذلك بمستوى دلالة قدرة $\propto$.

وهكذا نجد أنه يمكننا استخدام منطقة الثقة$R(1-\propto)$ لإجراء اختبار تحقق الفرضية$H_0\!: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ مقابل$H_1\!: \boldsymbol{\mu} \ne \boldsymbol{\mu_0}$ وباحتمال ثقة قدره $(1-\propto)$. وذلك من خلال دراسة فيما إذا كانت تلك المنطقة تحتوي على الشعاع المفروض $\boldsymbol{\mu_0}$. وهذا ما يقابل أن تكون نتيجة الاختبار T^2 قبولاً للفرضية H_0 مقابل H_1 وباحتمال ثقة $(1-\propto)$ [انظر العلاقة (24–4)]

والسؤال الآن: ما هو شكل منطقة الثقة $R(1-\propto)$ في الفضاء R^P ؟

والجواب هو: أنه إذا كان عدد المتحولات $(P > 3)$ فإنه يتعذر علينا رسم أو تجسيم تلك المنطقة في الفضاء R^P .

وبصورة عامة فإن منطقة الثقة المقابلة للاحتمال $(1-\propto)$ تتألف من جميع النقاط في الفضاء R^P، التي تحقق المتراجحة التالية:

$$n(\bar{\boldsymbol{X}}-\boldsymbol{\mu})' * \boldsymbol{S^{-1}} * (\bar{\boldsymbol{X}}-\boldsymbol{\mu}) \le \frac{P(n-1)}{n-P} F_{P,n-P}(\propto) = C^2(\propto) \qquad (29-4)$$

وعندما يكون p معلوماً و$\propto$ محدداً، فإن الطرف الأيمن يكون معلوماً وموجباً . لذلك رمزنا له بـ$C^2(\propto)$.

ولتعريف حدود تلك المنطقة نضع المتراجحة على شكل مساواة كما يلي:

$$n(\bar{\boldsymbol{X}}-\boldsymbol{\mu})' * \boldsymbol{S^{-1}} * (\bar{\boldsymbol{X}}-\boldsymbol{\mu}) = C^2(\propto) \qquad (30-4)$$

ولنرمز الآن للشعاع الواصل من $\bar{\boldsymbol{X}}$ إلى $\boldsymbol{\mu}$ بالرمز $\boldsymbol{Z} = (\bar{\boldsymbol{X}}-\boldsymbol{\mu})$ ونكتب العلاقة (4−30) كما يلي :

$$n * \boldsymbol{Z}' * \boldsymbol{S^{-1}} * \boldsymbol{Z} = C^2(\propto) \qquad (31-4)$$

وهنا نلاحظ أن الطرف الأيسر ما هو إلا أحد الأشكال النموذجية لمعادلة القيم الذاتية للمصفوفات الشاذة، لأنه لو أخذنا المعادلة النموذجية للقيم الذاتية المعرفة بالعلاقة التالية:

$$\boldsymbol{SZ} = \lambda \boldsymbol{Z} \qquad (32-4)$$

فإننا نجد أن حلها هو عبارة عن الأشعة الذاتية $\boldsymbol{Z_i}$ المقابلة للقيم الذاتية λ للمصفوفة الأصلية S . وإذا قمنا بإعادة صياغة(4−32) بضرب طرفيها من اليسار بـ $\boldsymbol{Z}' * \boldsymbol{S^{-1}}$ ، سنحصل على أن:

$$\boldsymbol{Z}' * \boldsymbol{S^{-1}} \quad \boldsymbol{S} * \boldsymbol{Z} = \lambda \boldsymbol{Z}' * \boldsymbol{S^{-1}} * \boldsymbol{Z}$$

ومنها نحصل على العلاقة المميزة التالية:

$$\boldsymbol{Z}' * \boldsymbol{S^{-1}} * \boldsymbol{Z} = \frac{1}{\lambda}\boldsymbol{Z}'\boldsymbol{Z} = \frac{1}{\lambda} * \ell^2 \qquad (33-4)$$

حيث أن ℓ^2 هو مربع طول شعاع الحل الذاتي $\mathbf{Z}$،(وذلك لأن $(\ell^2 = \boldsymbol{Z}' * \boldsymbol{Z})$) ، المقابل للقيم الذاتية λ للمصفوفة الأصلية $\mathbf{S}$.أو المقابل للقيمة الذاتية $\boldsymbol{\mu}$ للمصفوفة $\boldsymbol{S^{-1}}$ (حيث $\mu = \frac{1}{\lambda}$).

والآن نعوض $\boldsymbol{Z}'\boldsymbol{S^{-1}}\boldsymbol{Z}$ من العلاقة (4−33) في (4−31) فنحصل على أن :

$$\frac{C^2(\propto)}{n} = \frac{1}{\lambda} * \ell^2 \qquad (34-4)$$

ومنها ومن (4−29) نجد أن طول الشعاع Z يساوي:

$$\ell = \pm\sqrt{\frac{\lambda}{n}} * C(\propto) = \sqrt{\frac{\lambda}{n}} * \sqrt{\frac{P(n-1)}{n-P} F_{P,n-P}(\propto)} \qquad (35-4)$$

حيث أن λ هي أحد القيم الذاتية للمصفوفة $\mathbf{S}$ ، وهي تأخذ p قيمة . وعندما نرتب هذه القيم تنازلياً يكون لدينا:

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \cdots \lambda_P \ge 0 \qquad (36-4)$$

ومن العلاقة (4−35) نحصل على أطوال الأشعة الذاتية المقابلة لها وهي:

$$\ell_1 \ge \ell_2 \ge \ell_3 \ge \cdots \ell_P \ge 0 \qquad (37-4)$$

وهو عبارة عن أنصاف أقطار مجسم منطقة الثقة $R(1-\propto)$ على محاوره الأساسية في الفضاء R^{p} .

حالة خاصة: عندما يكون $P = 2$ ، فإنه يكون لدينا متحولان فقط $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ ، وأن منطقة الثقة في هذه الحالة تشكل قطوع ناقصة في المستوى مركزها رأس الشعاع $\overline{\boldsymbol{X}}$. وترتبط مساحة كل منها بقيمة الثابت $C(\propto)$. أما محاورها فتحسب من الأشعة الذاتية المقابلة للقيم الذاتية للمصفوفة **S**، والتي تحسب من معادلة المحدد التالي :

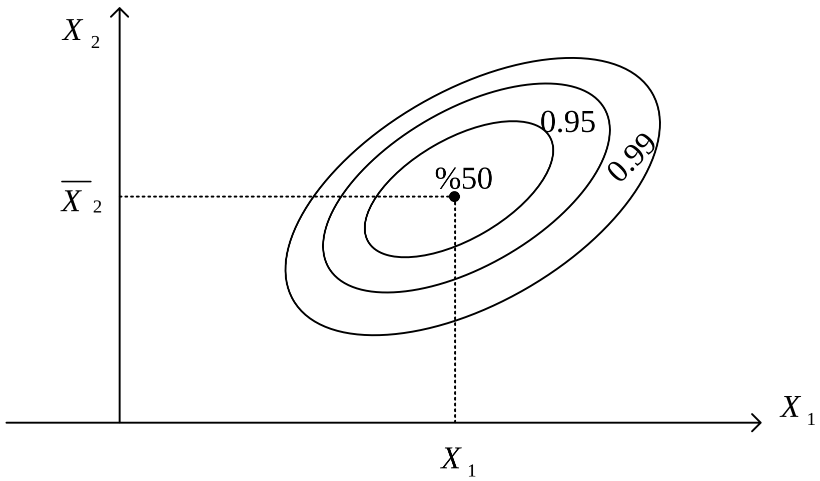
$$\begin{vmatrix} S_{11} - \lambda & S_{12} \\ S_{21} & S_{22} - \lambda \end{vmatrix} = 0 \qquad (38-4)$$

ومنها نحصل على قيمتين ذاتيتين هما: $\lambda_1 \geq \lambda_2$ ، ثم نقوم بحساب طول المحور الكبير بتعويض λ_1 في (4−35) فنجد أن :

$$\ell_1 = \sqrt{\frac{\lambda_1}{n}} * C(\propto) \qquad (39-4)$$

ثم نقوم بحساب طول المحور الصغير بتعويض λ_2 في العلاقة (4−35) فنجد أن:

$$\ell_2 = \sqrt{\frac{\lambda_2}{n}} * C(\propto) \qquad (40-4)$$



الشكل (4−4): مناطق الثقة في R^2

وهنا نلاحظ أن ℓ_1 و ℓ_2 يرتبطان بقيمة$C(\propto)$ المرتبطة بمستوى الدلالة $\propto$. ولهذا فإن العلاقتين تعطيانا قطوع متعددة، تزداد مساحتها كلما تناقص مستوى الدلالة $\propto$ [لأن$C^2(\propto)$ مرتبط طرداً بـ$F(\propto)$ المرتبط عكساً مع$\propto$. انظر الشكل (4−3)] وبذلك نحصل على القطوع الأساسية المذكورة في الفصل الثالث (الفقرة (3−3−6)) وهي:

القطع الأول: ويتضمن 50% من قيم المشاهدات .

القطع الثاني: ويتضمن 95% من قيم المشاهدات .

القطع الثالث: ويتضمن 99% من قيم المشاهدات .

وأخيراً يمكننا أن نقوم بحساب درجة التناسب بين طولي المحورين من العلاقة التالية:

$$Q=\frac{2\ell_1}{2\ell_2}=\frac{\sqrt{\frac{\lambda_1}{n}*C(\propto)}}{\sqrt{\frac{\lambda_2}{n}*C(\propto)}}=\frac{\sqrt{\frac{\lambda_1}{n}}\sqrt{\frac{P(n-1)}{n-P}*F_{P,n-P}(\propto)}}{\sqrt{\frac{\lambda_2}{n}}\sqrt{\frac{P(n-1)}{n-P}*F_{P,n-P}(\propto)}}$$

$$Q=\frac{\sqrt{\frac{\lambda_1}{n}}}{\sqrt{\frac{\lambda_2}{n}}}=\sqrt{\frac{\lambda_1}{\lambda_2}} \qquad (41-4)$$

مثال (4-4): لدراسة الاشعاع المتسرب من أجهزة (الميكرو وايف) حسب حالة أبوابها المفتوحة أو المغلقة ، قمنا بإجراء $n=42$ تجربة لكل حالة (مفتوحة أو مغلقة) وسجلنا المعلومات اللازمة لكل منهما، ثم حولناها لتتقارب من التوزيع الطبيعي من خلال التحويلين التاليين [Johnson, Wichern, P.184 بتصرف] :

$$X_1=\sqrt[4]{\text{كمية الإشعاع في حالة الأبواب المغلقة}}$$

$$X_2=\sqrt[4]{\text{كمية الإشعاع في حالة الأبواب المفتوحة}}$$

فكان شعاع متوسطيهما ومصفوفة تباينيهما المشترك كما يلي:

$$\overline{\boldsymbol{X}}=\begin{bmatrix}0.564\\0.603\end{bmatrix} \qquad \boldsymbol{S}=\begin{bmatrix}0.0144 & 0.0117\\0.0117 & 0.0140\end{bmatrix}$$

والمطلوب إيجاد منطقة الثقة المقابلة لـ $1-\propto=0.95$.

الحل: لإيجاد هذه المنطقة نعتمد على العلاقة (4-28) ، ولذلك يجب علينا حساب $\boldsymbol{S^{-1}}$ ، فنجد أن:

$$\boldsymbol{S^{-1}}=\begin{bmatrix}203.018 & -16.0391\\-16.0391 & 200.228\end{bmatrix}$$

ثم نقوم بحساب منطقة الثقة من العلاقة (4-29) فنجد أن:

$$T^2=n(\overline{\boldsymbol{X}}-\boldsymbol{\mu_0})'*\boldsymbol{S^{-1}}*(\overline{\boldsymbol{X}}-\boldsymbol{\mu_0})\le\frac{P(n-1)}{n-P}F_{P\,n-P}(\propto)=C^2(\propto) \qquad (29-4)$$

$$42[(0.564-\mu_1),(0.603-\mu_2)]*\begin{bmatrix}203.018 & -16.0391\\-16.0391 & 200.228\end{bmatrix}*\begin{bmatrix}(0.564-\mu_1)\\(0.603-\mu_2)\end{bmatrix}$$

$$\le\frac{2(42)}{40}F_{2,40}(0.05)=2.1*3.23=\ 6.62$$

وذلك لأن $F_{2,40}(0.05)=3.23$ ، والمتراجحة السابقة تأخذ الشكل التالي:

$$42(203.018)(0.564-\mu_1)^2+42(200.228)(0.603-\mu_2)^2$$
$$-84(163.39)(0.564-\mu_1)(0.603-\mu_2)\le 6.62$$

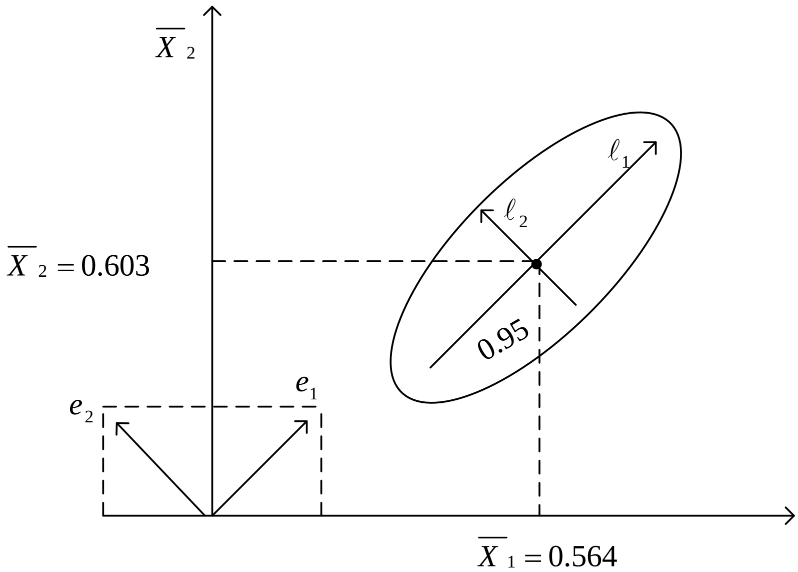
وهي العلاقة التي تعطينا معادلة القطع الناقص الثاني ذي احتمال الثقة $1-\propto=0.95$

ولنفرض الآن أنه يطلب مننا أن نختبر فيما إذا كان الشعاع $\boldsymbol{\mu_0} = \begin{bmatrix} 0.562 \\ 0.580 \end{bmatrix}$ يقع ضمن منطقة الثقة السابقة وذلك باحتمال ثقة 0.95 .

لذلك نعوض ذلك الشعاع في معادلة القطع السابق فنجد أن:

$$42(203.018)(0.564-0.562)^2 + 42(200.228)(0.603-0.580)^2$$
$$-84(163.39)(0.564-0.562)(0.603-0.580) = 1.30 \leq 6.62$$

أي أن الشعاع $\boldsymbol{\mu_0} = \begin{bmatrix} 0.562 \\ 0.580 \end{bmatrix}$ يقع ضمن منطقة الثقة السابقة . وبناء على ذلك نقبل فرضية العدم $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ باحتمال ثقة $(1 - \propto)$، أي أنه لا يمكن رفض H_0 مقابل $H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$ وبمستوى دلالة $\propto$. وإن القطع الناقص الثاني الذي يمثل منطقة الثقة السابقة يرسم في المستوى الشكل التالي:



الشكل (4-5): منطقة الثقة ذات الاحتمال 0.95

ولحساب طولي محوري ذلك القطع علينا أن نطبق العلاقتين (4-39) و(4-40) ، ولذلك نحسب أولاً القيم الذاتية للمصفوفة S من المعادلة التالية :

$$\begin{vmatrix} S_{11}-\lambda & S_{12} \\ S_{21} & S_{22}-\lambda \end{vmatrix} = \begin{vmatrix} 0.0144-\lambda & 0.0117 \\ 0.0117 & 0.0146-\lambda \end{vmatrix} = 0$$

ومنها نحصل على المعادلة المميزة التالية:

$$(0.0144-\lambda)(0.0146-\lambda) - (0.0117)^2 = 0$$

وبحلها نحصل على القيمتين الذاتيتين التاليتين:

$$\lambda_1 = 0.026 \qquad \lambda_2 = 0.002$$

ثم نقوم بحساب الشعاعين الذاتين الواحديين المقابلين لكل منهما فنجد أنهما يساويان :

$$\boldsymbol{e_1} = \begin{bmatrix} 0.704 \\ 0.710 \end{bmatrix} \qquad \boldsymbol{e_2} = \begin{bmatrix} -0.710 \\ 0.704 \end{bmatrix}$$

واعتماداً على هذه البيانات نقوم بحساب نصف طول المحور الكبير من العلاقة التالية:

$$\ell_1 = \sqrt{\frac{\lambda_1}{n}} * \sqrt{\frac{P(n-1)}{n-P} F_{P\ n-P}(0.05)} = \sqrt{\frac{0.026}{42}} \sqrt{\frac{2(41)}{40}(3.23)} = 0.064$$

ثم نحسب نصف طول المحور الصغير فنجد أن:

$$\ell_2 = \sqrt{\frac{\lambda_2}{n}} \sqrt{\frac{P(n-1)}{n-P} F_{P\ n-P}(\propto)} = \sqrt{\frac{0.002}{42}} \sqrt{\frac{2(41)}{40}(3.23)} = 0.018$$

وإن هذين المحورين يوازيان الشعاعين e_1 و e_2 المرسومين من النقطة المركزية 0 كما هو مبين على الشكل (4-5) .

وأخيراً نقوم بحساب نسبة طولي المحورين من العلاقة التالية :

$$Q = \frac{2\ell_1}{2\ell_2} = \sqrt{\frac{\lambda_1}{\lambda_2}} = \sqrt{\frac{0.026}{0.002}} = \frac{1.61}{0.045} = 3.6$$

أي أن طول المحور الكبير أكبر من طول المحور الصغير بـ 3.6 مرة .

4-4: مجالات الثقة المنفردة والمتزامنة:

لقد رأينا أن منطقة الثقة للشعاع X المؤلف من المتحولات $X_1\ X_2\ \ldots\ X_P$ تعطى بواسطة العلاقة التالية:

$$n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \le C^2(\propto) = \frac{P(n-1)}{n-\mathrm{P}} F_{P\ n-P}(\propto)$$

والآن سنبحث في إيجاد منطقة الثقة لأي تركيب خطي لتلك المتحولات ، لذلك نفترض أن الشعاع X يخضع للتوزيع الطبيعي المشترك $N(\boldsymbol{\mu}, \boldsymbol{V})$ ، وأن Z هو تركيب خطي من تلك المتحولات معرف بالعلاقة التالية :

$$Z = \ell_1 X_1 + \ell_2 X_2 + \cdots + \ell_P X_P = [\ell_1 \ell_2 \ldots \ell_P] \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix} = \boldsymbol{\ell}' \boldsymbol{X} \qquad (42-4)$$

وبذلك نجد أن التوقع الحقيقي لـ Z في المجتمع يساوي:

$$\mu_Z = E(Z) = E(\ell' * \boldsymbol{X}) = \ell' E(\boldsymbol{X}) = \boldsymbol{\ell}' * \boldsymbol{\mu} \qquad (43-4)$$

وإن تباينه في المجتمع σ_Z^2 يساوي:

$$\sigma_Z^2 = Vor\ (Z) = Vor(\boldsymbol{\ell}'\boldsymbol{X}) = \boldsymbol{\ell}' * Vor(\boldsymbol{X}) * \boldsymbol{\ell} = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} \qquad (44-4)$$

أي أن Z يخضع للتوزيع الطبيعي $N(\ell'\boldsymbol{\mu}, \ell'\boldsymbol{V}\ \ell)$.

وإذا سحبنا عينة بحجم n من ذلك المجتمع وأخذنا قياسات المتحولات **X** منها وحسبنا شعاع متوسطاتها $\overline{\boldsymbol{X}}$ ومصفوفة تباينتها المشتركة **S**، فإنه يمكننا تقدير متوسط وتباين **Z** من العلاقتين التاليتين :

$$\bar{Z} = \boldsymbol{\ell}' * \overline{\boldsymbol{X}} \qquad S_Z^2 = \boldsymbol{\ell}' * \boldsymbol{S} * \boldsymbol{\ell} \qquad (45-4)$$

حيث $\bar{X}$ هو شعاع متوسطات X المحسوب من العينة، و S هي مصفوفة التباينات المشتركة المحسوبة من العينة .

وهنا نلاحظ أن المتوسط $\bar{Z}$ هو متحول عشوائي مرتبط بـ$\bar{\boldsymbol{X}}$ من خلال التركيب الخطي (4−45) المتعلق أيضاً بقيم الشعاع $\boldsymbol{\ell}$. وهنا نميز بين الحالات التالية:

4−4−1: الحالة الأولى : إذا كان الشعاع $\boldsymbol{\ell}$ ثابتاً وكان σ_Z^2 مقدراً بواسطة S_Z^2

إن مجال الثقة ذي الاحتمال $(1 - \propto)$ للتوقع $\mu_Z = \boldsymbol{\ell}'\boldsymbol{\mu}$ يحسب وكأنه متحول واحد بالاعتماد على مؤشر اختبار (ستودينت) التالي:

$$|t| = \left|\frac{\bar{Z} - \mu_Z}{S_Z/\sqrt{n}}\right| \leq t_{n-1}\left(\frac{\propto}{2}\right) \qquad (46-4)$$

وهذا يقودنا إلى أن مجال الثقة لـ μ_Z يحسب من العلاقة المشابهة لـ (4−6) التالية :

$$\bar{Z} - t_{n-1}\left(\frac{\propto}{2}\right)\frac{S_Z}{\sqrt{n}} \leq \mu_Z \leq \bar{Z} + t_{n-1}\left(\frac{\propto}{2}\right) * \frac{S_Z}{\sqrt{n}} \qquad (47-4)$$

وبتعويض $\bar{Z}$ و S_Z من العلاقة (4−45) نحصل على مجال الثقة لـ μ_Z أو لـ $\ell'\mu$ التالي:

$$\boldsymbol{\ell}'\bar{\boldsymbol{X}} - t_{n-1}\left(\frac{\propto}{2}\right) * \frac{\sqrt{\ell' S\, \ell}}{\sqrt{n}} \leq \boldsymbol{\ell}'\boldsymbol{\mu} \leq \boldsymbol{\ell}'\bar{\boldsymbol{X}} + t_{n-1}\left(\frac{\propto}{2}\right)\frac{\sqrt{\ell' S\, \ell}}{\sqrt{n}} \qquad (48-4)$$

حيث أن $t_{n-1}\left(\frac{\propto}{2}\right)$ هي القيمة الحرجة لمتحول (ستودينت) عند $n-1$ درجة حرية ، التي تترك على يمينها نصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$.

ويمكننا الاستفادة من العلاقة (4−48) للحصول على مجالات ثقة لكل من التوقعات المنفردة μ_1 و μ_2 ... ثم μ_P الموجودة في مركبات الشعاع $\boldsymbol{\mu}$ ، وذلك بإعطاء إحدى مركبات $\boldsymbol{\ell}$ (أو $\boldsymbol{\ell}'$) قيمة تساوي الواحد وإعطاء بقية المركبات قيماً صفرية على التوالي:

فمثلاً: نجد أنه إذا كنا نريد إيجاد مجال الثقة للتوقع المنفرد μ_1 فإننا نضع $\boldsymbol{\ell}'(1,0,0,0,0,00)$ وعندها نجد أن:

$$\boldsymbol{\ell}'\boldsymbol{\mu} = (1,0,0,0,0,0,0) * \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} = \mu_1$$

$$\boldsymbol{\ell}'\bar{\boldsymbol{X}} = (1,0,0 \ldots 0,0) * \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_3 \end{bmatrix} = \bar{X}_1 \qquad (49-4)$$

$$\boldsymbol{\ell}'\boldsymbol{S}\,\boldsymbol{\ell} = (1,0,0 \ldots 0,0) * \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1P} \\ S_{21} & S_{21} & \cdots & S_{2P} \\ \vdots & \vdots & \ldots & \vdots \\ S_{P1} & S_{P2} & \cdots & S_{PP} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = S_{11}$$

وبذلك تأخذ العلاقة المركبة (4−48) الشكل الخاص بـ μ_1 التالي:

$$\bar{X}_1 - t_{n-1}\left(\frac{\propto}{2}\right) * \sqrt{\frac{S_{11}}{n}} \le \mu_1 \le \bar{X}_1 + t_{n-1}\left(\frac{\propto}{2}\right) * \sqrt{\frac{S_{11}}{n}} \qquad (50-4)$$

وهي نفس الصيغة المعرفة في (4–6) لمجال الثقة لتوقع المتحول $\bar{X}_1$ بمفرده .

وهكذا يمكننا أن نحصل على مجالات الثقة المنفردة لتوقعات المتحولات X_2 X_3 ... X_P الأخرى ، وذلك بوضع $\ell = (0,1,0,0,0,0)$ ثم $\ell = (0,0,1,0,0,0)$ ...و $\ell = (0,0,0,0,0,1)$ على الترتيب ، ثم تعويضها في العلاقة (4–48) ، فنحصل على مجالات الثقة للتوقعات μ_1 μ_2 ... μ_P على التوالي :

$$\begin{aligned}
&\bar{X}_1 - t_{n-1}\left(\frac{\propto}{2}\right) \sqrt{\frac{S_{11}}{n}} \le \mu_1 \le \bar{X}_1 + t_{n-1}\left(\frac{\propto}{2}\right) * \sqrt{\frac{S_{11}}{n}} \\
&\bar{X}_2 - t_{n-1}\left(\frac{\propto}{2}\right) \sqrt{\frac{S_{22}}{n}} \le \mu_2 \le \bar{X}_2 + t_{n-1}\left(\frac{\propto}{2}\right) * \sqrt{\frac{S_{22}}{n}} \qquad (51-4) \\
&\cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\
&\bar{X}_P - t_{n-1}\left(\frac{\propto}{2}\right) \sqrt{\frac{S_{PP}}{n}} \le \mu_P \le \bar{X}_P + t_{n-1}\left(\frac{\propto}{2}\right) * \sqrt{\frac{S_{PP}}{n}}
\end{aligned}$$

وهنا نلاحظ أن مجال الثقة لكل توقع μ_j يتضمن ذلك التوقع μ_j باحتمال ثقة قدره $(1 - \propto)$، ولكن هل احتمال تحققها معاً يساوي $(1 - \propto)$ ؟.

وهنا يبرز أمامنا السؤال التالي: ما هو احتمال الثقة الكلي الذي يضمن لنا أن تحتوي جميع مجالات الثقة السابقة معاً على جميع التوقعات المنفردة μ_j ؟

ولتسهيل الإجابة على هذا السؤال نسمي حادث تحقق جميع المتراجحات (4–51) معاً بحادث تحقق المجالات المتزامنة، ولنفترض الآن أن المتحولات X_1 X_2 ... X_P مستقلة عن بعضها البعض، فعندها يمكننا حساب احتمال الثقة الكلي من خلال تطبيق قاعدة ضرب احتمال الثقة $(1 - \propto)$، المقابلة لمجالات الثقة المنفردة لها .

ونكتب ذلك كما يلي:

احتمال [أن تحتوي جميع المجالات (4–51) قيم التوقعات μ_j]= جداء احتمالاتها(لأنها مستقلة)

أي أن:

$$P\left[All\ intervolls\ in\ (51-4) contain\ \mu_j, s\right] = (1-\propto)(1-\propto)\ldots(1-\propto) \text{ مرة } P$$

$$P(All) = (1-\propto)^P \qquad (52-4)$$

وكمثال على ذلك : لنفترض أن احتمال الثقة لكل من المجالات المنفردة كان: $1 - \propto = 0.95$ وأن عدد المتحولات المدروسة $P = 5$ ، فعندها نجد أن احتمال الثقة لتحقق جميع المتراجحات (4–51) معاً يساوي:

$$P(All) = (1-\propto)^P = (0.95)^5 = 0.77$$

وهو أصغر من 0.95 =∝− 1 بكثير. ويسمى الاحتمال الأخير $P(ALL)$ باحتمال الثقة الكلي لحادث تحقق المجالات (4−51) معاً وبشكل متزامن، وحتى نجعل ذلك الاحتمال الكلي المتزامن $P(ALL)$ مساوياً لاحتمال الثقة المعتمد (0.95 =∝− 1) أو أكبر منه، علينا أن نعمل على توسيع المجالات في (4−51) بمقادير مناسبة، ولكن بما أننا اعتبرنا – في هذه الحالة− أن الشعاع ℓ يأخذ قيمة ثابتة، فإن عملية التوسيع تقتضي استبدال القيمة الحرجة $t\left(\frac{\propto}{2}\right)$ فيها بقيمة أكبر منها في كل من هذه المجالات .

وحتى نعالج هذه المسألة بشكلها العام، نلاحظ أن سعات مجالات الثقة في (4−51) لا ترتبط بـ $t\left(\frac{\propto}{2}\right)$ فقط ، بل إن العلاقة (4−48) تشير بوضوح إلى أنها ترتبط أيضاً بشعاع أمثال التركيب الخطي ℓ . وكذلك بعدد المتحولات p وبحجم العينة n [وذلك من خلال المصفوفة S]. ولذلك سنأخذ الحالة العامة التي تعتبر أن ℓ تأخذ قيماً متغيرة متعددة ومحددة ونعالجها كما يلي:

4−4−2: الحالة الثانية : الحالة التي يكون فيها $\boldsymbol{\ell}$ متغيراً ويأخذ قيماً متعددة ومحددة

وعندها يكون لدينا عدد كبير من التراكيب الخطية مثل (4−42)، وإن كل منها يؤثر بطريقته الخاصة على سعات مجالات الثقة في (4−48). وسنقوم بطريقة مشابهة لما سبق بمعالجة تأثير هذه التراكيب الخطية على مجالات الثقة المتزامنة لتوقعات المتحولات $X_1\, X_2 \dots X_P$ ، وبحيث يبقى احتمال الثقة الكلي المتزامن لها مساوياً لـ (∝− 1). وهذا يعني أنه لو كانت لدينا عينة من البيانات عن المتحولات $X_1\, X_2 \dots X_P$ ، وكان $\boldsymbol{\ell}$ يأخذ أية قيمة عملية فإن حدود مجالات الثقة في العلاقة المركبة (4−48) تكون مرتبطة بقيم ℓ، وإن مراكزها تتشكل من القيم $\ell'\boldsymbol{\mu}$ التي تحقق العلاقة (4−46) التي تأخذ الشكل التالي:

$$|t| = \left|\frac{\bar{Z} - \mu_Z}{S_Z/\sqrt{n}}\right| = \left|\frac{\sqrt{n}(\boldsymbol{\ell}'\overline{\boldsymbol{X}} - \boldsymbol{\ell}'\boldsymbol{\mu})}{\sqrt{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}}\right| \le t_{n-1}\left(\frac{\propto}{2}\right) \qquad (53-4)$$

وبتربيع الطرفين نحصل على أن:

$$t^2 = \frac{n(\boldsymbol{\ell}'\overline{\boldsymbol{X}} - \boldsymbol{\ell}'\boldsymbol{\mu})^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}} = \frac{n[\boldsymbol{\ell}'(\overline{\boldsymbol{X}} - \boldsymbol{\mu})]^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}} \le t_{n-1}^2\left(\frac{\propto}{2}\right) \qquad (54-4)$$

وهذا يعني أن منطقة الثقة المتزامنة تعطى من خلال مجموعة قيم $\boldsymbol{\ell}'\boldsymbol{\mu}$ ، التي تجعل المقدار t^2 صغيراً نسبياً وأصغر من المقدار $t_{n-1}^2\left(\frac{\propto}{2}\right)$، وذلك من أجل جميع القيم الممكنة لـ $\boldsymbol{\ell}$ ، وإذا أخذنا الحالة الحدية فإنه يكون لدينا ما يلي: $max\, t^2 \le max\, t_{n-1}^2\left(\frac{\propto}{2}\right)$.

وهذا يجعلنا نتوقع أنه إذا استبدلنا القيمة الحرجة الثابتة $t_{n-1}^2\left(\frac{\propto}{2}\right)$ الواردة في العلاقة الأخيرة (4−54) بأكبر قيمة لها ، والتي نرمز لها بـ $C^2(\propto)$ ، وإذا أخذنا قيم $\boldsymbol{\ell}$ التي تحقق المتراجحة (4−54)، فإنه يكون لدينا ما يلي:

$$\max_{\ell} t^2 \le C^2(\propto) \qquad (55-4)$$

وبالتالي فإننا سنتوصل إلى أن:

$$\max_{\ell}\ t^2 = \max_{\ell} \frac{n[\boldsymbol{\ell}'(\overline{\boldsymbol{X}}-\boldsymbol{\mu})]^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}} \leq C^2(\propto) \qquad \text{من أجل جميع قيم } \ell \qquad (56-4)$$

وبناءً على نظرية (كوشي– شوارز)(*): نجد أنه عندما نضع $\boldsymbol{X}=\boldsymbol{\ell}$ و $\boldsymbol{d}=\overline{\boldsymbol{X}}-\boldsymbol{\mu}$ و $\boldsymbol{B}=\boldsymbol{S}$ نجد أن:

$$\max_{\ell}\ t^2 = n\max_{\ell} \frac{[\boldsymbol{\ell}'(\overline{\boldsymbol{X}}-\boldsymbol{\mu})]^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}} = n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})' * \boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu}) = T^2 \qquad (57-4)$$

وهكذا نجد أن القيمة العظمى لـ t^2 من أجل جميع القيم الممكنة لـ ℓ تساوي مربع المسافة الإحصائية T^2. وإن حلها يتناسب مع الشعاع $\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})$ ، أي أن حلها $\boldsymbol{\ell}$ يساوي $\boldsymbol{\ell}=\boldsymbol{C}\,\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})$.

ومما سبق نجد أنه يمكننا أن نكتب العلاقة (4–56) على الشكل التالي:

$$\max \frac{\boldsymbol{n}[\boldsymbol{\ell}'(\overline{\boldsymbol{X}}-\boldsymbol{\mu})]^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}} \leq C^2(\propto) \qquad \text{من أجل جميع قيم } \ell \qquad (58-4)$$

أو على الشكل التالي:

$$T^2 = n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \leq C^2(\propto) \qquad (59-4)$$

ومنها نجد أن احتمال الثقة الكلي لمنطقة الثقة المتزامنة يحسب كما يلي :

$$P[T^2 \leq C^2(\propto)] = 1-\propto \qquad (60-4)$$

ولكن بما أن T^2 تخضع لتوزيع Hotelling $(T_{P\ n-1})$ المرتبط بتوزيع فيشر F بواسطة العلاقة:

$$T^2_{P\ n-1} = \frac{P(n-1)}{n-P} F_{P\ n-P} \qquad (61-4)$$

فإنه يمكننا استبدال $C^2(\propto)$ بالقيمة الحرجة $T^2_{P\ n-1}(\propto)$ أو بالقيمة الحرجة المقابلة لها F، المحسوبة من العلاقة (4–61) ، ونضع ذلك كما يلي :

$$C^2(\propto) = T^2_{P\ n-1}(\propto) = \frac{P(n-1)}{n-P} F_{P\ n-P}(\propto) \qquad (62-4)$$

وبتعويض ذلك في العلاقة (4–60) نجد أنها تأخذ الشكل التالي:

$$P\left[T^2 \leq \frac{P(n-1)}{n-P} F_{P\ n-P}(\propto)\right] = 1-\propto \qquad (63-4)$$

وبالعودة إلى الرموز السابقة يمكننا أن نكتب مجال الثقة لـ T^2 من أجل جميع قيم ℓ كما يلي:

$$P\left[\frac{n(\boldsymbol{\ell}'\overline{\boldsymbol{X}}-\boldsymbol{\ell}'\boldsymbol{\mu})^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}} \leq C^2(\propto)\right] = 1-\propto \qquad : \ (For\ All\ \ell) \qquad (64-4)$$

وبعد معالجتها وجذرها نحصل على أن مجال الثقة للمقدار $(\boldsymbol{\ell}'\overline{\boldsymbol{X}}-\boldsymbol{\ell}'\boldsymbol{\mu})$ يساوي :

$$-C(\propto)\sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \leq (\boldsymbol{\ell}'\overline{\boldsymbol{X}}-\boldsymbol{\ell}'\boldsymbol{\mu}) \leq C(\propto)\sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \qquad (65-4)$$

(*) : إن خلاصة نظرية (كوشي– شوارز) هي أن: $\max\left[\frac{(X'*d)^2}{X'*B*X}\right] = d'*B^{-1}*d$

وإن متراجحة (كوشي– شوارز) هي: $\frac{(X'*d)^2}{X'*B*X} \leq d'B^{-1}d$

وإن حلها بالنسبة لـ X يساوي: $X = CB^{-1}*d$

وأخيراً نحصل على مجال الثقة لـ $\boldsymbol{\ell'\mu}$ من أجل جميع قيم $\boldsymbol{\ell}$، الذي يأخذ الشكل التالي:

$$\boldsymbol{\ell'\bar{X}} - C(\propto)\sqrt{\frac{\boldsymbol{\ell'S\ell}}{n}} \leq \boldsymbol{\ell'\mu} \leq \boldsymbol{\ell'\bar{X}} + C(\propto)\sqrt{\frac{\boldsymbol{\ell'S\ell}}{n}} \qquad (66-4)$$

وباستبدال $C(\propto)$ بما تساويه من (4–62) نحصل على أن مجال الثقة لـ $\boldsymbol{\ell'\mu}$ يساوي:

$$\boldsymbol{\ell'\bar{X}} - \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{\boldsymbol{\ell'S\ell}}{n}} \leq \boldsymbol{\ell'\mu} \leq \boldsymbol{\ell'\bar{X}} + \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{\boldsymbol{\ell'S\ell}}{n}} \quad (\mathbf{67-4})$$

وهكذا نكون قد حصلنا على مجالات الثقة المتزامنة لـ $\boldsymbol{\ell'\mu}$ ولجميع القيم الممكنة لـ $\boldsymbol{\ell}$ من خلال العلاقة (4–67) . وتسمى هذه المجالات بمجالاتT^2، لأنه يتم تحديد معاملات الثقة لها من التوزيع $(T_{P\ n-1})$ أو من خلال علاقته مع $(F_{P\ n-P})$، وهي تضمن لنا أن تحتوي على $\boldsymbol{\ell'\mu}$ باحتمال ثقة $(1-\propto)$ ومن أجل جميع قيم $\boldsymbol{\ell}$.

ويستفاد من العلاقة (4–67) في استنتاج مجالات الثقة المتزامنة للتوقعات الفردية $\mu_1\ \mu_2\ ...\ \mu_P$ للمتحولات $X_1\ X_2\ ...\ X_P$ ، وذلك بوضع قيم اختيارية لـ $\boldsymbol{\ell}$ على التوالي:

$$\boldsymbol{\ell'_1} = (1,0,0,0,0)\ , \qquad \boldsymbol{\ell'_2} = (0,1,0\ 0\ 0\ 0)\ , \quad ..... \quad \boldsymbol{\ell'_p} = (0,0,0,0,0,1)$$

فنجد أنه عندما نعوض هذه القيم لـ ℓ في العلاقة (4–67) نحصل على ما يلي :

$$\begin{array}{lllll} \boldsymbol{\ell'_1\mu} = \mu_1 & \boldsymbol{\ell'_2\mu} = \mu_2 & \boldsymbol{\ell'_3\mu} = \mu_3 & ... & \boldsymbol{\ell'_P\mu} = \mu_P \\ \boldsymbol{\ell'_1\bar{X}} = \bar{X}_1 & \boldsymbol{\ell'_2\bar{X}} = \bar{X}_2 & \boldsymbol{\ell'_3\bar{X}} = \bar{X}_3 & ... & \boldsymbol{\ell'_P\bar{X}} = \bar{X}_P \\ \boldsymbol{\ell'_1S\ell_1} = S_{11} & \boldsymbol{\ell'_2S\ell_2} = S_{22} & \boldsymbol{\ell'_3S\ell_3} = S_{33} & ... & \boldsymbol{\ell'_PS\ell_P} = S_{PP} \end{array} \qquad (68-4)$$

وبذلك نجد أن مجالات الثقة المتزامنة للتوقعات $\mu_1\ \mu_2\ ...\ \mu_P$ تأخذ الأشكال التالية :

$$\begin{array}{l} \bar{X}_1 - \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{11}}{n}} \leq \mu_1 \leq \bar{X}_1 + \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{11}}{n}} \\ \bar{X}_2 - \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{22}}{n}} \leq \mu_2 \leq \bar{X}_2 + \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{22}}{n}} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \bar{X}_P - \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{PP}}{n}} \leq \mu_P \leq \bar{X}_P + \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{PP}}{n}} \end{array} \qquad (69-4)$$

علماً بأن احتمال الثقة في كل من هذه المجالات أكبر أو يساوي $(1-\propto)$ ، وإن احتمال تحققها بشكل متزامن يساوي $(1-\propto)$.

كما يمكن الاستفادة من العلاقة (4–67) لاستخراج مجال الثقة للفرق بين أي توقعين $(\mu_h - \mu_K)$ ، ولذلك نضع قيم اختيارية مناسبة لـ $\boldsymbol{\ell}$ ، بحيث يكون لها الشكل التالي:

$$\boldsymbol{\ell'} = (0,0,\ell_h 0,0,0,\ell_K 0,0)$$

فعندما نضع $\ell_K = -1$ و$\ell_h = 1$ فإننا نحصل على أن :

$$\begin{array}{ll} \boldsymbol{\ell'\mu} = \mu_h - \mu_K & \boldsymbol{\ell'\bar{X}} = \bar{X}_h - \bar{X}_K \\ \boldsymbol{\ell'S\ell} = S_{hh} + S_{KK} - 2S_{hK} & \end{array} \qquad (70-4)$$

وبالتعويض في العلاقة (4−67) نحصل على المجال التالي:

$$(\bar{X}_h - \bar{X}_K) - \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{hh}+S_{KK}-2S_{hK}}{n}} \leq \mu_h - \mu_k \leq (\bar{X}_h - \bar{X}_K) +$$
$$+\sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{hh}+S_{KK}-2S_{hK}}{n}} \qquad (71-4)$$

وهنا نشير إلى أن مجالات الثقة T^2 المتزامنة تكون مجالات مثالية للبيانات الاستطلاعية (Snooping). وإن احتمال الثقة $(1-\propto)$ يبقى دون تغير من أجل أي خيار لـ $\boldsymbol{\ell}$، لأنه يمكننا تقدير معاملات التركيب الخطي للمركبات μ_j بكل جدارة من خلال الانحدار المتعدد .

مثال (4−5): [مأخوذ من: Johnson, Wichern P. 185 بتصرف] لتقدير مستويات ومجالات الثقة للنتائج الامتحانية في الشهادة الثانوية، سحبنا عينة من الطلاب بحجم $n = 87$ واعتمدنا على ثلاثة متحولات هي: المجموع العام X_1 – علامة الرياضيات X_2 – علامة العلوم X_3 . ثم حسبنا شعاع المتوسطات ومصفوفة التباينات المشتركة لها فكانت كما يلي:

$$\bar{\boldsymbol{X}} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix} \qquad \boldsymbol{S} = \begin{bmatrix} 5691.34 & 600.51 & 217.25 \\ 600.51 & 126.05 & 23.37 \\ 217.25 & 23.37 & 23.11 \end{bmatrix}$$

والمطلوب : حساب مجالات الثقة المتزامنة لـ μ_1 و μ_2 و μ_3 بمستوى دلالة $\propto = 0.05$.

الحل: لحساب المجالات المطلوبة نطبق العلاقات (4−69)، لذلك نقوم أولاً بحساب القيمة الحرجة المشتركة فيها وهي:

$$\frac{P(n-1)}{n-P}F_{P\ n-P}(0.05) = \frac{3(87-1)}{(87-3)} * F_{3,84}(0.05) = \frac{3(86)}{84}(2.71) = 8.32$$

وبتعويض ذلك في المعادلات (4−69) نحصل على مجالات الثقة المتزامنة التالية:

$$For \mu_1: 527.74 - \sqrt{8.32}\sqrt{\frac{5691.34}{87}} \leq \mu_1 \leq 527.74 + \sqrt{8.32}\sqrt{\frac{5691.34}{87}}$$

أي أن : $504.41 \leq \mu_1 \leq 551.07$

$$For \mu_2: 54.69 - \sqrt{8.32}\sqrt{\frac{126.05}{87}} \leq \mu_2 \leq 54.69 + \sqrt{8.32}\sqrt{\frac{126.05}{87}}$$

أي أن : $51.22 \leq \mu_2 \leq 58.16$

$$For \mu_3: 25.13 - \sqrt{8.32}\sqrt{\frac{23.11}{87}} \leq \mu_3 \leq 25.13 + \sqrt{8.32}\sqrt{\frac{23.11}{87}}$$

أي أن : $23.64 \leq \mu_3 \leq 26.62$

وهنا يجب أن نشير إلى أن نتائج تحليل بيانات المؤشر X_2 بواسطة الرسم الهامشي $Q-Q$ في المستوى، لم تعطينا أي مبرر جدي لاعتبار X_2 متحولاً خاضعاً للتوزيع الطبيعي . ولكننا مع ذلك، نجد أن حجم العينة الكبير $n = 87$ يعتبر حجماً كافياً للحكم على أن هذه البيانات لا تخل بشرط التوزيع الطبيعي .

ملاحظة: حتى نضمن تحقق قيمة الاحتمال $(1-\propto)$ لجميع المجالات المتزامنة لمركبات المتوسطات نلاحظ أن المجالات المتزامنة T^2 التي تعتمد على على p و n معاً ، تكون أطوالها أكبر من أطوال المجالات المنفردة t ، الي تعتمد على n فقط ، مع إن كل منهما يحقق الاحتمال $(1-\propto)$.

فمثلاً نجد أنه عندما يكون لدينا $1-\propto= 0.95$ و $n = 15$ و $p = 4$ فإن مضاريب المقادير $\sqrt{\frac{S_{jj}}{n}}$ في العلاقة (4−51) تكون مساوية لما يلي:

$$t_{n-1}\left(\frac{\propto}{2}\right) = t_{14}(0.025) = 2.145$$

بينما نجد أن مضاريب المقدار $\sqrt{\frac{S_{jj}}{n}}$ في العلاقة (4−69) تكون مساوية لما يلي:

$$\sqrt{\frac{P(n-1)}{n-P}F_{P\ \ n-P}(\propto)} = \sqrt{\frac{4(14)}{11}F_{4\ \ 11}(0.05)} = \sqrt{\frac{4(14)}{11}(3.36)4} = 4.14$$

وهكذا نجد أن المجالات في العلاقة (4−69) تكون أوسع منها في العلاقة (4−51) . أي إن المجالات المتزامنة تكون أوسع من المجالات المنفردة بنسبة تساوي [في هذا المثال] ما يلي:

$$\frac{(4{,}14 - 2{,}145)}{2{,}145} * 100 = 93\%$$

ونورد فيما يلي جدولاً لبعض القيم الحرجة للمجالات المنفردة t والمجالات المتزامنة T^2 حسب قيم n و P وعندما يكون $1-\propto= 0.95$.

جدول (4−1): أطوال المجالات المنفردة والمتزامنة حسب قيم n و P **وعندما يكون** $\propto= 0.05$.

حجم العينة n	القيمة الحرجة $t_{n-1}(0.025)$	القيمة الحرجة لـ $\sqrt{\frac{P(n-1)}{n-P}}F_{P\ \ n-P}(\propto)$	
		$P = 4$	$P = 10$
15	**2.145**	**4.14**	**11.52**
25	**2.064**	**3.60**	**6.39**
50	**2.010**	**3.31**	**5.05**
100	**1.970**	**3.19**	**4.61**
∞	**1.960**	**3.08**	**4.28**

المصدر: [Johnson wichern P. 189]

ومن الجدول السابق نلاحظ أن سعة المجالات المتزامنة T^2 بالنسبة للمجالات المنفردة t ، تزداد مع زيادة P (إذا كانت n ثابتة) وتتناقص مع ازدياد n (إذا كانت P ثابتة) .

4-4-3: طريقة (بونفيروني **Bonferroni**) للمقارنات المتعددة :

وهي تطبق عندما يكون عدد المتحولات p وبالتالي عدد عبارات الثقة المنفردة قليلاً ، ومن الممكن أن تكون أفضل من طريقة المجالات المتزامنة، أي أنه إذا كان p عدد مركبات شعاع المتوسط $\boldsymbol{\mu}$ قليلاً، أو كان m عدد التراكيب الخطية قليلاً، فإنه يمكن تطوير مجالات الثقة المتزامنة وجعلها أصغر مما هي في عبارات T^2 المذكورة في (4-69) ، والطريقة البديلة للمقارنات المتعددة تدعى طريقة: (بونفيروني) لأنها تعتمد على متراجحة (بونفيروني) في الاحتمالات . ولشرح هذه المتراجحة، نفترض مسبقاً- قبل سحب العينة- إنه يطلب منا إنشاء عبارات الثقة لـ $\boldsymbol{\ell'_i} * \boldsymbol{\mu}$ في m تركيباً خطياً معطياً كما يلي:

$$\begin{aligned}
\boldsymbol{\ell'_1} * \boldsymbol{\mu} &= \ell_{11}\mu_1 + \ell_{12}\mu_2 + \cdots + \ell_{1P}\mu_P \\
\boldsymbol{\ell'_2} * \boldsymbol{\mu} &= \ell_{21}\mu_1 + \ell_{22}\mu_2 + \cdots + \ell_{2P}\mu_P \qquad (72-4) \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\boldsymbol{\ell'_m} * \boldsymbol{\mu} &= \ell_{m1}\mu_1 + \ell_{m2}\mu_2 + \cdots + \ell_{mP}\mu_P
\end{aligned}$$

ولنفترض أن الحادث C_i هو حادث تحقق انتماء القيمة $\boldsymbol{\ell'_i} * \boldsymbol{\mu}$ إلى عبارة الثقة المطلوبة، وإن احتمال تحقق هذا الحادث فيها يساوي:

$$P[C_i true] = 1 - \propto_i \qquad i : 1\ 2\ 3 \dots\ m \qquad (73-4)$$

وبذلك نجد أن احتمال عدم تحقق الحادث C_i يساوي :

$$P[C_i false] = \propto_i \qquad (74-4)$$

ومن جهة أخرى نجد أن الاحتمال الكلي لتحقيق جميع الحوادث C_i معاً (بشكل متزامن) يساوي:

$$P[ALL\ C_i\ true] = 1 - P\begin{bmatrix} at\ Last\ one\ C_i\ false \\ \text{عدم تحقق واحد على الاقل من } C_i \end{bmatrix}$$

$$P[ALL\ C_i\ true] \geq 1 - \sum_{i=1}^{m} P[C_i\ false]$$

$$P[ALL\ C_i\ true] \geq 1 - (\propto_1 + \propto_2 + \cdots + \propto_m] \qquad (75-4)$$

وبذلك نحصل على حالة خاصة لمتراجحة (بونفيروني)، التي تعطينا وسيلة للتحكم في احتمال الخطأ الكلي $[\propto_1 + \propto_2 + \cdots + \propto_m]$ بغض النظر عن الارتباطات الهيكلية بين عبارات الثقة .

وإن هذه الوسيلة تكون مفيدة للتحكم في معدل الخطأ من أجل مجموعة من العبارات الهامة وتحقيق التوازن بينها بواسطة خيار آخر من العبارات غير الهامة .

والآن سنقوم بتطوير تقديرات المجالات المتزامنة من أجل المجموعة المختصرة والمؤلفة من m تركيباً فقط ، التي تتضمن المركبات μ_i من $\boldsymbol{\mu}$. واعتماداً على الأهمية النسبية لهذه المركبات، نجد أن المجال المنفرد t للمركبة μ_i وباحتمال دلالة خاصة $\propto_i$ يعطى من خلال (4-51) التالية :

$$P\left[\bar{x}_i - t_{n-1}\left(\frac{\propto_i}{2}\right)\sqrt{\frac{S_{ii}}{n}} \le \mu_i \le \bar{x}_i + t_{n-1}\left(\frac{\propto_i}{2}\right)\sqrt{\frac{S_{ii}}{n}}\right] = 1 - \propto_i \qquad i: 1\ 2\ 3 \dots m \qquad \mathbf{(76-4)}$$

وإذا جعلنا (أو افترضنا) أن قيم $\propto_i$ متساوية وتساوي : $\propto_i = \frac{\propto}{m}$ وعوضنا ذلك في (4–76)، فإننا نجد أن المجال السابق لـ μ_i يأخذ الشكل التالي:

$$P\left[\bar{x}_i - t_{n-1}\left(\frac{\propto}{2m}\right)\sqrt{\frac{S_{ii}}{n}} \le \mu_i \le \bar{x}_i + t_{n-1}\left(\frac{\propto}{2m}\right)\sqrt{\frac{S_{ii}}{n}}\right] = 1 - \frac{\propto}{m} \qquad i: 1\ 2\ 3 \dots m \qquad \mathbf{(77-4)}$$

وبناء على متراجحة (بونفيروني) (4–75) يمكننا أن نستخلص مباشرة أن الاحتمال الكلي لأن تتضمن جميع المجالات في (4–77) التوقعات μ_i يساوي :

$$P\left[ALL : \bar{x}_i \pm t_{n-1}\left(\frac{\propto}{2m}\right)\sqrt{\frac{S_{ii}}{n}} contains\ \mu_i : For\ All_i\right] \ge 1 - \left[\frac{\propto}{m} + \frac{\propto}{m} + \cdots \frac{\propto}{m}\right]$$

أي أن

$$P\left[ALL\ contains\ \mu_i\right] \ge 1 - \propto \qquad (78-4)$$

وبذلك نجد أن العلاقة (4–78) تعطينا مجالات منفردة لـ μ_i باحتمال ثقة أكبر أو يساوي $(1 - \propto)$.

ويفضل في كثير من الحالات أن نجعل أو نضع $(m = P)$ أي نجعل عدد التراكيب الخطية m مساوياً لعدد المتحولات p ، وعندها نجد أن العلاقة (4–78) تعطينا المجالات المتزامنة التالية:

$$\begin{aligned}
&\bar{x}_1 - t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{11}}{n}} \le \mu_1 \le \bar{x}_1 + t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{11}}{n}} \\
&\bar{x}_2 - t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{22}}{n}} \le \mu_2 \le \bar{x}_2 + t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{22}}{n}} \qquad (79-4) \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&\bar{x}_P - t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{PP}}{n}} \le \mu_P \le \bar{x}_P + t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{PP}}{n}}
\end{aligned}$$

وهنا نلاحظ أن احتمال الثقة الكلي لهذه المجالات المتزامنة أكبر أو يساوي: $(1 - \propto)$، وإن احتمال الثقة المنفرد لكل منها يساوي: $\propto_i = 1 - \frac{\propto}{P}$.

كما نلاحظ أن المجالات (4–79) تشابه هيكلياً المجالات المتزامنة في (4–69) ، حيث تم استبدال المضاريب $\frac{P(n-1)}{n-P} F_{P\ n-P}(\propto)$ بالمضاريب $t_{n-1}\left(\frac{\propto}{2P}\right)$ ، أي أن أطوال (سعات) المجالات قد تغيرت (تقلصت) ولكنها حافظت على نفس الصيغة الهيكلية في (4–67) وهي:

$$\boldsymbol{\ell}'\overline{\boldsymbol{X}} \pm (Critical\ Value)\sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \qquad (80-4)$$

ولحساب معدل التقلص الذي طرأ على المجالات المتزامنة T^2 بعد حساب المجالات المتزامنة بطريقة (بونفيروني)، نحسب نسبة طول مجال (بونفيروني) على طول مجال T^2 من العلاقة :

$$Q = \frac{\text{طول مجال (بونفيروني)}}{\text{طول مجال } T^2} = \frac{L_P}{L_{T^2}} = \frac{t_{n-1}\left(\frac{\propto}{2P}\right)}{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} \qquad (81-4)$$

ونلاحظ أن هذه النسبة لا تتعلق بالكميات العشوائية $\bar{\boldsymbol{X}}$ و $\boldsymbol{S}$ ، وبذلك نجد أن معدل التقلص يساوي :

$$R = 1 - Q \qquad (82-4)$$

مثال (4–6): لنتابع معالجة المثال (4–3) عن التعرق عند النساء . ولنقم بحساب المجالات المنفردة والمتزامنة بالطرائق المختلفة، طريقة المجالات المنفردة t، طريقة المجالات المتزامنة T^2 وطريقة (بونفيروني) فنجد أن :

أولاً: طريقة المجالات المنفردة t :وفيها نطبق العلاقة :

$$\bar{x}_i - t_{n-1}\left(\frac{\propto}{2}\right)\sqrt{\frac{S_{ii}}{n}} \le \mu_i \le \bar{x}_i + t_{n-1}\left(\frac{\propto}{2}\right)\sqrt{\frac{S_{ii}}{n}} \qquad (a)$$

وباعتبار أن $\propto = 0.05$ وأن $n = 20$ فإن $t_{n-1}\left(\frac{\propto}{2}\right) = t_{19}(0.025) = 2.093$ وبما أن:

$$\bar{\boldsymbol{X}} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix} \qquad \boldsymbol{S} = \begin{bmatrix} 2.879 & 100.002 & -1.810 \\ 10.002 & 199.798 & -5.627 \\ -1.810 & -5.627 & 3.628 \end{bmatrix}$$

نعوض ذلك في العلاقة (a) السابقة لحساب مجالات التوقعات μ_1 و μ_2 و μ_3 فنحصل على المجالات المنفردة التالية :

$$4.640 - (2.093)\sqrt{\frac{2.879}{20}} \le \mu_1 \le 4.640 + (2.093)\sqrt{\frac{2.879}{20}}$$

$$3.8459 \le \mu_1 \le 5.4341 \Longrightarrow L_1 = 1.5881$$

$$45.400 - (2.093)\sqrt{\frac{199.798}{20}} \le \mu_2 \le 45.400 + (2.093)\sqrt{\frac{199.798}{20}}$$

$$38.784 \le \mu_2 \le 52.0153 \Longrightarrow L_2 = 13.2306$$

$$9.965 - (2.093)\sqrt{\frac{3.628}{20}} \le \mu_3 \le 9.965 + (2.093)\sqrt{\frac{3.628}{20}}$$

$$9.0736 \le \mu_3 \le 10.8564 \Longrightarrow L_3 = 1.27828$$

ثانياً: طريقة المجالات المتزامنة T^2 : وفيها نطبق العلاقة:

$$\bar{x}_i - \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{11}}{n}} \le \mu_i \le \bar{x}_i + \sqrt{\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)}\sqrt{\frac{S_{11}}{n}} \qquad (b)$$

نقوم أولاً بحساب معامل القيمة الحرجة فنجد أن:

$$\frac{P(n-1)}{n-P}F_{P\ n-P}(\propto)=\frac{3(19)}{17}F_{3,17}(0.05)=\frac{3(19)}{17}(3,1968)=10,7187$$

ثم نعوض ذلك في العلاقة (b) السابقة لحساب مجالات التوقعات μ_1 و μ_2 و μ_3 ، فنحصل على المجالات المتزامنة التالية :

$$4.640-\sqrt{10.7187}*\sqrt{\frac{2.879}{20}}\le\mu_1\le 4.640+\sqrt{10.7187}*\sqrt{\frac{2.879}{20}}$$

$$3.3978\le\mu_1\le 5.8822\Longrightarrow L_1=2.4844$$

$$45.400-\sqrt{10.7187}*\sqrt{\frac{199.798}{20}}\le\mu_2\le 45.400+\sqrt{10.7187}*\sqrt{\frac{199.798}{20}}$$

$$35.0521\le\mu_2\le 55.7479\Longrightarrow L_2=20.6958$$

$$9.965-\sqrt{10.7187}*\sqrt{\frac{3.628}{20}}\le\mu_3\le 9.965+\sqrt{10.7187}*\sqrt{\frac{3.628}{20}}$$

$$8.5706\le\mu_3\le 11.3594\Longrightarrow L_3=2.7888$$

ثالثاً: طريقة (بونفيروني) : وفيها نطبق العلاقة:

$$\bar{x}_i-t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{ii}}{n}}\le\mu_i\le\bar{x}_i+t_{n-1}\left(\frac{\propto}{2p}\right)\sqrt{\frac{S_{ii}}{n}}\qquad (c)$$

لذلك نقوم بحساب القيمة الحرجة $t_{n-1}\left(\frac{\propto}{2P}\right)$ فنجد أن:

$$t_{n-1}\left(\frac{\propto}{2p}\right)=t_{19}\left(\frac{0.05}{2*3}\right)=t_{19}(0.00833)=2.625$$

ثم نعوض ذلك في العلاقة (c) السابقة لحساب مجالات التوقعات μ_1 و μ_2 و μ_3 فنجد أن:

$$4.640-(2.625)\sqrt{\frac{2.879}{20}}\le\mu_1\le 4.640+(2.625)\sqrt{\frac{2.879}{20}}$$

$$3.6441\le\mu_1\le 5.6359\Longrightarrow L_1=1.9918$$

$$45.400-(2.625)\sqrt{\frac{199.798}{20}}\le\mu_2\le 45.400+(2.625)\sqrt{\frac{199.798}{20}}$$

$$37.1032\le\mu_2\le 53.6968\Longrightarrow L_2=16.5936$$

$$9.965-(2.625)\sqrt{\frac{3.628}{20}}\le\mu_3\le 9.965+(2.625)\sqrt{\frac{3.628}{20}}$$

$$8.8470\le\mu_3\le 11.0830\Longrightarrow L_3=2.2360$$

ومن أجل مقارنة أطوال هذه المجالات حسب الطرائق الثلاثة نضعها في الجدول التالي:

جدول (4-2): أطوال مجالات الثقة حسب الطرائق الثلاثة لـ μ_1 و μ_2 و μ_3

النسبة Q	طريقة (بونفيروني)	طريقة المجالات المتزامنة T^2	طريقة المجالات المنفردة t	التوقع
0.8017	1.9918	2.4844	1.5881	μ_1
0.8017	16.5936	20.6958	13.2306	μ_2
0.8017	2.2360	2.7888	1.27828	μ_3

ومن الجدول السابق نلاحظ كيف أن أطوال (سعات) المجالات تتغير من طريقة لأخرى، ونلاحظ أن أطوال المجالات حسب المجالات المتزامنة T^2 أكبر من أطوالها في الطريقتين الأخيرتين . كما نلاحظ أن أطوال المجالات حسب طريقة (بونفيروني) تأخذ وضعاً وسطاً بين الطريقتين الأولى والثانية .

وأخيراً نلاحظ أن نسبة طول كل من مجالات (بونفيروني) على طول المجال المقابل له من المجالات المتزامنة T^2 تساوي :

$$Q = \frac{1.9918}{2.4844} = \frac{16.5936}{20.6958} = \frac{2.2360}{2.7888} = 0.8017$$

وإن معدل التقلص يساوي: $1 - Q = 1 - 0.8017 = 0.1983 = (19.83\%)$

وبصورة عامة فإن طريقة (بونفيروني) تعطينا مجالات متزامنة أقل طولاً من المجالات المتزامنة T^2 ، وللاطلاع على بعض تغيرات نسبة التقلص، نورد الجدول (4-3) من أجل قيم مختارة لـ p و n ولحالة m=p .

جدول (4-3): نسبة أطوال مجالات (بونفيروني) على أطول مجالات T^2 $\left(Q = \frac{L_P}{L_{T^2}}\right)$ لقيم مختارة لـ n و p، وذلك عندما يكون $\propto = 0.05$ $1 - \propto$، ويكون $\propto_i = \frac{0.05}{m}$ ، وعندما يكون $m = p$

حجم العينة n	$P = 2$	$P = 4$	$P = 10$
15	0.88	0.69	0.29
25	0.90	0.75	0.48
50	0.91	0.78	0.58
100	0.91	0.80	0.62
∞	0.91	0.81	0.66

المصدر : [Johnson, Wichern P. 190]

ويتضح من هذا الجدول أن طريقة (بونفيروني) تعطينا مجالات أقصر من مجالات T^2 . وهي طريقة سهلة التطبيق وتعطينا مجالات ثقة قصيرة نسبياً تلبي عمليات الاستدلال في التحليل متعدد المتغيرات . وهنا نلاحظ أن نسبة الأطوال ($Q = \frac{L_P}{L_{T^2}}$) تتزايد مع حجم العينة n، وإنها تستقر عند القيم المبينة في الجدول عندما يصبح حجم العينة n لانهائياً .

4-4-4 : الاستدلال حول شعاع المتوسطات في حالة العينات الكبيرة :

عندما يكون حجم العينة n كبيراً، فإن اختبارات الفروض ومناطق الثقة للمتوسط $\boldsymbol{\mu}$ ، يمكن أن تعمم بدون افتراض أو اشتراط أن يكون المجتمع طبيعياً. كما يمكننا الاستدلال حول متوسط المجتمع حتى عندما يكون التوزيع منقطعاً. وذلك لأن حجم العينة الكبير يغطي معظم الخواص العملية للمجتمع الطبيعي، ويجعل اختبارات الفرضيات وعبارات الثقة المتزامنة محققة (تقريباً) لعلاقة احتواء قيمها الأسمية. ولإنشاء مناطق الثقة المنفردة أو المتزامنة في هذه الحالة لنفترض أنه لدينا المتحولات (غير المشروطة) $X_1 \ X_2 \ ... \ X_P$ ، والتي شعاع توقعاتها ومصفوفة تبايناتها المشتركة في المجتمع كما يلي:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix} \quad \boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} \quad \boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & \cdots & V_{1P} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2P} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ V_{P1} & V_{P2} & V_{P3} & \cdots & V_{PP} \end{bmatrix} \quad (82-4)$$

ويقابلها في العينة شعاع المتوسطات $\overline{\boldsymbol{X}}$ ومصفوفة التباينات المشتركة $\mathbf{S}$:

$$\overline{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_P \end{bmatrix} \qquad \boldsymbol{S} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1P} \\ S_{21} & S_{21} & \cdots & S_{2P} \\ \vdots & \vdots & \dots & \vdots \\ S_{P1} & S_{P2} & \cdots & S_{PP} \end{bmatrix} \qquad (83-4)$$

ومن الفقرة (4-4) والعلاقة (4-59) وجدنا أن مجال الثقة المتزامن للشعاع $\boldsymbol{\mu}$ بمستوى دلالة $\propto$ يحسب من العلاقة التالية:

$$T^2 = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-\mathbf{1}}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq C^2(\propto) \qquad (84-4)$$

والتي يمكن كتابتها على الشكل التالي:

$$T^2 = (\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\left(\frac{\boldsymbol{S}}{\boldsymbol{n}}\right)^{-\mathbf{1}}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq C^2(\propto) \qquad (85-4)$$

وبذلك نجد أن الطرف الأيسر T^2 يعبر عن مربع المسافة الإحصائية للمتوسط $\overline{\boldsymbol{X}}$ عن التوقع $\boldsymbol{\mu}$ في الفضاء R^P . وهو يخضع تقاربياً لتوزيع χ_P^2 (عندما يكون n كبيراً) . أي أنه يمكننا استبدال الحد الأعلى له (الطرف الأيمن $C^2(\propto)$) بالقيمة الحرجة لمتحول χ^2 المقابلة لمستوى الدلالة $\propto$ ولدرجة الحرية p . وبذلك يمكننا كتابة العلاقة (4-85) كما يلي:

$$T^2 = (\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\left(\frac{\boldsymbol{S}}{\boldsymbol{n}}\right)^{-\mathbf{1}}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \chi_P^2(\propto) \qquad (86-4)$$

وهذا يعني أن احتمال الثقة لهذه المنطقة يساوي:

$$P[n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'(\boldsymbol{S})^{-\mathbf{1}}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \chi_P^2(\propto)] = 1 - \propto \qquad (87-4)$$

وهنا نلاحظ أن العلاقة (4-87) تعطينا منطقة الثقة لشعاع التوقع $\boldsymbol{\mu}$ باحتمال ثقة $(1 - \propto)$.

فإذا كان عدد المتحولات $p = 2$ فإن العلاقة (4-87) تعطينا القطوع الناقصة التالية :

$$n\left[(\bar{X}_1 - \mu_1), (\bar{X}_2 - \mu_2)\right] * \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{21} \end{bmatrix}^{-1} * \begin{bmatrix} (\bar{X}_1 - \mu_1) \\ (\bar{X}_2 - \mu_2) \end{bmatrix} \leq \chi_P^2(\propto) \qquad (88-4)$$

وإذا أخذنا أي تركيب خطي للمتحولات $X_1 \ \ X_2 \ \ \dots \ \ X_P$ نجد أن :

$$Z = \boldsymbol{\ell}'\boldsymbol{X} \Longrightarrow \mu_Z = \boldsymbol{\ell}'\boldsymbol{\mu} \quad , \qquad \bar{Z} = \boldsymbol{\ell}'\overline{\boldsymbol{X}} \quad , \quad \tilde{\sigma}_Z^2 = \boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell} \qquad (89-4)$$

وعندها فإن العلاقة (4–54) تكتب على الشكل التالي :

$$t^2 = \left(\frac{\bar{Z} - \mu_Z}{\sigma_z/\sqrt{n}}\right)^2 = \left[\frac{n(\boldsymbol{\ell}'\overline{\boldsymbol{X}} - \boldsymbol{\ell}'\boldsymbol{\mu})^2}{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}\right] \leq \chi_P^2(\propto) \qquad (90-4)$$

وهكذا نجد أن وأن:

$$(\boldsymbol{\ell}'\overline{\boldsymbol{X}} - \boldsymbol{\ell}'\boldsymbol{\mu})^2 \leq \chi_P^2(\propto) * \frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n} \qquad (91-4)$$

$$-\sqrt{\chi_P^2(\propto)} * \sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \leq (\boldsymbol{\ell}'\overline{\boldsymbol{X}} - \boldsymbol{\ell}'\boldsymbol{\mu}) \leq +\sqrt{\chi_P^2(\propto)}\sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \qquad (92-4)$$

وبطرح $\boldsymbol{\ell}'\overline{\boldsymbol{X}}$ من أطرافها الثلاثة وضربها بـ (-1) مع تغيير الاتجاه، نحصل على أن مجالات الثقة المتزامنة لـ $\boldsymbol{\ell}'\boldsymbol{\mu}$ في العينات الكبيرة تعطى من خلال العلاقة التالية:

$$\boldsymbol{\ell}'\overline{\boldsymbol{X}} - \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \leq \boldsymbol{\ell}'\boldsymbol{\mu} \leq \boldsymbol{\ell}'\overline{\boldsymbol{X}} + \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \qquad (93-4)$$

وللحصول على عبارات الثقة المتزامنة للتوقعات $\mu_1 \ \ \mu_2 \ \ \dots \ \ \mu_P$ نعطي منقول شعاع الأمثال $\boldsymbol{\ell}'$ القيم المتتالية:

$$\boldsymbol{\ell}_1' = (1,0,0,0,0) \ \ , \boldsymbol{\ell}_2' = (0,1,0,0,0) \, , \dots \dots , \boldsymbol{\ell}_P' = (0,0,0,0,1)$$

فنحصل على عبارات المجالات المتزامنة T^2 التالية [كما فعلنا في (4–51) أو في (4–69)]

$$\begin{aligned}
&\bar{x}_1 - \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{S_{11}}{n}} \leq \mu_1 \leq \bar{x}_1 + \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{S_{11}}{n}} \\
&\bar{x}_2 - \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{S_{22}}{n}} \leq \mu_2 \leq \bar{x}_2 + \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{S_{22}}{n}} \qquad (94-4) \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&\bar{x}_P - \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{S_{PP}}{n}} \leq \mu_P \leq \bar{x}_P + \sqrt{\chi_P^2(\propto)} * \sqrt{\frac{S_{PP}}{n}}
\end{aligned}$$

وإن احتمال الثقة الكلي لهذه المجالات أكبر أو يساوي $(1 -\propto)$، وذلك من أجل جميع العبارات السابقة. وهذا الاحتمال ما هو إلا نتيجة لنظرية تقارب التوزيعات في حالة العينات الكبيرة .

مثال (4–7): لدراسة قدرات الطلاب الناجحين في الثانوية (علمي) في مدينة ما، سحبنا عينة عشوائية منهم بحجم $n = 96$ طالباً. وحسبنا متوسطات علاماتهم في 7 مقررات $X_1 \ \ X_2 \ \ X_3 \dots X_7$ فكانت مع انحرافاتها المعيارية كما يلي:

جدول (4-4): المتوسطات والانحرافات المعيارية لعلامات أفراد العينة في المقررات

رمز المقرر	المتوسط $\bar{x}_i$	الانحراف المعياري $\sqrt{S_{ii}} = S_i$
X_1	28.1	5.76
X_2	26.6	5.85
X_3	35.4	3.82
X_4	34.2	5.12
X_5	23.6	3.76
X_6	22.0	3.93
X_7	22.7	4.03

المصدر: [Johnson, Wichern P. 193]

والمطلوب إيجاد مجالات الثقة المتزامنة المقابلة لاحتمال ثقة كلي 90% للتوقعات $\mu_1\ \mu_2\ \dots\ \mu_7$.

الحل: لإيجاد تلك المجالات المتزامنة لتوقعات هذه العينة الكبيرة نطبق العلاقة (4-94)، ولذلك نقوم أولاً بحساب القيمة الحرجة $\chi^2_P(\propto)$ المشتركة فيها ، فنجد أنها $\chi^2_P(\propto) = \chi^2_7(0.10) = 12.02$ ، ثم نعوض بيانات الجدول (4-4) في العلاقة (4-94)، فنحصل على المجالات T^2 التالية:

$$28.1 - \sqrt{12.02} * \frac{5.76}{\sqrt{96}} \le \mu_1 \le 28.1 + \sqrt{12.02} * \frac{5.76}{\sqrt{96}}$$

أي أن: $26.06 \le \mu_1 \le 30.14$

$$26.6 - \sqrt{12.02} * \frac{5.85}{\sqrt{96}} \le \mu_2 \le 26.6 + \sqrt{12.02} * \frac{5.85}{\sqrt{96}}$$

أي أن: $24.53 \le \mu_2 \le 28.67$

$$35.4 - \sqrt{12.02} * \frac{3.82}{\sqrt{96}} \le \mu_3 \le 35.4 + \sqrt{12.02} * \frac{3.82}{\sqrt{96}}$$

أي أن: $34.05 \le \mu_3 \le 36.75$

$$34.2 - \sqrt{12.02} * \frac{5.12}{\sqrt{96}} \le \mu_4 \le 34.2 + \sqrt{12.02} * \frac{5.12}{\sqrt{96}}$$

أي أن: $32.39 \le \mu_4 \le 36.01$

$$23.6 - \sqrt{12.02} * \frac{3.76}{\sqrt{96}} \le \mu_5 \le 33.6 + \sqrt{12.02} * \frac{3.76}{\sqrt{96}}$$

أي أن: $22.27 \le \mu_5 \le 24.93$

$$22.0 - \sqrt{12.02} * \frac{3.93}{\sqrt{96}} \le \mu_6 \le 22.0 + \sqrt{12.02} * \frac{3.93}{\sqrt{96}}$$

أي أن: $20.61 \le \mu_6 \le 23.39$

$$22.7 - \sqrt{12.02} * \frac{4.03}{\sqrt{96}} \le \mu_7 \le 22.7 + \sqrt{12.02} * \frac{4.03}{\sqrt{96}}$$

أي أن: $21.27 \le \mu_7 \le 24.41$

ملاحظة: إن تطبيق طريقة (بونفيروني) على العينات الكبيرة، يقتضي تعديل العلاقة (4–79) باستبدال القيمة الحرجة $t_{n-1}\left(\frac{\propto}{2P}\right)$ بالقيمة الحرجة للتوزيع الطبيعي المعياري $Z\left(\frac{\propto}{2P}\right)$ المقابلة لمستوى الدلالة $\left(\frac{\propto}{2P}\right)$ وتصبح العلاقة العامة لها كما يلي:

$$\bar{x}_i - Z\left(\frac{\propto}{2P}\right)\frac{S_i}{\sqrt{n}} \le \mu_i \le \bar{x}_i - Z\left(\frac{\propto}{2P}\right)\frac{S_i}{\sqrt{n}} \qquad (95-4)$$

مثال (4–8): لإيجاد مجالات (بونفيروني) للمثال السابق باحتمال ثقة 90% ، نقوم بحساب القيمة الحرجة المعيارية $Z\left(\frac{\propto}{2P}\right)$ ، فنجد أن:

$$Z\left(\frac{\propto}{2P}\right) = Z\left(\frac{0.10}{2*7}\right) = Z(0.00714) = 2.45$$

وبذلك فإن مجالات (بونفيروني) للتوقعات $\mu_1 \;\; \mu_2 \;\; \dots \;\; \mu_7$ حسب بيانات الجدول(4–4) تساوي ما يلي:

$$28.1 - (2.45)\frac{5.76}{\sqrt{96}} \le \mu_1 \le 28.1 + (2.45)\frac{5.76}{\sqrt{96}}$$

أي أن: $$26.66 \le \mu_1 \le 29.54$$

$$26.6 - (2.45)\frac{5.85}{\sqrt{96}} \le \mu_2 \le 26.6 + (2.45)\frac{5.85}{\sqrt{96}}$$

أي أن: $$25.137 \le \mu_2 \le 28.063$$

$$35.4 - (2.45)\frac{3.82}{\sqrt{96}} \le \mu_3 \le 35.4 + (2.45)\frac{3.82}{\sqrt{96}}$$

أي أن: $$34.445 \le \mu_3 \le 36.355$$

$$34.2 - (2.45)\frac{5.12}{\sqrt{96}} \le \mu_4 \le 34.2 + (2.45)\frac{5.12}{\sqrt{96}}$$

أي أن: $$32.920 \le \mu_4 \le 35.480$$

$$23.6 - (2.45)\frac{3.76}{\sqrt{96}} \le \mu_5 \le 23.6 + (2.45)\frac{3.76}{\sqrt{96}}$$

أي أن: $$22.660 \le \mu_5 \le 24.540$$

$$22.0 - (2.45)\frac{3.93}{\sqrt{96}} \le \mu_6 \le 22.0 + (2.45)\frac{3.93}{\sqrt{96}}$$

أي أن: $$21.017 \le \mu_6 \le 22.983$$

$$22.7 - (2.45)\frac{4.03}{\sqrt{96}} \le \mu_7 \le 22.7 + (2.45)\frac{4.03}{\sqrt{96}}$$

أي أن: $$21.692 \le \mu_7 \le 23.708$$

وهنا نلاحظ أن أطوال هذه المجالات أصغر من أطوال المجالات T^2 ، وإن قيمة النسبة Q بينهما تساوي:

$$Q = \frac{L_B}{L_{T^2}} = 0.7069 = (70.69\,\%)$$

5-4-4: الاستدلال حول نسبة الفئات المتكاملة في حالة العينات الكبيرة :

لمعالجة هذه المسألة، يمكننا حساب وتقدير نسبة اي فئة في المجتمع أو العينة من خلال تعريف متحولات ثنائية مقابلة لكل منها .

فمثلاً لحساب نسبة أو تقدير نسبة الناجحين P في مقرر ما. نعرف متحولاً ثنائياً X ، بحيث يأخذ X القيمة (1) إذا كان الطالب ناجحاً ويأخذ القيمة (0) إذا كان الطالب راسباً . ونكتب ذلك كما يلي:

$$X = \begin{cases} 1 & : \quad \text{إذا كان ناجحاً} \\ 0 & : \quad \text{إذا كان راسباً} \end{cases} \qquad (96-4)$$

وعندما نسحب عينة من المجتمع بحجم n، فإن X يأخذ احدى القيمتين 1 أو 0. مقابل كل طالب فيها. ونحصل منها على أن (مثلاً) .

$$X: \; 1,0,1,1,0, \dots ,0,1 \qquad \left(\text{قيمة } n\right)$$

فإذا أخذنا مجموع قيم X فإننا سنحصل على عدد الناجحين m ونكتب ذلك كما يلي:

$$\sum_{i=1}^{n} x_i = (1+0+1+1+0 \; \dots 1) = m$$

وإذا أخذنا المتوسط $\bar{X}$ فإننا سنحصل على النسبة في العينة $\tilde{P}$ لأن:

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{m}{n} = \tilde{P} \qquad (97-2)$$

أي أنه يمكننا النظر إلى نسبة أي خاصة في المجتمع أو العينة P، على أنها متوسط المتحول الثنائي X المقابل لها $\bar{x}$، وانطلاقاً من ذلك يمكننا تطبيق أساليب الاستدلال التي استخدمناها حول المتوسطات على الاستدلال حول النسب . مع الانتباه إلى ضرورة حساب مصفوفة التباين المشترك بدلالة تلك النسب كما سنرى لاحقاً .

والآن لنفترض أننا نريد تقدير شعاع النسب لعدة خواص منفصلة ومتكاملة في المجتمع [مثل نسب الطلاب حسب المحافظات، أو نسب الموظفين حسب الحالة الاجتماعية] وهذا يعني أن المجتمع يكون مقسماً إلى عدة فئات منفصلة ومتكاملة . وسنرمز لعددها في المجتمع بالرمز q، (وإذا لم تكن متكاملة نشكل من جميع الفئات المتبقية فئة خاصة نرمز لها بـ $q+1$)، وتسهيلاً لقضايا التحليل نفترض أن فئات المجتمع منفصلة ومتكاملة ، ولنرمز لنسب هذه الفئات في المجتمع بالشعاع **P** ونكتبه كما يلي :

$$\boldsymbol{P} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_q \end{bmatrix} \qquad (98-4)$$

وحتى نقدر هذه النسب علينا أن نسحب عينة عشوائية من المجتمع بحجم n، ونعرف مقابل كل فئة k في المجتمع، متحولاً ثنائياً X_K، ثم نجمع المعلومات عنها ونحسب شعاع المتوسطات $\overline{\boldsymbol{X}}$، فنحصل منها على شعاع النسب $\widetilde{\boldsymbol{P}}$ في العينة n . والذي يساوي :

$$\widetilde{\boldsymbol{P}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_q \end{bmatrix} = \begin{bmatrix} \frac{m_1}{n} \\ \frac{m_2}{n} \\ \vdots \\ \frac{m_q}{n} \end{bmatrix} = \begin{bmatrix} \tilde{P}_1 \\ \tilde{P}_2 \\ \vdots \\ \tilde{P}_q \end{bmatrix} \qquad (99-4)$$

وبما أن الفئات المدروسة منفصلة ومتكاملة فإن كل عنصر في المجتمع (أو في العينة) ينتمي إلى فئة واحدة فقط . أي أن مجموع هذه النسب في المجتمع (أو في العينة) يجب أن يساوي الواحد . أي أن:

$$\sum_{K=1}^{q} \tilde{P}_K = 1 \qquad \text{أو} \qquad \sum_{K=1}^{q} P_k = 1 \qquad (100-4)$$

أي أن درجة الحرية لها ورتبة المصفوفة (4–99) تساوي $q-1$، ويمكن البرهان على أن توقع الشعاع $\widetilde{\boldsymbol{P}}$ يساوي الشعاع الأصلي $\mathbf{P}$ ، أي أن :

$$E(\widetilde{\boldsymbol{P}}) = \boldsymbol{P} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_q \end{bmatrix} \qquad (101-4)$$

وإن مصفوفة التباينات المشتركة لها في المجتمع تقدر من بيانات العينة ومن خلال العلاقة التالية :

$$\boldsymbol{S}_{\widetilde{\mathbf{P}}} = \boldsymbol{cov}(\widetilde{\boldsymbol{P}}) = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1q} \\ S_{21} & S_{22} & \cdots & S_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ S_{q1} & S_{q2} & \cdots & S_{qq} \end{bmatrix} = \begin{bmatrix} \tilde{P}_1(1-\tilde{P}_1) - \tilde{P}_1\tilde{P}_2 - \tilde{P}_1\tilde{P}_q \\ -\tilde{P}_2\tilde{P}_1 \quad \tilde{P}_2(1-\tilde{P}_2) - \tilde{P}_2\tilde{P}_q \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ -\tilde{P}_q\tilde{P}_1 - \tilde{P}_q\tilde{P}_2 \ldots \tilde{P}_q(1-\tilde{P}_q) \end{bmatrix} \qquad (102-4)$$

وذلك لأن تباين كل متحول ثنائي X_k يساوي : $\sigma_{kk}^2 = var(X_{ki}) = P_k(1-P_k)$

وإن التباين المشترك لأي متحولين ثنائيين X_k و X_h يساوي : $cov(X_h \ , X_k\ ) = -P_hP_k$

وللبرهان على ذلك يكفي أن نتذكر أن X_k يأخذ القيمة (1) باحتمال P_k، ويأخذ القيمة (0) باحتمال $(1-P_k)$، وعندها نجد أن توقعه وتباينه يساويان ما يلي :

$$\mu_k = P_K = E(X_{Ki}) = 1 * P_K + 0(1-P_K) = P_K \qquad (103-4)$$

$$\sigma_{KK}^2 = var(X_{Ki}) = E(X_K^2) - E^2(X_K) = 1^2 * P_K + 0^2(1-P_K) - P_K^2$$
$$= P_K - P_K^2 = P_K(1-P_K) \qquad (104-4)$$

أما التباينات المشتركة فنحسبها من العلاقة:

$$\sigma_{hK} = cov(X_{hi}\, X_{Ki}\ ) = E(X_{hi} * X_{Ki}) - E(X_{hi})E(X_{Ki}) \quad : \ h \neq k \qquad (105-4)$$

ولكن بما أن المتحولين X_{hi} و X_{Ki} مستقلان، فإن الجداء $(X_{hi} * X_{Ki}\ ) = 0$ ، وبذلك نجد أن :

$$\sigma_{hK} = E(0) - P_h * P_K = -P_h P_K \qquad : \ h \neq k \qquad (106-4)$$

وهي تأخذ قيماً سالبة .

والآن لنأخذ أي تركيب خطي لشعاع النسب P في المجتمع وليكن :

$$Z = \boldsymbol{\ell}' * \boldsymbol{P} = \ell_1 P_1 + \ell_2 P_2 + \cdots \ell_q P_q \qquad (107-4)$$

فإننا نجد أنه يقدر بواسطة التركيب الخطي لشعاع النسب في العينة :

$$\tilde{Z} = \boldsymbol{\ell}' * \tilde{\boldsymbol{P}} = \ell_1 \tilde{P}_1 + \ell_2 \tilde{P}_2 + \cdots \ell_q \tilde{P}_q \qquad (108-4)$$

وإننا نجد أن توقع وتباين ذلك التركيب الخطي يساويان :

$$\mu_Z = E\left(\boldsymbol{\ell}'\tilde{\boldsymbol{P}}\right) = \boldsymbol{\ell}' * E\left(\tilde{\boldsymbol{P}}\right) = \boldsymbol{\ell}' * \boldsymbol{P} \qquad (109-4)$$

$$S_Z^2 = var\left(\boldsymbol{\ell}'\tilde{\boldsymbol{P}}\right) = \boldsymbol{\ell}' * var\left(\tilde{\boldsymbol{P}}\right) * \boldsymbol{\ell} = \boldsymbol{\ell}' * \boldsymbol{S} * \boldsymbol{\ell} \qquad (110-4)$$

وبناء على المعادلات المشابهة لشعاع المتوسطات $\overline{\boldsymbol{X}}$ ، نجد أن منطقة الثقة المتزامنة لـ $\boldsymbol{\ell}'\boldsymbol{P}$. والمقابلة لاحتمال الثقة $(1-\propto)$ تعطى بالعلاقة العامة التالية :

$$\boldsymbol{\ell}'\tilde{\boldsymbol{P}} - \sqrt{\chi^2_{q-1}(\propto)} * \sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \leq \boldsymbol{\ell}'\boldsymbol{P} \leq \boldsymbol{\ell}'\tilde{\boldsymbol{P}} + \sqrt{\chi^2_{q-1}(\propto)} * \sqrt{\frac{\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}}{n}} \qquad (111-4)$$

حيث أن $\tilde{\boldsymbol{P}}$ هو تقدير **P** من العينة وأن **S** هي مصفوفة التباينات المشتركة للشعاع $\tilde{\boldsymbol{P}}$ ، وهي مصفوفة مربعة من المرتبة $q * q$ وإن عناصرها: $S_{KK} = P_K(1-P_K)$ وإن: $S_{hK} = -P_h P_K$

أما $\chi^2_{q-1}(\propto)$ فهي القيمةالحرجة لمتحول التوزيع χ^2_{q-1} عند مستوى دلالة $\propto$ ودرجة حرية $q-1$.

ولتطبيق العلاقة (111-4) يشترط أن يكون $(n-q-1)$ كبيراً بشكل يجعل عدد المفردات في كل فئة $(n\tilde{P})$ أكبر أو يساوي العدد 20: وهذا الشرط لا يتحقق إلا إذا كان n أكبر من $q * 20$ بشكل ملحوظ، وإذا أردنا أن نحصل على مجالات الثقة المنفردة لكل P_k فإننا نجعل المركبة k في $\boldsymbol{\ell}'$ مساوية للواحد، ونضع بقية المركبات أصفاراً مثل $\boldsymbol{\ell}' = (0,0,1,0,0,0)$ فنحصل بطريقة مشابهة للعلاقة (94-4) على ما يلي:

$$\tilde{P}_K - \sqrt{\chi^2_{q-1}(\propto)} * \sqrt{\frac{\tilde{P}_K\left(1-\tilde{P}_K\right)}{n}} \leq P_K \leq \tilde{P}_K + \sqrt{\chi^2_{q-1}(\propto)} * \sqrt{\frac{\tilde{P}_K\left(1-\tilde{P}_K\right)}{n}} \qquad (112-4)$$

وإذا أردنا أن نحصل على مجال الثقة للفرق بين أي نسبتين $(P_h - P_K)$، نجعل في الشعاع $\boldsymbol{\ell}'$ المركبة h مساوية لـ (1) والمركبة k مساوية (1−)، ونضع بقية المركبات أصفاراً ، فنحصل على المجال التالي:

$$\left(\tilde{P}_h - \tilde{P}_K\right) - \sqrt{\chi^2_{q-1}(\propto)}\sqrt{\frac{\tilde{P}_h\left(1-\tilde{P}_h\right) + \tilde{P}_k\left(1-\tilde{P}_k\right) + 2\tilde{P}_h\tilde{P}_K}{n}} \leq (P_h - P_K) \leq \left(\tilde{P}_k - \tilde{P}_K\right) +$$

$$+\sqrt{\chi^2_{q-1}(\propto)} * \sqrt{\frac{\tilde{P}_h\left(1-\tilde{P}_h\right) + \tilde{P}_K\left(1-\tilde{P}_K\right) + 2\tilde{P}_h\tilde{P}_K}{n}} \qquad (113-4)$$

وذلك لأنه لو جعلنا $\ell' = (1,-1,0,0,0)$ لوجدنا أن:

$$\boldsymbol{\ell' S \ell} = S_{11} + S_{22} - 2S_{12} = \widetilde{P}_1(1-\widetilde{P}_1) + \widetilde{P}_2(1-\widetilde{P}_2) + 2\widetilde{P}_1\widetilde{P}_2 \qquad \mathbf{(114-3)}$$

ولكن عندما يكون اهتمامنا مركزاً على حساب مجالات (بونفيروني) المتزامنة والمقابلة لمستوى الدلالة للفروقات بين النسبتين P_K و P_h في العينات الكبيرة، فإننا نحسب تلك المجالات من خلال العلاقة التالية: [Johnson, Wichern P.197]

$$(\tilde{P}_h - \tilde{P}_K) - Z\left(\frac{\propto}{(q-1)q}\right)\sqrt{\frac{\tilde{P}_h(1-\tilde{P}_h) + \tilde{P}_K(1-\tilde{P}_K) + 2\tilde{P}_h\tilde{P}_K}{n}} \le (P_h - P_K) \le (\tilde{P}_h - \tilde{P}_K) +$$

$$+Z\left(\frac{\propto}{(q-1)q}\right)\sqrt{\frac{\tilde{P}_h(1-\tilde{P}_h) + \tilde{P}_K(1-\tilde{P}_K) + 2\tilde{P}_h\tilde{P}_K}{n}} \qquad (115-4)$$

حيث أن $Z\left(\frac{\propto}{(q-1)q}\right)$ هي القيمة الحرجة لمتحول التوزيع الطبيعي المعياري المقابلة لـ $\left(\frac{\propto}{(q-1)q}\right)$ ، علماً بأن q هو عدد الفئات المتكاملة في المجتمع ، وإن عدد الأزواج الممكنة يساوي: q/ 2 (q-1).

مثال (4-9): في دراسة لنسب توزع المسافرين على شركات النقل العاملة بين دمشق واللاذقية، والتي عددها (4) شركات معتمدة هي: A وB وC وD وهناك وسائل أخرى E . سحبنا عينة عشوائية من المسافرين خلال يوم ما بحجم $n = 355$ فكانت النتائج كما يلي :

جدول (4- 5): نتائج بيانات العينة علماً بأن $q = 4 + 1 = 5$ (لأننا اعتبرنا الوسائل الأخرى فئة خامسة)

اسم الشركة	شركة A	شركة B	شركة C	شركة D	وسائل أخرى E	المجموع
عدد المسافرين عليها	105	119	56	25	50	355
النسبة في المجتمع P_j	P_1	P_2	P_3	P_4	$P_5 = 1-(P_1+P_2+P_3+P_4)$	
النسبة في العينة $\tilde{P}_j$	$\tilde{P}_1 = \frac{105}{355} = 0.30$	$\tilde{P}_2 = 0.33$	$\tilde{P}_3 = 0.16$	$\tilde{P}_4 = 0.07$	$\tilde{P}_5 = 0.14$	

والمطلوب إيجاد مجالات الثقة المتزامنة للنسب P_1, P_2, P_3, P_4, P_5 في المجتمع باحتمال ثقة 0.95 .

الحل: لإيجاد هذه المجالات نقوم بتطبيق العلاقة (4-112)، ولذلك نقوم أولاً بإيجاد القيمة الحرجة $\chi^2_{q-1}(\propto)$ من جداول χ^2 فنجد أن:

$$\chi^2_{q-1}(\propto) = \chi^2_4(0.05) = 9.49$$

ثم نعوض هذه البيانات على التوالي في (4-112) فنحصل على المجالات التالية :

$$For\ P_1 : 0.30 - \sqrt{9.49}\sqrt{\frac{(0.30)(0.70)}{355}} \le P_1 \le 0.30 + \sqrt{9.49}\sqrt{\frac{(0.30)(0.70)}{355}}$$

$$0.23 \le P_1 \le 0.37$$

$$For\ P_2: 0.33 - \sqrt{9.49}\sqrt{\frac{(0.33)(0.67)}{355}} \le P_2 \le 0.30 + \sqrt{9.49}\sqrt{\frac{(0.33)(0.67)}{355}}$$

$$0.25 \le P_2 \le 0.41$$

$$For\ P_3: 0.16 - \sqrt{9.49}\sqrt{\frac{(0.16)(0.84)}{355}} \le P_3 \le 0.16 + \sqrt{9.49}\sqrt{\frac{(0.16)(0.84)}{355}}$$

$$0.10 \le P_3 \le 0.22$$

$$For\ P_4: 0.07 - \sqrt{9.49}\sqrt{\frac{(0.07)(0.93)}{355}} \le P_4 \le 0.07 + \sqrt{9.49}\sqrt{\frac{(0.07)(0.93)}{355}}$$

$$0.03 \le P_4 \le 0.11$$

$$For\ P_5: 0.14 - \sqrt{9.49}\sqrt{\frac{(0.14)(0.86)}{355}} \le P_5 \le 0.14 + \sqrt{9.49}\sqrt{\frac{(0.14)(0.86)}{355}}$$

$$0.08 \le P_5 \le 0.20$$

ولإيجاد مجال الثقة للفرق بين النسبتين في الشركتين A و B نطبق العلاقة (4-113)، فنجد أن:

$$(0.30 - 0.33) - \sqrt{9.49}\sqrt{\frac{0.30(0.70) + (0.33)(0.67) + 2(0.30)(0.33)}{355}} \le P_1 - P_2 \le$$

$$\le (0.30 - 0.33) + \sqrt{9.49}\sqrt{\frac{0.30(0.70) + (0.33)(0.67) + 2(0.30)(0.33)}{355}}$$

$$-0.16 \le P_1 - P_2 < 0.10$$

وإذا أردنا المزيد من الدقة نطبق العلاقة (4-115) لذلك نبحث عن القيمة الحرجة لـ Z فنجد أن:

$$Z\left(\frac{\propto}{(q-1)q}\right) = Z\left(\frac{0.05}{4 * 5}\right) = Z(0.0025) = 2.807$$

$$(0.30 - 0.33) - (2.807)\sqrt{\frac{0.30(0.70) + 0.33(0.67) + 2(0.30)(0.33)}{355}} \le (P_1 - P_2) \le$$

$$\le (0.30 - 0.33) + (2.807)\sqrt{\frac{0.30(0.70) + (0.33)(0.67) + 2(0.30)(0.33)}{355}}$$

$-0.14 \le (P_1 - P_2) < 0.088$ وهو أصغر من المجال السابق.

وهكذا نحسب المجالات المختلفة للفروقات الأخرى.

5-4: علاقة T^2 مع نسبة الإمكانية العظمى $\wedge$

نعلم – أو يُفترض أننا نعلم– أن تابع الإمكانية العظمى للتوزيع الطبيعي المتعدد للمتحولات الطبيعية $X_1\ X_2\ ...\ X_P$، التي توقعها الشعاع $\boldsymbol{\mu}$ ومصفوفة التباينات المشتركة لها هي $\mathbf{V}$ (محددة إيجابياً)، يساوي جداء توزيعاتها الطبيعية ويعطى بالعلاقة التالية:

$$L(\boldsymbol{\mu}\,,\boldsymbol{V}) = \frac{1}{(2\pi)^{\frac{np}{2}} * |V|^{\frac{n}{2}}} EXP\left[-\frac{1}{2}\sum_{j}^{\mathrm{n}}(\boldsymbol{X_j}-\boldsymbol{\mu})'\boldsymbol{V^{-1}}(\boldsymbol{X_j}-\boldsymbol{\mu})\right] \qquad (116-4)$$

ولتقدير المعلمين $\boldsymbol{\mu}$ و$\mathbf{V}$ في (4-116) نسحب عينة بحجم n من ذلك المجتمع ونحسب تقديرهما الأعظمي من العلاقتين:

$$\widetilde{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{\boldsymbol{X}} = \begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_P\end{bmatrix} \qquad (117-4)$$

$$\widetilde{\boldsymbol{V}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X_j}-\overline{\boldsymbol{X}})'(\boldsymbol{X_j}-\overline{\boldsymbol{X}}) = [S_{ij}] \qquad (118-4)$$

ولكن بما أن هذين التقديرين يختلفان من عينة لأخرى، فإن التابع $L(\boldsymbol{\mu}\ \boldsymbol{V})$ يأخذ قيماً مختلفة في الفضاء R^P، ومعلوم من طريقة الإمكانية العظمى أن أكبر قيمة للتابع $\mathrm{L}(\boldsymbol{\mu}\ \boldsymbol{V})$ – عندما يتغير $\boldsymbol{\mu}$ و $\boldsymbol{V}$ عبر قيمها الممكنة – تعطى بواسطة العلاقة التالية :

$$\max_{\mu,V} L(\boldsymbol{\mu},\boldsymbol{V}) = \frac{1}{(2\pi)^{\frac{nP}{2}}\left|\widetilde{\boldsymbol{V}}\right|^{\frac{n}{2}}} * e^{-\frac{nP}{2}} \qquad (119-4)$$

وتسمى هذه القيمة بالقيمة العظمى غير المشروطة بأي قيد عليها .

ولكن عندما نريد أن نختبر فرضية العدم التالية $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ فإننا نكون قد وضعنا شرطاً على تغيرات $\boldsymbol{\mu}$ هو أن $\boldsymbol{\mu}$ يجب أن تساوي $\boldsymbol{\mu_0}$، وعندها فإن مصفوفة التباين المشترك $\boldsymbol{V}$ لا تقدر من العلاقة (4-118) بل من العلاقة التالية :

$$\widetilde{\boldsymbol{V}}_{\boldsymbol{0}} = \frac{1}{n}\sum(\boldsymbol{X_i}-\boldsymbol{\mu_0})'(\boldsymbol{X_i}-\boldsymbol{\mu_0}) \qquad (120-4)$$

وعندها فإن أكبر قيمة للتابع $L(\boldsymbol{\mu_0}\,,\boldsymbol{V})$ ، الذي فيه $\boldsymbol{\mu_0}$ ثابتاً و$\mathbf{V}$ متغيرة، تحسب من العلاقة التالية:

$$\max_{V} L(\boldsymbol{\mu_0},\boldsymbol{V}) = \frac{1}{(2\pi)^{\frac{nP}{2}}\left|\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}\right|^{\frac{n}{2}}} * e^{-\frac{nP}{2}} \qquad (121-4)$$

وتسمى هذه القيمة بالقيمة العظمى المشروطة بالفرضية $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$

وبناء على ما تقدم تعرف نسبة الإمكانية العظمى $\wedge$ بالعلاقة التالية :

$$\wedge = \frac{\text{القيمة العظمى المشروطة بـ}\ \boldsymbol{\mu_0}}{\text{القيمة العظمى غير المشروطة لأي من}\ \boldsymbol{V}\ \text{و}\ \boldsymbol{\mu}} \qquad (122-4)$$

$$\wedge = \frac{max\, L(\boldsymbol{\mu_0}, \boldsymbol{V})}{max\, L(\boldsymbol{\mu}, \boldsymbol{V})} = \frac{\frac{1}{(2\pi)^{\frac{nP}{2}}|\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}|^{\frac{n}{2}}} * \bar{e}^{\frac{nP}{2}}}{\frac{1}{(2\pi)^{\frac{nP}{2}}|\widetilde{\boldsymbol{V}}|^{\frac{n}{2}}} * \bar{e}^{\frac{nP}{2}}} = \frac{\frac{1}{|\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}|^{\frac{n}{2}}}}{\frac{1}{|\widetilde{\boldsymbol{V}}|^{\frac{n}{2}}}}$$

$$\wedge = \left[\frac{|\widetilde{\boldsymbol{V}}|}{|\widetilde{\boldsymbol{V}}_0|}\right]^{\frac{n}{2}} = \text{نسبة الإمكانية العظمى} \qquad (123-4)$$

ويشتق من هذه النسبة مؤشر إحصائي يسمى مؤشر (Wilk's Lamada) ويحسب من العلاقة التالية:

$$\wedge^{\frac{2}{n}} = \frac{|\widetilde{\boldsymbol{V}}|}{|\widetilde{\boldsymbol{V}}_0|} = \frac{\text{محدد } \widetilde{V} \text{ غير المشروط}}{\text{محدد } \widetilde{\boldsymbol{V}}_{\boldsymbol{0}} \text{ المشروط}} \qquad (124-4)$$

وهنا نلاحظ أنه إذا كانت الفرضية $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ صحيحة فإن قيمة المحدد $|\widetilde{\boldsymbol{V}}|$ تكون قريبة من قيمة المحدد $\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}$ [أي يكون $\widetilde{\boldsymbol{V}} \approx \widetilde{\boldsymbol{V}}_{\boldsymbol{0}}$] ، وبذلك تأخذ $\left(\wedge^{\frac{n}{2}}\right)$ أو النسبة $(\wedge)$ قيمة قريبة من الواحد، ومنه نستنتج أن : $0 \leq \wedge \leq 1$.

وبناء على ذلك يمكننا الحكم على الفرضية $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ مقابل $H_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$ كما يلي:

إذا كانت قيمة $\wedge$ أو $\left[\wedge^{\frac{n}{2}}\right]$ قريبة من الواحد نقبل الفرضية H_0 .

أما إذا كانت قيمة $\wedge$ أو $\left[\wedge^{\frac{n}{2}}\right]$ قريبة من الصفر أو صغيرة جداً فإننا نرفض الفرضية H_0 .

وهذا يعني إننا سنرفض $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu_0}$ عند مستوى الدلالة $\propto$ إذا كانت :

$$\wedge < C(\propto) \qquad (125-4)$$

حيث أن: $C(\propto)$ هي القيمة الحرجة لمتحول التوزيع الاحتمالي لـ $\wedge$ [لاحظ أن نسبة الإمكانية العظمى $\wedge$ هي عبارة عن قوة لنسبة محددي مصفوفتي التباين $\widetilde{\boldsymbol{V}}$ و $\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}$] ، ولكن بما أن التوزيع الإحتمالي لـ $\wedge$ يرتبط مع التوزيع الاحتمالي لـ T^2 ، فإننا سوف لن نكون بحاجة إلى توزيع $\wedge$ لاستخدامه في حساب $C(\propto)$ المذكورة في العلاقة (4–125) .

ولاستخراج العلاقة التي تربط $\left(\wedge^{\frac{2}{n}}\right)$ مع T^2 [انظر Johnston, Wichern P.177]. ومن هناك نجد أنهما يرتبطان بالعلاقة التالية :

$$\wedge^{\frac{2}{n}} = \left[1 + \frac{T^2}{n-1}\right]^{-1} = \frac{|\widetilde{\boldsymbol{V}}|}{|\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}|} \qquad (126-4)$$

وبعبارة أخرى نجد أن:

$$1 + \frac{T^2}{n-1} = \frac{|\widetilde{\boldsymbol{V}}_{\boldsymbol{0}}|}{|\widetilde{\boldsymbol{V}}|} = \left[\frac{1}{\wedge}\right]^{\frac{2}{n}} = \wedge^{-\frac{2}{n}} \qquad (127-4)$$

وبذلك نجد أن T^2 ترتبط عكساً مع $\wedge$ ، أي أنه إذا كانت قيمة T^2 كبيرة، فإن نسبة الإمكانية العظمى$\wedge$ تكون صغيرة ، وفي كلتا هاتين الحالتين نرفض فرضية العدم $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$، ويمكننا من (4−127) أن نحسب قيمة T^2 فنجد أن:

$$T^2 = \left[\frac{|\widetilde{\boldsymbol{V}}_{\mathbf{0}}|}{|\widetilde{\boldsymbol{V}}|} - 1\right](n-1) \qquad (128-4)$$

وهكذا نكون قد توصلنا إلى طريقة أخرى لحساب T^2 بدون الحاجة لحساب مقلوب مصفوفة التباين V^{-1} (أو S^{-1}) .

وعندما تكون الفرضية $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ صحيحة فإن التوزيع الاحتمالي لـ $\wedge$ يحسب من العلاقة (4−123) السابقة . وهو يرتبط مع توزيع F بالعلاقة التالية :

$$T^2 = \left[\frac{|\widetilde{\boldsymbol{V}}_{\mathbf{0}}|}{|\widetilde{\boldsymbol{V}}|} - 1\right](n-1) \;\sim\; \frac{P(n-1)}{n-P}F_{P*n-P} \qquad (129-4)$$

واعتماداً على هذه العلاقة يمكننا استبدال القيمة الحرجة $C(\propto)$ من العلاقة (4-125) بالقيمة الحرجة التالية $\frac{P(n-1)}{n-P}F_{P*n-P}(\propto)$ المأخوذة من جداول F مقابل مستوى الدلالة $\propto$ ودرجتي الحرية $v_1 = P$ و $v_2 = n-P$. ويستفاد من العلاقة (4−128) في حسابT^2 من محددي المصفوفتين $|\widetilde{\boldsymbol{V}}_{\mathbf{0}}|$ و $|\widetilde{\boldsymbol{V}}|$ (بدون حساب $\boldsymbol{V^{-1}}$)، ثم مقارنة قيمتها المحسوبةT^2 مع القيمة الحرجة $\frac{P(n-1)}{n-P}F_{P*n-P}$ ، ونتخذ قرار القبول أو الرفض للفرضية $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ كما يلي:

$$\begin{aligned} &\text{إذا كانت} \quad T^2 \le \frac{P(n-1)}{n-P}F_{P*n-P} \quad \text{نقبل} \quad H_0 \\ &\text{أماإذا كانت} \quad T^2 > \frac{P(n-1)}{n-P}F_{P*n-P} \quad \text{نرفض} \quad H_0 \quad \text{ونقبل} \quad H_1 \end{aligned} \qquad (130-4)$$

ملاحظة هامة:

عندما يكون حجم العينة n كبيراً وتكون الشروط الأخرى محققة، فإن توزيع المعاينة للمقدار $(-2\, ln\wedge)$ يكون خاضعاً تقاربياً للتوزيع $\chi^2_{v-v_0}$ بدرجة حرية $(v-v_0)$، حيث أن:

$$v = P + \frac{P(P+1)}{2} \quad \text{و} \quad v_0 = 0 + \frac{P(P+1)}{2}$$

وأخيراً نكتب ذلك كما يلي : عندما يكون n كبيراً فإن:

$$-2\, ln\wedge = -2\, ln\left[\frac{max\, L(\boldsymbol{\mu_0}, \boldsymbol{V})}{max\, L(\boldsymbol{\mu}, \boldsymbol{V})}\right] \;\sim\; \chi^2_{v-v_0} \qquad (131-4)$$

حيث أن v درجة أبعاد (μ, V) و v_0 درجة أبعاد $(\mu_0\,, V)$، وبعد حساب قيمة $-2\, ln\wedge$ نقارنها مع القيمة الحرجة $\chi^2_{v-v_0}(\propto)$ ونقبل H_0 إذا كانت $-2\, ln\wedge \le \chi^2_{v-v_0}$ والعكس بالعكس .

الفصل الخامس

الاستدلال حول شعاعين لمتوسطات عدة متحولات طبيعية

5-1: تمهيد:

سنبدأ هذا الفصل باستعراض بعض خواص التراكيب الخطية المنتشرة في التحليل متعدد المتحولات . لذلك نفترض أنه لدينا P متحولاً $X_1 \ X_2 \ ... \ X_P$ تعكس بعض خواص المجتمع المدروس، وترتبط مع بعضها بواسطة علاقة خطية من الشكل التالي:

$$Z = \boldsymbol{a}' * \boldsymbol{X} = a_1X_1 + a_2X_2 + \cdots + a_PX_P \qquad (1-5)$$

حيث $\boldsymbol{a}'$: هو منقول الشعاع $\boldsymbol{a}$ و$\mathbf{X}$ هو شعاع المتحولات $X_1 \ X_2 \ ... \ X_P$

ولنفترض أننا سحبنا من ذلك المجتمع عينة عشوائية بحجم n، وعندها نجد أن كل مشاهدة i تقابلها علاقة خطية من الشكل (5-1) كما يلي:

$$Z_i = a' * X_i = a_1X_{i1} + a_2X_{i2} + \cdots + a_PX_{iP} \qquad (2-5)$$

حيث أن : $i : 1\ 2\ 3 \dots n$

ولحساب القيمة العددية لمتوسط هذا التركيب الخطي في العينة نجد أنها تساوي :

$$\bar{Z} = \frac{1}{n}\sum Z_i = \frac{1}{n}\sum^{n} a' * X_i = \boldsymbol{a}' \left[\frac{X_1 + \ X_2 \ ... + X_n}{n}\right] = \boldsymbol{a}' * \bar{\boldsymbol{X}} \qquad (3-5)$$

وكذلك نجد أن القيمة العددية لتباين هذا التركيب الخطي في العينة تساوي:

$$\sigma_Z^2 = var(Z) = \boldsymbol{a}' * \boldsymbol{S} * \boldsymbol{a} \qquad (4-5)$$

حيث أن: $\mathbf{S}$ هي مصفوفة التباينات المشتركة لتلك المتحولات $\mathbf{X}$.

وكذلك نجد أنه إذا كان لدينا تركيب خطي آخر لهذه المتحولات مثل:

$$Y = \boldsymbol{b}' * \boldsymbol{X} = b_1X_1 + b_2X_2 + \cdots + b_PX_P \qquad (5-5)$$

فإن متوسطه وتباينه في العينة يساويان :

$$\bar{Y} = \boldsymbol{b}' * \bar{\boldsymbol{X}} \qquad (6-5)$$

$$\sigma_Y^2 = \boldsymbol{b}' * \boldsymbol{S} * \boldsymbol{b} \qquad (7-5)$$

أما التباين المشترك للتركيبين الخطيين Z وY فهو يساوي:

$$cov(Z,Y) = cov\ (\boldsymbol{a}'\boldsymbol{X}\ , \boldsymbol{b}'\boldsymbol{X}\ ) = \boldsymbol{b}' * \boldsymbol{S} * \boldsymbol{a} = \boldsymbol{a}' * \boldsymbol{S} * \boldsymbol{b} \qquad (8-5)$$

مثال (5−1): لنفترض أنه لدينا تركيبين خطيين معرفين على ثلاثة متحولات $X_1 \;\; X_2 \;\; X_3$ كما يلي:

$$Z = a'X = [2\,,2\,,-1]\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = 2X_1 + 2X_2 - X_3$$

$$Y = b'X = [1,-1\,,3]\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = X_1 - X_2 + 3X_3$$

ولنفترض أننا سحبنا عينة من $n = 3$ عناصر، وأخذنا قياسات تلك المتحولات المقابلة لهذه العناصر فكانت كما يلي :

$$\boldsymbol{X} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \end{array} = \begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix}$$

نقوم بحساب شعاع متوسطات المتحولات X (حسب كل متحول X_{i}) فنجد أن:

$$\overline{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \end{bmatrix} = \begin{bmatrix} \dfrac{1+4+4}{3} \\ \dfrac{2+1+0}{3} \\ \dfrac{5+6+4}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$$

وهكذا نجد أن القيمة العددية لمتوسط التركيب Z تساوي :

$$\bar{Z} = \boldsymbol{a'} * \overline{\boldsymbol{X}} = [2\,,2\,,-1]\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} = 3$$

وأن القيمة العددية لمتوسط التركيب الثاني Y يساوي :

$$\bar{Y} = \boldsymbol{b'} * \overline{\boldsymbol{X}} = [1\,,-1\,,3]\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} = 17$$

ولحساب تبايني التركيبين الخطيين Z وY يجب علينا أن نطبق العلاقتين :

$$\sigma_Z^2 = \boldsymbol{a'} * \boldsymbol{S} * \boldsymbol{a}$$

$$\sigma_Y^2 = \boldsymbol{b'} * \boldsymbol{S} * \boldsymbol{b}$$

ولذلك علينا أن نقوم بحساب مصفوفة التباينات المشتركة S فنجد من البيانات السابقة أنها تساوي :

$$\boldsymbol{S} = \begin{bmatrix} 3 & -\dfrac{3}{2} & 0 \\ -\dfrac{3}{2} & 1 & \dfrac{1}{2} \\ 0 & \dfrac{1}{2} & 1 \end{bmatrix} \qquad \text{(قم بحسابها)}$$

وبذلك نجد أن تباين التركيب الخطي Z يساوي :

$$\sigma_Z^2 = \boldsymbol{a'} * \boldsymbol{S} * \boldsymbol{a} = [2\,,2\,,-1]\begin{bmatrix} 3 & -\frac{3}{2} & 0 \\ -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{bmatrix} * \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix} = [2\,,2\,,-1]\begin{bmatrix} 3 \\ -\frac{3}{2} \\ 0 \end{bmatrix} = 3$$

وكذلك نجد أن تباين التركيب الخطي Y يساوي:

$$\sigma_Y^2 = \boldsymbol{b'} * \boldsymbol{S} * \boldsymbol{b} = [1\,,-1\,,3]\begin{bmatrix} 3 & -\frac{3}{2} & 0 \\ -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} = [1\,,-1\,,3]\begin{bmatrix} \frac{9}{2} \\ -1 \\ \frac{5}{2} \end{bmatrix} = 13$$

ولحساب التباين المشترك للتركيبين Z و Y نطبق العلاقة :

$$cov\,(Z\,,Y) = cov\,(\boldsymbol{a'X}\,,\boldsymbol{b'X}\,) = \boldsymbol{a'} * \boldsymbol{S} * \boldsymbol{b} = \boldsymbol{b'} * \boldsymbol{S} * \boldsymbol{a}$$

$$cov\,(\boldsymbol{a'X}\,,\boldsymbol{b'X}\,) = [\mathbf{2}\,,\mathbf{2}\,,-\mathbf{1}]\begin{bmatrix} \mathbf{3} & -\frac{\mathbf{3}}{\mathbf{2}} & \mathbf{0} \\ -\frac{\mathbf{3}}{\mathbf{2}} & \mathbf{1} & \frac{\mathbf{1}}{\mathbf{2}} \\ \mathbf{0} & \frac{\mathbf{1}}{\mathbf{2}} & \mathbf{1} \end{bmatrix}\begin{bmatrix} \mathbf{1} \\ -\mathbf{1} \\ \mathbf{3} \end{bmatrix} = [\mathbf{2}\,,\mathbf{2}\,,-\mathbf{1}]\begin{bmatrix} \mathbf{1} \\ -\mathbf{1} \\ \mathbf{1} \end{bmatrix} = \frac{\mathbf{9}}{\mathbf{2}}$$

ونستنتج مما سبق أنه يمكننا حساب مؤشرات التراكيب الخطية المعرفة على المتحولات $\boldsymbol{X_1}\ \ \boldsymbol{X_2}\ \ldots\ \boldsymbol{X_P}$ من بيانات تلك المتحولات مباشرة وذلك باستخدام $\boldsymbol{\bar{X}}$ و $\boldsymbol{S}$.

كما يمكننا تعميم هذه القاعدة على عدة تراكيب خطية معرفة على تلك المتحولات، وليكن عددها q تركيباً خطياً، وتأخذ الشكل التالي:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \cdots + a_{1P}X_P \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \cdots + a_{2P}X_P \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mathrm{Z}_q &= a_{q1}X_1 + a_{q2}X_2 + a_{q3}X_3 + \cdots + a_{qP}X_P \end{aligned} \qquad (9-5)$$

وعندها نجد أنه يمكننا أن نكتب هذه التراكيب الخطية بشكل مصفوفي كما يلي:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_q \end{bmatrix} = \begin{bmatrix} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1P}X_P \\ a_{21}X_1 + a_{22}X_2 + \cdots + a_{2P}X_P \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{q1}X_1 + a_{q2}X_2 + \cdots + a_{qP}X_P \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1P} \\ a_{21} & a_{22} & \ldots & a_{2P} \\ \ldots & \ldots & \ldots & \ldots \\ a_{q1} & a_{q2} & \ldots & a_{qP} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix} = \boldsymbol{A} * \boldsymbol{X} = \boldsymbol{Z} \quad (10-5)$$

ومن (5−10) يمكننا الحصول على التركيب الخطي Z_K بضرب السطر K من المصفوفة A بعمود المتحولات X ، ثم حساب متوسطه من $\boldsymbol{a'_K}\ \boldsymbol{\bar{X}}$ وتباينه من $\boldsymbol{a'_K}\ \boldsymbol{S}\ \boldsymbol{a_K}$ ، وبصورة عامة يمكننا حساب شعاع المتوسطات $\bar{Z}$ من العلاقة :

$$\boldsymbol{\bar{Z}} = \boldsymbol{A} * \boldsymbol{\bar{X}} \qquad (11-5)$$

ثم حساب مصفوفة التباين المشترك من العلاقة :

$$cov\,(Z_K Z_h) = \boldsymbol{A'} * \boldsymbol{S} * \boldsymbol{A} \qquad (12-5)$$

5-2: المقارنات الزوجية المترابطة (لتجربتين على عينة واحدة: قبلية (1) وبعدية (2))

إن قياسات التجارب غالباً ما تؤخذ تحت جملة من الشروط التجريبية . وذلك لبيان فيما إذا كانت قيم الاستجابة تختلف معنوياً عن شروط تلك الجملة، ولإجراء المقارنات بين هذه القياسات سنميز بين عدة حالات هي:

5-2-1: حالة متحول واحد X:

لنفترض أننا نريد دراسة تأثير أحد الأدوية على مستويات السكر في الدم X لدى مرضى السكري (2)، لذلك نسحب عينة عشوائية من مرضى السكري (2) بحجم n مريضاً، ونرقمهم من 1 حتى n، ثم نقوم بقياس مستويات السكر على الريق، لأفراد هذه العينة على الترتيب قبل إعطائهم الدواء، ونطلق على هذه القياسات مصطلح القياسات القبلية (أو قياسات المعالجة 1) ونرمز لها بالرموز X_{1i} ونكتبها كما يلي:

$$X_{1i}\ :\ x_{11}\ , x_{12}\ \ , x_{13} \dots \dots , x_{1n} \qquad (13-5)$$

ثم نقوم بإعطاء ذلك الدواء لجميع ونفس أفراد تلك العينة، وبعد مرور ساعة صيام ، نقوم بأخذ قياسات السكر لهؤلاء الأفراد على الترتيب ، فنحصل على قياسات جديدة تسمى القياسات البعدية (أو قياسات المعالجة 2) . ونرمز لها بـ X_{2i} ونكتبها كمايلي :

$$X_{2i}\ :\ x_{21},\ \ x_{22}\ ,\ \ x_{23}, \dots \dots , x_{2n} \qquad (14-5)$$

ثم نضع نتائج القياسات القبلية لكل فرد مقابل القياسات البعدية له في جدول مناسب فنحصل على جدول كالتالي :

جدول (5-1): القياسات القبلية والبعدية لمستويات السكر X لأفراد العينة :

رقم الفرد i	1	2	3	….	i	….	n	
القياسات القبلية X_{1i}	x_{11}	x_{12}	x_{13}	….	x_{1i}	….	x_{1n}	
القياسات البعدية X_{2i}	x_{21}	x_{22}	x_{23}	….	x_{2i}	….	x_{2n}	
الفروقات d_i	d_1	d_2	d_3	….	d_i	….	d_n	$\bar{d}$
$(d_i-\bar{d})^2$	$(d_1-\bar{d})^2$	$(d_2-\bar{d})^2$	$(d_3-\bar{d})^2$	….	$(d_i-\bar{d})^2$	….	$(d_n-\bar{d})^2$	$S_{\bar{d}}^2$

ثم نقوم بحساب الفروقات بين كل زوج من القياسات المتقابلة من العلاقة :

$$d_i = x_{1i} - x_{2i} \qquad i: 1\ 2\ 3\ \dots n \qquad (15-5)$$

ونضعها في سطر مستقل ثم نحسب متوسطها $\bar{d}$ من العلاقة :

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad (16-5)$$

ثم نحسب تباينها المصحح S^2 (من السطر الأخير) باستخدام العلاقة :

$$S_{\bar{d}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathrm{d}_i - \bar{d}\right)^2 \qquad (17-5)$$

ولدراسة تأثير ذلك الدواء على مستويات السكر عند المرضى، نفترض أن الفروقات في (5−15) هي عبارة عن مشاهدات مستقلة وتخضع للتوزيع الطبيعي في المجتمع المدروس $N(\bar{D}\,,\sigma_d^2)$ الذي توقعها $E(d_i)=\bar{D}$ وتباينها $var(d_i)=\sigma_d^2$، ثم نضع الفرضيتين كما يلي:

- **فرضية العدم (توقع الفروقات في المجتمع يساوي الصفر)** $H_0 : \bar{D}=0$
- **فرضية البديلة (توقع الفروقات في المجتمع يختلف عن الصفر)** $H_1 : \bar{D}\neq 0$

ولاختبار الفرضية H_0 نستخدم مؤشر اختبار (ستودينت) التالي:

$$t=\frac{\bar{d}-\bar{D}_0}{S_d/\sqrt{n}}=\frac{\bar{d}-0}{S_d/\sqrt{n}} \qquad \mathbf{(18-5)}$$

وذلك تحت شرط تحقق فرضية العدم H_0، وهو يخضع لتوزيع (ستودينت) بـ $(n-1)$ درجة حرية . وإذا حددنا مستوى الدلالة بـ ∝ ، فإننا نقبل فرضية العدم، $H_0 : \bar{D}=0$ المقابلة لـ $H_1 : \bar{D}\neq 0$ إذا تحققت المتراجحة التالية:

$$|t|\leq t_{n-1}\left(\frac{\propto}{2}\right) \qquad \mathbf{(19-5)}$$

حيث $t_{n-1}\left(\frac{\propto}{2}\right)$ هي القيمة الحرجة لمتحول (ستودينت) عند $(n-1)$ درجة حرية، والتي تترك على يمينها احتمالاً يساوي $\left(\frac{\propto}{2}\right)$.

أما إذا كانت $|t|>t_{n-1}\left(\frac{\propto}{2}\right)$ فإننا نرفض H_0 ونقبل H_1 بمستوى دلالة ∝ .

ثم نقوم بإنشاء مجال الثقة لتوقع الفروقات $\bar{D}$ من العلاقة التالية:

$$\bar{d}-t_{n-1}\left(\frac{\propto}{2}\right)*\frac{S_d}{\sqrt{n}}\leq \bar{D}\leq \bar{d}+t_{n-1}\left(\frac{\propto}{2}\right)*\frac{S_d}{\sqrt{n}} \qquad \mathbf{(20-5)}$$

مثال (5−2): لدراسة تأثير (عصير الليمون) على مستويات السكر عند مرضى السكري (2)، سحبنا عينة عشوائية منهم بحجم $(n=10)$، ثم أخذنا قياسات السكر لهم على الريق قبل إعطائهم العصير، فحصلنا على القياسات القبلية X_{1i}، ثم أعطيناهم عصير الليمون، ومنعنا عنهم أي طعام أو شراب آخر، وبعد مرور ساعة على تناول ذلك العصير، أخذنا قياسات مستويات السكر البعدية X_{2i} ، ثم وضعنا النتائج المتقابلة لكل فرد i في جدول مناسب كالتالي:

جدول (5−2): القياسات القبلية والبعدية لمستويات السكر

المتوسط	10	9	8	7	6	5	4	3	2	1	رقم الفرد i
	160	150	130	140	160	130	150	170	140	130	القياسات القبلية X_{1i}
	140	120	110	120	140	110	120	150	130	120	القياسات البعدية X_{2i}
$\bar{d}=20$	20	30	20	20	20	20	30	20	10	10	الفروقات $d_i=X_{1i}-X_{2i}$
$S_d^2=\frac{400}{9}$	0	100	0	0	0	0	100	0	100	100	$(d_i-\bar{d})^2$

والمطلوب: اختبار تأثير ذلك العصير على مستوى السكر في الدم بمستوى دلالة 0.05 .

الحل: ومن الجدول (5−2) نجد أن متوسط الفروقات $\bar{d}$ يساوي:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{200}{10} = 20$$

وإن تباين هذه الفروقات S_d^2 يساوي :

$$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{10}(d_i - \bar{d})^2 = \frac{400}{9}$$

وإن انحرافها المعياري يساوي

$$S_{\mathrm{d}} = \sqrt{\frac{400}{9}} = \frac{20}{3}$$

ولإجراء الاختبار المطلوب بمستوى الدلالة $\propto = 0.05$ ، نضع الفرضيتين في المجتمع كما يلي:

$$H_0 : \bar{D} = 0 \qquad\qquad H_1 : \bar{D} \neq 0$$

ثم نحسب قيمة مؤشر الاختبار من العلاقة :

$$t = \frac{\bar{d} - \bar{D}_0}{S_d/\sqrt{n}} = \frac{20 - 0}{\frac{20}{3}/\sqrt{10}} = 9.487$$

ثم نقارن القيمة المحسوبة t مع القيمة الحرجة $t_{n-1}\left(\frac{\propto}{2}\right)$ والتي نأخذها من جدول (ستودينت) كما يلي :

$$t_{\mathrm{n}-1}\left(\frac{\propto}{2}\right) = t_9\left(\frac{0.05}{2}\right) = t_9(0.025) = 2.262$$

وعند المقارنة نجد أن $(t = 9.487) > t_{\left(\frac{\propto}{2}\right)} = 2.262$ ، لذلك نرفض فرضية العدم H_0 ونقبل H_1 ، التي تقول أنه يوجد فروقات معنوية بين القياسات البعدية والقياسات القبلية . وإن هذه الفروقات هي فروقات معنوية باحتمال دلالة أصغر من 0.05 . وهذا يعني أن عصير الليمون المستخدم في المعالجة (2)، يؤثر على مستويات السكر في الدم، ويؤدي إلى انخفاض تلك المستويات بشكل معنوي .

ولإنشاء مجال الثقة لتوقع الفروقات $\bar{D}$ في المجتمع نطبق العلاقة (5−20) فنجد أن:

$$20 - (2.262)\frac{\frac{20}{3}}{\sqrt{10}} \leq \bar{D} \leq 20 - (2.262)\frac{\frac{20}{3}}{\sqrt{10}}$$

$$15.23 \leq \bar{D} \leq 24.79$$

أي أن ذلك العصير يؤدي إلى تخفيض مستوى السكر في الدم من 15 إلى 25 درجة تقريباً . وذلك باحتمال ثقة 95% .

5−2−2: حالة متحولين X_1 و X_2 :

نفترض في هذه الحالة أننا نريد دراسة تأثير أحد الأدوية أو العوامل على متحولين X_1 و X_2 معرفين على مجتمع ما، كتأثير أحد الأدوية على مستوى السكر X_1 وعلى مستوى الكوليسترول X_2 عند مرض السكري (2)، ولذلك نسحب عينة عشوائية منهم بحجم n مريضاً، ثم نأخذ قياسات المتحولين X_1 و X_2

قبل إعطائهم الدواء، فنحصل على القياسات القبلية x_{11i} و x_{12i} ....لكل منهما، ثم نعطيهم الدواء ونمنع عنهم أي طعام ، وبعد ساعة نأخذ قياسات X_1 و X_2 لكل فرد منهم فنحصل على القياسات البعدية X_{21i} و X_{22i} ...لكل منهما .

ثم نضع هذه القياسات المتقابلة ضمن جدول واحد(القبلية في الأعلى والبعدية في الأسفل) كما يلي :

$$\begin{array}{r c c c} \text{القبلية لـ } X_1 \text{ و } X_2 & & & \begin{array}{ccccc} 1 & 2 & 3 & & n \end{array} \\ & \begin{bmatrix} X_{11} \\ X_{12} \\ \hline X_{21} \\ X_{22} \end{bmatrix} & = & \begin{bmatrix} x_{111} & x_{112} & x_{113} & \cdots & x_{11n} \\ x_{121} & x_{122} & x_{123} & \cdots & x_{12n} \\ \hline x_{211} & x_{212} & x_{213} & \cdots & x_{21n} \\ x_{221} & x_{222} & x_{223} & \cdots & x_{22n} \end{bmatrix} \begin{matrix} d_{X1i} \\ \\ d_{X2i} \end{matrix} \\ \text{البعدية لـ } X_1 \text{ و } X_2 & & & \end{array} \qquad (21-5)$$

ثم نقوم بحساب الفروقات بين القياسات القبلية والبعدية المتقابلة لكل من هذين المتحولين من العلاقات التالية (حسب الاسهم):

$$\begin{aligned} d_{X1i} &= X_{11i} - X_{21i} \qquad & i : 1\ 2\ 3\ \dots\ n \\ d_{X2i} &= X_{12i} - X_{22i} \qquad & i : 1\ 2\ 3\ \dots\ n \end{aligned} \qquad \mathbf{(22-5)}$$

ثم نقوم بحساب متوسطات هذه الفروقات لكل من X_1 و X_2 فنحصل على أن:

$$\begin{aligned} \bar{d}_{X_1} &= \frac{1}{n}\sum_{i=1}^{n} d_{X_1 i} \\ \bar{d}_{X_2} &= \frac{1}{n}\sum_{i=1}^{n} d_{X_2 i} \end{aligned} \quad \Rightarrow \quad \bar{\boldsymbol{d}} = \begin{bmatrix} \bar{d}_{X_1} \\ \bar{d}_{X_2} \end{bmatrix} \qquad \mathbf{(23-5)}$$

حيث رمزنا بـ $\bar{d}$ لشعاع متوسطات تلك الفروقات $\bar{d}_{X_1}$ و $\bar{d}_{X_2}$.

والآن لنفترض أن الفروقات $\bar{d}_{X_1}$ و $\bar{d}_{X_2}$ مستقلة، وتخضع للتوزيع الطبيعي المتعدد $N(\bar{D}\,,V)$ الذي توقعه $\bar{D}$ يساوي :

$$\bar{\boldsymbol{D}} = E\begin{bmatrix} \bar{d}_{x_1} \\ \bar{d}_{x_2} \end{bmatrix} = \begin{bmatrix} \bar{D}_{X_1} \\ \bar{D}_{X_2} \end{bmatrix} \qquad (24-5)$$

والذي مصفوفة تبايناته المشتركة تساوي V_d ، والتي تقدر من بيانات العينة من خلال مصفوفة التباينات المشتركة لها S_d ، وهي تحسب من العلاقة :

$$\boldsymbol{S_d} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{d_{X_1 i}} - \bar{\boldsymbol{d}}_{\boldsymbol{X_1}})(\boldsymbol{d_{X_2 i}} - \bar{\boldsymbol{d}}_{\boldsymbol{X_2}})' \qquad (25-5)$$

$$\boldsymbol{S_d} = \frac{1}{n-1}\sum_{i=1}^{n}\begin{bmatrix} d_{X_1 i} - \bar{d}_{X_1} \\ d_{X_2 i} - \bar{d}_{X_2} \end{bmatrix}\left[(d_{X_1 i} - \bar{d}_{X_1}),(d_{X_2 i} - \bar{d}_{X_2})\right]$$

$$\boldsymbol{S_d} = \begin{bmatrix} S_{X_1}^2 & S_{X_1 X_2} \\ S_{X_2 X_1} & S_{X_2}^2 \end{bmatrix} \qquad (26-5)$$

وعندها نجد أن مسألة الاستدلال حول شعاع توقع الفروقات $\overline{D}$ تعتمد على المسافة الإحصائية T^2، التي تحسب من العلاقة التالية:

$$T^2 = n(\overline{\boldsymbol{d}} - \overline{\boldsymbol{D}})' * \boldsymbol{S}_{\boldsymbol{d}}^{-\boldsymbol{1}} * (\overline{\boldsymbol{d}} - \overline{\boldsymbol{D}}) \qquad (27-5)$$

واعتماداً على خواص T^2 نجد أنه إذا كانت الفروقات $d_{X_1 i}$ و $d_{X_2 i}$ تشكل عينة عشوائية مسحوبة من مجتمع يخضع للتوزيع الطبيعي $N(\overline{D}\,, V_d)$. فإن مسافة الإحصائية T^2 تكون خاضعة تقاربياً لتوزيع Hotelling، المرتبط بالتوزيع F وفق العلاقة :

$$T^2 \sim \frac{P(n-1)}{P-n}\, F \qquad (28-5)$$

حيث P هو عدد المتحولات المستخدمة في التحليل المتعدد .

وبناءً على ذلك يمكننا أن نجري الاختبار حول معنوية الفروقات بين الأزواج d_{X_1} و d_{X_2} وفق التالي:

نفترض أن منقولا شعاعي الفروقات المشاهدة هما:

$$\begin{aligned} \boldsymbol{d}'_{\boldsymbol{X_1}} &= (d_{X_{11}}, d_{X_{12}}, \dots, d_{X_{1n}}) \\ \boldsymbol{d}'_{\boldsymbol{X_2}} &= (d_{X_{21}}, d_{X_{22}}, \dots, d_{X_{2n}}) \end{aligned} \qquad (29-5)$$

ثم نحدد مستوى الدلالة ∝ ، ونضع الفرضيتين كما يلي :

$$H_0 : \overline{\boldsymbol{D}} = 0 \qquad H_1 : \overline{\boldsymbol{D}} \neq 0 \qquad (30-5)$$

ثم نقوم بحساب قيمة T^2 من العلاقة (5−27) ونقارنها من القيمة الحرجة $T^2(\propto)$ المحسوبة من العلاقة (5−28) . ونتخذ القرار ضمن شروط الفرضية H_0 كما يلي: نقبل الفرضية H_0 إذا كانت :

$$T^2 = n(\overline{\boldsymbol{d}} - \boldsymbol{0})' * \boldsymbol{S}_{\boldsymbol{d}}^{-\boldsymbol{1}} * (\overline{\boldsymbol{d}} - \boldsymbol{0}) \leq \frac{P(n-1)}{P-n} F_{P\ n-P}(\propto) \qquad \mathbf{(31-5)}$$

ونرفض الفرضية H_0 إذا كان العكس .

حيث أن: $F_{P\ n-P}(\propto)$ هي القيمة الحرجة لمتحول التوزيع F ذي درجتي الحرية (P) و $(n-P)$. والتي تترك على يمينها احتمالاً قدره ∝،

وبناءً على ما ورد في الفقرة (4− 4)، فإن مجالات الثقة المتزامنة لتوقيع الفروقات الفردية تعطى بالعلاقة:

$$\bar{d}_{X_1} - \sqrt{\frac{P(n-1)}{P-n} F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{X_1}^2}{n}} \leq \overline{D}_{X_1} \leq \bar{d}_{X_1} + \sqrt{\frac{P(n-1)}{P-n} F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{X_1}^2}{n}} \qquad \mathbf{(32-5)}$$

$$\bar{d}_{X_2} - \sqrt{\frac{P(n-1)}{P-n} F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{X_2}^2}{n}} \leq \overline{D}_{X_2} \leq \bar{d}_{X_2} + \sqrt{\frac{P(n-1)}{P-n} F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{X_2}^2}{n}} \qquad \mathbf{(33-5)}$$

ملاحظة: عندما يكون العدد $(n-P)$ كبيراً، فإن توزيع المقدار T^2 يصبح مقاربا لتوزيع χ^2 بـ p درجة حرية ، أي أن:

$$T^2 = \frac{P(n-1)}{P-n} F_{P\ n-P} \approx \chi_p^2 \qquad \mathbf{(34-5)}$$

وعندها لا نحتاج إلى افتراض أن يكون المجتمع طبيعياً. ونستبدل ذلك في العلاقتين (5−32) و(5−33) لحساب مجالات الثقة المتزامنة.

ولإنشاء مجالات الثقة المتزامنة حسب طريقة (بونفيروني) لتوقعات الفروقات الفردية وباحتمال ثقة $(1-\propto)$ نطبق العلاقتين التاليتين :

$$\bar{d}_{X_1} - t_{n-p}\left(\frac{\propto}{P\ (p-1)}\right) * \sqrt{\frac{S_{X_1}^2}{n}} \le \bar{D}_{X_1} \le \bar{d}_{X_1} + t_{n-p}\left(\frac{\propto}{P\ (p-1)}\right) * \sqrt{\frac{S_{X_1}^2}{n}} \qquad \mathbf{(35-5)}$$

$$\bar{d}_{X_2} - t_{n-p}\left(\frac{\propto}{P\ (p-1)}\right) * \sqrt{\frac{S_{X_1}^2}{n}} \le \bar{D}_{X_2} \le \bar{d}_{X_2} + t_{n-p}\left(\frac{\propto}{P\ (p-1)}\right) * \sqrt{\frac{S_{X_2}^2}{n}} \qquad \mathbf{(36-5)}$$

حيث أن $t_{n-p}\left(\frac{\propto}{p\ (p-1)}\right)$: هي القيمة الحرجة لمتحول (ستودينت) المقابلة لمستوى دلالة $\left(\frac{\propto}{p\ (p-1)}\right)$ و $(n-p)$ درجة حرية .

مثال (5−3): [من Johnson ,Wichern P. 213 بتصرف]

لدراسة تأثير مياه الصرف الصحي على مياه والأنهار والجداول، قامت مصلحة حماية البيئة بسحب (11) جرعة من مياه الأنهار ومن مناطق مختلفة، وبعد تسجيلها وترتيبها وترقيمها، قامت المصلحة بتقسيم كل جرعة إلى قسمين فحصلت على عينتين متكافئتين بحجمين $n_1 = n_2 = 11$ جرعة . ولتحليل هاتين العينتين، أرسلت إحدى العينتين إلى مخبر خاص، وأرسلت العينة الأخرى إلى مخبر حكومي، وطلبت منهما تحديد كمية الأوكسجين الحيوي فيها(BOD) وكمية الأملاح العالقة بها (SS)، فكانت نتائج التحليل في هذين المخبرين كما يلي:

جدول (5−3): نتائج التحليل في المخبرين لـ (BOD) و(SS):

رقم الجرعة		رمز المتحول	1	2	3	4	5	6	7	8	9	10	11
نتائج المختبر الخاص	BOD	X_{11}	6	6	18	8	11	34	28	71	43	33	20
	SS	X_{21}	27	23	64	44	30	75	26	124	54	30	14
نتائج المختبر الحكومي	BOD	X_{12}	25	28	36	35	15	44	42	54	34	29	39
	SS	X_{22}	15	13	22	29	31	64	30	64	56	20	21

والسؤال الآن: هل نتائج التحليل في هذين المخبرين متوافقة أو مختلفة ؟

وإذا ظهرت فروقات، فما هي طبيعة تلك الفروقات ؟

الحل: للإجابة على ذلك علينا أن نقوم باختبار الفروقات بين نتائجي هذين التحليلين (الخاص والحكومي)، ودراسة فيما إذا كانت تلك الفروقات معنوية أم عشوائية ؟

وهنا يجب علينا أن نطبق اختبار T^2 على شعاعي الفروقات للمتحولين: BOD والذي سنرمز له بـX_1 وSS والذي سنرمز له بـ X_2 . لذلك نحدد مستوى الدلالة بـ $\propto = 0.05$ ونضع الفرضيتين حول توقع الفروقات في المجتمع كما يلي:

$$H_0 \,:\, \bar{\boldsymbol{D}} = 0 \Longrightarrow H_0 \,:\, \bar{D} = \begin{bmatrix} \bar{D}_{X_1} \\ \bar{D}_{X_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H_1 \,:\, \bar{\boldsymbol{D}} \neq 0 \Longrightarrow H_1 \,:\, \bar{D} = \begin{bmatrix} \bar{D}_{X_1} \\ \bar{D}_{X_2} \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

ثم نقوم بحساب الفروقات من بيانات العينة فنجد أنها تساوي:

جدول (5–4): الفروقات d_{x_i} و d_{yi} محسوبة من الجدول (5–3) :

رقم الجرحة	1	2	3	4	5	6	7	8	9	10	11	المتوسط
$d_{X_{1i}} = X_{11i} - X_{12i}$	−19	−22	−18	−27	−4	−10	−14	17	9	4	−19	−9.36
$d_{X_{2i}} = X_{21i} - X_{22i}$	12	10	42	15	−1	11	−4	60	−2	10	7	13.27

ومنها نحسب شعاع متوسط الفروقات في هذه العينة من العلاقة (5–23) فنجد من الجدول (5–4) أنه يساوي :

$$\bar{\boldsymbol{d}} = \begin{bmatrix} \bar{d}_{X_1} \\ \bar{d}_{X_2} \end{bmatrix} = \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix}$$

ثم نقوم بحساب مصفوفة التباينات المشتركة لهما من العلاقة (5–26) فنجد أنها تساوي :

$$\boldsymbol{S_d} = \begin{bmatrix} S_{X_1}^2 & S_{X_1X_2} \\ S_{X_2X_1} & S_{X_2}^2 \end{bmatrix} = \begin{bmatrix} 199.36 & 88.38 \\ 88.38 & 418.61 \end{bmatrix}$$

وإن مقلوبها S_d^{-1} يساوي:

$$\boldsymbol{S_d^{-1}} = \begin{bmatrix} 0.0055 & -0.0012 \\ -0.0012 & 0.0026 \end{bmatrix}$$

ثم نقوم بحساب قيمة مؤشر الاختبار T^2 من العلاقة (5–27) فنجد أنه حسب الفرضية H_0 يساوي :

$$T^2 = n\left[(\bar{d}_{X_1} - 0), (\bar{d}_{X_2} - 0)\right] * S_d^{-1} * \begin{bmatrix} (\bar{d}_{X_1} - 0) \\ (\bar{d}_{X_2} - 0) \end{bmatrix}$$

أي أن:

$$T^2 = 11[-9.36\,, 13.27] \begin{bmatrix} 0.0055 & -0.0012 \\ -0.0012 & 0.0026 \end{bmatrix} * \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix}$$

وبعد إجراء الحسابات نجد أن: $T^2 = 13.6$

ولمقارنة هذه القيمة المحسوبة T^2 مع القيمة الحرجة $T^2(\propto)$ نستعين بجداول التوزيع F فنجد من العلاقة (5–28) أن:

$$T^2(\propto) = \frac{P(n-1)}{n-P} F_{P\;n-P}(\propto) = \frac{2(11-1)}{11-2}\, F_{2\;9}(0.05) = \frac{20}{9}(4.26) = 9.47$$

وبمقارنة T^2 مع $T^2(\propto)$ نجد أن:

$$(T^2 = 13.6) > (T^2(\propto) = 9.47)$$

لذلك نرفض فرضية العدم H_0 حول عدم وجود فروقات بين نتائج التحليل من المخبرين الخاص والحكومي: ونستنتج أنه يوجد فروقات معنوية (غير معدومة) بين نتائج التحليل في المخبرين المذكورين .

وفي الحقيقة أن تلك الفروقات تظهر فوراً من الجدول (5−3) حيث نلاحظ أن نتائج المتحول X_1 (أي BOD) في المخبر الخاص بشكل عام أقل مما عليه في المخبر الحكومي، وإن نتائج المتحول X_2 (أي SS) في المخبر الخاص بشكل عام أكبر من المخبر الحكومي .

ولإنشاء مجالات الثقة المتزامنة لتوقعات الفروقات $\overline{D}_{X_1}$ و $\overline{D}_{X_2}$ بينهما نطبق العلاقتين (5−35) و(5−36) فنجد أن :

$$\bar{d}_{X_1} - \sqrt{\frac{P(n-1)}{P-n} F_{P\ \ n-P}(\propto)} * \sqrt{\frac{S^2_{X_1}}{n}} \le \overline{D}_{X_1} \le \bar{d}_{X_1} - \sqrt{\frac{P(n-1)}{P-n} F_{P\ \ n-P}(\propto)} * \sqrt{\frac{S^2_{X_1}}{n}}$$

$$-9.36 - \sqrt{9.47} * \sqrt{\frac{199.26}{11}} \le \overline{D}_{X_1} \le -9.36 - \sqrt{9.47} * \sqrt{\frac{199.26}{11}}$$

وبذلك نجد أن:
$$-22.46 \le \overline{D}_{X_1} \le 3.74$$

وكذلك نجد أن مجال الثقة لـ $\overline{D}_{X_2}$ يساوي:

$$\bar{d}_{X_2} - \sqrt{\frac{P(n-1)}{n-P} F_{P\ \ n-P}(\propto)} * \sqrt{\frac{S^2_{X_2}}{n}} \le \overline{D}_{X_2} \le \bar{d}_{X_2} - \sqrt{\frac{P(n-1)}{n-P} F_{P\ \ \mathrm{n}-P}(\propto)} * \sqrt{\frac{S^2_{X_2}}{n}}$$

$$13.27 - \sqrt{9.47} * \sqrt{\frac{418.61}{11}} \le \overline{D}_{X_2} \le 13.27\sqrt{9.47} * \sqrt{\frac{418.61}{11}}$$

وبذلك نجد أن:
$$-5.71 \le \overline{D}_{X_2} \le 32.25$$

وهنا نلاحظ أن هذين المجالين المتزامنين يحتويان التوقعين $\overline{D}_{X_1}$ و $\overline{D}_{X_2}$ باحتمال ثقة 95% . وإن كل منهما يتضمن قيمة الصفر، التي رفضناها عندما رفضنا الفرضية $H_0\ :\ \overline{D} = 0$ بمستوى دلالة 5% . فكيف سنفسر ذلك ؟

إن مثل هذه الحالات قد تحدث عندما نرفض الفرضية H_0 . وذلك لأن تصميم المجالات المتزامنة يشمل جميع التراكيب الخطية للتوقعين $\overline{D}_X$ و $\overline{D}_y$ [مثل : $\ell_1\overline{D}_{X_1}$ و $\ell_2\overline{D}_{X_2}$] ، ولكن مجالات الحالة العملية، التي صادفتنا في هذا المثال، تقابل الحالتين التاليتين فقط :

$$(\ell_1 = 1\,, \ell_2 = 0\,)\ :\ \ For\ \overline{D}_{X_1}$$

$$(\ell_1 = 0\,, \ell_2 = 1\,)\ :\ \ For\ \overline{D}_{X_2}$$

أما بقية الحالات الممكنة لـ ℓ_1 و ℓ_2 فإنها ستعطينا مجالات متزامنة لا تتضمن قيمة الصفر . وكذلك الأمر في مجالات (بونفيروني). وعدا عن ذلك فإننا نفترض عند التحليل أن الفروقات d_{X_1} و d_{X_2} تخضع للتوزيع الطبيعي، ولكن في الحالات العملية قد لا يكون هذا الافتراض محققاً بالنسبة لواحد أو أكثر من الفروقات . كما أن تلك الفروقات قد تعود لأسباب تنظيمية أخرى .

وأخيراً نشير إلى أنه إذا كانت نتيجة الاختبار قبول فرضية العدم H_0 . فهذا يقتضي بالضرورة أن تتضمن جميع المجالات المتزامنة نقطة الصفر .

5-2-3: حالة عدة متحولات $X_1\ X_2\ X_3 \dots X_P$:

في هذه الحالة نفترض أنه لدينا P متحولاً معرفاً على مجتمع ما، ونريد دراسة تأثر احد العوامل على هذة المتحولات $X_1\ X_2\ X_3 \dots X_P$، مثل تأثير أحد الأدوية على مستوى السكر X_1 وعلى مستوى الكوليسترول X_2، وعلى مستوى الضغط الانبساطي X_3 ، وعلى مستوى الضغط الانقباضي X_4 وعلى عدد نبضات القلب في الدقيقة X_5 ...الخ .

لذلك فإننا نقوم بسحب عينة عشوائية من أفراد ذلك المجتمع بحجم n، ثم نأخذ قياسات هذه المتحولات قبل إعطاء الدواء فنحصل على القياسات القبلية لها . نرمز لها بـ x_{1ji} ونكتبها مع متوسطاتها كما يلي:

$$\begin{bmatrix} X_{11i} \\ X_{12i} \\ X_{13i} \\ \vdots \\ X_{1Pi} \end{bmatrix} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & \dots & n \end{array} \\ \begin{bmatrix} x_{111} & x_{112} & x_{113} & \cdots & x_{11n} \\ x_{121} & x_{122} & x_{123} & \cdots & x_{12n} \\ x_{131} & x_{132} & x_{133} & \cdots & x_{13n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1P1} & x_{1P2} & x_{1P3} & \cdots & x_{1Pn} \end{bmatrix} \end{array} \Rightarrow \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \\ \vdots \\ \bar{X}_{1P} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{11i} \\ \frac{1}{n}\sum_{i=1}^{n} x_{12i} \\ \frac{1}{n}\sum_{i=1}^{n} x_{13i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{1Pi} \end{bmatrix}$$

ثم نقوم بإعطاء الدواء لجميع أفراد العينة (أو نطبق العامل المدروس عليها). ثم ننتظر ساعة ونأخذ قياسات تلك المتحولات، فنحصل على القياسات البعدية لها . ونرمز لها بـ x_{2ji} ونكتبها مع متوسطاتها كما يلي:

$$\begin{bmatrix} X_{21i} \\ X_{22i} \\ X_{23i} \\ \vdots \\ X_{2Pi} \end{bmatrix} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} x_{211} & x_{212} & \cdots & x_{21n} \\ x_{221} & x_{222} & \cdots & x_{22n} \\ x_{231} & x_{232} & \cdots & x_{23n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{2P1} & x_{2P2} & \cdots & x_{2Pn} \end{bmatrix} \end{array} \Rightarrow \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \\ \vdots \\ \bar{X}_{2P} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum^{n} x_{21i} \\ \frac{1}{n}\sum^{n} x_{22i} \\ \frac{1}{n}\sum^{n} x_{23i} \\ \vdots \\ \frac{1}{n}\sum^{n} x_{2Pi} \end{bmatrix}$$

وبناء على ذلك يمكننا كتابة نتائج هذه القياسات ومتوسطاتها ضمن عمود واحد (شعاع واحد) كما يلي:

$$\begin{array}{c} d_{1i} \\ \\ d_{2i} \end{array} \begin{bmatrix} X_{11i} \\ X_{12i} \\ \vdots \\ X_{1Pi} \\ \hline X_{2Pi} \\ X_{2Pi} \\ \vdots \\ X_{2Pi} \end{bmatrix} \frac{\text{القياسات القبلية للمتحولات}\ X_1\ X_2 \dots X_P}{\text{القياسات البعدية للمتحولات}\ X_1\ X_2 \dots X_P} \quad \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \vdots \\ \bar{X}_{1P} \\ \hline \bar{X}_{2P} \\ \bar{X}_{2P} \\ \vdots \\ \bar{X}_{2P} \end{bmatrix} \frac{\text{متوسطات القياسات القبلية}}{\text{متوسطات القياسات البعدية}} \qquad (39-5)$$

وبناء على ذلك يمكننا حساب شعاع الفروقات بين القياسات القبلية والبعدية ومتوسطاتها من العلاقات التالية:

$$\boldsymbol{d_i} = \begin{bmatrix} d_{1i} \\ d_{2i} \\ \vdots \\ d_{Pi} \end{bmatrix} = \begin{bmatrix} x_{11i} - x_{21i} \\ x_{12i} - x_{22i} \\ \vdots \\ x_{1Pi} - x_{2Pi} \end{bmatrix} \Rightarrow \boldsymbol{\bar{d}} = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \\ \vdots \\ \bar{d}_P \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum d_{1i} \\ \frac{1}{n}\sum d_{2i} \\ \vdots \\ \frac{1}{n}\sum d_{Pi} \end{bmatrix} \qquad \boldsymbol{(40-5)}$$

حيث أن $i : 1\ 2\ 3 \ldots n$

والآن نفترض أن توقع شعاع هذه الفروقات ومصفوفة تبايناتها المشتركة في المجتمع يساويان :

$$\boldsymbol{\bar{D}} = E(\boldsymbol{d_i}) = \begin{bmatrix} \bar{D}_1 \\ \bar{D}_2 \\ \vdots \\ \bar{D}_P \end{bmatrix} \qquad \boldsymbol{cov}\ (\boldsymbol{d_i}) = \boldsymbol{V_d} \qquad (41-5)$$

كما نفترض أن هذه الفروقات مستقلة وخاضعة للتوزيع الطبيعي $N(\boldsymbol{\bar{D}}\ ,\boldsymbol{V_d})$، وعندها فإن مسألة الاستدلال حول شعاع متوسطات الفروقات $\bar{d}$ ضمن الفرضية $H_0 : \boldsymbol{\bar{D}} = 0$، تعالج بواسطة مؤشر الاختبار T^2 الذي يعطى من العلاقة التالية:

$$T^2 = n(\boldsymbol{\bar{d}} - \boldsymbol{\bar{D}})' * \boldsymbol{S_d^{-1}} * (\boldsymbol{\bar{d}} - \boldsymbol{\bar{D}}) \qquad (42-5)$$

حيث أن: $\boldsymbol{S_d}$ هي تقدير مصفوفة التباينات المشتركة $\boldsymbol{V_d}$ للفروقات d_j وتحسب من العلاقة:

$$\boldsymbol{S_d} = \frac{1}{n-1}\sum(\boldsymbol{\bar{d}_i} - \boldsymbol{\bar{D}})(\boldsymbol{\bar{d}_i} - \boldsymbol{\bar{D}})' = \boldsymbol{\tilde{V}_d}$$

وأن $\bar{d}_i$ هو شعاع متوسطات الفروقات المعرف بالعلاقة (5−40)، أما قيمة $\bar{D}$ فتأخذ من فرضية العدم $H_0 : \boldsymbol{\bar{D}} = 0$، ويتم اتخاذ القرار حول معنوية الفروقات d_i حسب العلاقة (5−31)، أي أننا نقبل H_0 باحتمال ثقة $(1 - \propto)$ إذا كانت:

$$T^2 \leq \frac{P(n-1)}{n-P} F_{P\ n-P}(\propto) \qquad (43-5)$$

ونشكل مجالات المتزامنة لها باحتمال ثقة $(1 - \propto)$ من العلاقة:

$$\bar{d}_{ii} \pm \sqrt{\frac{P(n-1)}{n-P} F_{P\ n-P}(\propto)} * \sqrt{\frac{S_{ii}}{n}} \qquad (44-5)$$

- **طريقة المصفوفة المتعارضة (Contrast) للاستدلال حول شعاع المتوسطات لـ p متحولاً :**

إن المصفوفة المتعارضة هي مصفوفة مستطيلة، تتميز بأن أسطرها مستقلة، وإن مجموع عناصر كل سطر فيها يساوي الصفر، ويرمز لها بالحرف C، مثل المصفوفات التالية :

$$\boldsymbol{C_{4*8}} = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{array}\right] \quad \boldsymbol{C_{4*5}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \qquad \boldsymbol{(45-5)}$$

$$\boldsymbol{C_{3*4}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \qquad \boldsymbol{C_{3*4}} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$

وتستخدم هذه المصفوفات للمساعدة في حساب قيمة المؤشر T^2 من القياسات القبلية والبعدية مباشرة، أي دون المرور بحساب شعاع الفروقات $\boldsymbol{d_i}$ ودون حساب المصفوفة $\boldsymbol{S_d}$ لها، وذلك باتباع التعليمات التالية :

نأخذ القياسات القبلية X_{1ji} والقياسات البعدية X_{2ji} للمفردة i عن المتحول j ونضعها في شعاع عمود واحد ثم نحسب متوسطاتها فنحصل على الشعاعين التاليين :

$$\underset{\boldsymbol{2P*1}}{\boldsymbol{X_i}} = \begin{bmatrix} X_{11i} \\ X_{12i} \\ \vdots \\ X_{1Pi} \\ \hline X_{21i} \\ X_{22i} \\ \vdots \\ X_{2Pi} \end{bmatrix} \frac{\text{القياسات القبلية للمتحولات للمفردة } i}{\text{القياسات البعدية للمتحولات للمفردة } i}, \underset{\boldsymbol{2P*1}}{\boldsymbol{\bar{X}}} = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \vdots \\ \bar{X}_{1P} \\ \hline \bar{X}_{2P} \\ \bar{X}_{2P} \\ \vdots \\ \bar{X}_{2P} \end{bmatrix} \frac{\text{المتوسطات القبلية للمتحولات}}{\text{المتوسطات البعدية للمتحولات}} = \begin{bmatrix} \frac{1}{n}\sum^{n} X_{11i} \\ \frac{1}{n}\sum^{n} X_{12i} \\ \vdots \\ \frac{1}{n}\sum^{n} X_{1Pi} \\ \hline \frac{1}{n}\sum^{n} X_{21i} \\ \frac{1}{n}\sum^{n} X_{22i} \\ \vdots \\ \frac{1}{n}\sum^{n} X_{2Pi} \end{bmatrix} \qquad \boldsymbol{(46-5)}$$

ثم نعرف المصفوفة المتعارضة C_{P*2P} من المرتبة $(P*2P)$ على الشكل التالي :

$$\boldsymbol{C_{P*2P}} = \begin{array}{c} \left[\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{array}\right] \\ \left(P \text{ عموداً}\right) \qquad \left(P \text{ عموداً}\right) \end{array} \left(P \text{ سطراً}\right) \qquad (47-5)$$

وعندها نجد أنه يمكننا التعبير عن شعاع الفروقات d_i لكل مفردة i والمعرفة بالعلاقة (5−40) من خلال العلاقة التالية:

$$\boldsymbol{d_i = CX_i} = \left[\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{array}\right] \begin{bmatrix} X_{11i} \\ X_{12i} \\ \vdots \\ X_{1Pi} \\ \hline X_{21i} \\ X_{22i} \\ \vdots \\ X_{2Pi} \end{bmatrix} = \begin{bmatrix} x_{11i} - x_{21i} \\ x_{12i} - x_{22i} \\ \vdots \\ x_{1Pi} - x_{2Pi} \end{bmatrix} \qquad (48-5)$$

حيث أن: $i : 1\ 2\ 3 \dots n$

وكذلك نجد أن شعاع متوسطاتها $\bar{d}$ يساوي:

$$\bar{\boldsymbol{d}} = \boldsymbol{C}\bar{\boldsymbol{X}} \qquad (49-5)$$

وأخيراً نجد أن مصفوفة التباينات المشتركة S_d تساوي :

$$\boldsymbol{S_d} = \boldsymbol{C} * \boldsymbol{S} * \boldsymbol{C'} \qquad (50-5)$$

حيث أن **S** هي مصفوفات التباينات المشتركة للقياسات القبلية X_{1ji} وللقياسات البعدية X_{2ji} الواردة في الشعاع X_i ، وتكون من المرتبة $2P * 2P$ ومؤلفة من أربع مصفوفات مربعة كما يلي:

$$\boldsymbol{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \qquad (51-5)$$

وحيث أن S_{11} هي مصفوفة التباينات المشتركة للقياسات القبلية X_{1ji} فقط ، وأن S_{22} هي مصفوفة التباينات المشتركة للقياسات البعدية X_{2ji} فقط . أما المصفوفتين S_{12} و S_{21} فهما مصفوفة التباينات المشتركة المتبادلة بين القياسات القبلية X_{1ji} والقياسات البعدية X_{2ji} . علماً بأنهما مرتبطتان بالعلاقة $S_{21} = S'_{12}$.

وبتعويض هذه المصفوفات في عبارة T^2 المعرفة في العلاقة (5−42)، وتحت فرضية العدم $(H_0: \bar{D} = 0)$ نجد أن T^2 تصبح مساوية كما يلي:

$$T^2 = n\bar{\boldsymbol{X}}'\boldsymbol{C}' * (\boldsymbol{CSC}')^{-1} * \boldsymbol{C}\bar{\boldsymbol{X}} \qquad (52-5)$$

وهي عبارة لا تتضمن الفروقات d_i، وهذا يعني أنه يمكننا حساب T^2 من العلاقة (5−52)، وذلك بدلالة المصفوفة المتعارضة C ومصفوفة التباينات المشتركة المزدوجة **S** للقياسات القبلية والبعدية معاً، وباستخدام شعاع المتوسطات المزدوج $\bar{X}_{2P}$ وضمن الفرضية: $H_0: \bar{\boldsymbol{D}} = 0$.

وبعد حساب T^2 نقارنها مع القيمة الحرجة لها : $T^2(\propto) = \frac{P(n-1)}{n-P} F_{P\ n-P}(\propto)$. فإذا كانت:

$$T^2 \leq \frac{P(n-1)}{n-P} F_{P\ n-P}(\propto) \qquad (53-5)$$

فإننا نقبل فرضية العدم $H_0: \bar{\boldsymbol{D}} = 0$ وذلك باحتمال ثقة $(1-\propto)$ والعكس بالعكس .

كما إن عملية إنشاء مجالات الثقة المتزامنة أو المنفردة لفروقات توقعات تلك المتحولات $\boldsymbol{C\mu_i}$ تعطى بالعلاقة :

$$\boldsymbol{C\mu_i} = \boldsymbol{C}\bar{\boldsymbol{X}}_{\boldsymbol{i}} \pm \sqrt{\frac{P(n-1)}{n-P} F_{P\ n-P}(\propto)} \sqrt{\frac{S_{ii}}{n}} \qquad (54-5)$$

5−2−4: حالة التجارب المكررة (على عينة واحدة ولمتحول واحد) :

هي تعميم آخر للحالة (5−2−1) للمقارنات الزوجية على متحول واحد باستخدام t (ستودينت). وذلك عندما نقوم بتكرار التجربة q مرة، على نفس العينة لقياس قيم أحد المتحولات X خلال فترات متتالية من الزمن . مثل متابعة أوزان عينة من المواليد بعد كل شهر . أو متابعة أوزان عينة من الأشخاص الخاضعين لنظام التنحيف بعد كل أسبوع ...الخ .

وعندها فإن كل مفردة i تعطينا شعاعاً مؤلفاً من q قيمة عددية نرمز لها بـ X_{qi} ونكتبها كما يلي :

$$\boldsymbol{X_i} = \begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{qi} \end{bmatrix} \qquad (55-5)$$

حيث أن X_{ji} هي نتيجة التجربة j على المفردة i

ولمقارنة هذه القياسات المستهدفة نفترض أن توقعاتها في المجتمع تشكل الشعاع $\boldsymbol{\mu}$ المعرف بالعلاقة التالية :

$$\boldsymbol{\mu} = E(\boldsymbol{X_i}) = \begin{bmatrix} E(X_{1i}) \\ E(X_{2i}) \\ \vdots \\ E(X_{qi}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} \qquad (56-5)$$

ثم نعمل على إجراء المقارنات الزوجية بين كل مركبتين من مركبات $\boldsymbol{\mu}$. ونتبع أحد الأسلوبين التاليين :

- مقارنة المركبة الأولى μ_1 مع إحدى المركبات الأخرى μ_j فنحصل على $(q-1)$ زوجاً . ولإجراء ذلك نستخدم المصفوفة المتعارضة C_1 ونكتبها كما يلي :

$$\begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_q \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \boldsymbol{C_1} * \boldsymbol{\mu} \qquad (57-5)$$

- مقارنة كل مركبة μ_j مع المركبة التي تسبقها μ_{j-1} فنحصل أيضاً على $(q-1)$ زوجاً . ولإجراء ذلك نستخدم المصفوفة المتعارضة C_2 ونكتبها كما يلي :

$$\begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \\ \vdots \\ \mu_q - \mu_{q-1} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \boldsymbol{C_2} * \boldsymbol{\mu} \qquad (58-5)$$

إن طبيعة هذا التصميم تعكس لنا الفروقات بين مركبات التجارب μ لكل واحدة مقابل أخرى . وهنا نشير إلى أنه يمكن للباحث أن يقوم بترتيب التجارب عشوائياً في كل موضوع حسب ما يراه مناسباً .

وعندما تكون توقعات تلك التجارب متساوية، فإن :

$$\boldsymbol{C_1\mu} = \boldsymbol{C_2\mu} = 0 \qquad (59-5)$$

ولذلك فإن فرضية العدم تكون حول عدم وجود فروقات بين أزواج قياسات التجارب من أجل أي خيار للمصفوفة المتعارضة C . أي من أجل أي تشكيلة لهذه الأزواج . وبصورة عامة نكتب الفرضيتين كما يلي:

$$H_0: \boldsymbol{C\mu} = 0 \qquad H_1: \boldsymbol{C\mu} \neq 0 \qquad (60-5)$$

وذلك من أجل أية مصفوفة متعارضة C .

ثم نجري التجارب ونحسب الفروقات بينهما باستخدام المصفوفة المتعارضة C كما يلي:

$$\boldsymbol{CX_i} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{qi} \end{bmatrix} = Y \qquad (61-5)$$

ثم نحسب متوسطاتها ومصفوفة تبايناتها المشتركة فنجد أن:

$$\boldsymbol{\overline{Y}_d = C\overline{X}} \qquad : \ \boldsymbol{\overline{X}} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum \boldsymbol{X_i} \qquad (62-5)$$

$$\boldsymbol{S_d = C * S * C' = C' * S * C} \qquad (63-5)$$

حيث أن **S:** هي مصفوفة التباينات المشتركة للقياسات X_{ji} . ولإجراء الاختبار حول فرضية العدم $H_0\!: \boldsymbol{C\mu} = 0$ نستخدم المؤشر T^2 المعرف بالعلاقة (5−52) والذي يأخذ الشكل التالي :

$$T^2 = n(\mathbf{C}\boldsymbol{\overline{X}})' * (\boldsymbol{CSC'})^{-\mathbf{1}} * \boldsymbol{C\overline{X}} \qquad (64-5)$$

وهو شكل لا يتضمن الفروقات d_i، بل يعطي ويحسب T^2 من القياسات المباشرة X_{ji} للتجارب المكررة . وهكذا نجد أنه إذا كانت القياسات X_{ji} خاضعة للتوزيع الطبيعي المتعدد $N(\boldsymbol{\mu}\,, \boldsymbol{V})$ وكانت فرضية العدم H_0 والفرضية البديلة H_1 كما يلي:

$$H_0\!: \boldsymbol{C\mu} = 0 \qquad\qquad H_1\!: \boldsymbol{C\mu} \neq 0 \qquad (65-5)$$

فإننا نقبل الفرضية H_0 إذا كانت قيمة T^2 تحقق المتراجحة التالية المعدلة من (5−53) باستبدال p بعدد أزواج الفروقات $(q-1)$، أي باعتبار عدد المتحولات p مساوياً لعدد الأزواجَ $(q-1)$، فنجد أن :

$$\mathrm{T}^2 = n(\boldsymbol{C\overline{X}})' * (\boldsymbol{CSC'})^{-\mathbf{1}} * \boldsymbol{C\overline{X}} \leq \frac{(q-1)(n-1)}{n-q+1} * F_{q-1\ n-q+1}(\propto) \qquad (66-5)$$

ولإنشاء مجالات الثقة المتزامنة باحتمال ثقة$(1-\propto)$ ولكل تجربة بمفردها $\left(C_j'\mu\right)$ نستخدم العلاقة التالية:

$$For\ \boldsymbol{C'\mu_j} :\ \boldsymbol{C_j'\overline{X}} \pm \sqrt{\frac{(q-1)(n-1)}{n-q+1} * F_{q-1\ n-q+1}(\propto)} * \sqrt{\frac{\boldsymbol{C_j' * S * C_j}}{n}} \qquad (67-5)$$

حيث أن C_j' هو السطر j من C .

ملاحظة: يمكن تعميم فكرة التجارب المكررة على متحولين X_1 وX_2 مؤثرين على مفردات العينة، ثم الاستعانة بالمصفوفات المتعارضة C لحساب T^2 ثم مقارنتها مع $T^2(\propto)$، أو لإنشاء مجالات الثقة المتزامنة . كما سنرى من خلال المثال التالي :

مثال (5−4): [مأخوذ من Johnston ,Wichern P. 218 بتصرف]

لدراسة تأثير المواد المخدرة على المدة الفاصلة بين نبضات القلب عند الكلاب . تم سحب عينة عشوائية منهم بحجم $n = 19$ كلباً . وبعد تحضيرهم تم تعريضهم لغاز ثاني أوكسيد الكربون (CO_2) بضغط عالٍ ثم بضغط منخفض، ثم تم إعطاءهم مادة (الهالوثين : H) وإعادة التجربة عليهم وتعريضهم ثانية

لغاز (CO_2) بضغط عالٍ، ثم بضغط منخفض، وتم تسجيل المدد الفاصلة بين نبضات القلب بعد كل تجربة (بالميلى ثانية) وتم تصميم جدول خاص لعرض نتائج هذه التجارب المتقاطعة كما يلي:

جدول (5-5): تصميم نتائج التجارب المتقاطعة (4 حالات)

CO_2 \ حالة H	حالات التعريض لـ (CO_2)		المجموع
	بضغط عال	بضغط منخفض	
قبل تناول H	التجربة الأولى وتوقعها μ_1	التجربة الثانية وتوقعها μ_2	$\mu_1+\mu_2$
بعد تناول H	التجربة الثالثة وتوقعها μ_3	التجربة الرابعة وتوقعها μ_4	$\mu_3+\mu_4$
المجموع	$\mu_1+\mu_3$	$\mu_2+\mu_4$	

وهكذا يكون شعاع التوقعات للمدد الفاصلة بين نبضات القلب حسب هذه التجارب الأربعة 1 , 2 , 3 , 4 كما يلي :

$$\boldsymbol{\mu}=\begin{bmatrix}\mu_1\\\mu_2\\\mu_3\\\mu_4\end{bmatrix}=\begin{bmatrix}\text{التجربة 1 : ضغط عالٍ لــ } CO2 \text{ بدون } H\\\text{التجربة 2 : ضغط منخفض لــ } CO2 \text{ بدون } H\\\text{التجربة 3 : ضغط عالٍ لــ } CO2 \text{ مع } H\\\text{التجربة 4 : ضغط منخفض لــ } CO2 \text{ مع } H\end{bmatrix}$$

وهنا نلاحظ أن هذا التصميم يعطينا ثلاثة أزواج للفروقات هي:

1- الفرق بين وجود (الهالوثين H) وعدم وجوده في كلتا حالتي ضغط CO_2 ويمثل بالفرق بين المجموعتين $(\mu_3+\mu_4)-(\mu_1+\mu_2)$.

2-الفرق بين التعريض بضغط عالٍ وضغط منخفض، في كلتا حالتي الدواء H ويمثل بالفرق بين المجموتين: $(\mu_1+\mu_3)-(\mu_2+\mu_4)$.

3-الفرق بين تداخل حالتي (الهالوثين) مع حالتي ضغط التعريض (CO_2) ويمثل بالفرق بين المجموعتين: $(\mu_1+\mu_4)-(\mu_2+\mu_3)$.

ويمكن التعبير عن هذه الفروقات باستخدام المصفوفة المتعارضة C_{3*4} التالية :

$$\boldsymbol{C}_{3*4}=\begin{bmatrix}-1&-1&1&1\\1&-1&1&-1\\1&-1&-1&1\end{bmatrix}$$

حيث نجد أنه يمكننا كتابة تلك الفروقات على الشكل التالي :

$$\boldsymbol{C\mu}=\begin{bmatrix}-1&-1&1&1\\1&-1&1&-1\\1&-1&-1&1\end{bmatrix}*\begin{bmatrix}\mu_1\\\mu_2\\\mu_3\\\mu_4\end{bmatrix}=\begin{bmatrix}(\mu_3+\mu_4)-(\mu_1+\mu_2)\\(\mu_1+\mu_3)-(\mu_2+\mu_4)\\(\mu_1+\mu_4)-(\mu_2+\mu_3)\end{bmatrix}$$

وبذلك تكون فرضية العدم على الشكل : $H_0\colon \boldsymbol{C\mu}=0$ مقابل $H_1\colon \boldsymbol{C\mu}\neq 0$

وبعد اعتماد هذا التصميم تم إجراء تلك التجارب على كلاب العينة المذكورة (التي حجمها $n = 19$ كلباً) وأخذوا قياسات المدد الفاصلة بين نبضات القلب (بالميلي ثانية) فكانت كما يلي:

جدول (5-6): نتائج قياس المدد الفاصلة بين نبضات القلب (بالميلي ثانية) وحسب التجارب الأربعة

رقم الكلب	نتائج التجربة 1 High CO_2-H	نتائج التجربة 2 Low CO_2-H	نتائج التجربة 3 High CO_2+H	نتائج التجربة 4 Low CO_2+H
1	426	609	558	600
2	253	236	392	395
3	359	433	349	357
4	432	431	522	600
5	405	426	513	513
6	324	438	507	539
7	410	312	410	426
8	326	326	350	504
9	375	447	547	548
10	286	286	403	422
11	349	382	473	497
12	429	410	488	547
13	348	377	447	514
14	412	473	472	446
15	347	326	455	468
16	434	458	637	524
17	364	367	432	469
18	420	395	508	531
19	397	556	645	625
المتوسطات	368.21	404.63	479.26	502.89
التباينات	2819.29	7963.14	6851.32	4878.99

ومن هذه البيانات نجد أن شعاع المتوسطات في العينة يساوي:

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \\ \bar{X}_4 \end{bmatrix} = \begin{bmatrix} 368.21 \\ 404.63 \\ 479.26 \\ 502.89 \end{bmatrix}$$

كما نجد أن مصفوفة التباينات المشتركة لهذه المتحولات الأربعة $X_1 \;\; X_2 \;\; X_3 \;\; X_4$ تساوي:

$$\boldsymbol{S} = \begin{bmatrix} 2819.29 & 3568.42 & 2943.49 & 2295.35 \\ 3568.42 & 7963.14 & 5303.98 & 4065.44 \\ 2943.49 & 5303.98 & 6851.32 & 4499.63 \\ 2295.35 & 4065.44 & 4499.63 & 4878.99 \end{bmatrix}$$

ومنها نجد أن الجداء $\boldsymbol{C\overline{X}}$ يساوي:

$$\boldsymbol{C\overline{X}} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} * \begin{bmatrix} 268.21 \\ 404.63 \\ 479.26 \\ 502.89 \end{bmatrix} = \begin{bmatrix} 209.31 \\ -60.05 \\ -12.79 \end{bmatrix}$$

وأن

$$\boldsymbol{C} * \boldsymbol{S} * \boldsymbol{C}' = \begin{bmatrix} 9432.32 & 1098.92 & 927.62 \\ 1098.92 & 5195.84 & 914.54 \\ 927.62 & 914.54 & 7557.44 \end{bmatrix}$$

وبالتعويض في (5−52) وإجراء الحسابات اللازمة نجد أن T^2 تساوي :

$$T^2 = n(\boldsymbol{C\overline{X}})' * (\boldsymbol{C} * \boldsymbol{S} * \boldsymbol{C}')^{-\mathbf{1}} * \boldsymbol{C\overline{X}} = 19(6.11) = 116$$

وبمقارنة هذه القيمة لـ T^2 مع القيمة الحرجة $T^2(\propto)$ التي تساوي عندما $(\propto = 0.05)$ ما يلي:

$$T^2(\propto) = \frac{(q-1)(n-1)}{n-q+1} * F_{q-1\ n-q+1}(\propto) = \frac{3*18}{16} F_{3\,,16}(0.05) = \frac{3*18}{16}(3.24) = 10.94$$

نجد أن $(T^2 = 116) > (T^2(\propto) = 10.94)$ لذلك نرفض فرضية العدم $H_0\colon \boldsymbol{C\mu} = 0$ ونقبل الفرضية H_1 التي تقول بأنه يوجد فروقات معنوية بين نتائج تلك التجارب الأربعة .

ولبيان أي الأسطر المتعارضة أدى إلى رفض الفرضية H_0 . نقوم بإنشاء مجالات الثقة المتزامنة لهذه الأسطر باحتمال ثقة قدره 0.95، فنجد أن السطر الأول يعطينا أن مجال الثقة للفرق الأول يساوي :

$$C_1'\mu = (\mu_3 + \mu_4) - (\mu_1 + \mu_2) = \text{تأثير}\left(\text{الهالوثين}\right)$$

وإن مجال الثقة له يقدر بما يلي:

$$[(\bar{x}_3 + \bar{x}_4) - (\bar{x}_1 + \bar{x}_2)] \pm \sqrt{\frac{3*18}{16} F_{3\,,16}(0.05)} * \sqrt{\frac{C_1'SC_1}{9}} =$$

$$= 209.31 \pm \sqrt{10.94}\sqrt{\frac{9432.32}{9}} =$$

$$I_1 = [135.61 \;\; , 283.01]$$

وهو مجال لا يتضمن النقطة صفر، أي أنه يوجد تأثير معنوي (للهالوثين) على المدد الفاصلة بين نبضات القلب وإن وجوده يجعلها أطول .

وبصورة مشابهة نجد أن مجال الثقة للفرق الثاني يساوي:

$$[(\bar{x}_1 + \bar{x}_3) - (\bar{x}_2 + \bar{x}_4)] \pm \sqrt{\frac{3*18}{16} F_{3\,,16}(0.05)} * \sqrt{\frac{C_2'SC_2}{n}} =$$

$$= -60.05 \pm \sqrt{10.94}\sqrt{\frac{5195.84}{9}} =$$

$$I_2 = [-5.35 \;\; , 104.5]$$

وهو مجال يتضمن النقطة صفر، أي أنه لا يوجد تأثير معنوي لمستوى ضغط CO_2 على المدد الفاصلة بين نبضات القلب . مع إن الضغط المنخفض لـ CO_2 يجعلها أطول قليلاً .

وكذلك الأمر بالنسبة للفرق الثالث حيث نجد أن:

$$[(\bar{x}_1+\bar{x}_4)-(\bar{x}_2+x_3)] \pm \sqrt{\frac{3*18}{16}F_{3,16}(0.05)} * \sqrt{\frac{C_3'SC_3}{n}} =$$

$$= -12.79 \pm \sqrt{10.94}\sqrt{\frac{7557.44}{9}} =$$

$$I_3 = [-146.73 \ \ ,54.18]$$

وهو يتضمن نقطة الصفر، أي أنه لا يوجد فرق معنوي لتداخل تأثير(الهالوثين) مع حالتي الضغط لـCO_2.

5-3: مقارنة شعاعي المتوسطات في مجتمعين طبيعيين (من عينتين مستقلتين):

5-3-1: تمهيد :

سنقوم الآن بتطوير فكرة المؤشر T^2 لمقارنة قياسات عدة متحولات، مأخوذة من مجتمعين طبيعيين وذلك دون أن نقوم بحساب الفروقات المتقابلة، كما فعلنا في حالة المقارنات الزوجية .

لذلك لنفترض أنه لدينا مجتمعين (مجتمع أول، ومجتمع ثاني) ومعرف عليهما p متحولاً هي: $X_1\ X_2\ X_3 \dots X_P$ ونريد مقارنة تغيرات هذه المتحولات في المجتمعين المذكورين، وذلك من خلال مقارنة شعاعي متوسطات تلك المتحولات في هذين المجتمعين والاستدلال على تجانسها أو اختلافهما . وحتى نستطيع القيام بعملية الاستدلال نفترض أن شعاعي توقعات هذه المتحولات في المجتمعين (1) و(2) هما: $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$. وإن الفرق بينهما هو: $(\mu_1-\mu_2)$ ونكتب ذلك كمايلي :

$$\boldsymbol{\mu_1} = \begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \vdots\\ \mu_{1P}\end{bmatrix} \quad \boldsymbol{\mu_2} = \begin{bmatrix}\mu_{21}\\ \mu_{22}\\ \vdots\\ \mu_{2P}\end{bmatrix} \quad , \quad (\boldsymbol{\mu_1}-\boldsymbol{\mu_2}) = \begin{bmatrix}\mu_{11}-\mu_{21}\\ \mu_{12}-\mu_{22}\\ \vdots\\ \mu_{1P}-\mu_{2P}\end{bmatrix} \qquad (68-5)$$

كما نفترض أن مصفوفتا التباينات المشتركة لتلك المتحولات في هذين المجتمعين هما $\boldsymbol{V_1}$ و $\boldsymbol{V_2}$ على الترتيب .

ونفترض أيضاً أن قياسات تلك المتحولات في المجتمع الأول يخضع للتوزيع الطبيعي المتعدد $N(\boldsymbol{\mu_1}, \boldsymbol{V_1})$ ، وأن قياسات تلك المتحولات في المجتمع الثاني تخضع للتوزيع الطبيعي المتعدد $N(\boldsymbol{\mu_2}, \boldsymbol{V_2})$.

وعلينا الآن أن نجيب على السؤال التالي :

هل $\boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ ؟، أو بعبارة أخرى هل $\boldsymbol{\mu_1} - \boldsymbol{\mu_2} = 0$ ؟، وإذا كان $\boldsymbol{\mu_1} - \boldsymbol{\mu_2} \neq 0$ فأي المركبات المتوسطة هي سبب هذا الاختلاف ؟

وحتى نجيب على هذه الأسئلة لابد من أن نميز- كما فعلنا في حالة المتحول الواحد- بين الحالتين التاليتين :

أ- الحالة التي يكون فيها $\boldsymbol{V_1 = V_2}$: حالة تجانس عناصر المصفوفتين .

ب-الحالة التي يكون فيها $\boldsymbol{V_1 \neq V_2}$: حالة عدم تجانس عنصر واحد أو أكثر منهما .

كما علينا أن نميز بين حالتي العينات الصغيرة والعينات الكبيرة . لذلك نفترض أيضاً أننا سحبنا من المجتمع الأول عينة عشوائية بحجم n_1 عنصراً، وسحبنا من المجتمع الثاني عينة عشوائية بحجم n_2 عنصراً . ثم أخذنا قياسات المتحولات $X_1\ X_2\ \dots\ X_P$ من عناصر هاتين العينتين، فحصلنا على القياسات الفعلية لها ثم حسبنا شعاعي متوسطاتها ومصفوفتي التباينات المشترك لكل منهما، فكانت كمايلي:

1-**قياسات العينة الأولى** : ونرمز لها بـ x_{1jn_1} ونكتبها كما يلي :

$$\begin{array}{lccccc} & 1 & 2 & 3 & \dots & n_1 \\ X_{11}: & x_{111} & x_{112} & x_{113} & \dots & x_{11n_1} \\ X_{12}: & x_{121} & x_{122} & x_{123} & \dots & x_{12n_1} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ X_{1P}: & x_{1P1} & x_{1P2} & x_{1P3} & \dots & x_{1Pn_1} \end{array} \qquad (69-5)$$

ونرمز لمتوسط المتحول X_j في العينة الأولى بـ $\bar{X}_{1j}$ ولشعاع المتوسطات فيها بـ $\bar{X}_1$ ونحسبهما كما يلي:

$$\bar{X}_{1j} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1ji} \implies \bar{\boldsymbol{X}}_{\boldsymbol{1}} = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \vdots \\ \bar{X}_{1P} \end{bmatrix} \qquad (70-5)$$

كما نحسب مصفوفة التباينات المشتركة لقياسات المتحولات $X_1\ X_2\ \dots\ X_P$ من العينة الأولى من العلاقة:

$$\boldsymbol{S_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\boldsymbol{X_{1ji}} - \bar{\boldsymbol{X}}_{\boldsymbol{1j}}\right)\left(\boldsymbol{X_{1ji}} - \bar{\boldsymbol{X}}_{\boldsymbol{1j}}\right)' \qquad (71-5)$$

2-**قياسات العينة الثانية**: ونرمز لها بـ x_{2ji} ونكتبها كما يلي :

$$\begin{array}{lccccc} X_{21}: & x_{211} & x_{212} & x_{213} & \dots & x_{21n_2} \\ X_{22}: & x_{221} & x_{222} & x_{223} & \dots & x_{22n_2} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ X_{2P}: & x_{2P1} & x_{2P2} & x_{2P3} & \dots & x_{2Pn_2} \end{array} \qquad (72-5)$$

ونرمز لمتوسط المتحول X_j في العينة الثانية بـ $\bar{X}_{2j}$ ولشعاع المتوسطات فيها $\bar{X}_2$ ونحسبهما كما يلي :

$$\bar{X}_{2j} = \frac{1}{n_2}\sum_{i=1}^{n_1} x_{2ji} \implies \bar{\boldsymbol{X}}_{\boldsymbol{2}} = \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \vdots \\ \bar{X}_{2P} \end{bmatrix} \qquad (73-5)$$

كما نحسب مصفوفة التباينات المشتركة لقياسات المتحولات $X_1\ X_2\ \dots\ X_P$ من العينة الثانية من العلاقة:

$$\boldsymbol{S_2} = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}\left(\boldsymbol{X_{2ji}} - \bar{\boldsymbol{X}}_{\boldsymbol{2j}}\right)\left(\boldsymbol{X_{2ji}} - \bar{\boldsymbol{X}}_{\boldsymbol{2j}}\right)' \qquad (74-5)$$

وأخيراً نشير إلى أنه إذا كانت العينتان المسحوبتان من المجتمعين مستقلتين، فإن هذا يعني أن القياسات الناتجة عنهما تكون مستقلة عن بعضهما البعض، أي أن قياسات كل متحول X_j في العينة الأولى تكون مستقلة عن قياساته في العينة الثانية، وهذا يعني أن القياسات X_{1ji} تكون مستقلة عن X_{2ji} لذلك سنعمل ضمن الفرضيات التالية :

1- إن قياسات العينة الأولى X_{1ji} خاضعة للتوزيع الطبيعي المتعدد الذي توقعه μ_1 ومصفوفة تبايناته المشتركة $\boldsymbol{V_1}$.

2- إن قياسات العينة الثانية X_{2ji} خاضعة للتوزيع الطبيعي المتعدد الذي توقعه μ_2 ومصفوفة تبايناته المشتركة $\boldsymbol{V_2}$.

3- أن تكون العينتان المسحوبتان من المجتمعين مستقلتان عن بعضهما البعض أي أن القياسات (X_{1ji}) مستقلة عن (X_{2ji}) . أما عندما يكون حجما العينيتين n_1 و n_2 صغيراً . فإننا نحتاج إلى وضع فرضية جديدة هي .

4- أن يكون $\boldsymbol{V_1} = \boldsymbol{V_2}$ أي أن المجتمعين لهما نفس مصفوفة التباينات المشتركة . وهذا يعني أن المجتمعين متجانسين كما في حالة المتحول الواحد .

5-3-2: مقارنة شعاعي المتوسطين عندما $\boldsymbol{V_1} = \boldsymbol{V_2}$:

في هذه الحالة نضع لهما رمز موحد $\mathbf{V}$ ونكتب ذلك كما يلي:

$$\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V} \qquad (75-5)$$

وبما أن V_1 تقدر بواسطة المصفوفة S_1 من العينة الأولى

وأن V_2 تقدر بواسطة المصفوفة S_2 من العينة الثانية

فإن التقدير المشترك وغير المتحيز لـ $\boldsymbol{V}$ هو التقدير الذي يدمج المصفوفتين $\boldsymbol{S_1}$ و $\boldsymbol{S_2}$ المتعلقتين بدرجتي الحرية (n_1-1) و (n_2-1)، بمصفوفة جديدة S_P، تسمى مصفوفة التباينات المشتركة المدمجة (Pooled) وتحسب من العلاقة التالية :

$$\boldsymbol{S_P} = \frac{(n_1-1)\boldsymbol{S}_1 + (n_2-1)\boldsymbol{S_2}}{n_1+n_2-2} \qquad (76-5)$$

ولإجراء المقارنة نضع الفرضيتين كما يلي:

$$H_0 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = \boldsymbol{\delta_0} \qquad : \boldsymbol{\delta_0} = \begin{bmatrix} \delta_{01} \\ \delta_{02} \\ \vdots \\ \delta_{0P} \end{bmatrix} \qquad (77-5)$$

$$H_1 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} \neq \boldsymbol{\delta_0}$$ من أجل δ_{0j} واحدة على الأقل

وبما أن الفرق $(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ يقدر بالفرق $(\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2})$ وإن القياسات في العينتين مستقلة عن بعضها البعض، فإن توقع الفرق $(\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2})$ يساوي حسب H_0 ما يلي :

$$E(\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2}) = E(\boldsymbol{\bar{X}_1}) - E(\boldsymbol{\bar{X}_2}) = \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = \boldsymbol{\delta_0} \qquad (78-5)$$

وإن التباين المشترك لهما $cov(\overline{\boldsymbol{X}}_{\boldsymbol{1}} , \overline{\boldsymbol{X}}_{\boldsymbol{2}}) = 0$، وبذلك نجد أن مصفوفة التباينات المشتركة للفرق $(\bar{X}_1 - \bar{X}_2)$ تساوي :

$$cov(\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) = cov(\overline{\boldsymbol{X}}_{\boldsymbol{1}}) + cov(\overline{\boldsymbol{X}}_2) - 2cov(\overline{\boldsymbol{X}}_{\boldsymbol{1}} , \overline{\boldsymbol{X}}_{\boldsymbol{2}}) \qquad (79-5)$$

وبما أن $V_1 = V_2 = V$ وأن $cov(\bar{X}_1 , \bar{X}_2) = 0$ فإننا نجد أن :

$$cov(\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) = \frac{\boldsymbol{V_1}}{n_1} + \frac{\boldsymbol{V_2}}{n_2} + 0 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{V} \qquad (80-5)$$

وإن هذه المصفوفة تقدر بواسطة المصفوفة S_P كما يلي :

$$cov(\widetilde{\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}}) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S_P} \qquad (81-5)$$

وبطريقة مشابهة لـ (27-5) نجد أن اختبار نسبة الامكانية العظمى للفرضية $H_0 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = \boldsymbol{\delta_0}$ مقابل $H_1 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} \neq \boldsymbol{\delta_0}$ يعتمد على مؤشر مربع المسافة الإحصائية T^2 بين الشعاعين $(X_1 - X_2)$ و δ_0 والمعرف بالعلاقة التالية :

$$T^2 = [(\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) - \boldsymbol{\delta_0}]' * \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S_P}\right]^{-1} [(\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) - \boldsymbol{\delta_0}] \qquad (82-5)$$

ونتخذ القرار حول قبول H_0 إذا كان :

$$T^2 \leq C^2(\ltimes) \qquad (83-5)$$

حيث أن $C^2(\ltimes)$ هي مربع المسافة الحرجة المقابلة لـ $\ltimes$ والمحددة من توزيع T^2 في العينتين . وهو توزيع مرتبط بالتوزيع F بواسطة العلاقة التالية :

$$T^2_{p\ n_1+n_2-1} = \frac{p(n_1 + n_2 - 2)}{n_1 + n_2 - p - 1} F_{p\ ,n_1+n_2-p-1} \qquad (84-5)$$

وهي تعديل للعلاقة (5−43) وذلك باستبدال كل n بـ $(n_1 + n_2 - 1)$ لأنه تم ربط المصفوفتين $\boldsymbol{S_1}$ و $\boldsymbol{S_2}$ بالعلاقة (5−76) للحصول على مصفوفة واحدة $\boldsymbol{S_P}$.

ولذلك فإننا نتخذ قرار قبول الفرضية H_0 إذا تحققت المتراجحة التالية :

$$T^2 \leq \frac{p(n_1 + n_2 - 2)}{n_1 + n_2 - p - 1} F_{p\ ,n_1+n_2-p-1}(\ltimes) = C^2(\ltimes) \qquad (\mathbf{85-5})$$

ونرفضها إذا كانت حالة العكس .

وأخيراً يمكننا أن نحدد منطقة الثقة من العلاقة:

$$P\left[\left((\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})_0\right)' * \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S_P}\right]^{-1} * \left((\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})_0\right) \leq C^2(\ltimes)\right] = 1 - \ltimes \ (\mathbf{86-5})$$

حيث أن:

$$C^2(\ltimes) = \frac{P(n_1 + n_2 - 2)}{n_1 + n_2 - P - 1} F_{P\ ,n_1+n_2-P-1}(\ltimes) \qquad (87-5)$$

وأن $F(\ltimes)$ تؤخذ من جداول F المقابلة لـ p و $n_1 + n_2 - p - 1$ درجتي حرية .

إن منطقة الثقة المحددة بالعلاقة (5–86) هي عبارة عن قطع ناقص متمركز في نقطة الفرق المشاهدة $(\bar{X}_1-\bar{X}_2)$، ويتم تحديد محوريه بواسطة القيم الذاتية والأشعة الذاتية للمصفوفة المدمجة S_P (أو لـ S_P^{-1}).

ولإيجاد مجالات الثقة المتزامنة للفرق $(\mu_1-\mu_2)$ نشكل التراكيب الخطية للفروقات $(\bar{X}_1-\bar{X}_2)$ ونكتبها كما يلي :

$$\bar{Z}=\ell'(\bar{X}_1-\bar{X}_2) \qquad (88-5)$$

وعندها نجد أن التوقع الرياضي للمتحول $\bar{Z}$ يساوي :

$$E(\bar{Z})=\ell'E(\bar{X}_1-\bar{X}_2)=\ell'\big((\mu_1-\mu_2)\big) \qquad (89-5)$$

وإن التباين المدمج لـ $\bar{Z}$ يساوي :

$$Var(\bar{Z})=S_{\bar{Z}}^2=\ell'\,S_P\ell \qquad (90-5)$$

ولاختبار فرضية العدم لهذه التراكيب الخطية نضع :

$$H_0:\ell'(\mu_1-\mu_2)=\ell'\delta_0 \qquad (91-5)$$

ونختبر H_0 باستخدام مربع مؤشر (ستودينت) t^2 التالي :

$$t_\ell^2=\frac{[\ell'(\bar{X}_1-\bar{X}_2)-\ell'(\mu_1-\mu_2)]^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{\bar{Z}}^2}=\frac{[\ell'((\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2))]^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\ell'Sp\ell} \qquad (92-5)$$

واعتماداً على متراجحة (كوشي– شوارز) لتعظيم الطرف الأيمن من أجل جميع قيم ℓ نضع $d=(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)$ و $B=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_P$. فنجد أن :

$$Max\ t_\ell^2=\big((\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)\big)'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_P\right]^{-1}\big((\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)\big)=T^2 \quad (93-5)$$

أي أن القيمة العظمى لـ t^2 تساويT^2 من أجل جميع قيم ℓ، وهكذا نجد أنه من أجل جميع قيم $\ell\neq 0$ أن احتمال الثقة يساوي:

$$P\big(T^2\leq C^2(\propto)\big)=P\big[t_\ell^2\leq C^2(\propto)\big]=1-\propto \qquad (94-5)$$

أي أن المتراجحة التالية :

$$\frac{[\ell'(\bar{X}_1-\bar{X}_2)-\ell'(\mu_1-\mu_2)]^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\ell'*S_p\ell}\leq C^2(\propto) \qquad (95-5)$$

تكون محققة باحتمال ثقة $(1-\propto)$ من أجل جميع قيم $\ell\neq 0$. ومن (3–95) نستخلص أن:

$$[\ell'(\bar{X}_1-\bar{X}_2)-\ell'(\mu_1-\mu_2)]^2\leq C^2(\propto)*\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\ell'S_p\ell \qquad (96-5)$$

وبعد جذر الطرفين نجد أن:

$$\ell'(\bar{X}_1-\bar{X}_2)-\ell'(\mu_1-\mu_2)\leq\pm C(\propto)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\ell'S_p\ell} \qquad (97-5)$$

أي أن مجالات الثقة المتزامنة للمقدار $\boldsymbol{\ell}'(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})$ تعطى ومن أجل جميع قيم $\boldsymbol{\ell}$ بالعلاقة العامة التالية :

$$\boldsymbol{\ell}'(\boldsymbol{\mu_1}-\boldsymbol{\mu_2}) = \boldsymbol{\ell}'(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2) \pm C(\propto)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\boldsymbol{\ell}'\boldsymbol{S}\boldsymbol{\ell}_p} \qquad (98-5)$$

حيث أن $C(\propto)$ تساوي :

$$C(\propto) = \sqrt{\frac{P(n_1+n_2-2)}{n_1+n_2-P-1}F_{P\,,n_1+n_2-P-1}\,(\propto)} \qquad (99-5)$$

ومن العلاقة (5-98) يمكننا استخلاص مجالات الثقة المنفردة، وذلك بوضع أحد مركبات $\boldsymbol{\ell}'$ مساوية للواحد والأخرى معدومة. فمثلاً لو وضعنا $\boldsymbol{\ell}(1\,,00000)$ نحصل على مجال الثقة المنفرد للفرق الأول $(\mu_{11}-\mu_{21})$ حيث نجد أنه يساوي :

$$(\mu_{11}-\mu_{21}) = (\bar{X}_{11}-\bar{X}_{21}) \pm C(\propto)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{11}} \qquad (100-5)$$

وهكذا نجد أنه إذا وضعنا$\ell'(010000)$ فإننا نحصل على مجال الثقة المنفرد للفرق الثاني$(\mu_{12}-\mu_{22})$ التالي :

$$(\mu_{11}-\mu_{21}) = (\bar{X}_{12}-\bar{X}_{22}) \pm C(\propto)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{22}} \qquad (101-5)$$

وإذا أردنا أن ننشأ مجالات الثقة المتزامنة حسب طريقة (بونفيروني) للفرق j, وهو $(\mu_{1j}-\mu_{2j})$ ، فإننا نستخدم العلاقة :

$$(\mu_{1j}-\mu_{1j}) : (\bar{X}_{1j}-\bar{X}_{2j}) \pm t_{n_1+n_2-P-1}\left(\frac{\propto}{2P}\right)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{jjP}} \qquad (102-5)$$

حيث أن $t\left(\frac{\propto}{2P}\right)$ هي القيمة الحرجة لمتحول (ستودينت) عند مستوى الدلالة $\left(\frac{\propto}{2P}\right)$ والمقابل لدرجة الحرية (n_1+n_2-P-1) .

مثال (5–5): لدراسة الفروقات بين متوسطات استهلاك الكهرباء (كيلو واط/ ساعة) في المنازل غير المكيفة (المجتمع1) وفي المنازل المكيفة (المجتمع2) خلال شهر تموز لعام 1977، في مدينة (ويسكونسين) الأمريكية. [مأخوذ من Johnson, Wichern P. 227 بتصرف] .

تم سحب عينتين عشوائيتين من هذين المجتمعين بحجمين $n_1 = 45$ و$n_2 = 55$ على الترتيب .

ولقياس الفروقات بين متوسطات الاستهلاك في هذين المجتمعين تم اعتماد المتحولين التاليين:

X_1 – كمية استهلاك الكهرباء في ساعات الذروة .

X_2 – كمية استهلاك الكهرباء في الساعات العادية .

وبعد أخذ القياسات المطلوبة لهذين المتحولين من عناصر العينيتن المسحوبتين من المجتمعين المذكورين. تم حساب شعاع المتوسط في العينة الأولى $\bar{\boldsymbol{X}}_{\mathbf{1}}$ وشعاع المتوسط في العينة الثانية $\bar{\boldsymbol{X}}_{\mathbf{2}}$، ثم تم حساب مصفوفة التباينات المشتركة لقياسات العينة الأولى $\boldsymbol{S}_{\mathbf{1}}$ وحساب مصفوفة التباينات المشتركة لقياسات العينة الثانية $\boldsymbol{S}_{\mathbf{2}}$، فكانت كما يلي(المركبة الأولى للمنازل غير المكيفة والثانية للمنازل المكيفة) :

$$\bar{\boldsymbol{X}}_{\mathbf{1}} = \begin{bmatrix}\bar{X}_{11}\\ \bar{X}_{12}\end{bmatrix} = \begin{bmatrix}204.4\\ 556.6\end{bmatrix} \quad , \quad \boldsymbol{S}_{\mathbf{1}} = \begin{bmatrix}13825.3 & 23823.4\\ 23823.4 & 73107.4\end{bmatrix} \quad , \quad n_1 = 45$$

$$\bar{\boldsymbol{X}}_{\mathbf{2}} = \begin{bmatrix}\bar{X}_{21}\\ \bar{X}_{22}\end{bmatrix} = \begin{bmatrix}130.0\\ 355.0\end{bmatrix} \quad , \quad \boldsymbol{S}_{\mathbf{2}} = \begin{bmatrix}8632.0 & 19616.7\\ 19616.7 & 55964.5\end{bmatrix} \quad , \quad n_2 = 55$$

والمطلوب: اختبار فرضية العدم : $H_0 \,:\, \boldsymbol{\mu}_{\mathbf{1}} - \boldsymbol{\mu}_{\mathbf{2}} = 0$

مقابل الفرضية البديلة : $H_1 \,:\, \boldsymbol{\mu}_{\mathbf{1}} - \boldsymbol{\mu}_{\mathbf{2}} \neq 0$

وذلك بمستوى دلالة $\propto = 0.05$ وبفرض أن $\boldsymbol{V}_{\mathbf{1}} = \boldsymbol{V}_{\mathbf{2}}$. ثم إنشاء مجالات الثقة المتزامنة للفروقات بين مركبات الشعاعين المتوسطين . وذلك باحتمال ثقة 95% .

الحل: لاختبار الفرضية H_0 علينا أن نقوم بحساب قيمة T^2 ثم مقارنتها مع القيمة الحرجة $T^2(\propto)$. وهكذا نجد من العلاقة (5−82) إنه لحساب T^2 نحتاج إلى حساب المصفوفة المدمجة $\boldsymbol{S}_{\boldsymbol{P}}$ فنجد مما سبق أن:

$$\boldsymbol{S}_{\boldsymbol{P}} = \frac{(n_1 - 1)\boldsymbol{S}_{\mathbf{1}} + (n_2 - 1)\boldsymbol{S}_{\mathbf{2}}}{n_1 + n_2 - 2} = \begin{bmatrix}10963.7 & 21505.5\\ 21505.5 & 63661.3\end{bmatrix}$$

وإن المقلوب $\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_P\right]^{-1}$ يساوي :

$$\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_{\boldsymbol{P}}\right]^{-1} = \left[\left(\frac{1}{45} + \frac{1}{55}\right)\boldsymbol{S}_{\boldsymbol{P}}\right]^{-1} = \begin{bmatrix}0.00669 & -0.00225\\ -0.00225 & 0.00115\end{bmatrix}$$

وبذلك نجد أن T^2 تحسب (ضمن شروط الفرضية H_0) كما يلي:

$$T^2 = [(\bar{X}_{11} - \bar{X}_{21}) - 0\,,(\bar{X}_{12} - \bar{X}_{22}) - 0\,] * \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_{\boldsymbol{P}}\right]^{-1} * \begin{bmatrix}(\bar{X}_{11} - \bar{X}_{21}) - 0\\ (\bar{X}_{12} - \bar{X}_{22}) - 0\end{bmatrix}$$

$$T^2 = \left[(204.4 - 130.0), (556.6 - 355.0)\begin{bmatrix}0.00669 & -0.00225\\ -0.00225 & 0.00115\end{bmatrix} * \begin{bmatrix}(204.4 - 130.0)\\ (556.6 - 355.0)\end{bmatrix}\right.$$

$$T^2 = [74.4\,, 201.6]\begin{bmatrix}0.00669 & -0.00225\\ -0.00225 & 0.00115\end{bmatrix} * \begin{bmatrix}74.4\\ 201.6\end{bmatrix}$$

$$T^2 = [74.4\,, 201.6] * \begin{bmatrix}0.044136\\ 0.06444\end{bmatrix} = 162.728$$

ثم نقوم بإيجاد القيمة الحرجة $T^2(\propto)$ من العلاقة :

$$T^2(\propto) = \frac{p(n_1 + \mathrm{n}_2 - 2)}{n_1 + n_2 - p - 1} F_{p\,,n_1+n_2-p-1}\,(\propto) = \frac{2 * 98}{97} * F_{2\,,97}\,(0.05) = 2.02 * (3.1)$$

$$T^2(\propto) = 6.26$$

وبمقارنة T^2 المحسوبة مع $T^2(\propto)$ الحرجة نجد أن :

$$(T^2 = 16.2728) > (T^2(\propto) = 6.26)$$

لذلك نرفض فرضية العدم $H_0 : \boldsymbol{\mu_1 - \mu_2} = 0$ ونقبل الفرضية البديلة $H_1 : \boldsymbol{\mu_1 - \mu_2} \neq 0$ ، أي أنه يوجد فروقات بين الأزواج المتقابلة لمركبات الشعاعين المتوسطين $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ (من أجل زوج واحد على الأقل). ولإنشاء مجالات الثقة المتزامنة للفروقات $(\boldsymbol{\mu_1 - \mu_2})$ بالنسبة لكل متحول من هذين المتحولين، نجد بالنسبة للمتحول الأول $\boldsymbol{\bar{X}_1}$ إن مجال الثقة المطلوب يحسب كما يلي:

$$(\mu_{11} - \mu_{21}) = (\bar{X}_{11} - \bar{X}_{21}) \pm C(\propto) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{11P}}$$

ومن الحسابات السابقة نجد أن:

$$(\mu_{11} - \mu_{21}) = (204.4 - 130.0) \pm \sqrt{6.26} \sqrt{\left(\frac{1}{45} + \frac{1}{55}\right) 10963.7}$$

$$= 74.4 \pm 52.66$$

أي أن مجال الثقة المتزامن ذي الاحتمال 0.95 للفرق الأول هو:

$$21.7 \leq (\mu_{11} - \mu_{21}) \leq 127.1$$

وهو مجال لا يضم النقطة صفر، وهذا ما يؤكد صحة القرار السابق حول رفض الفرضية H_0 .

أما بالنسبة للفرق الثاني $(\boldsymbol{\mu_{12} - \mu_{22}})$ فنجد أن مجال الثقة المتزامن ذي الاحتمال 0.95 يحسب كما يلي:

$$(\mu_{12} - \mu_{22}) = (\bar{X}_{12} - \bar{X}_{22}) \pm C(\propto) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{22P}}$$

$$= (556.6 - 355.0) \pm \sqrt{6.26} \sqrt{\left(\frac{1}{45} + \frac{1}{55}\right) 63661.3}$$

$$= 201.6 \pm 126.89$$

أي أن مجال الثقة المتزامن ذي الاحتمال 0.95 للفرق الثاني هو:

$$74.7 \leq (\mu_{12} - \mu_{22}) \leq 328.5$$

وهو أيضاً لا يتضمن النقطة صفر ، وهذا ما يؤكد صحة القرار السابق حول رفض الفرضية H_0 .

ونستخلص مما سبق بأنه يوجد فرق معنوي بين مركبتي شعاعي المتوسطين $\bar{X}_1$ و $\bar{X}_2$ لاستهلاك الكهرباء بين المنازل العادية والمنازل المكيفة في كلا التوقيتين ،في وقت الذروة أو في الوقت العادي .

ولإيجاد منطقة الثقة نبحث عن قطع الثقة ذي الاحتمال 0.95 من خلال حساب القيم الذاتية لـ $\boldsymbol{S_P}$ والأشعة الذاتية للمصفوفة المدمجة $\boldsymbol{S_P}$. لذلك نحسب القيم الذاتية من المعادلة التالية :

$$|\boldsymbol{S_P} - \lambda I| = \begin{vmatrix} 10963.7 - \lambda & 21505.5 \\ 21505.5 & 63661.3 - \lambda \end{vmatrix} = 0$$

ومنها نجد أن:

$$(10963.7 - \lambda)(63661.3 - \lambda) - (21505.5)^2 = 0$$

$$\lambda^2 - 74625\lambda + 235476864.6 = 0$$

وبحلها نجد أن :

$$\lambda_1 = 71323.5 \qquad\qquad \lambda_2 = 3301.5$$

ثم نقوم بإيجاد الأشعة الذاتية الممعيرة المقابلة لهاتين القيمتين فنجد أن:

$$\boldsymbol{e_1} = \begin{bmatrix} 0.336 \\ 0.942 \end{bmatrix} \qquad\qquad \boldsymbol{e_2} = \begin{bmatrix} 0.942 \\ -0.336 \end{bmatrix}$$

وهكذا نجد أن طول المحور الأول (الكبير) يساوي :

$$\ell_1 = \lambda_1 \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) C^2(\propto)} = 71323.5 \sqrt{\left(\frac{1}{45} + \frac{1}{55}\right)(6.26)} = 134.3$$

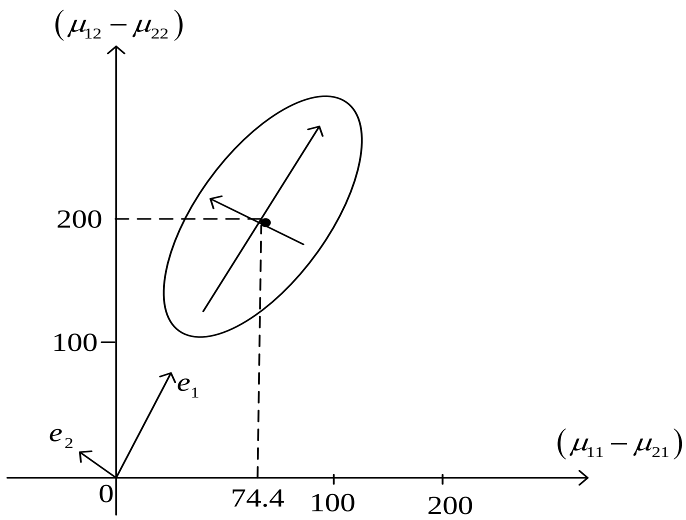
وإن طول المحور الثاني (الصغير) يساوي :

$$\ell_2 = \lambda_2 \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) C^2(\propto)} = 3301.5 \sqrt{\left(\frac{1}{45} + \frac{1}{55}\right)(6.26)} = 28.9$$

ولإيجاد مركز ذلك القطع نحسب شعاع الفروقات $(\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2})$:

$$(\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2}) = \begin{bmatrix} (\bar{X}_{11} - \bar{X}_{21}) \\ (\bar{X}_{12} - \bar{X}_{22}) \end{bmatrix} = \begin{bmatrix} (204.4 - 130.0) \\ 556.6 - 355.0 \end{bmatrix} = \begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

نقوم برسم هذا القطع على المستوى كما يلي:



الشكل (5-1): قطع الثقة 0.95

وهذا نلاحظ أن هذا القطع لا يشمل النقطة (صفر) وهذا ما جعل الاختبار T^2 يرفض فرضية العدم H_0 بمستوى دلالة $\propto = 0.05$.

وإذا أردنا حساب مجالات الثقة المتزامنة حسب (بونفيروني) نطبق العلاقة (5-102) فنجد أن مجال الثقة للفرق الأول $(\mu_{11} - \mu_{21})$ يساوي :

$$(\bar{X}_{11} - \bar{X}_{21}) \pm t_{n_1+n_2-P-1}\left(\frac{\propto}{2P}\right) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{11P}}$$

$$74.4 \pm 2.2767 \sqrt{\left(\frac{1}{45}+\frac{1}{55}\right)10963.7}$$

$$26.48 \leq (\mu_{11}-\mu_{21}) \leq 122.32$$

وكذلك نجد أن مجال الثقة للفرق الثاني $(\mu_{12}-\mu_{22})$ يساوي :

$$201.6 \pm 2.2767 \sqrt{\left(\frac{1}{45}+\frac{1}{55}\right)63661.3}$$

$$86.13 \leq (\mu_{12}-\mu_{22}) \leq 317.07$$

5-3-3: حالة مجتمعين فيهما $\boldsymbol{V_1} \neq \boldsymbol{V_2}$:

عندما تكون $\boldsymbol{V_1} \neq \boldsymbol{V_2}$ في المجتمعين المدروسين . فإنه لا يمكننا قياس مربع المسافة ضمن مؤشر مثل T^2، وذلك لأن التوزيع الاحتمالي له لا يعتمد على المصفوفتين المجهولتين $\boldsymbol{V_1}$ و$\boldsymbol{V_2}$. ويزداد الأمر سوءاً إذا كان التوزيع غير طبيعي، لأن التقاء حالة عدم تساوي المصفوفتين $\boldsymbol{V_1}$ و$\boldsymbol{V_2}$ مع حالة التوزيع غير الطبيعي لا تعالج إلا بواسطة اختبار (بارليت). ولكن هناك تحويلات تحسن من هذه الأمور عندما تكون التباينات الهامشية مختلفة تماماً، كما أنه عندما يكون n_1 و n_2 كبيرين فإنه يمكننا صياغة تراكيب مشابهة لمصفوفات التباينات المتشابهة .

ولمعالجة الحالة التي تكون فيها $\boldsymbol{V_1} \neq \boldsymbol{V_2}$ ، فإننا نصيغ تركيباً للفروقات بين المتوسطات بواسطة علاقة خطية كما يلي :

$$\bar{Z} = \boldsymbol{\ell}'(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2}) \qquad (103-5)$$

وبذلك نجد أن التوقع الرياضي لهذه التراكيب الخطية من العينتين يساوي :

$$E(\bar{Z}) = E[\boldsymbol{\ell}'(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})] = \boldsymbol{\ell}'[E(\boldsymbol{\bar{X}}_1)-E(\boldsymbol{\bar{X}_2})] = \boldsymbol{\ell}'(\boldsymbol{\mu_1}-\boldsymbol{\mu_2}) \qquad (104-5)$$

ولكن تباين هذه التراكيب يساوي :

$$Var(\bar{Z}) = \boldsymbol{\ell}' Var(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})\boldsymbol{\ell} = \boldsymbol{\ell}'\left[\frac{\boldsymbol{V_1}}{n_1}+\frac{\boldsymbol{V_2}}{n_2}\right]\boldsymbol{\ell} \qquad (105-5)$$

وحسب نظرية النهاية المركزية فإن $(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})$ يكون خاضعاً للتوزيع الطبيعي المتعدد التالي :

$$N\left[\boldsymbol{\ell}'(\boldsymbol{\mu_1}-\boldsymbol{\mu_2}), \boldsymbol{\ell}'\left[\frac{\boldsymbol{V_1}}{n_1}+\frac{\boldsymbol{V_2}}{n_2}\right]\boldsymbol{\ell}\right] \qquad (106-5)$$

وعندما تكون $\boldsymbol{V_1}$ و$\boldsymbol{V_2}$ معلومتين فإن مربع المسافة من $(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})$ إلى $(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})$ يساوي :

$$D^2 = [(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})-(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})]' * \left[\frac{\boldsymbol{V_1}}{n_1}+\frac{\boldsymbol{V_2}}{n_2}\right]^{-1} * [(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})-(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})] \qquad \mathbf{(107-5)}$$

وحسب نظرية الاحتمالات فإن المسافة D^2 تخضع تقاربياً إلى توزيع χ^2_P بـ p درجة حرية، حيث p عدد المتحولات المدروسة .

وعندما يكون $(n_1 - P)$ و $(n_2 - P)$ كبيرين، فإن $\boldsymbol{S_1}$ تقترب من $\boldsymbol{V_1}$ ، و $\boldsymbol{S_2}$ تقترب من $\boldsymbol{V_2}$ باحتمال كبير، فإنه يمكننا استبدال المصفوفتين $\boldsymbol{V_1}$ و $\boldsymbol{V_2}$ بالمصفوفتين $\boldsymbol{S_1}$ و $\boldsymbol{S_2}$.

ولإجراء الاختبار في هذه الحالة نضع الفرضيتين كما يلي:

$$H_0 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = 0 \qquad \text{مقابل} \qquad H_1 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} \neq 0 \qquad \mathbf{(108-5)}$$

وبذلك نجد أنه يمكننا قبول الفرضية H_0 باحتمال ثقة $(1 - \propto)$ إذا كانت المتراجحة التالية محققة .

$$[(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})]' * \left[\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\right]^{-1} * [(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})] \leq \chi_P^2(\propto) \qquad \mathbf{(109-5)}$$

حيث $\chi_P^2(\propto)$ هي القيمة الحرجة لمتحول التوزيع χ_P^2 ذي p درجة حرية، والتي تترك على يمينها احتمالاً قدرة $\propto$، ومنها نحدد قطع منطقة القبول .

وكذلك يمكننا إنشاء مجالات الثقة المتزامنة لتراكيب $\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ من العلاقة :

$$\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = \boldsymbol{\ell}'(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) \pm \sqrt{\chi_P^2(\propto)} * \sqrt{\boldsymbol{\ell}'\left(\frac{\boldsymbol{V_1}}{n_1} + \frac{\boldsymbol{V_2}}{n_2}\right)\boldsymbol{\ell}} \qquad (110-5)$$

وهذه المجالات تحتوي على $\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ باحتمال ثقة $(1 - \propto)$ وذلك من أجل جميع $\boldsymbol{\ell}$ الممكنة .

ملاحظة: إذا كان $n_1 = n_2 = n$ فإنه يكون لدينا :

$$\frac{n-1}{n+n-2} = \frac{n-1}{(n-1)+(n-1)} = \frac{1}{2} \qquad (111-5)$$

وعندها نجد أن:

$$\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2} = \frac{1}{n}(\boldsymbol{S_1} + \boldsymbol{S_2}) = \frac{2}{n}\left[\frac{\boldsymbol{S_1}}{\mathbf{2}} + \frac{\boldsymbol{S_2}}{\mathbf{2}}\right] = \left(\frac{1}{n} + \frac{1}{n}\right)\left[\frac{\boldsymbol{S_1}}{\mathbf{2}} + \frac{\boldsymbol{S_2}}{\mathbf{2}}\right] \qquad (112-5)$$

وباستبدال $\frac{1}{2}$ بما تساويه من (5−111) في (5−112)، يمكننا كتابة ذلك كما يلي:

$$\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2} = \left(\frac{1}{n} + \frac{1}{n}\right)\left[\frac{(n-1)\boldsymbol{S_1}}{n+n-2} + \frac{(n-1)\boldsymbol{S_2}}{n+n-2}\right] = \left(\frac{1}{n} + \frac{1}{n}\right)\boldsymbol{S_P} \qquad (113-5)$$

حيث أن: $\boldsymbol{S_P}$ هي المصفوفة المدمجة المشابهة للمصفوفة S_P السابقة .

ومما سبق نستنتج أنه إذا كان حجما العينتين متساويين $n_1 = n_2 = n$ ، فإن أسلوب العينات الكبيرة يكون مشابهاً لأسلوب المصفوفة المدمجة في حالة $\boldsymbol{V_1} = \boldsymbol{V_2}$.

ومعلوم لدينا أنه في حالة المتحول الواحد يكون تأثير التباينات غير المتساوية $\boldsymbol{V_1} \neq \boldsymbol{V_2}$ ، أقل ما يمكن عندما يكون حجما العينتين متساويين $n_1 = n_2$. ويكون أكبر ما يمكن عندما يكون n_1 أكبر بكثير من n_2 . أو يكون n_2 أكبر بكثير من n_1 .

مثال (5−6): لنتابع تحليل بيانات المثال (5−5) وذلك بافتراض أن $\boldsymbol{V_1} \neq \boldsymbol{V_2}$. ولذلك نقوم أولاً بحساب ما يلي:

$$\frac{1}{n_1}\boldsymbol{S_1} + \frac{1}{n_2}\boldsymbol{S_2} = \frac{1}{45}\begin{bmatrix}13825.3 & 23823.4\\23823.4 & 73107.4\end{bmatrix} + \frac{1}{55}\begin{bmatrix}8632.0 & 19616.7\\19616.7 & 55964.5\end{bmatrix}$$
$$= \begin{bmatrix}464.17 & 886.08\\886.08 & 2642.15\end{bmatrix}$$

وحتى نحصل على مجالي الثقة المتزامنين للتركيبين الخطيين $\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ نضع $\boldsymbol{\ell}' = (1\,,0)$ ثم $\boldsymbol{\ell}' = (0\,,1)$ فنجد على الترتيب أن:

$$\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = [1\,,0]\begin{bmatrix}(\mu_{11} - \mu_{21})\\(\mu_{21} - \mu_{22})\end{bmatrix} = [\mu_{11} - \mu_{21}]$$
$$\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = [0\,,1]\begin{bmatrix}(\mu_{11} - \mu_{21})\\(\mu_{21} - \mu_{22})\end{bmatrix} = [\mu_{21} - \mu_{22}]$$

من أجل اختبار الفرضية $H_0 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = 0$ نقوم بحساب D^2 من العلاقة :

$$D^2 = (\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2})'\left[\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\right]^{-1}(\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2})$$

$$D^2 = [(204.4 - 130.0)][(556.6 - 355.6)]\begin{bmatrix}464.17 & 886.08\\886.08 & 2642.15\end{bmatrix}^{-1}\begin{bmatrix}(204.4 - 130.0)\\(556.6 - 355.6)\end{bmatrix}$$
$$= (74.4\,, 201.6) * (10^{-4})\begin{bmatrix}59.874 & -20.080\\-20.080 & 10.519\end{bmatrix}\begin{bmatrix}74.4\\201.6\end{bmatrix} = 15.66$$

وباعتبار أن $\propto = 0.05$ نقوم بحساب القيمة الحرجة $\chi^2(\propto)$ فنجد أن: $\chi_2^2(0.05) = 5.99$

وبالمقارنة نجد أن: $(D^2 = 15.66) > \chi^2(0.05) = 5.99$، لذلك نرفض فرضية العدم H_0 .

وحتى نحصل على مجالي الثقة المتزامنين لمركبتي الفرق $\boldsymbol{\ell}'\,(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ نضع $\boldsymbol{\ell}'\,(1\,,0)$ ثم $\boldsymbol{\ell}'(0\,,1)$ ، فنجد على الترتيب أن:

$$\boldsymbol{\ell}'\,(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = (1\,,0)\begin{bmatrix}\mu_{11} - \mu_{21}\\\mu_{21} - \mu_{22}\end{bmatrix} = \mu_{11} - \mu_{21}\ ,\quad \boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = (0\,,1)\begin{bmatrix}\mu_{11} - \mu_{21}\\\mu_{21} - \mu_{22}\end{bmatrix} = \mu_{21} - \mu_{22}$$

وهكذا نجد أن مجال الثقة ذي الاحتمال 0.95 للفرق الأول $(\mu_{11} - \mu_{21})$ يساوي:

$$(\mu_{11} - \mu_{21}) = (1\,,0)[\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2}] \pm \sqrt{\chi_2^2(\propto)}\sqrt{(1\,,0)\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)\begin{bmatrix}1\\0\end{bmatrix}}$$

$$74.4 \pm \sqrt{5.99}\,\sqrt{464.7}$$

أي أن:

$$21.7 \leq \mu_{11} - \mu_{21} \leq 127.1$$

وكذلك نجد أن مجال الثقة ذي الاحتمال 0.95 للفرق الثاني $(\mu_{21} - \mu_{22})$ يساوي :

$$\left(\mu_{21}-\mu_{22}\right)=(0\,,1)[\bar{\boldsymbol{X}}_{\mathbf{1}}-\bar{\boldsymbol{X}}_{\mathbf{2}}]\pm\sqrt{\chi_2^2(\propto)}\sqrt{(0\,,1)\left(\frac{\boldsymbol{S_1}}{n_1}+\frac{\boldsymbol{S_2}}{n_2}\right)\begin{bmatrix}0\\1\end{bmatrix}}$$

$$201.6+\sqrt{5.99}\,\sqrt{2642.15}$$

أي أن:

$$75.8\le\left(\mu_{21}-\mu_{22}\right)\le 327.4$$

وهنا نلاحظ أن الاختلاف بين هذين المجالين والمجالين المحسوبين في المثال السابق المحلول حسب أسلوب المصفوفة المدمجة، اختلاف بسيط ولا يستحق الاهتمام، وهو يعود لطريقة الحساب.

ولإيجاد مركبة الشعاع $\boldsymbol{\ell}$ الحرجة، التي أدت إلى رفض الفرضية H_0 نقوم بحساب $\boldsymbol{\ell}$ (حسب كوشي–شوارز) كما يلي :

$$\boldsymbol{\ell}\approx\left(\frac{\boldsymbol{S_1}}{n_1}+\frac{\boldsymbol{S_2}}{n_2}\right)^{-1}(\bar{\boldsymbol{X}}_{\mathbf{1}}-\bar{\boldsymbol{X}}_{\mathbf{2}})=(10^{-4})\begin{bmatrix}59.874 & -20.080\\-20.080 & 10.519\end{bmatrix}*\begin{bmatrix}74.4\\201.6\end{bmatrix}$$

$$\boldsymbol{\ell}=\begin{bmatrix}0.041\\0.063\end{bmatrix}$$

أي أن مساهمة فرق الاستهلاك الكهربائي للمنازل المكيفة خلال الفترتين المذكورتين، كانت أكبر من مساهمة فرق الاستهلاك للمنازل غير المكيفة المقابل لها، في رفض الفرضية $H_0\ :\ \boldsymbol{\mu_1}-\boldsymbol{\mu_2}=0$.

وذلك لأن قيمة المركبة الثانية لـ $\boldsymbol{\ell}$ أكبر من قيمة المركبة الأولى .

الفصل السادس

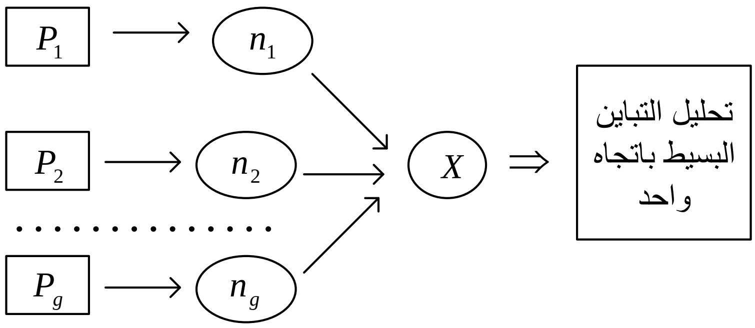
الاستدلال بواسطة تحليل التباين البسيط (ANOVA)

سنتناول في هذا الفصل عدة أنواع من تحليل التباين هي:

- تحليل التباين باتجاه واحد (ANOVA one way)
- تحليل التباين باتجاهين (ANOVA tow way)
- تحليل التباين بثلاث اتجاهات (ANOVA three way)
- تحليل المربع اللاتيني (LATIN SQUARE)
- تحليل التباين المشترك باتجاه واحد (ANCOVA one way)

6-1: تحليل التباين البسيط باتجاه واحد (ANOVA one way):

يتناول تحليل التباين البسيط باتجاه واحد دراسة تغيرات متحول واحد X (يسمى بالتابع)، الناتجة عن عدة مجتمعات (أو معالجات) نرمز لها بـ $P_1\ P_2\ ...\ P_g$. وهنا يشترط أن يكون عدد المجتمعات $g > 2$ (لأنه إذا كان $g = 2$ فإننا نستخدم اختبار (ستودينت) t لمقارنة متوسطي المجتمعين)، ويمكن تمثيل تأثير هذه المجتمعات على X كما في الشكل التالي:



الشكل (6-1): تمثيل ANOVA باتجاه واحد

ولدراسة تغيرات X الناتجة عن تأثيرات هذه المجتمعات نسحب من كل مجتمع k، عينة عشوائية بحجم n_k ، ثم نأخذ قياسات X من عناصر هذه العينات، ونضعها في جدول مناسب، يتضمن قياسات X من كل عينة n_k ومتوسطها $\bar{X}_k$ وتوقعها الرياضي في المجتمع μ_k ، كما يمكن أن يتضمن تباينها S_k^2 كالجدول التالي :

جدول (6-1): قيم X حسب عينات المجتمعات

تباينات العينات	التوقعات في المجتمع	متوسطات العينات	قياسات X من عناصر العينات	حجوم العينات	المجتمعات
S_1^2	μ_1	$\bar{X}_1$	$x_{11} \quad x_{12} \quad x_{13} \quad \ldots \quad x_{1n_1}$	n_1	P_1
S_2^2	μ_2	$\bar{X}_2$	$x_{21} \quad x_{22} \quad x_{23} \quad \ldots \quad x_{2n_2}$	n_2	P_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S_K^2	μ_k	$\bar{X}_K$	$x_{k1} \quad x_{k2} \quad x_{k3} \quad \ldots \quad x_{kn_k}$	n_k	P_K
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S_k^2	μ_g	$\bar{X}_g$	$x_{g1} \quad x_{g2} \quad x_{g3} \quad \ldots \quad x_{gn_g}$	n_g	P_g

ويشترط في هذه العينات والبيانات أن تحقق الشروط أو الافتراضات التالية :

1- أن تكون العينات المسحوبة من المجتمعات عشوائية ومستقلة عن بعضها البعض .

2- أن يكون تباين X في جميع هذه المجتمعات موحداً ويساوي σ^2 .

3- أن تكون قيم X في كل مجتمع k خاضعة للتوزيع الطبيعي $N(\mu_K, \sigma^2)$. الذي توقعه μ_k وتباينه ثابت ويساوي σ^2 في كل المجتمعات .

ويمكننا تجاهل الشرط الثالث (حول الطبيعية) عندما تكون حجوم العينات كبيرة، والاستناد على مفعول نظرية النهاية المركزية في الاحتمالات، والتي تنص على أن توزيعات X في تلك المجتمعات تنتهي إلى التوزيع الطبيعي .

وعندها فإن فرضية العدم والفرضية البديلة حول توقعات هذه المجتمعات تأخذان الشكل التالي:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu_g \qquad (1-6)$$

$$H_1 : \mu_k \neq \mu_\ell$$

وذلك من أجل زوج واحد على الأقل $(K \neq \ell)$

وإذا رمزنا للمتوسط المثقل لهذه التوقعات بالرمز $\bar{\mu}$ والذي يحسب من العلاقة التالية :

$$\bar{\mu} = \frac{n_1\mu_1 + n_2\mu_2 + \cdots + n_g\mu_g}{n_1 + n_2 + \cdots + n_g} = \frac{\sum^g n_k\mu_k}{\sum_{k=1} n_k} = \frac{\sum^g n_k\mu_k}{n} \qquad (2-6)$$

حيث أن n_K هو حجم العينة المسحوبة من المجتمع k، وإن مجموعها = n، أي أن: $n = \sum_{K=1}^{g} n_k$

ويطلق على المتوسط $\bar{\mu}$ مصطلح المتوسط أو التوقع الكلي (Grand mean)، وبناء على ذلك يمكننا التعبير عن قيمة أي توقع μ_k بدلالة التوقع الكلي $\bar{\mu}$ كما يلي :

$$\mu_k = \bar{\mu} + (\mu_k - \bar{\mu}) \qquad (3-6)$$

أو على الشكل التالي :

$$\mu_k = \bar{\mu} + \tau_k \qquad (4-6)$$

حيث أن:

$$\tau_K = \mu_k - \bar{\mu} \qquad (5-6)$$

ونعبر عن ذلك لفظياً كما يلي:

(تأثير المجتمع k (المعالجة k)) + (التوقع الكلي) = (توقع X في المجتمع k)

وهذا يقودنا إلى تعديل فرضية العدم حول التوقعات μ_k إلى فرضية عدم جديدة H_0 مقابل H_1 كما يلي:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \cdots = \tau_g = 0 \qquad (6-6)$$

وذلك من أجل مجتمع واحد على الأقل : $H_1 : \tau_k \neq 0$

وهذا يجعلنا نعتبر أن قياسات X في المجتمع k ، والتي سنرمز لها بـ x_{ki} ، خاضعة للتوزيع الطبيعي $N[(\bar{\mu} + \tau_k) , \sigma^2]$ وبالتالي يمكننا كتابة كل قياس منها x_{ki} كما يلي :

$$x_{ki} = \bar{\mu} + (\mu_k - \bar{\mu}) + (x_{ki} - \mu_k) = \bar{\mu} + \tau_k + \varepsilon_{ki} \qquad (7-6)$$

والتي يمكن كتابتها على الشكل التالي:

(حد الخطأ العشوائي) + (تأثير المجتمع k) + (التوقع الكلي) = (القياس x_{ki})

حيث أن: $\varepsilon_{ki} = x_{ki} - \mu_k$ ، وهو حد الخطأ العشوائي للقياس x_{ki} (أو البواقي)، وهي حدود مستقلة ويفترض أن تكون خاضعة للتوزيع الطبيعي $N(0 , \sigma^2)$. ولكن التأثيرات المجتمعية τ_k المعرفة في العلاقة (6−7) مرتبطة مع بعضها البعض ، وذلك لأنه اعتماداً على العلاقة (6−2) نجد أن:

$$\sum_{k=1}^{g} n_k\tau_k = \sum_{k=1}^{g} n_k(\mu_k - \bar{\mu}) = \sum n_k\mu_k - \bar{\mu}\sum n_k = \sum n_k\mu_k - n\mu = 0$$

وبالتالي نحصل على أن:

$$\sum_{k=1}^{g} n_k * \tau_k = 0 \qquad (8-6)$$

وهذا يعني أن مجموع قيم τ_k المثقلة بأحجام العينات n_k يساوي الصفر. وبالتالي يكون متوسطها المثقل $\bar{\tau} = 0$

وبناء على التركيب (6−7)، فإن تحليل التباين (ANOVA) في العينات يستخدم نموذجاً مشابهاً لـ (6−7)، للتعبير عن القياسات x_{ki} المشاهدة في العينات المسحوبة من تلك المجتمعات وذلك كما يلي:

$$x_{ki} = \bar{X} + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k) = \bar{X} + \tilde{\tau}_k + e_{ki} \qquad (9-6)$$

والذي يمكن كتابته في العينة على الشكل التالي:

(10−6) **(البواقي) + (تقدير تأثير المجتمع k) + (المتوسط الكلي في العينة) = (القياس x_{Ki} المشاهد)**

حيث أن $\bar{X}$ هو المتوسط الكلي في العينة ويحسب من $\bar{X} = \frac{1}{n}\sum n_k\bar{X}_k$ ويعتبر $\bar{X}$ تقديراً غير متحيز للتوقع الكلي في المجتمع $\bar{\mu}$.

وأن $\tilde{\tau}_k = (\bar{X}_k - \bar{X})$ هو تقدير لحد التأثير τ_k للمجتمع k، علماً بأن هذه الحدود يجب أن تحقق الشرط (6−8) التالي: $\sum n_k\tau_k = 0$.

وأن $e_{ki} = (\bar{x}_{ki} - \bar{X}_k)$ هو تقدير لحد الخطأ العشوائي ε_{ki} ، ويسمى بحد البواقي(residual) في العينة k ، وهذه الحدود تشكل متحولات عشوائية، توقعاتها تساوي الصفر وتخضع لـ $N(0 , \sigma^2)$.

مثال: (6–1): لنفترض أننا نريد دراسة تغيرات أحد المتحولات X في 3 مجتمعات، فسحبنا منها ثلاث عينات بحجوم مختلفة هي: $n_1 = 3\ ,\ n_2 = 2\ ,\ n_3 = 3$. وبعد أخذ قياسات X من عناصر هذه العينات حصلنا على القياسات التالية :

$(n_1 = 3)$: المجتمع 1 : X_{1i} 9 , 6 , 9

$(n_2 = 2)$: المجتمع 2 : X_{2i} 0 , 2 ,

$(n_3 = 3)$: المجتمع 3 : X_{3i} 3 , 1 , 2

وعند حساب متوسطات X في هذه العينات نجد أنها تساوي ما يلي :

$$\bar{X}_1 = \frac{9+6+9}{3} = 8$$
$$\bar{X}_2 = \frac{0+2}{2} = 1$$
$$\bar{X}_3 = \frac{3+1+2}{3} = 2$$

وكذلك نجد أن المتوسط الكلي لها يساوي :

$$\bar{X} = \frac{1}{n}\sum n_k\bar{X}_k = \frac{3*8+2*1+3*2}{3+2+3} = 4$$

وللتحقق من صحة العلاقة (6–9) نحسب قيمتي القياسين x_{11} و x_{31} فنجد أن:

$$9 = x_{11} = \bar{X} + (\bar{X}_1 - \bar{X}) + (x_{11} - \bar{X}_1) = 4 + (8-4) + (9-8) = 4+4+1 = 9$$

....

$$3 = x_{31} = \bar{X} + (\bar{X}_3 - \bar{X}) + (x_{31} - \bar{X}_3) = 4 + (2-4) + (3-2) = 4-2+1 = 3$$

وهكذا يتم حساب بقية القيم x_{ki} .

وإذا قمنا بتطبيق مثل تلك الحسابات على كل مشاهدة من المشاهدات السابقة، نحصل على التصفيفة (arrays) التالية (وليس المصفوفة) :

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & - \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & - \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & - \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & - \\ 1 & -1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \text{تصفيفة} \\ \text{المشاهدات} \end{pmatrix} = \begin{pmatrix} \text{تصفيفة} \\ \text{المتوسط الكلي} \end{pmatrix} + \begin{pmatrix} \text{تأثير تصفيفة} \\ \text{المجتمعات (المعالجات)} \end{pmatrix} + \begin{pmatrix} \text{تصفيفة} \\ \text{البواقي} \end{pmatrix}$$

لأن : $(x_{ki}) = (\bar{X}) + (\bar{X}_K - \bar{X}) + (x_{Ki} - \bar{X}_K)$

وبذلك نجد أن السؤال عن تساوي المتوسطات يتحول إلى تحديد فيما إذا كانت مساهمة تصفيفة تأثير المجتمعات (المعالجات) أكبر نسبياً من تصفيفة البواقي . علماً بأن التقديرات $\tilde{\tau}_k = (\bar{X}_k - \bar{X})$ للتأثيرات τ_k يجب أن تحقق دائماً الشرط (6–8) التالي :

$$\sum n_k\tilde{\tau}_k = 0$$

وضمن فرضية العدم H_0 ، يكون كل من $\tilde{\tau}_k$ تقديراً للعدد صفر، وهذا يعني أن تأثيرات المجتمعات تكون صغيرة .

أما إذا كانت تأثيرات المجتمعات (تأثيرات المعالجات) كبيرة . فإن ذلك سيؤدي إلى رفض الفرضية H_0 ، ولقياس مقدار مساهمة كل تصفيفة نكتب سطورها على شكل شعاع واحد، ثم نحسب مربع طول ذلك الشعاع، ويسمى هذا المقدار الجديد بمجموع المربعات (Sum of Squares) ونرمز له بالرمز SS .
فمثلاً نجد بالنسبة لتصفيفة القياسات المشاهدة (Observations) أن منقول شعاع عناصرها يكتب كما يلي :

$$Y' = [9\ \ 6\ \ 9\ \ 0\ \ 2\ \ 3\ \ 1\ 2]$$

وهو شعاع في الفضاء R^8 ، وإن مربع طوله يساوي $||Y||^2$ ويحسب كما يلي :

$$||Y||^2 = SS_{obs} = 9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2 = 216$$

وكذلك نجد أن مربع طول شعاع تصفيفة المتوسط الكلي يساوي:

$$SS_{means} = 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 = 128$$

وإن مربع طول شعاع تصفيفة تأثيرات المجتمعات أو (المعالجات treatments) يساوي :

$$SS_{tr} = 4^2 + 4^2 + 4^2 + (-3)^2 + (-3)^2 + (-2)^2 + (-2)^2 + (-2)^2 = 78$$

وإن مربع طول شعاع تصفيفة البواقي (residual) يساوي :

$$SS_{res} = 1^2 + (-2)^2 + 1^2 + (-1)^2 + 1^2 + 1^2 + (-1)^2 + 0^2 = 10$$

وبذلك نجد أن هذه المجاميع للمربعات ترتبط بنفس التركيب المعرف في (6−9) للمشاهدات . وهي تساوي :

$$SS_{obs} = SS_{means} + SS_{tr} + SS_{res}$$

حيث نلاحظ أن قيمها العددية تساوي :

$$216 = 128 + 78 + 10$$

أي أن تغيرات X الناتجة عن هذه المجتمعات تتوزع إلى ثلاث مركبات هي: تأثيرات المتوسطات والمعالجات والبواقي .

وإن تحليل التباين يقوم على مقارنة المقدار SS_{tr} (للمعالجات) مع المقدار SS_{res} (للبواقي) . فإذا كانت SS_{tr} أكبر بكثير من SS_{res} نرفض الفرضية H_0 ونقبل H_1 ونعتبر أن تغيرات X في هذه المجتمعات متباينة أو مختلفة .

والآن سنقوم باستخراج المعادلات الرياضية لهذه العلاقات وننطلق من العلاقة (6−9) التالية :

$$x_{ki} = \bar{X} + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k) \qquad (10-6)$$

ثم نطرح من الطرفين المتوسط الكلي $\bar{X}$ فنحصل على الانحرافات المصححة (أو الممعيرة) التالية :

$$(x_{ki} - \bar{X}) = 0 + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k)$$

وبذلك تختفي تصفيفة المتوسطات، ثم نقوم بتربيع الطرفين فنحصل على أن:

$$(x_{ki} - \bar{X})^2 = (\bar{X}_k - \bar{X})^2 + (x_{ki} - \bar{X}_k)^2 + 2(\bar{X}_k - \bar{X})(x_{ki} - \bar{X}_k) \qquad (11-6)$$

ثم نأخذ مجموع الحدود في الطرفين، المأخوذة على عناصر كل عينة k (أي $\sum_{i=1}^{n_k}$)، فنجد أنه ضمن كل عينة k يكون لدينا ما يلي :

$$\sum_{i=1}^{n_k}(x_{ki}-\bar{X})^2=\sum_{i=1}^{n_k}(\bar{X}_k-\bar{X})^2+\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2+2(\bar{X}_k-\bar{X})\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k) \qquad (12-6)$$

وهنا نلاحظ أن المجموع الأول في الطرف الأيمن ليس له علاقة بدليل القياسات i ، لذلك فهو يساوي:

$$\sum_{i=1}^{n_k}(\bar{X}_k-\bar{X})^2=n_k(\bar{X}_k-\bar{X})^2$$

كما نلاحظ أن المجموع الأخير يساوي الصفر: $\sum_{i=1}^{n_k}(\mathrm{x}_{ki}-\bar{X}_k)=0$

لأنه يمثل مجموع انحرافات قياسات X في العينة k عن متوسطها $\bar{X}_k$. وهذا يؤدي إلى انعدام الحد الأخير بكامله .

وبذلك تأخذ العلاقة (6−12) الشكل التالي :

$$\sum_{i=1}^{n_k}(x_{ki}-\bar{X})^2=n_k(\bar{X}_k-\bar{X})^2+\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2+0 \qquad (13-6)$$

وذلك ضمن كل عينة k مسحوبة من ذلك المجتمع .

ثم نقوم بأخذ مجموع الحدود في الطرفين، المأخوذ على جميع المجتمعات $\left(\text{أي } \sum_{k=1}^{g}\right)$ فنجد أن:

$$\sum_{k=1}^{g}\sum_{i=1}^{n_k}(x_{ki}-\bar{X})^2=\sum_{k=1}^{g}n_k(\bar{X}_k-\bar{X})^2+\sum_{k=1}^{g}\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2 \qquad (14-6)$$

وهذا يعني أن هذه الأطراف حسب المفاهيم السابقة تساوي ما يلي:

$$\begin{pmatrix}\text{اجمالي مجموع المربعات}\\ \text{المصحح بعد طرح } \bar{X}\end{pmatrix}=\begin{pmatrix}\text{مجموع المربعات}\\ \text{بين العينات المصحح}\end{pmatrix}+\begin{pmatrix}\text{مجموع مربعات البواقي}\\ \text{داخل العينات}\end{pmatrix}$$

ونرمز لهذه المجاميع بالرموز التالية :

$$SST=SSB+SSW \qquad (15-6)$$

وهنا نشير إلى أن درجة حرية الحد SSB تساوي $(g-1)$ لأن المتوسطات $\bar{X}_k$ ترتبط مع المتوسط العام $\bar{X}$ بالعلاقة $\left(\bar{X}=\frac{1}{n}\sum n_k\,\bar{X}_k\right)$، وهذا ما ينقص عدد درجات الحرية بمقدار واحد وتصبح $v_1=(g-1)$، ولإيجاد درجة حرية الحد الأخير SSW، نأخذ المجموع الداخلي منه $\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2$ فنجد أن درجة حريته تساوي (n_k-1) لأن قياسات كل عينة x_{ki} مرتبطة بمتوسطها $\bar{X}_k$ وفق العلاقة $\left(\bar{X}_k=\frac{1}{n_k}\sum x_{ki}\right)$، ومن ذلك نجد أن درجة حرية المجموع المضاعف $\sum_{k=1}^{g}\left[\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2\right]$ تساوي مجموع درجات الحرية لما بداخله ، أي أنها تساوي :

$$v_2=\sum_{k=1}^{g}(n_k-1)=\sum_{k=1}^{g}n_k-g=n-g \qquad (16-6)$$

ولحساب درجة حرية الطرف الأيسر SST ، نأخذ مجموع درجات الحرية لحرية الطرف الأيمن، وبذلك تكون درجة حرية SST مساوية لما يلي : $v = v_1 + v_2$ أي أن:

$$v = (g-1) + (n-g) = n-1 \qquad (17-6)$$

وأخيراً نضع نتائج هذه الحسابات في جدول منظم كالتالي :

جدول (6-2): الجدول النموذجي لـ ANOVA الأحادي One way .

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط مجموع المربعات	قيمة F
بين العينات (للمعالجات)	$SSB = \sum_{k=1}^{g} n_k(\bar{X}_k - \bar{X})^2$	$v_1 = g-1$	$M\ SSB = \frac{SSB}{(g-1)}$	$F = \frac{M\ SSB}{M\ SSW}$
داخل العينات (الخطأ)	$SSW = \sum_{k}^{g}\sum_{i}(x_{ki} - \bar{X}_k)^2$	$v_2 = \sum_{k=1}^{g} n_k - g$	$M\ SSW = \frac{SSW}{(\sum n_k - g)}$	
المجموع الكلي المصحح	$SST = \sum_{k}^{g}\sum_{i}^{n_k}(x_{ki} - \bar{X})^2$	$v = \sum_{k=1}^{g} n_k - 1$		

ولاختبار صحة الفرضية H_0 نستخدم المؤشر F المعرف بالعلاقة التالية :

$$F = \frac{SSB/(g-1)}{SSW/(\sum n_k - g)} = \frac{M\ SSB}{M\ SSW} \qquad (18-6)$$

ثم نقارن هذه القيمة المحسوبة مع القيمة الحرجة $F_{v_1 v_2}(\propto)$ لمتحول التوزيع F المقابلة لمستوى الدلالة $\propto$ ولدرجتي الحرية :

$$v_1 = g-1 \qquad\qquad v_2 = \left(\sum_{k=1}^{g} n_k - g\right) = n-g$$

ونتخذ القرار كما يلي :

(19 − 6) إذا كانت $F \le F_{v_1 v_2}(\propto)$ نقبل H_0 باحتمال ثقة قدره $(1-\propto)$.

(20 − 6) أما إذا كانت $F > F_{v_1 v_2}(\propto)$ نرفض H_0 ونقبل H_1 بمستوى دلالة قدره $\propto$.

مثال (6-2): لنأخذ بيانات المثال (6-1) السابق والتي تتناول قياسات متحول واحد X من (3) عينات مسحوبة من (3) مجتمعات والتي كانت كما يلي:

$$\begin{matrix} P_1: (n_1 = 3) \\ P_2: (n_2 = 2): \\ P_3: (n_3 = 3) \end{matrix} \begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & - \\ 3 & 1 & 2 \end{pmatrix} : \begin{matrix} \bar{x}_1 = 8 \\ \bar{x}_2 = 1 \\ \bar{x}_3 = 2 \end{matrix} \quad , \quad \bar{X} = 4$$

ومنها نجد أن :

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + n_3(\bar{X}_3 - \bar{X})^2$$
$$= 3(8-4)^2 + 2(1-4)^2 + 3(2-4)^2 = 78$$

$$SSW = \sum_{k=1}^{g}\sum_{i=1}^{n_g}(x_{ki} - \bar{X}_k)^2 = [(9-8)^2 + (6-8)^2 + (9-8)^2] +$$
$$+[(0-1)^2 + (2-1)^2] + [(3-2)^2 + (1-2)^2 + (2-2)^2] = 10$$

وبذلك نجد أن قيمة SST (المصححة) تساوي:

$$SST = 78 + 10 = 88$$

ثم ننظم جدولاً خاصاً بذلك كما يلي :

جدول (6-3): تحليل التباين ANOVA الأحادي One way .

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط مجموع المربعات	قيمة F
بين العينات (للمعالجات)	$SSB = 78$	$g - 1 = 2$	$M\ SSB = \frac{78}{2} = 39$	$F = \frac{M\ SSB}{M\ SSW} = \frac{39}{2}$
داخل العينات (الخطأ)	$SSW = 10$	$\sum n_k - g = 5$	$M\ SSW = \frac{10}{5} = 2$	
التباين الكلي (المصحح)	$SST = 88$	$\sum n_k - 1 = 7$		

ومنه نجد أن قيمة F المحسوبة تساوي :

$$F = \frac{M\ SSB}{M\ SSW} = \frac{\frac{SSB}{g-1}}{\frac{SSW}{\sum n_k - g}} = \frac{\frac{78}{2}}{\frac{10}{5}} = \frac{39}{2} = 19.5$$

ومن جداول التوزيع F نجد أن القيمة الحرجة لـ $F_{v_1 v_2}(\propto)$ عندما تكون $\propto = 0,05$ والمقابلة لدرجتي الحرية $v_1 = 2$ و $v_2 = 5$ ، تساوي: $F_{2,5}(0,05) = 5.786$: وبمقارنة F المحسوبة مع $F_{2,5}(\propto)$ الحرجة نجد أن :

$$(F = 19,5) > \left(F_{2,5}(0,05) = 5.786\right)$$

لذلك نرفض فرضية العدم H_0 التي تقول أن: $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ بمستوى دلالة 0.05، ونقبل H_1 التي تقول بوجود فروقات بين متوسطات تلك المجتمعات . وهنا يجب أن نتابع البحث عن مصدر تلك الفروقات .

ملاحظة 1: مما سبق نستنتج أنه يتم رفض H_0 عندما تأخذ النسبة $\frac{M\ SSB}{M\ SSW}$ قيمة كبيرة (أكبر من $F(\propto)$)، أو عندما تكون قيمة المقدار $\left(1 + \frac{M\ SSB}{M\ SSW}\right)$ كبيرة أيضاً، وهذا يكافئ القول التالي: إننا نرفض H_0 عندما تكون قيمة مقلوب المقدار السابق صغيرة . أي أننا نرفض H_0 عندما تكون قيمة المقدار التالي :

$$\lambda = \frac{1}{1 + \frac{M\ SSB}{M\ SSW}} = \frac{M\ SSW}{M\ SSB + M\ SSW} \qquad (20\,a - 6)$$

صغيرة بقدر كاف لتحقيق مستوى الدلالة $\propto$.

ويستفاد من هذه العلاقة في إيجاد العلاقة ، التي سوف نستخدمها في رفض H_0 في حالة عدة متحولات، كما سنرى لاحقاً .

ملاحظة 2: حول علاقة مؤشر تحليل التباين F باختبار (ستودينت) t : عندما يكون لدينا مجتمعان فقط $(g = 2)$ ، فإنه يمكننا كتابة الحد SSB كما يلي:

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 \qquad (20\,b-6)$$

علماً بأن المتوسط الكلي $\bar{X}$ يحسب من العلاقة التالية :

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

وبتعويض $\bar{X}$ في $(20\,b-6)$ نحصل على أن :

$$SSB = n_1\left(\bar{X}_1 - \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right)^2 + n_2\left(\bar{X}_2 - \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right)^2$$

$$SSB = \frac{n_1[(n_1 + \mathrm{n}_2)\bar{X}_1 - (n_1\bar{X}_1 + n_2\bar{X}_2)]^2 + n_2[(n_1 + n_2)\bar{X}_2 - (n_1\bar{X}_1 + n_2\bar{X}_2)]^2}{(n_1 + n_2)^2}$$

$$SSB = \frac{n_1\ n_2(\bar{X}_1 - \bar{X}_2)^2}{n_1 + n_2} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبذلك نجد أن المؤشر F (عندما $g = 2$) يأخذ الشكل التالي :

$$F = \frac{\frac{SSB}{g-1}}{\frac{SSE}{n-g}} = \frac{\frac{SSB}{1}}{\frac{SSE}{n_1 + n_2 - 2}} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{SSE_1 + SSE_2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبما أن $SSE_1 = (n_1 - 1)S_1^2$ و $SSE_2 = (n_2 - 1)S_2^2$ نجد أن:

$$F_{1,n-2}(\propto) = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = t_{n-2}^2\left(\frac{\propto}{2}\right): \qquad (20\,C-6)$$

وبمقارنة العبارة الأخيرة لـ F مع المؤشر t نجد: $F_{1,n-2}(\propto) = t_{n-2}^2\left(\frac{\propto}{2}\right)$ ، وهذا يعني أن مؤشر تحليل التباين F لمجتمعين يكافئ مربع اختبار (ستودينت) t، كما يمكن اعتبار الاختبار F تعميماً لاختبار (ستودينت) t .

2-6: تحليل التباين البسيط باتجاهين (وn مشاهدة) (ANOVA tow way) :

إن تحليل التباين البسيط باتجاهين يختلف جوهرياً عن تحليل التباين البسيط باتجاه واحد . وهو يستخدم لدراسة تغيرات متحول واحد (X) الناتجة عن تأثير عاملين نوعيين F_A و F_B (وليس عن مجتمعين كما في حالة الاتجاه الواحد)، وإن كل من هذين العاملين يأخذ عدة حالات، يدخلهما أو يتحكم بهما الباحث خلال مجربات التجارب، ثم يدرس تأثيرهما على المتحول التابع X. وهكذا يتم تنفيذ معظم الأبحاث العلمية.

فمثلاً: يمكن للباحث أن يدرس تغيرات أسعار طراز معين من الفساتين بتأثير عاملين: نوع القماش F_A وشكله الفني F_B . كما يمكنه أن يدرس تغيرات الجاذبية الأرضية بتأثير درجة الطول F_A ودرجة العرض F_B .

وفي هذه الحالة يتوجب على الباحث تحديد الحالات أو القيم التي يأخذها كل من F_A و F_B، وأن يرسم جدولاً خاصاً لتقاطعاتهما، ثم عليه أن يجري تجربة واحدة على الأقل مقابل كل حجرة لتقاطعهما . ثم عليه وضع نتائج تلك التجارب في جدول مناسب لحالات تقاطع F_A و F_B .

ولنفترض الآن أن F_A يأخذ g حالة منفصلة، وأن F_B يأخذ q حالة منفصلة . وان الباحث قد أجرى تجربة واحدة مقابل كل تقاطع لهما ووضع نتائجه في جدول كالتالي :

جدول (6–4): الحالات المتقاطعة لـ F_A و F_B

F_B \ F_A	1	2	3	⋯	g
1	x_{11}	x_{21}	x_{31}	⋯	x_{g1}
2	x_{12}	x_{22}	x_{32}	⋯	x_{g1}
⋮	⋮	⋮	⋮	⋯	⋮
q	x_{1q}	x_{2q}	x_{3q}	⋯	x_{gq}

ولكن الباحث يمكن أن يقوم بتكرار تجاربه مقابل كل حجرة (k,ℓ) عدداً من المرات . وليكن n مرة، فعندها سيحصل في كل حجرة على n نتيجة . وهذه النتائج تمثل عينة من القياسات مأخوذة من مجتمع التجارب الذي يقابل كل حجرة (k,ℓ) . وبذلك يكون لدينا $(g*q)$ مجتمعاً إحصائياً سحبت منها $(g*q)$ عينة عشوائية بحجوم متساوية n قياساً لـ X في كل منها .

ولنفترض أن توقع X في كل حجرة (k,ℓ) منها يساوي $(\mu_{k\ell})$، وأن التوقع الكلي لـ X لها يساوي μ حيث أن:

$$\mu = \frac{\sum_{k=1}^{g}\sum_{\ell=}^{q}\mu_{k\ell}}{g*q} \qquad (21-6)$$

وعندها يمكننا أن نعبر عن التوقع $\mu_{k\ell}$ في الحجرة (k,ℓ) بعلاقة مركبة كما يلي :

$$\mu_{k\ell} = \mu + (\mu_{k\ell} - \mu) \qquad (22-6)$$

وحتى نظهر تأثير العاملين F_A و F_B وتأثيرهما المشترك $(F_A\,F_B)$ على نتائج تلك التجارب نكتب التوقع $\mu_{K\ell}$ على الشكل التالي (وذلك بإضافة وطرح التوقعات الهامشية μ_k و μ_ℓ من الطرف الأيمن) :

$$\mu_{k\ell} = \mu + (\mu_k - \mu) + (\mu_\ell - \mu) + (\mu_{k\ell} - \mu_k - \mu_\ell + \mu) \qquad (23-6)$$

وهنا نلاحظ أن الأقواس تعكس تأثير العامل F_A والعامل F_B وتأثيرهما معاً . واختصاراً للرموز نكتب ذلك كما يلي :

$$\mu_{k\ell} = \mu + \propto_k + \beta_\ell + \gamma_{k\ell} \qquad (24-6)$$

حيث أن: $\propto_k = (\mu_k - \mu)$ و $\beta_\ell = (\mu_\ell - \mu)$ و $\gamma_{k\ell} = (\mu_{k\ell} - \mu_k - \mu_\ell + \mu)$

وهنا يشترط على المقادير $\propto_k$ و β_ℓ و $\gamma_{k\ell}$ أن تحقق الشروط التالية (يمكن البرهان على ذلك كما فعلنا في (6–8) مع ملاحظة أن حجوم العينات هنا متساوية وتساوي n) .

$$\sum_{k=1}^{g} \propto_k = 0 \qquad \sum_{\ell=1}^{q} \beta_\ell = 0$$
$$\sum_{k=1}^{g} \gamma_{k\ell} = 0 \qquad \sum_{\ell=1}^{q} \gamma_{k\ell} = 0 \qquad (25-6)$$

والآن نعود إلى العلاقة (6–24) ونكتبها كما يلي:

$$\mu_{k\ell} = E(x_{k\ell i}) = \mu + \propto_{\mathrm{k}} + \beta_\ell + \gamma_{k\ell} \qquad (26-6)$$

وفي العينات يمكننا كتابة قيمة أي قياس $x_{K\ell i}$ بخطأ $e_{K\ell i}$ كما يلي:

$$x_{k\ell i} = \bar{X} + \propto_k + \beta_\ell + \gamma_{k\ell} + e_{k\ell i} \qquad (27-6)$$

أو كما يلي :

$$x_{K\ell i} = \bar{X} + (\bar{X}_k - \bar{X}) + (\bar{X}_\ell - \bar{X}) + (\bar{X}_{K\ell} - \bar{X}_k - \bar{X}_\ell - \bar{X}) + (x_{K\ell i} - \bar{X}_{K\ell}) \qquad \mathbf{(28-6)}$$

$$\begin{pmatrix} \text{قيمة القياس} \\ x_{K\ell i} \end{pmatrix} = \begin{pmatrix} \text{المتوسط} \\ \text{الكلي} \end{pmatrix} + \begin{pmatrix} \text{تأثير العامل} \\ \text{الأول } F_A \end{pmatrix} + \begin{pmatrix} \text{تأثير العامل} \\ \text{الثاني } F_B \end{pmatrix} + \begin{pmatrix} \text{تأثير تداخل العاملين} \\ F_B \text{ و } F_A \text{ معا} \end{pmatrix} + \begin{pmatrix} \text{الخطأ العشوائي} \\ \text{أو البواقي} \end{pmatrix}$$

حيث أن $\bar{X}_k$ ترمز لمتوسطات العامل الأول F_A وأن $k : 1\ 2\ 3 \dots g$

وأن $\bar{X}_\ell$ ترمز لمتوسطات العامل الثاني F_B وأن $\ell : 1\ 2\ 3 \dots q$

وأن $\bar{X}_{k\ell}$ ترمز لمتوسطات القياسات في الحجرة (k,ℓ): $\bar{X}_{k\ell} = \frac{1}{n}\sum_{i=1}^{n} x_{k\ell i}$

وأن $e_{k\ell i}$ هي قيم البواقي أو الخطأ العشوائي، وهي عبارة عن متحولات عشوائية مستقلة ضمن كل حجرة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$ الذي توقعه يساوي الصفر ولها تباين موحد يساوي σ^2 .

وعندها فإن الفرضيات البحثية تأخذ الشكل التالي: فرضية العدم: وهي تتألف مما يلي :

$$H_0 : \begin{cases} \propto_k = 0 & k : 1\ 2\ 3 \dots g \\ \beta_\ell = 0 & \ell : 1\ 2\ 3 \dots q \\ \gamma_{k\ell} = 0 & k,\ell : 1\ 2\ 3 \dots \end{cases} \qquad (29-6)$$

أما الفرضية البديلة فتكون كما يلي:

$$H_1 : \begin{cases} \propto_K \neq 0 & \text{من أجل } k \text{ واحدة على الأقل :} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل :} \\ \gamma_{k\ell} \neq 0 & \text{من أجل زوج } (\ell\,,k) \text{ واحد على الأقل :} \end{cases} \qquad (30-6)$$

ولاستخراج مؤشرات الاختبار المناسبة نقوم بمعالجة العلاقة السابقة (6–28) كما فعلنا مع العلاقة السابقة (6–12) فنحصل على العلاقة التالية :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(x_{k\ell i} - \bar{X})^2 = qn\sum_{k=1}^{g}(\bar{X}_k - \bar{X})^2 + gn\sum_{\ell=1}^{q}(\bar{X}_\ell - \bar{X})^2 +$$
$$+ n\sum_{k=1}^{g}\sum_{\ell=1}^{q}(\bar{X}_{k\ell} - \bar{X}_k - \bar{X}_\ell + \bar{X})^2 + \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(x_{k\ell i} - \bar{X}_{k\ell})^2 \qquad (31-6)$$

والتي سنرمز لأطرافها اختصاراً كما يلي:

$$SST = SSA + SSB + SSAB + SSE \qquad (32-6)$$

أما بالنسبة لدرجات الحرية التي تقابل كل منها فهي تساوي:

$$(g*q*n-1) = (g-1)+(q-1)+(g-1)(q-1)+gq(n-1) \qquad (33-6)$$

ثم نعرف مؤشرات الاختبار المناسبة لكل من F_A و F_B وللتداخل بينهما كما يلي:

$$F_A = \frac{\frac{SSA}{g-1}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSA}{(g-1)SSE} \qquad (\text{للعامل } F_A) \qquad \mathbf{(34-6)}$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (g-1)$ و $v_2 = g*q(n-1)$.

$$F_B = \frac{\frac{SSB}{q-1}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSB}{(q-1)SSE} \qquad (\text{للعامل } F_B) \qquad \mathbf{(35-6)}$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (q-1)$ و $v_2 = g*q(n-1)$.

أما مؤشر اختبار التداخل بين F_A و F_B فيعرف كما يلي:

$$F_{AB} = \frac{\frac{SSAB}{(g-1)(q-1)}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSAB}{(g-1)(q-1)SSE} \qquad \mathbf{(36-6)}$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (g-1)(q-1)$ و $v_2 = g*q(n-1)$.

ثم نقوم بتنظيم جدول مناسب لتحليل التباين البسيط باتجاهين كما يلي:

جدول (6-5): جدول ANOVA باتجاهين tow way:

مصدر التباين	مجموع المربعات	درجة الحرية	المؤشر F
العامل F_A	$SSA = (\quad)$	$g-1$	$F_A = \frac{gq(n-1)SSA}{(g-1)SSE}$
العامل F_B	$SSB = (\quad)$	$q-1$	$F_B = \frac{gq(n-1)SSB}{(q-1)SSE}$
التداخل F_AF_B	$SSAB = (\quad)$	$(g-1)(q-1)$	$F_{AB} = \frac{gq(n-1)SSAB}{(g-1)(q-1)SSE}$
البواقي أو الخطأ العشوائي	$SSE = (\quad)$	$gq(n-1)$	———
الإجمالي (المصحح)	$SST= (\quad)$	$gq\,n-1$	———

وبعدها نقوم بمقارنة كل من F_A و F_B و F_{AB} بالقيم الحرجة المقابلة لها $F_{v_1,v_2}(\propto)$ بدرجتي الحرية v_1 و v_2 و نتخذ القرار حسب العامل المفروض كما يلي :

إذا كانت $F \le F_{v_1,v_2}(\propto)$ نقبل الفرضية H_0 حسب العامل المفروض

أما إذا كانت $F > F_{v_1,v_2}(\propto)$ نرفض الفرضية H_0 ونقبل H_1 $\qquad (37-6)$

ثم نستخلص النتائج الممكنة من هذه الاختبارات كما سنرى من خلال المثال التالي :

مثال (6-3): لنفترض أننا نريد دراسة تغيرات أسعار طراز معين من الفساتين في السوق، ومعرفة درجة تأثر أسعارها X بنوعية القماش F_A أو بشكل وزخرفة الفستان F_B . وبعد الدراسة تبين لنا أن هذه الفساتين تُصنع من (3) أنواع من القماش هي $A(A_1, A_2, A_3)$، وأن شكلها يأخذ شكلين أساسين من الزخرفة هما (B_1, B_2) . ثم قمنا بتتبع أسعار هذه الفساتين حسب كل تقاطع لحالات النوع والشكل، ولذلك أخذنا (4) أسواق (محلات) تبيع هذه الفساتين وسجلنا الأسعار فيها حسب النوع والشكل، فكانت كما يلي (حسبنا متوسطات الاسعار في كل حجرة ووضعناها ضمن مستطيلات):

جدول (6-6) بيانات المثال (فرضية):

أنواع القماش / أشكال الفساتين	نوع أول A_1	نوع ثاني A_2	نوع ثالث A_3		الاجمالي	المتوسطات $X\ell$
أحمر مزخرف B_1	430 450 $\boxed{467.5}$ 460 530	410 420 $\boxed{425}$ 430 440	420 440 $\boxed{450}$ 460 480		5370	447.5
أبيض مزخرف B_2	400 400 $\boxed{407.5}$ 400 430	350 370 $\boxed{380}$ 400 400	390 370 $\boxed{390}$ 400 400		4710	392.5
الاجمالي X_k	3500	3220	3360		10080	
المتوسطات $\bar{X}_k$	437.5	402.5	420			$\bar{X} = 420$

والمطلوب دراسة تأثير العاملين A وB على السعر X . وإجراء الاختبارات اللازمة بمستوى دلالة $(\propto = 0.05)$ ، علماً بأن الأرقام ضمن المستطيلات في كل حجرة هي متوسطات القياسات فيها $\bar{X}_{k\ell}$ وإن المتوسط الكلي لها $\bar{X} = 420$.

الحل: نقوم أولاً بحساب المجاميع التي في العلاقة (6-31) فنجد أن:

$$SST = \sum_{k=1}^{3}\sum_{\ell=1}^{2}\sum_{i=1}^{4}(x_{k\ell i} - 420)^2 = 34600 \qquad (31-6)$$

وذلك لأن:

$$SST = (100 + 900 + 1600 + 12100) + (100 + 0 - 100 + 400) + (0 + 400 + 1600 + 3600) \\ + (400 + 400 + 400 + 100) + (4900 + 2500 + 400 + 400) \\ + (900 + 2500 + 400 + 400) = 34600$$

ثم نقوم بحساب SSA من العلاقة :

$$SSA = qn\sum_{k=1}^{3}(\bar{X}_k - \bar{X})^2 =$$

$$SSA = 2 * 4[306.5 + 306.5 + 0] = 4900$$

ثم نقوم بحساب SSB من العلاقة :

$$SSB = gn\sum_{\ell=1}^{2}(\bar{X}_{\ell}-\bar{X})^2 =$$
$$SSB = 3*4[756.25+756.25] = 18150$$

ثم نقوم بحساب حد البواقي أو الخطأ العشوائي في جميع الحجر من العلاقة :

$$SSE = \sum_{k=1}^{3}\sum_{\ell=1}^{2}\sum_{i=1}^{4}(x_{k\ell i}-\bar{X}_{k\ell})^2 =$$
$$SSE = 5675+500+2000+675+1800+600 = 11250$$

وأخيراً نقوم بحساب حد التداخل SSAB من العلاقة :

$$SSAB = SST - SSA - SSB - SSE$$
$$SSAB = 34600 - 4900 - 18150 - 11250 = 300$$

ثم نقوم بتنظيم جدول التحليل التالي :

جدول (6–7) جدول تحليل ANOVA **باتجاهين** :

مصدر التباين	مجموع المربعات	درجة الحرية	متوسطات المربعات	قيم F
نوعية القماش A	$SSA = 4900$	$g-1=2$	$MSSA = 2450$	$FA = 3.92$
شكل الفستان B	$SSB = 18150$	$q-1=1$	$MSSB = 18150$	$F_B = 29.04$
تداخل A وB	$SSAB = 300$	$(g-1)(q-1)=2$	$MSSAB = 150$	$F_{AB} = 0.24$
الخطأ العشوائي	$SSE = 11250$	$gq(n-1)=18$	$MSSE = 625$	———
الاجمالي	$SST = 34600$	$gqn-1=19$	———	———

ولاختبار تأثير هذين العاملين وتأثير تداخلهما، علينا أن نقوم بإيجاد القيمة الحرجة لـ F المقابلة لكل حالة فنجد أن:

$$F_{2,18}(0.05) = F_{2,18}(0.05) = 3.55 \quad \text{و} \quad F_{1,18}(0.05) = 4.41$$

ثم نقوم بمقارنة قيم F_A مع $F_{2,18}(\propto)$ نجد أن: $(F_A = 3.92) > 3.55$ ، لذلك نرفض فرضية العدم H_0، التي تقول أنه لا يوجد تأثير لنوعية القماش على سعره ، ونقبل H_1 التي تقول أن أسعار الفساتين تتأثر بنوعية القماش (بنسبة $\left(\frac{4900}{34600}\right)*100\%$) .

ثم نقوم بمقارنة F_B مع $F_{1,18}(\propto)$ نجد أيضاً أن: $(F_B = 29.04) > 4.41,$ ، لذلك نرفض فرضية العدم التي تقول أنه لا يوجد تأثير لشكل الفستان على سعره ، ونقبل H_1 التي تقول أن أسعار الفساتين تتأثر كثيراً بشكل القماش (بنسبة $\left(\frac{18150}{34600}\right)*100\%$) .

ولاختبار تأثير التفاعل الداخلي للعاملين A وB نقارن قيمة F_{AB} مع القيمة الحرجة$(\propto)F_{2\ 18}$ فنجد أن: $(F_{AB} = 0.24) < 3.55$ ، لذلك نقبل H_0 ونعتبر أن أسعار الفساتين لا تتأثر بتداخل العاملين المدروسين وهما النوعية والشكل .

ملاحظة: يفضل في التطبيقات العملية أن نبدأ بإجراء اختبار التداخل . فإذا كان تأثير ذلك التداخل معنوياً (حالة رفض H_0) . فإن ذلك يعني أن تداخل العاملين F_A و F_B يؤثر كثيراً على تغيرات المتحول

التابع X . وهذا يجعل عملية اختبار تأثير كل من F_A و F_B على حدة على X غير واضحة وصعبة التفسير، وينصح بعدم متابعة دراسة تأثيراتهما المنفردة.

أما إذا كان تأثير التداخل مهملاً فإنه يمكننا متابعة التحليل وإجراء الاختبارين حول تأثير F_A و F_B واستخلاص النتائج الممكنة .

6-3: تحليل التباين البسيط بثلاث اتجاهات (n مشاهدات) :

إن تحليل التباين في هذه الحالة هو تعميم للحالة السابقة (6-2) . لذلك نفترض أننا نريد دراسة تغيرات أحد المتحولات X (المتحول التابع) الناتجة عن تأثير (3) عوامل نوعية: F_A و F_B و F_C . وإن لكل من هذه العوامل عدة حالات يرمز لها كما يلي:

$$\begin{aligned} F_A &: A_1, A_2, \dots, A_K, \dots, A_g \\ F_B &: B_1, B_2, \dots, B_\ell, \dots, B_q \\ F_C &: C_1, C_2, \dots, C_m, \dots, C_r \end{aligned} \qquad (38-6)$$

ولذلك ننشأ مكعب التقاطعات الممكنة لهذه الحالات، فنحصل على $(g*q*r)$ حجرة، ثم نأخذ من كل حجرة منه n قياساً للمتحول المدروس X . فنحصل على $(g*q*r)$ عينة مسحوبة من مجتمعات تلك الحجر وعلى $(g*q*r*n)$ قياساً .

وإذا رمزنا لتلك القياسات بالرموز $x_{k\ell mi}$ فإنه يمكننا كتابة النموذج الرياضي الموافق لهذا التحليل باستخدام نفس المفاهيم والرموز السابقة كما يلي :

$$x_{k\ell mi} = \bar{X} + \propto_k + \beta_\ell + \gamma_m + (\propto\beta)_{k\ell} + (\propto\gamma)_{km} + (\beta\gamma)_{\ell m} + (\propto\beta\gamma)_{k\ell m} + (e_{k\ell mi}) \quad \mathbf{(39-6)}$$

ويشترط على هذه الرموز أن تحقق شروط مشابهة للشروط السابقة المفروضة على تحليل التباين باتجاهين والمذكورة في (6-25) .

أما فرضية العدم H_0 (وهي عدم وجود تأثير لهذه العوامل على X) فتكتب كما يلي:

$$H_0: \begin{cases} \propto_1 = \propto_2 = \cdots = \propto_g = 0 \\ \beta_1 = \beta_2 = \cdots = \beta_q = 0 \\ \gamma_1 = \gamma_2 = \cdots = \gamma_r = 0 \end{cases} \qquad (40-6)$$

أما الفرضية البديلة H_1 فتكتب كما يلي:

$$H_1: \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_m \neq 0 & \text{من أجل } m \text{ واحدة على الأقل} \end{cases} \qquad (41-6)$$

ثم نقوم بمعالجة العلاقة (6-39) كما فعلنا في الفقرات السابقة فنحصل مجاميع المربعات المختلفة ونكتبها كما يلي:

$$SST = SSA + SSB + SSC + SS(AB) + SS(AC) + SS(BC) + SS(ABC) + SSE \quad (42-6)$$

وإن درجات الحرية المقابلة لهذه المجاميع هي كما يلي:

$$\begin{aligned} (gqrn) - 1 = {} & (g-1) + (q-1) + (r-1) + (g-1)(q-1) + (g-1)(r-1) \\ & + (q-1)(r-1) + (g-1)(q-1)(r-1) + gqr(n-1) \end{aligned} \qquad (43-6)$$

وسنقوم بتعريف هذه المجاميع وشرح كيفية حسابها لاحقاً من خلال المثال (6–4) :

أما بالنسبة لإجراء الاختبارات اللازمة نبدأ بإجراء اختبار تأثير التداخل الثلاثي (SS(ABC، ونحسب قيمة مؤشر الاختبار F_{ABC} المقابل له من العلاقة :

$$F_{ABC} = \frac{\dfrac{SS(ABC)}{(g-1)(q-1)(r-1)}}{\dfrac{SSE}{gqr(n-1)}} = \frac{\dfrac{SS(ABC)}{v_1}}{\dfrac{SSE}{v_2}} \qquad (44-6)$$

ثم نقارن هذه القيمة المحسوبة مع القيمة الحرجة لمتحول$F_{v_1,v_2}(\propto)$ المقابل لمستوى الدلالة $\propto$ ولدرجتي الحريةv_1 و v_2، ونتخذ القرار كما يلي:

$$\begin{array}{l} \text{إذا كانت } F \leq F_{v_1,v_2}(\propto) \text{ نقبل فرضية العدم } H_0 \\ \text{وإذا كانت } F > F_{v_1,v_2}(\propto) \text{ نرفض } H_0 \text{ ونقبل } H_1 \end{array} \qquad (45-6)$$

وإذا كانت النتيجة رفض H_0 وقبول H_1، فإن ذلك يعني أن تداخل العوامل الثلاثة يؤثر معنوياً على تغيرات المتحول المدروس X . لذلك فإن عملية البحث عن تأثير كل منها بمفرده تصبح غير واضحة وغير ممكنة التفسير .

أما إذا كانت نتيجة الاختبار السابق قبول H_0، فإن ذلك يعني أن تأثير تداخل العوامل الثلاثة غير معنوي ويمكن إهماله . وبعدها ننتقل إلى اختبارات تأثيرات التداخلات الثنائية . وإذا كانت غير معنوية، نقوم باختبار تأثيرات العوامل المنفردة وذلك بتطبيق نفس الإجراءات السابقة .

مثال (6–4): لدراسة مقاومة ألواح السيراميك للصدمات تم تحديد (3) عوامل مؤثرة عليها وهي :

- نوع المادة الأولية وتأخذ حالتين فقط (A_1 , A_2) .
- مساحة اللوح ويأخذ حالتين أيضاً (B_1 , B_2) .
- سماكة اللوح ويأخذ (3) حالات هي : (C_1 , C_2 , C_3) .

ثم تم اختبار (5) ألواح من كل حجرة لتقاطعاتها وأجريت عليها تجارب لقياس المقاومة (بالكيلوغرام) فكانت نتائج تلك القياسات كما في الجدول التالي :

جدول (6–7): بيانات المثال (فرضية)

نوع المادة		نوع المادة A_1						نوع المادة A_2					
المساحة		المساحة B_1			المساحة B_2			المساحة B_1			المساحة B_2		
السماكة		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
عناصر العينة i	$i=1$	67	72	78	67	79	78	54	52	63	54	57	60
	2	66	67	81	71	80	78	51	56	54	56	58	68
	3	62	75	67	72	81	77	47	52	65	58	61	61
	4	71	70	75	70	80	83	51	52	62	51	59	61
	5	69	71	75	81	85	79	59	53	60	57	55	67

والمطلوب دراسة تأثير هذه العوامل على مقاومة ألواح السيراميك وإجراء الاختبارات اللازمة عليها بمستوى دلالة $\propto = 0.05$

الحل: لإجراء هذه الاختبارات نحتاج إلى كثير من الحسابات المعقدة لإيجاد قيم حدود العلاقة (6–42) وسنستخدم لذلك علاقات رياضية مبسطة (بدون برهانها) . وحتى نستطيع تطبيق تلك العلاقات أعددنا الجدولين (6–8) و (6–9) واستخدمنا بياناتهما في عملية الحساب كما يلي:

نقوم بحساب SST من العلاقة التالية :

$$SST = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r}\sum_{i=1}^{n}(x_{k\ell mi}-\bar{X})^2 = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r}\sum_{i=1}^{n}x_{k\ell mi}^2 - \frac{X^2}{g*q*r*n} \qquad (45-6)$$

ومن الجدولين (6–8) و(6–9) نجد أن :

$$SST = 265174 - \frac{(3942)^2}{2*2*3*5} = 265174 - 258989.4 = 6184.6$$

ومن السطر الأخير في الجدول (6–8) نجد مباشرة أن :

$$SSE = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r} SSE_{k\ell mi} = 614$$

ثم نقوم بحساب SSA من بيانات الجدول (6–9) وباستخدام العلاقة التالية:

$$SSA = \frac{1}{qrn}\sum_{k=1}^{2}X_k^2 - \frac{X^2}{gqrn} = \frac{1}{30}[(2228)^2 + (2044)^2] - 258989.4 = 4403.3 \qquad (46-6)$$

جدول (6–8) نتائج الحسابات:

النوع K	نوع المادة A_1						نوع المادة A_2						
المساحة ℓ	المساحة B_1			المساحة B_2			المساحة B_1			المساحة B_2			
السماكة m	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	المجموع
i=1	67	72	78	67	79	78	54	52	63	54	57	60	
2	66	67	81	71	80	78	51	56	54	56	58	68	
3	62	75	67	72	81	77	47	52	65	58	61	61	
4	71	70	75	70	80	83	51	52	62	51	59	61	
5	69	71	75	81	85	79	59	53	60	57	55	67	
$X_{k\ell m}$ المجموع	335	355	377	361	405	395	362	265	304	276	290	311	3942
$\frac{X_{k\ell m}^2}{n}$	22445	25205	28425.6	26064.2	32803	31205	13728.8	14045	18483.2	15235.2	16820	200988	
$\sum x_{k\ell m}^2$	22491	25239	26535	26173	32827	31227	13808	14057	18554	15266	16840	20155	$T = 265174$
$SSE_{k\ell m}$	46	36	109.2	1196	22	22	79.2	12	70.8	30.8	20	57.2	$SSE = 614$

الجدول (6–9) مجاميع المجاميع حسب الحالات السابقة:

المجاميع	$m=1$	$m=2$	$m=3$	المجموع $X_{k\ell 1}$
X_{111}	335	355	377	= 1067
X_{121}	361	405	395	= 1161
X_{211}	362	265	304	= 831
X_{221}	276	290	317	= 883
$X_{k1}=\begin{cases}k=1\\k=2\end{cases}$	696 538	760 555	772 621	= 2228 = 2044
$X_{\ell=2}=\begin{cases}\ell=1\\\ell=2\end{cases}$	597 637	620 695	681 712	= 1898 = 2044
X_m	1234	1315	1393	$X=3942$

ثم نقوم بحساب SSB بناءً على الجدول (6–9) وباستخدام العلاقة التالية:

$$SSB=\frac{1}{grn}\sum_{\ell=1}^{2}X_\ell^2-\frac{X^2}{gqrn}=\frac{1}{30}[(1898)^2+(2044)^2]-258989.4=355.3 \qquad (47-6)$$

ثم نقوم بحساب SSC بناءً على الجدول (6–9) من العلاقة التالية :

$$SSC=\frac{1}{gqn}\sum X_m^2-\frac{X^2}{gqrn}=\frac{1}{20}[(1234)^2+(1315)^2-(1393)^2]-258989.4=632.1 \qquad (\mathbf{48-6})$$

ثم نقوم بحساب SS(AB) بناءً على العمود الأخير الجدول (6–9) من العلاقة التالية:

$$SS(AB)=\frac{1}{rn}\sum_{k=1}^{2}\sum_{\ell=1}^{2}X_{k\ell}^2-\frac{X^2}{gqrn}-SSA-SSB \qquad (49-6)$$

$$=\frac{1}{15}[(1067)^2+(1161)^2+(831)^2+(883)^2]-258989.4-4403.3-355.3=29.3$$

ثم نقوم بحساب SS(AC) بناءً على الجدول (6–9) من العلاقة التالية :

$$SS(AC)=\frac{1}{qn}\sum_{k=1}^{2}\sum_{m=1}^{2}X_{km}^2-\frac{X^2}{gqrn}-SSA-SSC \qquad (50-6)$$

$$=\frac{1}{10}[(696)^2+(760)^2+(772)^2+(538)^2+(555)^2+(621)^2]-258989.4-4403.3-632.1=68.2$$

ثم نقوم بحساب SS(BC) بناءً على الجدول (6–9) من العلاقة التالية:

$$SS(BC)=\frac{1}{gn}\sum_{\ell=1}^{2}\sum_{m=1}^{3}X_{\ell m}^2-\frac{X^2}{gqrn}-SSB-SSC \qquad (51-6)$$

$$=\frac{1}{10}[(597)^2+(620)^2+(681)^2+(637)^2+(695)^2+(712)^2]-258989.4-355.3-632.1=54$$

وأخيراً نحسب حد التداخل الثلاثي SS(ABC) من العلاقة (6–42) فنجد أن :

$$SS(ABC)=6184.6-614-4403.3-355.3-632.1-29.3-68.2-54=10.4$$

ثم نقوم بوضع نتائج هذه الحسابات في جدول تحليل التباين ذي الاتجاهات الثلاثة المرفقة بدرجات الحرية كما يلي:

جدول (6–10): جدول تحليل التباين بثلاث اتجاهات :

مصدر التباين	رمز المصدر	درجة الحرية	مجموع المربعات	متوسط المربعات	$\tilde{F}$	$F(\propto)$
F_A نوع المادة	SSA	$g-1=1$	4403.3	4403.3	344	4.04
F_B المساحة	SSB	$q-1=1$	355.3	355.3	27.8	4.04
F_C السماكة	SSC	$r-1=2$	632.1	316.0	24.7	3.19
أثر المادة والمساحة	$SS(AB)$	$(g-1)(q-1)=1$	29.3	29.3	2.29	4.04
أثر المادة والسماكة	$SS(AC)$	$(g-1)(r-1)=2$	86.2	43.1	3.37	3.19
أثر المساحة والسماكة	$SS(BC)$	$(q-1)(r-1)=2$	54.0	27.0	2.11	3.19
أثر العوامل الثلاثة المادة والمساحة والسماكة	$SS(ABC)$	$(g-1)(q-1)(r-1)=2$	10.4	5.2	0.41	3.19
الخطأ العشوائي (البواقي)	SSE	$gqr(n-1)=48$	614.0	12.79	–	–
الإجمالي	SST	$(gqrn)-1=59$	6184.6	–	–	–

وبعد حساب قيم F لجميع هذه المجاميع نقارنها مع قيمة $F(\propto)$ الحرجة المقابلة لها، والمبينة في العمود الأخير من الجدول (6–10) فنلاحظ ما يلي:

1– إن تأثير التداخل الثلاثي غير معنوي لأن $F_{ABC}=0.41<3.19$.

2– إن تأثير التداخل الثنائي SS(AC) معنوي وهذا يؤثر على تفسير النتائج .

3– إن تأثير التداخلين الثنائيين(SS(AB و SS(BC) غير معنويين .

4– إن تأثير العوامل الثلاثة F_A و F_B و F_C معنوية وأن العامل F_A هو الأكثر تأثيراً على تغيرات المقاومة .

6–4: تحليل المربع اللاتيني (بمشاهدة واحدة):

لقد لاحظنا في الفقرة السابقة أن تحليل التباين بثلاث اتجاهات (و n مشاهدة) يحتاج إلى حسابات معقدة وإلى عدد كبير من المشاهدات . ولكن يمكننا في بعض الأبحاث تنظيم وترتيب العوامل الداخلة في تحليل التباين الثلاثي وعرضها على شكل مربع يسمى المربع اللاتيني . وهو يستخدم كثيراً في الأبحاث العلمية كالأبحاث الزراعية والطبية والاقتصادية وغيرها . وهو يتألف من عدة عناصر هي :

- العامل الأول F_A ويأخذ g حالة منفصلة .
- العامل الثاني F_B ويأخذ أيضاً g حالة منفصلة .
- المربع اللاتيني الذي يتم تشكيله من تقاطع حالات العاملين F_A و F_B . وهو يتألف من g^2 خلية .

- جملة من طرائق المعالجة وعددها يساوي g طريقة أيضاً وذلك لتطبيقها على خلايا المربع اللاتيني، بحيث يتم تطبيق كل طريقة مرة واحدة في كل سطر، ومرة واحدة في كل عمود .
- متحول تابع X يقيس نتائج المعالجات السابقة في جميع الخلايا بواحدة قياس موحدة .

ولتمثيل أحد أشكال المربع اللاتيني ونفترض أن عدد حالات F_A و F_B يساوي $g = 3$ ، وأنه لدينا (3) طرائق هي $A\ B\ C$ لتطبيقها على تقاطعات تلك الحالات، مرة واحدة في كل سطر ومرة واحدة في كل عمود، وعندها يمكن أن يأخذ المربع اللاتيني الشكل التالي :

	F_{A1}	F_{A2}	F_{A3}
F_{B1}	A	B	C
F_{B2}	B	C	A
F_{B3}	C	A	B

الشكل (6- 2)

وهنا نلاحظ أن الشكل السابق (6- 2) للمربع اللاتيني هو أحد الأشكال الممكنة، وتم الحصول عليه بتطبيق تسلسل معين للطرائق المستخدمة هو $(A\ B\ C)$ على خلايا السطر الأول، ثم القيام بسحب عناصره إلى اليسار خطوة واحدة وتطبيقها على خلايا السطر الثاني (مع نقل A إلى الحجرة الأخيرة) . ثم سحب عناصره إلى اليسار خطوة واحدة وتطبيقها على خلايا السطر الثالث (مع نقل B إلى الحجرة الأخيرة) . وبذلك نحصل على ما يسمى الشكل القياسي للمربع اللاتيني، وهو المربع الذي يتم تشكيله بتدوير عناصر السطر الأول لتوزيعها على السطر الثاني ثم تدوير عناصر الثاني لتوزيعها على الثالث، وهكذا دواليك . وبذلك تكون عناصر الأسطر متناظرة مع عناصر الأعمدة. ولكن الشكل القياسي للمربع اللاتيني ليس وحيداً ،فهناك عدد من الأشكال التي تعطينا توزيعات مختلفة ، حيث نجد أنه يمكننا ترتيب عناصر السطر الأول عشوائياً بـ !3 طريقة ، ثم ترتيب العناصر المتبقية في العمود الأول عشوائياً (بعد السطر الأول) بـ !2 طريقة وتبقى الخلية الأخيرة والتي ترتب بطريقة واحدة فقط . وبذلك يكون عدد الأشكال الممكنة (حسب الأسطر أو الأعمدة) في الحالة التي يكون فيها $n = 3$ مساوياً لـ: لـ $m = 3! * 2! * 1 = 12$ ، وهذه الأشكال الممكنة تعطينا المربعات اللاتينية التالية :

$$\begin{bmatrix} A & B & C \\ B & C & A \\ C & A & B \end{bmatrix} \begin{bmatrix} A & C & B \\ C & B & A \\ B & A & C \end{bmatrix} \begin{bmatrix} B & A & C \\ A & C & B \\ C & B & A \end{bmatrix} \begin{bmatrix} B & C & A \\ C & A & B \\ A & B & C \end{bmatrix} \begin{bmatrix} C & A & B \\ A & B & C \\ B & C & A \end{bmatrix} \begin{bmatrix} C & B & A \\ B & A & C \\ A & C & B \end{bmatrix}$$

$$\begin{bmatrix} A & B & C \\ C & A & B \\ B & C & A \end{bmatrix} \begin{bmatrix} A & C & B \\ B & A & C \\ C & B & A \end{bmatrix} \begin{bmatrix} B & A & C \\ C & B & A \\ A & C & B \end{bmatrix} \begin{bmatrix} B & C & A \\ A & B & C \\ C & A & B \end{bmatrix} \begin{bmatrix} C & A & B \\ B & C & A \\ A & B & C \end{bmatrix} \begin{bmatrix} C & B & A \\ A & C & B \\ B & A & C \end{bmatrix}$$

ولكننا في الأبحاث العلمية لا نحتاج إلى جميع هذه الأشكال. بل نقوم باختيار أحدها عشوائياً . وبذلك تكون نتيجة التجارب (طرائق المعالجة) في كل خلية هي عبارة عن متحول عشوائي معرف عليها .

وهكذا نجد أن المربع اللاتيني هو تحليل ثلاثي الاتجاه (بمشاهدة واحدة) وإن العوامل المعتمدة فيه هي :

- عامل الأسطر ونرمز له بـ F_A وعدد حالاته g حالة .
- عامل الأعمدة ونرمز له بـ F_B وعدد حالاته g حالة أيضاً .
- طرائق المعالجة وعددها g حالة أيضاً. وهي تطبق على خلايا المربع اللاتيني، بحيث يتم تطبيق كل طريقة مرة واحدة في كل سطر ومرة واحدة في كل عمود، وحسب أحد الأشكال الممكنة للمربع.

ونتيجة لتطبيق هذه الطرائق على الخلايا نحصل على g^2 قياساً للمتحول التابع X، وسنرمز لقيمة المتحول التابع X في الخلية (k,ℓ) بالرمز $x_{k\ell}$ ونضعها إلى جانب رمز الطريقة المطبقة في تلك الخلية، ثم نحسب مجاميع ومتوسطات الأسطر والأعمدة ونضعها في الهوامش.

وإذا أخذنا الحالة التي يكون لدينا فيها: $(g=4)$ طرائق ويكون لكل من العاملين F_A و F_B أربعة حالات، فإننا سنحصل على الجدول التالي:

جدول (6-11): المربع اللاتيني 4×4

الأعمدة F_A / الأسطر F_B	1	2	3	4	المجموع X_k	المتوسط $\bar{X}_k$
1	A / x_{11}	B / x_{12}	C / x_{13}	D / x_{14}	X_1	$\bar{X}_1$
2	B / x_{21}	C / x_{22}	D / x_{23}	A / x_{24}	X_2	$\bar{X}_2$
3	C / x_{31}	D / x_{32}	A / x_{33}	B / x_{34}	X_3	$\bar{X}_3$
4	D / x_{41}	A / x_{42}	B / x_{43}	C / x_{44}	X_4	$\bar{X}_4$
المجموع X_ℓ	X'_1	X'_2	X'_3	X'_4	X	$\bar{X}$
المتوسط $\bar{X}_\ell$	$\bar{X}'_1$	$\bar{X}'_2$	$\bar{X}'_3$	$\bar{X}'_4$	$\bar{X}$	$\bar{X}$

حيث أن هذه المجاميع والمتوسطات تحسب من العلاقات التالية :

المتوسط في السطر k: $\bar{X}_k = \frac{1}{n}\sum_{\ell=1}^{g} x_{k\ell} = \frac{X_k}{g}$ اجمالي السطر k $X_k = \sum_{\ell=1}^{g} x_{k\ell}$

(52 المتوسط في العمود ℓ: $\bar{X}_\ell = \frac{1}{n}\sum_{k=1}^{g} x_{k\ell} = \frac{X_\ell}{g}$ اجمالي العمود ℓ $X_\ell = \sum_{k=1}^{g} x_{k\ell}$

المتوسط في الجدول ككل : $\bar{X} = \frac{X}{g*g} = \frac{\sum^g \sum^g x_{k\ell}}{g*g}$ ، الاجمالي الكلي $X = \sum_{k=1}^{g} X_k = \sum_{k=1}^{g} X_\ell = \sum_{k=1}^{g}\sum_{\ell=1}^{g} x_{k\ell}$

(6 −

وهناك متوسطات من نوع آخر هي متوسطات قياسات X حسب طرائق المعالجة $(A\ B\ C\ D)$ ويتم حساب هذه المتوسطات لكل طريقة على حدة، وذلك بتتبع قيم X حسب كل طريقة ضمن المربع

اللاتيني، ونرمز لهذه المتوسطات بالرموز $\bar{X}_t$، حيث t هو دليل الطريقة t، فمثلاً نجد أن متوسط قيم X المقابلة للطريقة A يحسب من الخلايا التي تطبق فيها الطريقة A وهو يساوي (حسب الجدول السابق (11-6)) مايلي:

$$\bar{X}_A = \frac{1}{4}[x_{11} + x_{24} + x_{33} + x_{42}] = \frac{X_A}{g}$$

وكذلك نجد أن متوسطات الطرائق الأخرى تساوي :

$$\bar{X}_B = \frac{1}{4}[x_{12} + x_{21} + x_{34} + x_{43}] = \frac{X_B}{g}$$

$$\bar{X}_C = \frac{1}{4}[x_{13} + x_{22} + x_{31} + x_{44}] = \frac{X_C}{g} \qquad (53-6)$$

$$\bar{X}_D = \frac{1}{4}[x_{14} + x_{23} + x_{32} + x_{41}] = \frac{X_D}{g}$$

ويمكن تسهيل حساب هذه المتوسطات بتنظيم جدول خاص بذلك، يتضمن تبويب قياسات X حسب طرائق المعالجة، وحسب الأسطر (أو حسب الأعمدة) كما يلي:

جدول (6-12): حساب المتوسطات حسب طرائق المعالجة

الطريقة t / الأسطر	A	B	C	D	X_i
1	x_{11}	x_{12}	x_{13}	x_{14}	X_1
2	x_{24}	x_{21}	x_{22}	x_{23}	X_2
3	x_{33}	x_{34}	x_{31}	x_{32}	X_3
4	x_{42}	x_{43}	x_{33}	x_{41}	X_4
المجموع	X_A	X_B	X_C	X_D	X
المتوسط $\bar{x}_t$	$\bar{X}_A$	$\bar{X}_B$	$\bar{X}_C$	$\bar{X}_D$	

وهنا يشترط على هذه القياسات في كل عمود (أو سطر) ان تكون مستقلة عن بعضها البعض ، وخاضعة للتوزيع الطبيعي ولها تباين موحد مساوٍ لـ σ^2 .

ولنفترض الآن أن التوقع الرياضي لقيم X في الخلية (k , ℓ) من المربع يساوي $\mu_{k\ell}$، وأن التوقع العام لقيم X في كامل المربع بالرمز μ . وبما أن القياسات مستقلة فإن التفاعلات الثنائية للعوامل تكون معدومة، كما أن التفاعل الثلاثي للعوامل الثلاثة معاً يكون معدوماً أيضاً .

وبناء على ذلك يمكننا أن نصيغ نموذج تحليل المربع اللاتيني لهذه العوامل الثلاثة المستقلة كما يلي:

$$\mu_{k\ell} = \mu + \propto_k + \beta_\ell + \gamma_t \qquad (54-6)$$

حيث أن: $\propto_k$ هو تأثير السطر k و β_ℓ تأثير العمود ℓ، و γ_t تأثير الطريقة t، وبما أن المتوسط العام في العينات هو تقدير لـ μ في المجتمع، فإنه يمكننا التعبير عن قيمة X في كل خلية كما يلي:

$$x_{k\ell} = \bar{X} + \propto_k + \beta_\ell + \gamma_t + e_{k\ell} \qquad (54\,a-6)$$

حيث أن: $e_{k\ell}$ هو حد الخطأ العشوائي ويخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$، وهكذا نجد أن مجموع مربعات الانحرافات عن μ يساوي :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\mu)^2 \qquad (55-6)$$

ولإظهار تأثير العوامل الثلاثة نستبدل μ بتقديره $\bar{X}$ ثم نضيف ونطرح المتوسطات $\bar{X}_k$ و $\bar{X}_\ell$ و $\bar{X}_\ell$ ونكتب المجموع السابق كما يلي ($\bar{X}_t$ متوسط الطريقة t) :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X})^2=\sum_{k=1}^{g}\sum_{\ell=1}^{g}[(x_{k\ell}-\bar{X}_k-\bar{X}_\ell-\bar{X}_t+2\bar{X})+(\bar{X}_k-\bar{X})+(\bar{X}_\ell-\bar{X})+(\bar{X}_t-\bar{X})]^2 \quad \mathbf{(56-6)}$$

ثم نقوم بتربيع الحدود التي داخل القوس المتوسط، ولكن نظراً لاستقلال العوامل المستخدمة في التحليل فإن مجاميع الجداءات الثنائية تكون معدومة (يمكن البرهان على ذلك لكل حد على حدة . انظر المرجع . (Dugue P.282

ونتيجة بعض الإصلاحات نحصل على أن مجموع مربعات الانحرافات الإجمالي يساوي :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X})^2=g\sum_{k=1}^{g}(\bar{X}_k-\bar{X})^2+g\sum_{\ell=1}^{g}(\bar{X}_\ell-\bar{X})^2+g\sum_{t=1}^{g}(\bar{X}_t-\bar{X})^2+$$
$$+\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X}_k-\bar{X}_\ell-\bar{X}_t+2\bar{X})^2 \qquad \mathbf{(57-6)}$$

وباستخدام نفس الرموز السابقة نكتب المجاميع السابقة كما يلي :

$$SST=SSA+SSB+SSt+SSE \qquad (58-6)$$

وإن درجات الحرية التي تقابلها تساوي ما يلي :

$$g^2-1=(g-1)+(g-1)+(g-1)+(g-1)\,(g-2) \qquad (59-6)$$

وبعد حساب هذه المجاميع نضعها في جدول تحليل التباين الثلاثي كالتالي :

جدول (6-13): تحليل التباين في المربع اللاتيني :

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط المربعات	F المحسوبة
F_A	$SSA=$	$g-1$	$MSA=\dfrac{SSA}{g-1}$	$F_A=\dfrac{MSA}{MSE}$
F_B	$SSB=$	$g-1$	$MSB=\dfrac{SSB}{g-1}$	$F_B=\dfrac{MSB}{MSE}$
طرائق المعالجة t_r	$SSt_r=$	$g-1$	$MSt=\dfrac{SSt_r}{g-1}$	$F_t=\dfrac{MSt}{MSE}$
الخطأ العشوائي	$SSE=$	$(g-1)\,(g-2)$	$MSE=\dfrac{SSE}{(g-1)\,(g-2)}$	
الاجمالي	SST	g^2-1	——	

أما بالنسبة للفرضيات فهي تنطلق من النموذج (6−54) ونكتبها كما يلي:

$$H_0: \begin{cases} \propto_k = 0 & : K : 1\ 2\ 3 \dots g \\ \beta_\ell = 0 & : \ell : 1\ 2\ 3 \dots g \\ \gamma_t = 0 & : t : 1\ 2\ 3 \dots g \end{cases} \qquad (60-6)$$

أما الفرضية البديلة H_1 فتكتب كما يلي:

$$H_1: \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_t \neq 0 & \text{من أجل } t \text{ واحدة على الأقل} \end{cases} \qquad (61-6)$$

وهنا نفترض أن تصميم المربع اللاتيني يعتبر أن تأثيرات الأسطر وتأثيرات الأعمدة ثابتة ، رغم أنها تستخدم للتحكم في مصادر الاختلاف . لذلك فإن تأثيراتها يجب أن تحقق العلاقتين التاليتين :

$$\sum_{k=1}^{g} \propto_k = 0 \qquad \sum_{\ell=1}^{g} \beta_\ell = 0 \qquad (62-6)$$

أما تأثيرات طرائق المعالجة γ_t فإما أن تكون ثابتة أو عشوائية

فإذا كانت γ_t ثابتة فإن تأثيراتها تقدر على أنها انحرافات عن المتوسط العام $\bar{X}$. ولذلك فإنها يجب أن تحقق العلاقة التالية :

$$\sum_{t=1}^{g} \gamma_t = 0 \qquad \left(\text{إذا كانت } \gamma_t \text{ ثابتة}\right) \qquad (63-6)$$

أما إذا كانت γ_t عشوائية فإننا نفترض ونتحقق من أنها خاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$، ثم نتعامل معها بأسلوب آخر (اختبار Tukey) . وهنا نشير إلى أن الفرضية الأساسية التي نريد اختبارها هي فرضية العدم التالية :

$$H_0: \gamma_1 = \gamma_2 = \cdots \gamma_g = 0 \qquad (64-6)$$

مقابل الفرضية البديلة :

$$H_1: \gamma_t \neq 0 \qquad \text{من أجل } t \text{ واحدة على الأقل}$$

ولاختبار الفرضية $H_0: \gamma_1 = \gamma_2 = \cdots \gamma_g = 0$ نقوم بحساب مجاميع مربعات الانحرافات ومتوسطاتها ونضعها في جدول تحليل التباين، ونبدأ بحساب مؤشر الاختبار لتأثير المعالجات من العلاقة :

$$F_t = \frac{\dfrac{SSt}{g-1}}{\dfrac{SSE}{(g-1)(g-2)}} = \frac{MSt_r}{MSE} \qquad (65-6)$$

ثم نقارن هذه القيمة المحسوبة F_t مع القيمة الحرجة لـ F والمقابلة لمستوى الدلالة $\propto = 0.05$ ولدرجتي الحرية $v_1 = (g-1)$ ، $v_2 = (g-1)(g-2)$ ونتخذ القرار كما يلي :

إذا كانت $F_t \leq F_{v_1\,v_2}(\propto)$ نقبل الفرضية H_0، والتي تقول أن تأثيرات طرائق المعالجة معدومة أو إنها غير معنوية والعكس بالعكس .

أما بالنسبة لاختبار تأثير العاملين F_A و F_B نقوم بحساب قيمتي المؤشرين :

$$F_A = \frac{MSA}{MSE} \qquad F_B = \frac{MSB}{MSE} \qquad (66-6)$$

ثم نقوم بمقارنة كل منهما مع القيمة الحرجة $F_{v_1\,v_2}(\propto)$، ونقبل الفرضية H_0 حول $\propto_k = 0$ إذا كان $F_A \leq F(\propto)$، كما نقبل الفرضية H_0 حول $\beta_\ell = 0$ إذا كان $F_B \leq F(\propto)$ والعكس بالعكس .

ولكن المؤشرين F_A و F_B يعتبران مؤشرين تقريبيين، لذلك نحسب الكفاءة النسبية للأسطر وللأعمدة من العلاقتين التاليتين :

$$RE\ (\text{للأسطر}) = \frac{MSA + (g-1)MSE}{g * MSE} 100 \qquad (67-6)$$

$$RE\ (\text{للأعمدة}) = \frac{MSB + (g-1)MSE}{g * MSE} 100 \qquad (68-6)$$

ولتقدير الكفاءة الكلية للمربع اللاتيني نستخدم العلاقة :

$$RE\ (\text{للمربع}) = \frac{MSA + MSB + (g-1)MSE}{(g+1)MSE} \qquad (69-6)$$

ملاحظة إذا كانت قيمة إحدى المشاهدات $x_{k\ell}$ مفقودة، فإنه لا يمكننا تطبيق أسلوب المربع للاتيني بدونها، وحتى نتابع العمل علينا أن نقوم بتقديرها من العلاقة :

$$\tilde{x}_{k\ell} = \frac{g(X_k + X_\ell + X_t) - 2X}{(g-1)(g-2)} \qquad (70-6)$$

وهنا علينا أن نخفض درجات الحرية للأخطاء بمقدار درجة واحدة وتصبح كما يلي : $(g-1)(g-2) - 1$

وأن نخفض درجات الحرية لإجمالي المربعات درجة واحدة أيضاً فتصبح : $(g^2 - 2)$

ولتقدير الخطأ المعياري لتقدير المتوسط $\bar{x}_t$ نستخدم العلاقة :

$$S_{\bar{x}_t} = \sqrt{\frac{MSE}{g}} \qquad (71-6)$$

ثم ننشأ مجال الثقة له ذي الاحتمال $(1-\propto)$ من العلاقة:

$$\bar{x}_t \pm t_{(g-1)(g-2)}\left(\frac{\propto}{2}\right) * S_{\bar{x}_k} \qquad (72-6)$$

كما يمكننا حساب الخطأ المعياري للفرق بين متوسطي طريقتين $(\bar{x}_{t1} - \bar{x}_{t2})$ من العلاقة :

$$S_{(\bar{x}_{t1} - \bar{x}_{t2})} = \sqrt{\frac{MSE}{g} + \frac{MSE}{g}} = \sqrt{\frac{2MSE}{g}} \qquad (73-6)$$

ثم إنشاء مجال الثقة له ذي الاحتمال $(1-\propto)$ من العلاقة :

$$(\bar{x}_{t1}-\bar{x}_{t2}) \pm t_{(g-1)(g-2)}\left(\frac{\propto}{2}\right) * S_{(\bar{x}_{t1}-\bar{x}_{t2})} \qquad (74-6)$$

مثال (6-5): لدراسة أداء (6) عمال من عمال شركة النسيج، نرمز لهم بالرموز $(A\,,B\,,C\,,D\,,E\,,F)$ تم اعتماد المتحول التابع X الذي يعبر عن إنتاجية العامل في اليوم من القماش (م^2 / يوم) . ولإجراء التجارب تم تجهيز (6) آلات نسيج مختلفة، على أن يتم توزيع هؤلاء العمال على تلك الآلات عشوائياً في كل يوم ولمدة (6) أيام (أيام العمل في الأسبوع)، وبحيث يعمل كل عامل مرة واحدة على آلة مختلفة في كل يوم من ايام العمل في الأسبوع (مرة واحدة في كل عمود ومرة واحدة في كل سطر) .

والمطلوب دراسة تأثير كل من الأيام D والآلات M والعمال L على إنتاجية هؤلاء العمال، إذا علمت أن توزيعات هؤلاء العمال وإنتاجياتهم اليومية كانت بعد تنفيذ الدراسة كما يلي :

جدول (6-14): بيانات المربع اللاتيني للمثال (فرضية) :

الآلات M_ℓ / الأيام D_k	M_1	M_2	M_3	M_4	M_5	M_6	المجموع X_k	المتوسط $\bar{X}_k$
السبت D_1	B 28.7	E 28.4	D 25.4	C 30.7	A 30.6	F 30.9	$X_1 = 174.7$	29.12
الأحد D_2	F 31.4	C 30.1	B 27.4	E 26.8	D 29.8	A 29.8	$X_2 = 175.3$	29.22
الاثنين D_3	D 29.4	F 29.7	A 30.4	B 22.0	E 24.1	C 32.9	$X_4 = 163.0$	28.08
الثلاثاء D_4	A 29.6	B 21.8	E 22.5	F 30.0	C 39.6	D 28.5	$X_3 = 168.5$	27.17
الأربعاء D_5	C 25.8	D 21.9	F 23.1	A 24.3	B 20.7	E 17.7	$X_5 = 133.5$	22.25
الخميس D_6	E 18.1	A 23.6	C 22.5	D 20.2	F 23.7	B 18.9	$X_6 = 127.0$	21.17
المجموع X_ℓ	163.0	155.3	151.3	154.0	159.5	158.7	$X = 942.0$	—
المتوسط $\bar{X}_\ell$	27.17	25.88	25.22	25.67	26.58	26.45	—	$\bar{X} = 26.17$

ولحساب إنتاجية كل عامل على حدة علينا أن نتبع مقدار إنتاجيته على كل آلة وفي كل يوم . لذلك نتبع إنتاجية هؤلاء العمال حسب الآلات ونصمم جدولاً خاصاً لتبويبها من جديد حسب العمال والآلات فنحصل من الجدول (6-10) السابق على الجدول التالي :

جدول (6−15) إنتاجية العمال حسب الآلات :

العمال / الآلات	A	B	C	D	E	F	المجموع X_ℓ	المتوسط $\bar{X}_\ell$
M_1	29.6	28.7	25.8	29.4	18.1	31.4	163.0	27.17
M_2	23.6	21.8	30.1	21.9	28.4	29.7	155.5	25.88
M_3	30.4	27.4	22.5	25.4	22.5	23.1	151.3	25.22
M_4	24.3	22.0	30.7	20.2	26.8	30.0	154.0	25.67
M_5	30.6	20.7	30.6	29.8	24.1	23.7	159.5	26.58
M_6	29.8	18.9	32.9	28.5	19.7	30.9	158.7	26.45
X_t	168.3	139.5	172.6	155.2	137.7	168.8	942.0	——
$\bar{X}_t$	28.05	23.25	28.77	25.87	22.95	28.13	——	26.17

ومن السطر الأخير لهذا الجدول نلاحظ أن إنتاجية هؤلاء العمال تختلف من عامل لآخر . وإن أحسنها هي إنتاجية العمال C ثم F ثم A . وإن أسوأها هي إنتاجية العمال E ثم B ثم D . وكذلك يمكننا أن نبوب إنتاجية هؤلاء العمال حسب الأيام فنحصل من الجدول (6−14) على جدول مشابه للجدول (6−15) السابق . إلا أن حساب الجدول (6−15) يغني عنه في الحسابات اللاحقة . والآن نعتمد النموذج النظري التالي :

$$x_{k\ell} = \bar{X} \pm \propto_k + B_\ell + \gamma_t + e_{k\ell}$$

ثم نضع الفرضيات كما يلي:

$$H_0 : \begin{cases} \propto_1 = \propto_2 = \cdots = \propto_6 = 0 \\ B_1 = B_2 = \cdots = B_6 = 0 \\ \gamma_1 = \gamma_2 = \cdots = \gamma_6 = 0 \end{cases}$$

$$H_1 : \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ B_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_t \neq 0 & \text{من أجل } t \text{ واحدة على الأقل} \end{cases}$$

ولاختبار هذه الفرضيات نقوم بحساب مجاميع مربعات الانحرافات SST وSSA وSSB وSSt وSSE من بيانات الجدولين (6−14) و(6−15) ونستخدم العلاقات التالية :

$$SST = \sum_{k=1}^{6}\sum_{\ell=1}^{6}(x_{k\ell} - \bar{X})^2 = \sum_{k=1}^{6}\sum_{\ell=1}^{6} x_{k\ell}^2 - \frac{X^2}{g^2}$$

$$SST = [(28.7)^2 + (28.4)^2 + \cdots (23.7)^2 + (18.9)^2] - \frac{(942)^2}{36}$$

$$SST = 25299.10 - 24649 = 650.10$$

ثم نقوم بحساب مجموع المربعات للأيام D فنجد من مجاميع الأسطر في الجدول (6−14) أن:

$$SSA = \sum_{k=1}^{6} X_k^2 - \frac{X^2}{g^2} = [(174.7)^2 + (175.3)^2 + \cdots (127.0)^2] - \frac{(942)^2}{36} = 378.11$$

ثم نقوم بحساب مجموع المربعات للآلات فنجد من مجاميع الأعمدة في الجدول (6–14) أن:

$$SSB = \sum_{\ell=1}^{6} X_{\ell}^{2} - \frac{X^{2}}{g^{2}} = [(163.3)^{2} + (155.5)^{2} + \cdots (158.7)^{2}] - \frac{(942)^{2}}{36} = 14.81$$

ثم نقوم بحساب مجموع المربعات للعمال اعتماداً على أعمدة الجدول (6-15) فنجد أن:

$$SSt = \sum_{t=1}^{6} X_{t}^{2} - \frac{X^{2}}{g^{2}} = [(168.3)^{2} + (139.5)^{2} + \cdots (168.8)^{2}] - \frac{(942)^{2}}{36} = 199.36$$

ثم نقوم بحساب SSE من العلاقة :

$$SSE = SST - SSA - SSB - SSt$$

$$SSE = 650.10 - 378.4 - 14.81 - 199.36 = 57.82$$

وأخيراً نضع نتائج هذه الحسابات مع درجات حرياتها في جدول تحليل التباين الثلاثي فنحصل على أن :

جدول (6–16): تحليل التباين للمربع اللاتيني :

مصدر التباين	مجموع المربعات	درجة حرية	متوسط المربعات	قيمة F المحسوبة
الأسطر (الأيام)	$SSA = 378.11$	$g - 1 = 5$	75.62	$F_A = 26.16$
الأعمدة (الآلات)	$SSB = 14.8$	$g - 1 = 5$	2.96	$F_B = 1.02$
المعالجات (العمال)	$SSt = 199.36$	$g - 1 = 5$	39.87	$F_t = 13.79$
الأخطاء	$SSE = 57.82$	$(g-1)(g-2) = 20$	2.89	ــــــ
الإجمالي	$SST = 650.10$	$g^2 - 1 = 35$	ــــــ	ــــــ

ثم نقوم بإيجاد قيمة F الحرجة الموافقة لدرجتي حرية $v_1 = 5$, $v_2 = 20$ ولمستوى دلالة $\propto = 0.05$ فنجد أن :

$$F_{v_1,v_2}(\propto) = F_{5,20}(0.05) = 2.71$$

وهي نفسها لجميع العوامل، لذلك نقارن F_A و F_B و F_t معها، فنجد ما يلي :

1-بالنسبة للعمال نجد أن: $F_t > F(\propto)$، لذلك نرفض فرضية العدم $H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_6 = 0$ ونقبل الفرضية البديلة H_1 . التي تقول أن الإنتاجية X تتأثر بالعمال (وهذا أمر واضح من السطر الأخير في الجدول (6–15) .

2-بالنسبة للأيام (الأسطر) نجد أن $F_A > F(\propto)$، لذلك نرفض فرضية العدم $H_0: \propto_1 = \propto_2 = \cdots = \propto_6 = 0$ ونقبل الفرضية البديلة H_1. التي تقول أن الإنتاجية X تتأثر بالأيام (وهذا أمر واضح من العمود الأخير في الجدول (6–10) .

3-بالنسبة للآلات (الأعمدة) نجد أن $F_B < F(\propto)$ لذلك نقبل فرضية العدم $H_0: B_1 = B_2 = \cdots = B_6 = 0$ ونستنتج أن الآلات لا تؤثر على إنتاجية العمال .

ولتقدير كفاءة هذا التصميم نقوم بحساب الكفاءة النسبية للمربع ككل فنجد أن :

$$RE(Squere) = \frac{MSA + MSB + (g-1)MSE}{(g+1)MSE} 100 = \frac{75.62 + 2.96 + 5 * 2.89}{7 * 2.89} 100 = 459.86\%$$

وهذا يدل على كفاءة عالية بالنسبة للقطاعات العشوائية التامة .

كما يمكننا حساب الكفاءة النسبية للأسطر (بدون الأعمدة) من العلاقة :

$$RE(row) = \frac{MSA + (g-1)MSE}{g * MSE}100 = \frac{75.62 + 5 * 2.89}{6 * 2.89}100 = 519.43\%$$

وهذا يدل على كفاءة عالية بالنسبة للقطاعات العشوائية التامة .

وكذلك نقوم بحساب الكفاءة النسبية للأعمدة (بدون الأسطر) من العلاقة :

$$RE(column) = \frac{MSB + (g-1)MSE}{g * MSE}100 = \frac{2.96 + 5 * 2.89}{6 * 2.89}100 = 100.40\%$$

وهذا يدل على أن كفاءة الأعمدة (الآلات) بقيت تساوي نفسها 100% ولم تؤثر على إنتاجية العمال .

6-5: تحليل التباين المشترك (تحليل التغايرANCOVA):

6-5-1: تمهيد:

لقد استعرضنا في النماذج السابقة لتحليل التباين النماذج التي تدرس تغيرات متحول تابع X (Dependent variable)، الناتجة عن متغير أو متغيرات وصفية مستقلة (Independent)، مثل المجتمعات أو المعالجات أو القطاعات .

ولكن إذا كان التابع X مرتبطاً بمتحول كمي آخر Y غير مرغوب به، ولكنه ملازم لـ X ولا نستطيع التحكم فيه، ولكننا نريد التخلص من تأثيره على التابع X . فإننا نلجأ إلى إجراء تحليل التباين المشترك (التغاير ANCOVA) لفصل تأثيرات Y غير المرغوب فيها عن تغيرات X ونتبع في ذلك إحدى الطريقتين :

1- إزالة اختلافات X المرتبطة بـ Y من قياسات X (إذا كان ذلك ممكناً) وعندها نحصل على قياسات صافية لـ X . ثم نقوم بإجراء تحليل التباين على القياسات الصافية فنحصل على اختبارات قوية (ولكن هذه الطريقة قد لا تكون ممكنة في معظم الحالات) .

2- تعديل متوسطات المعالجات في المجتمعات، بحيث يتم طرح قيم موحدة للمتغير المستقل Y منها، وبالتالي نحصل على طريقة عادلة لمقارنة قيم متوسطات X في تلك المجتمعات المختلفة .

فمثلاً لدراسة تغيرات الانفاق الشهري X لطلاب الجامعة حسب الجنس (مجتمع الذكور ومجتمع الإناث). نلاحظ أن ذلك الانفاق لا يتأثر بنوع الجنس فقط، بل يتأثر أيضاً بمقدار الدخل الشهري المخصص للطالب والذي سنرمز له بـ Y . لذلك علينا أن نقوم بعزل تأثير الدخل Y على X . وذلك حتى نتمكن من إجراء مقارنة عادلة لمتوسطي الإنفاق X حسب الجنس دون تأثير الدخل Y عليهما .

وللتخلص من تأثيرات المتحول الإضافي Y على قيم X نلجأ إلى تحليل التباين المشترك (التغاير ANCOVA)، وإذا اجرينا تحليل التباين البسيط (ANOVA) بدون عزل تأثير Y، فإن ذلك سيؤدي

إلى تضخيم الخطأ التجريبي، ويصبح من الصعب اكتشاف الفروقات الحقيقية بين المجتمعات . وبذلك نجد أن المهمة الرئيسية لتحليل التغاير هي تصغير قيمة الخطأ التجريبي .

وتتضمن المراجع المختصة عدة أنواع لتحليل التغاير هي:

- تحليل التغاير في تصميم العشوائية التامة (CRD) .
- تحليل التغاير في تصميم القطاعات العشوائية (RCBD) .
- تحليل التغاير في تصميم التجارب العاملية 2^k (2^kFD) .

وسنقتصر في منشورنا هذا على النوع الأول لأنه الأكثر انتشاراً والأسهل تطبيقاً .

6-5-2: تحليل التغاير في تصميم العشوائية التامة (CRD) باتجاه واحد (ANCOVA one way) .

يمكن النظر إلى تحليل التغاير على أنه تعديل لتحليل التباين البسيط باتجاه واحد (ANOVA)، وذلك بعد إضافة المتحول Y إلى النموذج الرياضي ثم إعادة صياغة النموذج بحيث يتم عزل تأثير Y منه .

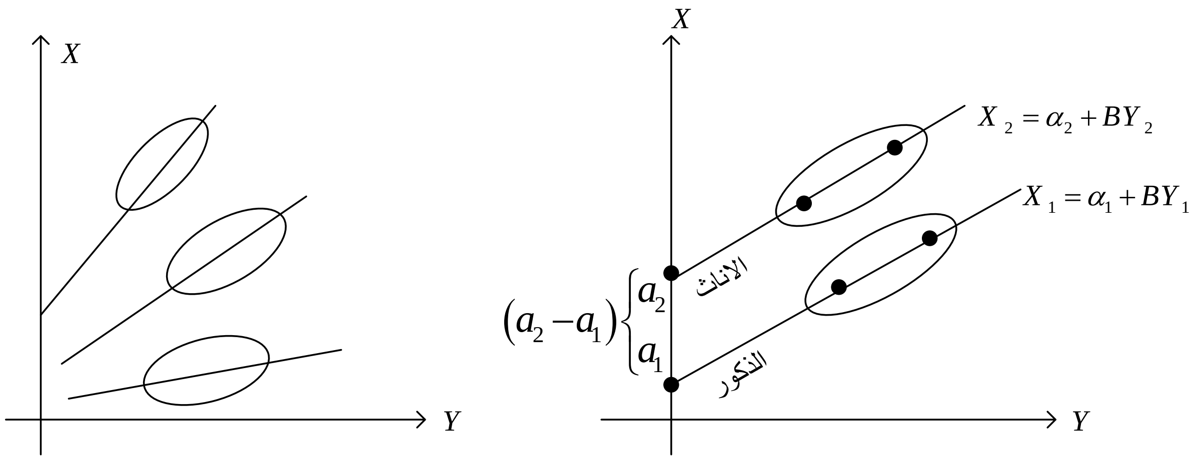
وبصورة عامة نفترض أنه لدينا g مجتمعاً، تؤثر على متحول تابع X ، وأن X يترافق مع متحول إضافي كمي Y . وأن X يرتبط مع Y في كل مجتمع k بعلاقة انحدار خطية من الشكل:

$$X_k = a_k + \beta_k Y_k = a_k + \beta Y_k \qquad : k: 1\ 2\ 3 \ldots g \qquad (75-6)$$

ولتسهيل صياغة النموذج الرياضي المشترك نفترض أن قيم الأمثال β_k متساوية في جميع تلك المجتمعات، أي إننا نفترض أو نشترط أن يكون :

$$\beta_1 = \beta_2 \ldots = \beta_g = \beta \qquad (75\,a-6)$$

وإن هذا يعني أن ميول مستقيمات هذه العلاقات متساوية في جميع المجتمعات المدروسة، وترسم الشكل اليميني التالي :



حالة الأمثال المتساوية $\beta_1 = \beta_2 = \beta$ حالة الأمثال غير متساوية

الشكل (6-3): حالات علاقة x بـ Y

كما نفترض أن حجوم العينات المسحوبة من تلك المجتمعات متساوية وتساوي $n_k = r$،ومن جهة ثانية نجد أنه يمكننا صياغة نموذج تحليل التباين البسيط (ANOVA) للمتحول التابع X حسب (6–7) كما يلي:

$$x_{ki} = \mu_x + \propto'_k + e'_{ki} \qquad (76-6)$$

حيث أن: $k = 1\ 2\ 3 \ldots g$ و $i = 1\ 2\ 3 \ldots r$

وحيث أن: μ_x هو توقع X، وإن: $\propto'_k$ هو مقدار تأثير المجتمع k على X، وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto'_k = 0$.

وأن: e'_{ki} هي حدود الخطأ العشوائي (البواقي) ويشترط فيها أن تكون مستقلة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$.

ومن جهة ثالثة نجد أنه يمكننا صياغة النموذج الرياضي لتحليل التباين (ANOVA) للمتحول الإضافي Y حسب (6–7) كما يلي:

$$y_{ki} = \mu_y + \propto''_k + e''_{ki} \qquad (77-6)$$

حيث أن: $k = 1\ 2\ 3 \ldots g$ و $i = 1\ 2\ 3 \ldots r$

وحيث أن: μ_y هو توقع Y، وإن: $\propto''_k$ هو مقدار تأثير المجتمع k على Y، وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto''_k = 0$.

وأن: e''_{ki} هي حدود الخطأ العشوائي (البواقي) ويشترط فيها أن تكون مستقلة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$.

وبعدها نضرب طرفي العلاقة (6–77) بقيمة الأمثال β (بعد حسابها من(6–75))، ونطرح أطرافها من أطراف العلاقة (6–76) فنحصل على ما يلي:

$$(x_{ki} - \beta y_{ki}) = \left(\mu_x - \beta\mu_y\right) + (\propto'_k - \beta \propto''_k) + (e'_{ki} - \beta e''_{ki}) \qquad (78-6)$$

وإذا رمزنا للطرف الأيسر بـ $Z_{ki} = \left(x_{ki} - \beta y_{ki}\right)$ وأخذنا توقعه نجد أن:

$$E(Z_{ki}) = E(x_{ki}) - \beta E(y_{ki}) = \mu_x - \beta\mu_y = \mu_z \qquad (79-6)$$

وهكذا يظهر لدينا متحول جديد هو: $Z = X - \beta Y$

وأن توقعه: $\mu_z = \mu_x - \beta\mu_y$

وإذا رمزنا للحدود الأخرى في (6–78) بالرموز التالية:

$\propto_k = \propto'_k - \beta \propto''_k$ و $e_{ki} = e'_{ki} - \beta e''_{ki}$ فإننا نحصل على النموذج التالي:

$$Z_{ki} = \mu_z + \propto_k + e_{ki} \qquad (80-6)$$

حيث أن: $k = 1\ 2\ 3 \ldots g$ و $i = 1\ 2\ 3 \ldots r$ وهو نموذج تحليل التباين البسيط لـ Z .

حيث أن: μ_z هو توقع المتحول Z، وأن: $\propto_k$ هو تأثير المجتمع k على Z وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto_k = 0$.

وأن e_{ki}: هي حدود الخطأ العشوائي الجديدة . وهي حدود مستقلة وتخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$.

وهكذا نكون قد توصلنا إلى نفس الصيغة الرياضية التي يأخذها نموذج ANOVA للمتحول الجديد Z.
ولإظهار المتحول Y في النموذج (6–80) نعيد الرموز إلى أصولها وندمج الحدود المتشابهة. فنجد أن العلاقة (6–80) تأخذ الشكل التالي:

$$(x_{ki} - \beta y_{ki}) = (\mu_x - \beta\mu_y) + \propto_k + e_{ki}$$

وبعد الإصلاح نحصل على الصيغة التي تعطينا أية قيمة x_{ki} كما يلي:

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki} \qquad (81-6)$$

حيث أن: $k = 1\ 2\ 3 \ldots g$ و $i = 1\ 2\ 3 \ldots r$.

وهي صيغة نموذج تحليل التغاير ANCOVA للمتحول X المرتبط بالمتحول الإضافي Y. وهي الصيغة الأكثر انتشاراً .

ويمكننا الحصول على صيغة أخرى لـ (6–81) إذا قمنا بكتابتها كما يلي:

$$x_{ki} = (\mu_x + \beta\mu_y) + \propto_k + \beta y_{ki} + e_{ki}$$

وإذا رمزنا للمقدار $\mu' = \mu_x + \beta\mu_y$ نحصل على الصيغة التالية :

$$x_{ki} = \mu' + \propto_k + \beta y_{ki} + e_{ki} \qquad (82-6)$$

حيث أن: $k = 1\ 2\ 3 \ldots g$ و $i = 1\ 2\ 3 \ldots r$ ، وحيث أن: μ' هو مقدار ثابت يعبر عن توقع متحول ثالث وليس عن توقع X، لأن توقع X يساوي $\mu_x = \mu' + \beta\mu_y$.

وهكذا نكون قد توصلنا إلى نموذج تحليل التغاير للمتحول X المرتبط بالمتحول الإضافي Y، وهو يأخذ إحدى الصيغتين المتكافئتين (6–81) و(6–82) .

وأخيراً نشير إلى أنه لتطبيق هذا النموذج يشترط على عناصره أن تحقق افتراضات تحليل التباين ANOVA وافتراضات الانحدار البسيط ونلخصها بما يلي:

- الافتراضات على نموذج ANCOVA :

1– أن تكون العينات المسحوبة عشوائية ومستقلة، وأن تكون حجومها متساوية وتساوي: $n_1 = n_2 = \cdots = n_g = r$.

2– أن يكون X وY متحولان كميان وخاضعان في كل مجتمع k للتوزيع الطبيعي بتوقعين μ_{kx} و μ_{ky} على الترتيب .

3– أن تكون تباينات التابع X في جميع المجتمعات متساوية وتساوي σ^2 .

4– أن تكون قياسات المتحول الإضافي Y نهائية، ولا تتضمن أخطاءً في القياس ولا تتأثر بالمجتمعات أو المعالجات .

5– أن تكون العلاقات بين X وY في جميع المجتمعات خطية وتأخذ الشكل التالي: $X_k = a_k + \beta_k Y_k$ وتكون $\beta_k \neq 0$.

6- أن تكون قيم جميع الأمثال β_k متساوية في جميع المجتمعات المدروسة أي $\beta_k = \beta$، أي أن تكون المستقيمات متوازية كما في الشكل (6-3) السابق .

7- أن يكون مجموع تأثيرات المجتمعات على X معدوماً . أي أن يكون: $\sum_{k=1}^{g} \propto_k = 0$.

8- أن تكون الأخطاء e_{ki} أو البواقي الناتجة عن النموذج مستقلة وخاضعة للتوزيع الطبيعي بتوقع معدوم وتباين ثابت σ^2 أي لـ $N(0\,,\sigma^2)$.

ونلاحظ من هذه الافتراضات أن الافتراض الرابع ينص على أن المتغير الإضافي Y لا يتأثر بالمجتمعات (المعالجات) وهو أهم افتراض في تحليل التغاير ، لأنه إذا كان Y يتأثر بالمجتمعات، فإن معاملات الانحدار β_k ستختلف من مجتمع لآخر، وهذا يؤدي إلى عدم ثبات قيم تلك المعاملات في المجتمعات المختلفة، وهذا يخل في الافتراض السادس $(\beta_k = \beta)$ ويجعل النموذج (6-81) غير صالح للتطبيق على تلك المجتمعات .

ونستنتج مما سبق عند تطبيق ANCOVA أنه يجب علينا قبل كل شيء التحقق من أن Y لا يتأثر بالمجتمعات، وذلك بإجراء تحليل التباين ANOVA على Y بمفرده، فإذا كانت النتيجة قبول فرضية العدم H_0 (لا توجد فروقات لـ Y بين المجتمعات)، فإننا نتابع العمل للتحقق من الافتراضين الخامس $(\beta_k \neq 0)$ والسادس $(\beta_k = \beta)$ ، ولذلك يجب علينا أن نقوم بإيجاد معاملات العلاقات الخطية بين X Y في جميع المجتمعات $X_k = a_k + \beta_k Y_k$ بطريقة المربعات الصغرى، ثم القيام باختبار عدم وجود فروقات معنوية بين المعاملات β_k وذلك بوضع الفرضيتين كما يلي:

$H_0 :\ \beta_k = \beta_0$ $\qquad : k = 1\ 2\ 3 \ldots g$

$H_1 :\ \beta_k \neq \beta_0$ $\qquad$ على الأقل من أجل قيمة واحدة لـ k

حيث أن β_0 هي قيمة افتراضية يضعها الباحث أو يحسبها من متوسط المعاملات β_k، وللتحقق من H_0 نقوم بحساب قيمة اختبار (ستودينت) t المعرف بالعلاقة :

$$t_k = \frac{\beta_k - \beta_0}{S(\beta_k)} \qquad : k = 1\ 2\ 3 \ldots g \qquad (83-6)$$

حيث $S(\beta_k)$ هو الانحراف المعياري للمعامل β_k .

ثم نقوم بمقارنة قيمة t_k المحسوبة مع قيمة t الحرجة والمقابلة لـ $(n-1)$ درجة حرية ولنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$.

فإذا كانت $|t_k| \leq t_{n-1}\left(\frac{\propto}{2}\right)$ فإننا نقبل H_0 ونعتبر المعاملات β_k متساوية باحتمال 0.95 وبذلك يكون الافتراض السادس محققاً .

ولاختبار تحقق الافتراض الخامس نطبق الاختبار t على الفرضيتين التاليتين كما يلي:

$H_0 :\ \beta_0 = 0 \qquad H_1 :\ \beta_0 \neq 0$

ونطبق نفس المؤشر المعرف في (6–83) على جميع β_k بطريقة إعادة الاختبار. فإذا كانت النتيجة هي رفض H_0 من أجل جميع β_k، فإننا نعتبر أن $\beta_0 \neq 0$ وأنه يوجد علاقة خطية معنوية بين X و Y في تلك المجتمعات، ويمكن التحقق من الافتراض الخامس عن طريق معامل الارتباط r أو معامل التحديد R^2 ...الخ .

وبناءً على ذلك يتم إجراء تحليل التغاير ANCOVA للتابع X المرتبط بمتحول إضافي Y. وذلك ضمن تحقق الافتراضات المذكورة أعلاه .

وبعد هذه المقدمة ننتقل إلى إجراء التحليل اللازم للنموذج (6–81) ونضع الفرضيتين الخاصتين به (العدم والبديلة) على الشكل التالي :

$$H_0 : \ltimes_1 = \ltimes_2 = \ltimes_3 = \cdots = \ltimes_g = 0$$

$$H_1 : \ltimes_k \neq 0 \quad \text{من أجل } k \text{ واحدة على الأقل} \qquad (84-6)$$

ولاختبار الفرضيات (6–84) علينا أن نقوم بحساب مجاميع مربعات الانحرافات لكل من المتحولين X و Y ولجدائهما $X * Y$ ، وذلك باستخدام العلاقات الرياضية التالية (مع الانتباه هنا إلى أن r هو الحجم الموحد للعينات المسحوبة من تلك المجتمعات و g عدد المجتمعات) :

بالنسبة للمتحول X نجد أن:

$$SST_x = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X})^2 = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki}^2 - \frac{X^2}{g * r} \qquad (85-6)$$

حيث أن: $X = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki}$ هو مجموع قيم X في جميع العينات .

$$SSA_x = \sum_{k=1}^{g}(\bar{X}_k - \bar{X})^2 = \sum_{k=1}^{g}\frac{X_k^2}{r} - \frac{X^2}{g * r} \qquad (86-6)$$

حيث أن: $X_k = \sum_{i=1}^{r} x_{ki}$ هو مجموع قيم X في العينة k فقط و $\bar{X}_k$ متوسطها .

$$SSE_x = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X}_k)^2 = SST_x - SSA_x \qquad (87-6)$$

ولها درجات الحرية : $g(r-1) \;\; = (gr-1) \;\; (g-1)$

أما بالنسبة للمتحول Y فنجد أن:

$$SST_y = \sum_{k=1}^{g}\sum_{i=1}^{r}(y_{ki} - \bar{Y}) = \sum_{k=1}^{g}\sum_{i=1}^{r} y_{ki}^2 - \frac{Y^2}{g * r} \qquad (88-6)$$

حيث أن: $Y = \sum\sum y_{ki}$ هو مجموع قيم Y في جميع العينات .

$$SSA_y = \sum_{k=1}^{g}(\bar{Y}_k - \bar{Y})^2 = \sum_{k=1}^{g}\frac{Y_k^2}{r} - \frac{Y^2}{g * r} \qquad (89-6)$$

حيث أن: $Y_k = \sum_{i=1}^{r} y_{ki}$ هو مجموع قيم Y في العينة k فقط ومتوسطها $\bar{Y}_k$.

$$SSE_y = \sum_{k=1}^{g}\sum_{i=1}^{r}(y_{ki} - \bar{Y}_k)^2 = SST_y - SSA_y \qquad (90-6)$$

أما بالنسبة لمجاميع الجداء $X * Y$ فنجد أن:

$$SPT_{xy} = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X})(y_{ki} - \bar{Y}) = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki} * y_{ki} - \frac{X * Y}{gr} \qquad (91-6)$$

حيث أن: $X = \sum\sum x_{ki}$ وأن $Y = \sum\sum y_{ki}$ وأن درجة حريته $(gr-1)$.

$$SPE_{xy} = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X}_k)(y_{ki} - \bar{Y}_k) = \sum_{k=1}^{g}\left[\sum_{i=1}^{r}(x_{ki} * y_{ki}) - \frac{X_k * Y_k}{gr}\right] = \sum_{k=1}^{g} SPE_k \qquad (\mathbf{92-6})$$

حيث أن: $X_k = \sum_{i=1}^{r} x_{ki}$ وأن: $Y_k = \sum_{i=1}^{r} y_{ki}$ وأن درجة حريته $= g(r-1)$.

$$SPA_{xy} = \sum_{k=1}^{g}(\bar{X}_k - \bar{X})(\bar{Y}_k - \bar{Y}) = \sum_{k=1}^{g}\frac{X_k Y_k}{r} - \frac{XY}{gr} = SPT_{xy} - SPE_{xy} \qquad (\mathbf{93-6})$$

وهنا نشير إلى أن المجاميع الجدائية SPT_{xy} و SPE_{xy} SPA_{xy} يمكن أن تكون سالبة على عكس مجاميع المربعات لـ X و Y الموجبة دائماً ، ولسهولة العرض نضع هذه المربعات والجداءات في جدول منظم كالتالي:

جدول (6−17): جدول التحليل قبل التعديل

مصدر التباين	درجة الحرية dk	مربعات المتحول X	مربعات المتحول Y	مجاميع الجداء $X * Y$
المجتمعات (المعالجات) SSA	$g-1$	SSA_x	SSA_y	SPA_{xy}
الخطأ العشوائي SSE	$g(r-1)$	SSE_x	SSE_y	SPE_{xy}
الاجمالي SST	$g\,r-1$	SST_x	SST_y	SPT_{xy}

وأخيراً نقوم بحساب مجاميع مربعات الانحرافات المعدلة بعد عزل تأثير المتحول الإضافي Y، من العلاقات التالية:

$$(SST)' = (SST)_x - \frac{(SPT)^2_{xy}}{(SST)_y} \qquad : df = gr-2 \qquad (94-6)$$

$$(SSE)' = (SSE)_x - \frac{(SPE)^2_{xy}}{(SSE)_y} \qquad : df = g(r-1)-1 \qquad (95-6)$$

$$(SSA)' = (SST)' - (SSE)' \qquad : df = g-1 \qquad (96-6)$$

وبناء على ذلك يمكننا أن نضع هذه النتائج في جدول تحليل التباين المشترك المعدل والذي يأخذ الشكل(6−18) التالي، ومنه نلاحظ أنه لقد تم تخفيض درجة حرية الاجمالي ودرجة حرية الخطأ التجريبي بمقدار درجة واحدة عما كانت عليه في الجدول (6−17)، وذلك لأننا استخدمنا بيانات العينة في تقدير β والذي يساوي: $\tilde{\beta} = \frac{(SSE)^2_{xy}}{SSE_y}$.

أما المقدار: $\frac{(SPT)_{xy}^{2}}{SST_y}$ فهو مجموع مربعات الانحدار للنموذج (6−81) بدون اعتبار تأثير المجتمعات .
والذي يأخذ الشكل التالي (بدون $\propto_k$) :

$$x_{ki} = \mu_x + \beta(y_{ki} - \bar{Y}) + e_{ki}$$

وهو يعبر عن كمية التباين في قيمة X الناجمة عن المتحول المستقل Y .

جدول (6−18): جدول تحليل التباين المشترك المعدل:جدول ANCOVA

قيمة F المحسوبة	متوسط مجموع المربعات المعدلة	درجة حرية	الرمز	مصدر التباين
$F = \frac{MSA'}{MSE'}$	$MSA' = \frac{SSA'}{g-1}$	$g-1$	$(SSA)'$	بين المجتمعات أو المعالجات
ــــــ	$MSE' = \frac{SSE'}{g(r-1)-1}$	$g(r-1)-1$	$(SSE)'$	الأخطاء العشوائية
ــــــ	ــــــ	$g*r-2$	$(SST)'$	الاجمالي (المعدل)
	حيث أن : r هو حجم العينة الموحد في المجتمعات			

ثم نقوم بإجراء الاختبار اللازم باستخدام المؤشر F المعرف بالعلاقة التالية :

$$F = \frac{\frac{SSA'}{g-1}}{\frac{SSE'}{(g(r-1)-1)}} = \frac{MSA'}{MSE'} \sim F_{(g-1),(g(r-1)-1)} \qquad (97-6)$$

وهو يخضع لتوزيع F بدرجتي حرية $v_1 = (g-1)$ و $v_2 = (g(r-1)-1)$، لذلك نقارن هذه القيمة المحسوبة مع القيمة الحرجة للمتحول F عند مستوى دلالة ∝ ونتخذ القرار كما يلي:

إذا كانت $F \leq F_{v_1,v_2}(\propto)$ نقبل فرضية العدم H_0

أما إذا كانت $F > F_{v_1,v_2}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 بمستوى دلالة ∝ .

ملاحظة: لا يجوز إجراء تحليل التباين المشترك ANCOVA ، قبل حساب معادلة الانحدار الخطي بين X و Y واختبار معنوية β ، والتي سيكون لها الصيغة التالية :

$$X = a + \beta Y \qquad (98-6)$$

فإذا كانت قيمة β غير معنوية (معدومة $\beta = 0$ في جميع المجتمعات) فإن ذلك يعني أن X غير مرتبط بـ Y ، ولا داعي لإدخال Y في النموذج وتعديله، وعندها نقوم بحساب تحليل التباين البسيط باتجاه واحد كالعادة على X فقط .

أما إذا كانت قيمة β معنوية (غير معدومة)، فإننا نختبر قيمها إذا كانت متساوية في جميع المجتمعات أم لا.

فإذا كانت قيم β متساوية في جميع المجتمعات نقوم بإجراء تحليل ANCOVA ونعدل النموذج كما سبق.

أما إذا كانت قيم β غير متساوية في المجتمعات المدروسة ، فإننا نقوم بحساب معادلات الانحدار في كل مجتمع على حدة ونتخذ القرار المناسب حول الأسلوب المناسب لتفسير أسباب تغيرات X .

ملاحظة 2: يمكن تعميم أسلوب تحليل ANCOVA على متحولين إضافيين أو أكثر Y_1 و Y_2 مرتبطين بالمتحول التابع المدروس X، وذلك باتباع نفس المعالجة وباستخدام نفس الفرضيات والاختبارات .

مثال (6-6): لنفترض أننا نريد دراسة تغيرات علامة الرياضيات للطلاب الدارسين في 3 مدارس محددة (3 مجتمعات) . لذلك نرمز لعلاماتهم في مقرر الرياضيات فيها بـ X، ونريد أن ندرس تغيرات X الناتجة على اختلاف تلك المدارس(أي دراسة تأثير المدارس على علامة الرياضيات) . إلا أن أحد المختصين لفت انتباهنا إلى أن علامة الرياضيات X مرتبطة أيضاً بدرجات ذكاء الطالب Y ، لذلك يجب عزل تأثيره. ولهذا علينا استخدام أسلوب تحليل ANCOVA وفق النموذج المعدل (6-81) التالي :

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki} \qquad (99-6)$$

ولذلك قمنا بسحب (3) عينات عشوائية من طلاب هذه المدارس بحجم موحد $n_k = r = 10$ ، وأخذنا منهم علامات الرياضيات وحددنا درجة الذكاء لكل طالب، فكانت كما في الجدول التالي (العلامات X والدرجة Y حسبت من 100) .

جدول (6-19): العلامات X والدرجات Y مع مجاميعها

رقم الطالب	المدرسة الاولى		المدرسة الثانية		المدرسة الثالثة		المجموع	
	Y_{1i}	X_{1i}	Y_{2i}	X_{2i}	Y_{3i}	X_{3i}	Y	X
1	73	55	50	76	82	62		
2	60	70	66	80	88	90		
3	45	30	90	86	90	82		
4	33	27	86	70	50	40		
5	90	89	91	85	70	42		
6	68	50	80	73	75	80		
7	77	60	50	40	80	90		
8	80	98	40	35	95	60		
9	85	79	47	25	40	30		
10	70	82	90	60	50	44		
$\sum$ المجموع	680	640	690	630	720	620	$Y = 2090$	$X = 1890$
معادلات الانحدار	$\tilde{X}_1 = -12.87 + 1.130Y$ $r_1 = 0.836$		$\tilde{X}_2 = 10.42 + 0.762Y_2$ $r_2 = 0.712$		$\tilde{X}_3 = -3.54 + 0.910Y_3$ $r_3 = 0.7745$			

ولدراسة تغيرات X بمعزل عن Y نقوم أولاً بحساب معادلات الانحدار الخطي لعلاقة X بـ Y في كل مدرسة على حدة فنجد أنها تساوي (انظر السطر الأخير من الجدول):

$$\tilde{X}_1 = a + bY_1 = -12.87 + 1.130\,Y_1 \quad (r_1 = 0.836)$$
$$\tilde{X}_2 = a + bY_2 = 10.42 + 0.762\,Y_2 \quad (r_2 = 0.712)$$
$$\tilde{X}_3 = a + bY_3 = -3.54 + 0.910\,Y_3 \quad (r_3 = 0.7745)$$

وبدراسة هذه المعادلات ومعاملات الارتباط فيها نلاحظ أن قيم b فيها ليست معدومة (وليست قريبة من الصفر) . ولذلك فإننا نتابع التحليل ANCOVA دون الخوض في مسألة البرهان على ذلك . ونترك مسألة البرهان على تساوي أو عدم تساوي قيم b في المجتمعات الثلاثة إلى القارئ على سبيل التمرين .

ولمتابعة تحليل ANCOVA نضع النموذج الرياضي كما يلي:

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki}$$

ثم نضع الفرضيتين كما يلي:

$$H_0 : \propto_1 = \propto_2 = \propto_3 = \cdots = \propto_g = 0 \qquad \left(\text{بشرط } \sum \propto_k = 0\right)$$

$H_1 : \propto_k \neq 0$ من أجل k واحدة على الأقل

وحتى نستطيع حساب مجاميع المربعات والجداءات السابقة، قمنا بحساب مجاميع قيم X ومجاميع قيم Y الكلية والهامشية، ووضعناها في الجدول (6–19) السابق، كما قمنا بحساب مجاميع المربعات والجداءات لـ X و Y ووضعناها في الجدول (6–20) ، وبناء على معطيات هذين الجدولين نجد أن:

جدول (6–20): جدول مساعد لحساب مربعات وجداءات Y و X من عناصر العينات

الرموز	المدرسة الأولى			المدرسة الثانية			المدرسة الثالثة			المجاميع		
	y_{1i}^2	x_{1i}^2	$y_{1i}x_{1i}$	y_{2i}^2	x_{2i}^2	$y_{2i}x_{2i}$	y_{3i}^2	x_{3i}^2	$y_{3i}x_{3i}$	Y^2	X^2	Y*X
1	5184	3025	3960	2500	5776	3800	6721	3844	5084			
2	3600	4900	4200	4356	6400	5280	7744	5100	7920			
3	2025	900	1350	8100	7396	7740	8100	6724	7380			
4	1089	729	891	7596	4900	6020	2500	1600	2000			
5	5100	7921	8010	8281	7225	7735	4900	1764	2940			
6	4624	2500	3400	6400	5329	5840	5625	6400	6000			
7	5929	3600	4620	2500	1600	2000	6400	8100	7200			
8	6400	9604	7840	1600	1225	1400	9025	3600	5700			
9	7225	6241	6715	2209	625	1175	1600	900	1200			
10	4900	6724	5740	8100	3600	5400	2500	1936	2200			
المجموع $\sum_{i=1}^{10}$	49067	46144	46726	51442	44067	46390	55118	72968	47624	155636	133188	140740

الرموز	المدرسة الأولى			المدرسة الثانية			المدرسة الثالثة			المجاميع		
	y_{1i}^2	x_{1i}^2	$y_{1i}x_{1i}$	y_{2i}^2	x_{2i}^2	$y_{2i}x_{2i}$	y_{3i}^2	x_{3i}^2	$y_{3i}x_{3i}$	Y^2	X^2	Y*X
متوسط [1(*)] مربعات المجاميع	46240	40960	—	47610	39690	—	51840	38440	—	—	—	—
متوسط [(**)] جداء المجاميع	—	—	43520	—	—	43470	—	—	44640	—	—	—
$(SSE)_k$	2836	5184	—	3832	4386	—	3278	4528	—	9946	14098	—
$(SPE)_k$	—	—	3206	—	—	2920	—	—	2984	—	—	9110

$$Y_1 = \sum_{i=1}^{10} y_{1i} = 680 \qquad X_1 = \sum_{i=1}^{10} x_{1i} = 640$$

$$Y_2 = \sum_{i=1}^{10} y_{2i} = 690 \qquad X_2 = \sum_{i=1}^{10} x_{2i} = 630$$

$$Y_3 = \sum_{i=1}^{10} y_{3i} = 720 \qquad X_3 = \sum_{i=1}^{10} x_{3i} = 620$$

$$Y = \sum_{k=1}^{3}\sum_{i=1}^{10} y_{ki} = 2090 \qquad X = \sum_{k=1}^{3}\sum_{i=1}^{10} x_{ki} = 1890$$

ومن الجدول (6–19) نجد أن:

$$T_y = \sum\sum y_{ki}^2 = 155636 \quad , T_x = \sum\sum x_{ki}^2 = 133188$$

وبالاعتماد على العلاقات من (6–84) إلى(6–88) السابقة وعلى الجدولين (6–19) و(6–20) ، قمنا بحساب مجاميع المربعات والجداءات لكل من Y وX كما يلي:

$$(SST)_y = T_y - \frac{Y^2}{gr} = 155636 - \frac{(2090)^2}{30} = 10032.67$$

$$(SSE)_y = \sum_{k=1}^{3} (SSE_k)_y = 2836 + 3832 + 3278 = 9946$$

$$(SSA)_y = (SST)_y - (SSE)_y = 10032.67 - 9946 = 86.67$$

(*) تم حساب متوسط مربعات المجاميع اعتماداً على الجدول (4–19) ومن العلاقة: $\frac{Y_k^2}{r}$ ثم $\frac{X_k^2}{r}$ مثل: $\frac{Y_1^2}{r} = \frac{(680)^2}{10} = 46240$

(**) كما تم حساب متوسط جداءات المجاميع اعتماداً على الجدول (5–19) ومن العلاقة : $\frac{Y_k X_k}{r}$ مثل: $\frac{Y_1 X_1}{r} = \frac{(680)(640)}{10} = 43520$

$$(SST)_x = T_x - \frac{X^2}{gr} = 133188 - \frac{(1890)^2}{30} = 14118$$

$$(SSE)_x = \sum_{k=1}^{3}(SSE_k)_x = 5184 + 4386 + 4528 = 14098$$

$$(SSA)_x = (SST)_x - (SSE)_x = 14118 - 14098 = 20$$

$$(SPT)_{xy} = T_{xy} - \frac{X * Y}{gr} = 140740 - \frac{(2090)(1890)}{30} = 74905$$

$$(SPE)_{xy} = \sum_{k=1}^{3}(SSE_k)_{xy} = 3206 + 2920 + 2984 = 9110$$

$$(SPA)_{xy} = (SPT)_{xy} - (SPE)_{xy} = 74905 - 9110 = 65795$$

ولتسهيل الإجراءات العملية نقوم بوضع نتائج هذه الحسابات في جدول مختصر كما يلي:

جدول (6–21): مربعات وجداءات X و Y

مصدر التباين	درجة الحرية	مجموع مربعات X	مجموع مربعات Y	مجموع الجداءات $X * Y$
المجتمعات (المعالجات) SSA	2	20	86.67	65795
الأخطاء (البواقي) SSE	27	14098	9946	9110
الاجمالي SST	29	14118	10032.67	74905

ثم نقوم بحساب المجاميع المعدلة للتابع X من العلاقات التالية :

$$(SST)' = (SST)_x - \frac{(SPT)_{xy}^2}{(SST)_y} = 14118 - \frac{(74905)^2}{10032.67} = 8525.5$$

$$(SSE)' = (SSE)_x - \frac{(SPE)_{xy}^2}{(SSE)_y} = 14098 - \frac{(9110)^2}{9946} = 5753.73$$

$$(SSA)' = (SST)' - (SSE)' = 8525.5 - 5753.73 = 2771.77$$

ثم نقوم بوضع المجاميع المعدلة الأخيرة في جدول تحليل ANCOVA فنحصل على أن:

جدول (6–22) جدول تحليل التغاير ANCOVA حيث أن: $g = 3$ و $r = 10$

مصدر التباين	درجة الحرية	مجاميع المربعات المعدلة	متوسط مجاميع المربعات المعدلة	قيمة F المحسوبة
المجتمعات أو المعالجات SSA	$g - 1 = 2$	2771.77	1385.885	$F = \frac{1385.885}{221.30} = 6.25$
الأخطاء العشوائية SSE'	$g(r-1) - 1 = 26$	5753.73	221.30	———
الاجمالي Totel	$gr - 2 = 28$	8525.5	———	———

ثم نقوم بحساب قيمة مؤشر الاختبار F من العلاقة :

$$F = \frac{MSSA'}{MSSE'} = \frac{1385.885}{221.30} = 6.25$$

ثم نبحث في جداول F عن القيمة الحرجة لمتحول F المقابلة لمستوى الدلالة ($\propto = 0.05$) ولدرجتي الحرية $v_1 = 2$ و $v_2 = 26$ فنجد أن :

$$F_{v_1,v_2}(\propto) = F_{2,26}(0.05) = 3.37$$

وبالمقارنة نجد أن: $(F = 6.25) > 3.37$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 ، التي تقول أنه يوجد تأثير للمجتمعات المدروسة على علامات الطلاب في مقرر الرياضيات، وذلك بعد عزل أثر المتحول Y المرتبط مع المتحول X والمتمثل بدرجة الذكاء لهؤلاء الطلاب .

ملاحظة: يستخدم تحليل التباين المشترك ANCOVA لتحقيق هدفين هما:

1- **لزيادة دقة التجربة** وذلك لأنه يعمل على عزل وإبعاد المتحول الإضافي Y المرتبط مع X. وحتى نظهر للقارئ معنى هذا الكلام ننشأ جدول تحليل التباين البسيط ANOVA قبل استبعاد المتحول Y وتعديل المجاميع . فنجد أنه كما يلي:

جدول (6-19): جدول ANOVA **قبل التعديل**

مصدر التباين		درجة الحرية	مجاميع مربعات الانحرافات		
			المربعات X	المتوسطات XY	F
المجتمعات	(SSA)	2	20	10	$F = 0.019$
الأخطاء	(SSE)	27	14096	522.07	
الإجمالي	(SST)	29	14118		

ومنه نحسب قيمة المؤشر F للمتحول X فنجد أن:

$$F_x = \frac{20/2}{14098/27} = \frac{10}{522.15} = 0.019$$

وبمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $F_{2,27}(0.05) = 3.37$ نجد أن $F < 3.37$ ، لذلك كان يمكن أن نقبل H_0 التي تقول أن $\propto_k = 0$، ونقول أن المجتمعات المدروسة لا تؤثر على علامات الرياضيات. وهكذا نجد أن هذه النتيجة تخالف النتيجة السابقة التي حصلنا عليها بعد عزل تأثير درجة الذكاء Y. وهكذا نجد أن تحليل التباين المشترك ANCOVA يكون مفيداً في تدقيق النتائج وتصويبها .

2- **لتقليل نسبة الخطأ وزيادة الكفاءة النسبية:** ففي مثالنا الحالي نجد أن نسبة الخطأ (من المجموع) قبل التعديل كانت تساوي:

$$P_1 = \frac{SSE_X}{SST_X} = \frac{14098}{14118} = 0.9986$$

أما نسبة الخطأ بعد التعديل (من المجموع المعدل) فتساوي :

$$P_2 = \frac{SSE'}{SST'} = \frac{5753.73}{8525.5} = 0.6749$$

ولتقدير الكفاءة النسبية لتحليل التغاير نقوم بحسابها من العلاقة التالية :

$$RE = \frac{MSSE}{MSSE'} = \frac{SSE/g(r-1)}{SSE'/(g(r-1)-1)} = \frac{1}{1-r^2}$$

$$RE = \frac{14098/27}{5758.73/26} = \frac{523.14}{221.30} = 2.36$$

أي أن تطبيق تحليل التباين المشترك ANCOVA قد أدى إلى زيادة كفاءة التحليل ودقة التجربة بمقدار 2.36 مرة .

ملاحظة: يمكن التوسع في تطبيقات تحليل التباين المشترك لزيادة حساسية التجربة، وذلك باستخدام معيارين وإجراء التحليل باتجاهين للتقليل من أثر المتحول الإضافي Y . وعندها نقوم بنفس الخطوات المذكورة في هذه الفقرة مع إجراء بعض التعديلات اللازمة عليها .

وسنترك ذلك للقارئ للتعمق بها ودراستها في المراجع المختصة .

الفصل السابع

الاستدلال بواسطة تحليل التباين المتعدد (MANOVA)

Multivariate Analysis of Variance

7-1: تمهيد:

إن تحليل التباين المتعدد يتناول دراسة تغيرات عدة متحولات طبيعية نرمز لها بـ $X_1\ X_2\ X_3\ \dots\ X_p$ ، في عدة مجتمعات منفصلة ومستقلة (g مجتمعاً و$g<2$) نعطيها الأرقام $1,2,3,\dots,g$، وذلك من خلال سحب عينات عشوائية (g عينة) من تلك المجتمعات بحجوم مختلفة أو متساوية $n_1\ n_2\ n_3\ \dots\ n_g$ ، ثم الحصول على قياسات تلك المتحولات من عناصر هذه العينات والعمل على تحليلها والاستدلال منها على بعض خواص تلك المتحولات في المجتمعات المدروسة .

ويستخدم تحليل التباين المتعدد (MANOVA) لاختبار فيما إذا كانت أشعة متوسطات تلك المتحولات في تلك المجتمعات متساوية أم مختلفة . وإذا كانت مختلفة يجب أن يتم تحديد المركبات المتوسطة التي تختلف عن مقابلاتها في المجتمعات الأخرى .

وإن كل ذلك يجب أن يتم تحت وضمن افتراضات معينة وهي :

الافتراضات المطلوبة لتحليل التباين المتعدد :

حتى نستطيع إجراء تحليل التباين المتعدد بشكل صحيح يجب أن تحقق المتحولات والعينات الافتراضات والشروط التالية :

1- أن تكون المتحولات $X_1\ X_2\ X_3\ \dots\ X_p$ كمية ومستمرة (أو شبه مستمرة) .

2- أن تكون العينات المسحوبة من المجتمعات عشوائية ومستقلة عن بعضها البعض .

3- أن يكون شعاع توقعات قياسات المتحولات في العينة المسحوبة من المجتمع k مساوياً لشعاع توقعاتها في ذلك المجتمع μ_k .

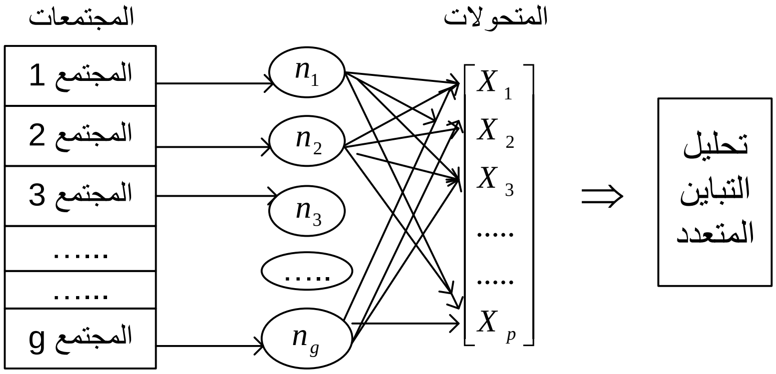
4- أن يكون للمتحولات $X_1\ X_2\ \dots\ X_p$ في جميع تلك المجتمعات مصفوفة تباين موحدة وتساوي $\mathbf{V}$. أي أن يكون:

$$\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V_3} = \cdots = \boldsymbol{V_g} = \boldsymbol{V}$$

5- أن تكون المتحولات $X_1\ X_2\ \dots\ X_p$ في كل مجتمع k خاضعة للتوزيع الطبيعي المتعدد .

ولكن الشرط الأخير (حول الطبيعية) لا يعتبر شرطاً ضرورياً إذا كانت حجوم العينات كبيرة، لأنه في مثل هذه الحالات (الحجوم الكبيرة) يمكننا الاستناد إلى نظرية النهاية المركزية وقانون الأعداد الكبيرة في نظرية الاحتمالات، والتي تنص على أن التوزيع المشترك لتلك المتحولات يتقارب من التوزيع الطبيعي المتعدد .

وللتوضيح نقوم بتمثيل عمليات تحليل التباين المتعدد من خلال الشكل التالي :



الشكل (7−1): تمثيل عمليات تحليل التباين المتعدد

ولتسهيل عمليات التحليل سنقوم أولاً باستعراض مفصل لتحليل التباين المتعدد باتجاه واحد ثم باتجاهين كما يلي:

7−2: تحليل التباين المتعدد باتجاه واحد (MANOVA one way):

وهو يتناول دراسة تغيرات عدة متحولات كمية ومستمرة في عدة مجتمعات منفصلة ومستقلة . لذلك نفترض أننا نريد دراسة تغيرات p متحولاً كمياً هي: $X_1\ X_2\ X_3\ \dots\ X_p$ في g مجتمعاً $(g > 2)$ نرمز لها بـ $1\,,2\,,3\,,\dots,g$، وذلك من خلال مقارنة أشعة متوسطاتها في تلك المجتمعات . ولتنفيذ هذه الدراسة علينا أن نسحب عينات عشوائية (g عينة) من تلك المجتمعات بحجوم مختلفة أو متساوية نرمز لها بـ :

$$n_1\ \ n_2\ \ n_3\ \ \dots\ \ n_g \qquad \text{(على الترتيب)}$$

ثم نقوم بأخذ القياسات اللازمة من هذه العينات في لحظة ما، ودراسة الفروقات بين أشعة متوسطاتها في تلك المجتمعات .

ولاختصار عمليات التحليل الرياضي نرمز لشعاع توقعات تلك المتحولات $X_1\ X_2\ \dots\ X_p$ في المجتمع k (كأحد المجتمعات من g مجتمعاً) بالرمز $\boldsymbol{\mu_k}$ ونكتبه كما يلي:

$$\boldsymbol{\mu_k} = \begin{bmatrix} E_k(X_1) \\ E_k(X_2) \\ \vdots \\ E_k(X_p) \end{bmatrix} = \begin{bmatrix} \mu_{k1} \\ \mu_{k2} \\ \vdots \\ \mu_{kp} \end{bmatrix} \qquad (1-7)$$

حيث أن: $E_k(X_j) = \mu_{kj}$ في المجتمع k.

وعندها يمكننا وضع فرضية العدم H_0 ، التي تنص على عدم وجود فروقات معنوية بين أشعة تلك التوقعات من مجتمع k إلى مجتمع آخر ℓ بدلالة الأشعة كما يلي :

$$H_0: \boldsymbol{\mu_1} = \boldsymbol{\mu_2} = \boldsymbol{\mu_3} = \cdots = \boldsymbol{\mu_g} \qquad (2-7)$$

أو بشكل مفصل كما يلي:

$$H_0: \begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \vdots\\ \mu_{1p}\end{bmatrix} = \begin{bmatrix}\mu_{21}\\ \mu_{22}\\ \vdots\\ \mu_{2p}\end{bmatrix} = \begin{bmatrix}\mu_{31}\\ \mu_{32}\\ \vdots\\ \mu_{3p}\end{bmatrix} = \cdots = \begin{bmatrix}\mu_{g1}\\ \mu_{g2}\\ \vdots\\ \mu_{gp}\end{bmatrix} \qquad (3-7)$$

أما الفرضية البديلة فنكتبها باختصار كما يلي :

$$H_1 : \boldsymbol{\mu_k} \neq \boldsymbol{\mu_\ell} \; : \; k \neq \ell \qquad (4-7)$$

وذلك من أجل زوج واحد على الأقل $(\boldsymbol{\mu_k}, \boldsymbol{\mu_\ell})$

كما نرمز لمصفوفة التباينات المشتركة لتلك المتحولات في المجتمع k بالرمز V_k وعندها يكون لدينا g مصفوفة هي:

$$\boldsymbol{V_k} = \begin{bmatrix}\sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1P}\\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2P}\\ \vdots & \vdots & \dots & \vdots\\ \sigma_{P1} & \sigma_{P2} & \dots & \sigma_{PP}^2\end{bmatrix}_K \qquad : k = 1\ 2\ 3 \dots .\ g \qquad (5-7)$$

وإذا كانت هذه المصفوفات متجانسة فإنه يكون لدينا :

$$\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V_3} = \cdots = \boldsymbol{V_g} = \boldsymbol{V} \qquad (6-7)$$

- **صياغة النموذج الرياضي وتحديد عناصره :**

إن النموذج الرياضي لتحليل التباين المتعدد يستند إلى علاقة مشابهة لـ (6-4) في تحليل التباين ANOVA ويمكن صياغتها مباشرة بدلالة الأشعة كما يلي:

1-نحسب شعاع التوقع الكلي (Grand mean) لهذه المتحولات في جميع المجتمعات من العلاقة الشعاعية التالية:

$$\bar{\boldsymbol{\mu}} = \frac{n_1\boldsymbol{\mu_1} + n_2\boldsymbol{\mu_2} + \cdots + n_k\boldsymbol{\mu_k}}{n_1 + n_2 + \cdots + n_K} = \frac{\sum_{k=1}^{g} n_k\boldsymbol{\mu_k}}{n} \qquad (7-7)$$

حيث أن: $n = \sum_{k=1}^{g} n_K$

2-نعرف النموذج الذي يعطينا شعاع التوقعات $\boldsymbol{\mu_k}$ في أي مجتمع k، بالعلاقة الشعاعية البسيطة التالية:

$$\boldsymbol{\mu_k} = \bar{\boldsymbol{\mu}} + (\boldsymbol{\mu_k} - \bar{\boldsymbol{\mu}}) \qquad (8-7)$$

والتي يمكن كتابتها على الشكل التالي :

$$\boldsymbol{\mu_k} = \bar{\boldsymbol{\mu}} + \boldsymbol{\tau_k} \qquad : \quad \boldsymbol{\tau_k} = \boldsymbol{\mu_k} - \bar{\boldsymbol{\mu}} \quad \text{حيث أن} \qquad (9-7)$$

$$\begin{pmatrix}\text{شعاع توقع}\\ \text{المتحولات في}\\ k \text{ المجتمع}\end{pmatrix} = \begin{pmatrix}\text{شعاع}\\ \text{التوقع الكلي}\\ \bar{\boldsymbol{\mu}}\end{pmatrix} + \begin{pmatrix}\text{شعاع تأثير}\\ k \text{ المجتمع}\\ k \text{ أو تأثير المعالجة}\end{pmatrix}$$

إن الرمز $\boldsymbol{\tau}$ يلفظ (تاو) :

وهذا يقودنا إلى تعديل صياغة فرضية العدم H_0 والفرضية البديلة H_1 لتصبحان كما يلي:

$$H_0 : \boldsymbol{\tau_1} = \boldsymbol{\tau_2} = \cdots = \boldsymbol{\tau_k} = \mathbf{0} \quad \text{(تأثير المجتمعات معدوم)}$$
$$H_1 : \boldsymbol{\tau_k} \neq \mathbf{0} \quad \text{من أجل مجتمع واحد على الأقل :} \qquad (10-7)$$

ولكن الأشعة $\boldsymbol{\tau_k}$ في النموذج (7−9) غير مستقلة عن بعضها البعض لأن :

$$\sum_{k=1}^{g} n_k \boldsymbol{\tau_k} = \sum_{k=1}^{g} n_k(\boldsymbol{\mu_k} - \overline{\boldsymbol{\mu}}) = \sum n_k \boldsymbol{\mu}_k - \overline{\boldsymbol{\mu}} \sum n_k = \sum_{k=1}^{g} n_k \boldsymbol{\mu_k} - \boldsymbol{\mu} * n = 0 \quad \text{(انظر العلاقة } \mathbf{7-7}\text{)}$$

أي أن الأشعة τ_K مرتبطة مع بعضها بالعلاقة التالية :

$$\sum n_k \boldsymbol{\tau_k} = 0 \qquad (11-7)$$

وهو شرط مفروض على الأشعة $\boldsymbol{\tau_K}$ المعرفة في النموذج (7−9) .

3− لاختبار الفرضية H_0 نسحب عينات عشوائية من كل مجتمع بحجوم $n_1\ n_2\ n_3 \dots n_g$، ثم نأخذ قياسات المتحولات $X_1\ X_2 \dots X_p$ من عناصر كل عينة من هذه العينات . ونرمز لتلك القياسات بـ (x_{kji}) حيث k ترمز للمجتمع التي سحبت من العينة . و j للمتحول X_j ، و i لعنصر العينة، ثم نحسب أشعة متوسطاتها في كل مجتمع k ونرمز لها بـ $\overline{\boldsymbol{X}}_{\boldsymbol{kj}}$ ، ثم نحسب مصفوفة التباين المشترك لها في كل مجتمع k ونرمز لها بـ $\boldsymbol{S_k}$ ، ثم ننظم نتائج هذه القياسات والحسابات في جدول مناسب كما يلي:

جدول (7−1): البيانات اللازمة لـ MANOVA

المجتمعات	حجم العينة	المتحولات	قياسات المتحولات من العينات 1 2 3 … n_k	أشعة المتوسطات $\overline{\boldsymbol{X}}_{\boldsymbol{kj}}$	مصفوفات التباين المشترك
المجتمع 1	n_1	X_1 X_2 $\vdots$ X_P	$x_{111}\ x_{112}\ x_{113} \dots x_{11}n_1$ $x_{121}\ x_{122}\ x_{123} \dots x_{12}n_1$ $\vdots\ \vdots\ \vdots \dots \vdots$ $x_{1P1}\ x_{1P2}\ x_{1P3} \dots x_{1P}n_1$	$\bar{X}_{11}$ $\bar{X}_{12}$ $\vdots$ $\bar{X}_{1P}$	$\boldsymbol{S_1}(p*p)$
المجتمع 2	n_2	X_1 X_2 $\vdots$ X_P	$x_{211}\ x_{212}\ x_{213} \dots x_{21}n_2$ $x_{221}\ x_{222}\ x_{223} \dots x_{22}n_2$ $\vdots\ \vdots\ \vdots \dots \vdots$ $x_{2P1}\ x_{2P2}\ x_{2P3} \dots x_{2P}n_2$	$\bar{X}_{21}$ $\bar{X}_{22}$ $\vdots$ $\bar{X}_{2P}$	$\boldsymbol{S_2}(p*p)$
المجتمع k	$\vdots$ n_K $\vdots$	X_1 X_2 $\vdots$ X_P	x_{kji}	$\vdots$ $\bar{X}_{kj}$ $\vdots$	$\vdots$ $\boldsymbol{S_K}(p*p)$ $\vdots$
المجتمع g	n_g	X_1 X_2 $\vdots$ X_P	$x_{g11}\ x_{g12}\ x_{g13} \dots x_{g1}n_g$ $x_{g21}\ x_{g22}\ x_{g23} \dots x_{g2}n_g$ $\vdots\ \vdots\ \vdots \dots \vdots$ $x_{gP1}\ x_{gP2}\ x_{gP3} \dots x_{gP}n_g$	$\bar{X}_{g1}$ $\bar{X}_{g2}$ $\vdots$ $\bar{X}_{gP}$	$\boldsymbol{S_g}(p*p)$

4-نرمز لشعاع المتوسطات في المجتمع k بالرمز $\bar{\boldsymbol{X}}_{\boldsymbol{k}}$ ونكتبه كما يلي :

$$\bar{\boldsymbol{X}}_{\boldsymbol{k}} = \begin{bmatrix} \bar{X}_{k1} \\ \bar{X}_{k2} \\ \vdots \\ \bar{X}_{kP} \end{bmatrix} \qquad k : 1\ 2\ 3 \dots g \qquad (12-7)$$

ويعتبر هذا الشعاع تقديراً للشعاع $\boldsymbol{\mu}_{\boldsymbol{k}}$ في المجتمع k ونكتبه ذلك كما يلي:

$$\tilde{\boldsymbol{\mu}}_{\boldsymbol{k}} = \bar{\boldsymbol{X}}_{\boldsymbol{k}} \qquad (13-7)$$

ثم نحسب شعاع المتوسط الكلي في العينة $\bar{\boldsymbol{X}}$ من العلاقة :

$$\bar{\boldsymbol{X}} = \frac{n_1\bar{\boldsymbol{X}}_{\boldsymbol{1}} + n_2\bar{\boldsymbol{X}}_{\boldsymbol{2}} + \cdots + n_g\bar{\boldsymbol{X}}_{\boldsymbol{g}}}{n_1 + n_2 + \cdots + n_g} = \frac{\sum n_k\bar{\boldsymbol{X}}_{\boldsymbol{k}}}{n} \qquad (14-7)$$

ويعتبر هذا الشعاع تقديراً للشعاع $\boldsymbol{\mu}$ الكلي: (Grand mean)، أي أن $\tilde{\boldsymbol{\mu}} = \bar{\boldsymbol{X}}$

5-نقوم بصياغة النموذج التقديري للنموذج (7-8) وبدلالة الأشعة $\bar{\boldsymbol{X}}_{\boldsymbol{k}}$ و $\bar{\boldsymbol{X}}$ على الشكل التالي :

$$\bar{\boldsymbol{X}}_{\boldsymbol{k}} = \bar{\boldsymbol{X}} + (\bar{\boldsymbol{X}}_{\boldsymbol{k}} - \bar{\boldsymbol{X}}) \qquad (15-7)$$

$$\bar{\boldsymbol{X}}_{\boldsymbol{k}} = \bar{\boldsymbol{X}} + \tilde{\boldsymbol{\tau}}_{\boldsymbol{k}} \qquad : \tilde{\boldsymbol{\tau}}_{\boldsymbol{k}} = \bar{\boldsymbol{X}}_{\boldsymbol{k}} - \boldsymbol{X} \quad : \quad \text{حيث أن}$$

والذي يمكن كتابته كما يلي :

$$\begin{bmatrix} \bar{X}_{k1} \\ \bar{X}_{k2} \\ \vdots \\ \bar{X}_{kP} \end{bmatrix} = \begin{bmatrix} \bar{X} \\ \bar{X} \\ \vdots \\ \bar{X} \end{bmatrix} + \begin{bmatrix} \tilde{\tau}_{k1} \\ \tilde{\tau}_{k2} \\ \vdots \\ \tilde{\tau}_{kg} \end{bmatrix} \Longrightarrow \bar{X}_{k_j} = \bar{X} + \tilde{\tau}_{k_j} \qquad (16-7)$$

وهكذا نجد أن المركبة $\bar{X}_{kj}$ من الشعاع $\bar{\mathbf{X}}_{\mathbf{k}}$ يمكن أن تكتب عددياً على الشكل التالي :

$$\bar{X}_{kj} = \bar{X} + \tilde{\tau}_{kj} = \bar{X} + \left(\bar{X}_{kj} - \bar{X}\right) \qquad (17-7)$$

وبناء على ذلك نجد أنه يمكننا كتابة أية قيمة عددية للمتحول X_j المأخوذة من عنصر العينة i المسحوبة من المجتمع k على الشكل التالي :

$$x_{kji} = \bar{X}_{kj} + \left(\mathrm{x}_{kji} - \bar{X}_{kj}\right) \qquad : \ j : 1\ 2\ 3 \dots P \qquad (18-7)$$

وبعد تعويض $\bar{X}_{kj}$ من (7-16) نحصل على أن القيمة العددية x_{kji} لأي متحول j في المجتمع k تأخذ الشكل التالي (بدلالة الأعداد) :

$$x_{kji} = \bar{X} + \tilde{\tau}_{kj} + \left(x_{kji} - \bar{X}_{kj}\right) \qquad j : 1\ 2\ 3 \dots P$$
$$\mathrm{x}_{kji} = \bar{X} + \tilde{\tau}_{kj} + e_{kji} \qquad : e_{kji} = \left(x_{kji} - \bar{X}_{kj}\right) \qquad (19-7)$$

ولتحويل هذه الكميات العددية إلى أشعة بدلالة جميع المتحولات $X_1\ X_2 \dots X_p$ نكتب جميع معادلات تلك المتحولات على الترتيب، فنحصل على المعادلات التالية :

$$\begin{aligned} x_{k1i} &= \bar{X} + \tilde{\tau}_{k1} + e_{k1i} \\ x_{k2i} &= \bar{X} + \tilde{\tau}_{k2} + e_{k2i} \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ x_{kpi} &= \bar{X} + \tilde{\tau}_{kp} + e_{kpi} \end{aligned} \qquad (20-7)$$

نعتبر كل عمود شعاعاً ، فنجد أنه يمكننا كتابة هذه المعادلات بدلالة أشعة تلك المتحولات كما يلي :

$$\boldsymbol{X_{ki}} = \boldsymbol{\bar{X}} + \boldsymbol{\tilde{\tau}_k} + \boldsymbol{e_{ki}} \qquad : \ \boldsymbol{e_{ki}} = \left(\boldsymbol{\bar{x}_{kj}} - \boldsymbol{\bar{X}_k}\right) \text{ حيث أن} \qquad (21-7)$$

ثم نعوض $\boldsymbol{\tilde{\tau}_k}$ و $\boldsymbol{e_{ki}}$ بما تساويانه فنحصل على العلاقة التالية :

$$\boldsymbol{X_{ki}} = \boldsymbol{\bar{X}} + (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}}) + (\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}_k}) \qquad (22-7)$$

أو التي يمكن كتابتها بدلالة الأشعة على الشكل التالي :

$$(\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}}) = (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}}) + (\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}_k}) \qquad (23-7)$$

6- لإيجاد مصفوفات مجاميع المربعات للطرفين . نقوم بضرب حدود الطرفين من اليمين بمنقول شعاع الطرف الأيسر $(\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}})'$. ثم نضيف ونطرح $\boldsymbol{\bar{X}_k}$ من كل قوس في الجداء فنجد أن :

$$(\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}})(\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}})' = [(\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}_k}) + (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})][(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k}) + (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})]' =$$

$$= (\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})' + (\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})' +$$

$$+(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})(\boldsymbol{X_{ki}} - \boldsymbol{\bar{X}_k})' + (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})' \qquad \mathbf{(24-7)}$$

ثم نأخذ مجموع حدود الطرفين المأخوذ على عناصر العينة k (أي $\sum_{i=1}^{n_k}$)، فنجد أن الحدين الثاني والثالث ينعدمان لأن:

$$\sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})' = \left[\sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})\right](\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})' = 0$$

$$\sum_{i=1}^{nk}(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})' = (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})\left[\sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})'\right] = 0 \qquad \mathbf{(25-7)}$$

وبذلك نجد أن العلاقة (7-24) تصبح مساوية لما يلي :

$$\sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}})' = \sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})' + \sum_{i=1}^{n_k}(\boldsymbol{X_k} - \boldsymbol{\bar{X}})(\boldsymbol{X_k} - \boldsymbol{\bar{X}})' =$$

$$= n_k(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})' + \sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{X_k})'$$

ثم نأخذ مجموع حدود الطرفين على جميع المجتمعات (أي $\sum_{k=1}^{g}$)، فنحصل على المجاميع التالية والتي كل منها يمثل مصفوفة من المرتبة $p * p$:

$$\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}})'$$

$$= \sum_{k=1}^{g} n_k(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})(\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})' + \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})(\boldsymbol{\bar{X}_{ki}} - \boldsymbol{\bar{X}_k})' \qquad \mathbf{(26-7)}$$

$$\begin{pmatrix} \text{مصفوفة تأثير البواقي} \\ residual \text{ أو داخل المجتمعات وهو} \\ withen \text{ مجموع مربعات الانحرافات} \end{pmatrix} + \begin{pmatrix} \text{مصفوفة التأثير بين المجتمعات} \\ between \text{ وهو مجموع مربعات} \\ \text{الانحرافات وجداءات تقاطعهما} \end{pmatrix} = \begin{pmatrix} \text{مصفوفة المجموع الاجمالي (المميز)} \\ \text{لمربعات الانحرافات وجداءات تقاطعهما} \end{pmatrix}$$

$$\begin{pmatrix} sum\ squere\ and\ cross \\ product\ of\ total \end{pmatrix} = \begin{pmatrix} sum\ squers\ and\ cros-product \\ of\ between\ populations \end{pmatrix} + \begin{pmatrix} \text{وجداءات تقاطعهما} \\ sum\ squers\ and\ cros \\ product\ of\ withen\ populations \end{pmatrix}$$

$$[SSCP(total)]_{P*P} = [SSCP(between)]_{P*P} + [SSCP(withen)]_{P*P}$$

وهنا نلاحظ أن مصفوفة البواقي، وهي عبارة عن مجموع المربعات والجداءات المتقاطعة داخل المجتمعات(withen) ، ونرمز لها اختصاراً بـ SSPW وأنه يمكن كتابتها على الشكل التالي :

$$SSPW = \sum_{k=1}^{g}\left[\sum_{i=1}^{n_k}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})'\right]$$

وبناءً على التعاريف في(5−25) نلاحظ أن ما داخل القوس هو $(n_k - 1)\boldsymbol{S_k}$ ، وبذلك نجد أن:

$$SSPW = \sum_{k=1}^{g}(n_k - 1)\,\boldsymbol{S_k} = (n_1 - 1)\boldsymbol{S_1} + (n_2 - 1)\boldsymbol{S_2} + \cdots + (n_g - 1)\boldsymbol{S_g} \qquad (27-7)$$

وذلك لأن$\boldsymbol{S_k}$ هي مصفوفة التباينات المشتركة لتلك المتحولات في عينة المجتمع k، وهي تساوي حسب التعريف مايلي:

$$\boldsymbol{S_k} = \frac{1}{n_k - 1}\sum_{i=1}^{n_k}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})' \qquad (28-7)$$

أما مصفوفة التأثيرات بين المجتمعات (Between) فنرمز لها بـ SSPB وهي تمثل مجموع مربعات الانحرافات وجداءات تقاطعاتها بين المجتمعات . وبذلك نجد أن:

$$SSPB = \sum_{k=1}^{g} n_k(\bar{\boldsymbol{X}}_{\boldsymbol{k}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{k}} - \bar{\boldsymbol{X}})' \qquad (29-7)$$

ثم نأخذ مصفوفة المجموع الاجمالي (المصحح)، التي في الطرف الأيسر، ونرمز لها بالرمز SSPT وهي تمثل مجموع مربعات الانحرافات عن المتوسط الكلي $\bar{\boldsymbol{X}}$ وجداءات تقاطعاتها على جميع عناصر العينات من كل المجتمعات . وبذلك يكون لدينا :

$$SSPT = \sum_{\boldsymbol{k=1}}^{\boldsymbol{g}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n_k}}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}})' \qquad (30-7)$$

ومما سبق نستنتج العلاقة التي ترتبط بين هذه المصفوفات (SSPW وSSPB وSSPT) وهي العلاقة التالية :

$$SSPT_{P*P} = SSPB_{P*P} + SSPW_{P*P} \qquad (31-7)$$

أما بالنسبة لدرجات الحرية المقابلة لكل من هذه المصفوفات فنحسبها بنفس الطريقة التي اتبعناها في ANOVA ، فنجد أن:

$$df:\ n-1=(g-1)+(n-g) \qquad (32-7)$$

حيث أن : $n=\sum_k n_k$ = مجموع حجوم العينات n_k .

وأخيراً ننظم جدولاً لتحليل MAVOVA مشابهاً لجدول تحليل ANOVA من حيث الشكل، ولكنه يختلف عنه من حيث المضمون . حيث أن العناصر التي ضمن جدولMANOVA هي مصفوفات من المرتبة $(P*P)$. بينما عناصر جدول ANOVA هي عبارة عن أعداد حقيقية .

وبذلك يأخذ جدول تحليل MANOVA الشكل التالي :

جدول (7–2) جدول تحليل MANOVA **باتجاه واحد**

مصدر التباين	مصفوفات مربعات الانحرافات وجداءات تقاطعاتها ورموزها	درجة الحرية	المؤشر F
بين المجتمعات أو المعالجات	$SSPB=\sum_{k=1}^{g} n_k(\bar{\boldsymbol{X}}_{\boldsymbol{k}}-\bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{k}}-\bar{\boldsymbol{X}})'$	$g-1$	$F=\dfrac{\frac{\lvert SSPB\rvert}{g-1}}{\frac{\lvert SSPW\rvert}{n-g}}$
داخل المجتمعات أو البواقي (الأخطاء)	$SSPW=\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}}-\bar{\boldsymbol{X}}_{\boldsymbol{k}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}}-\bar{\boldsymbol{X}}_{\boldsymbol{k}})'$ $=\sum_{k=1}^{g}(n_k-1)\boldsymbol{S}_{\boldsymbol{k}}$	$n-g$	
الاجمالي (المصحح من $\bar{X}$)	$SSPT=\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}}-\bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}}-\bar{\boldsymbol{X}})'$	$n-1$	حيث أن: $n=\sum_{k=1}^{g} n_k$

ملاحظة: نلاحظ أن هناك تشابهاً بين جدول MANOVA وجدول ANOVA من حيث الشكل، ولكننا استبدلنا المجموع العددي للمربعات $\sum_{k=1}^{g} n_k(\bar{X}_k-\bar{X})^2$ في جدول ANOVA بالمصفوفة SSPB التي يتألف كل عنصر فيها من مجموع مربعات الانحرافات وجداءات تقاطعاتها .

كما نلاحظ أن مصفوفة البواقي SSPW ترتبط بمصفوفة التباين المدمجة S_P والتي تساوي في حالة g مجتمعاً ما يلي:

$$\boldsymbol{S_P}=\frac{(n_1-1)\boldsymbol{S_1}+(n_2-1)\boldsymbol{S_2}+(n_3-1)\boldsymbol{S_3}+\cdots+(n_g-1)\boldsymbol{S_g}}{(n_1-1)+(n_2-1)+(n_3-1)+\cdots+(n_g-1)}=\widetilde{\boldsymbol{V}}$$

$$\boldsymbol{S_P}=\frac{SSPW}{\sum_{k=1} n_k-g}=\frac{SSPW}{n-g}=\widetilde{\boldsymbol{V}} \qquad (33-7)$$

ولاختبار الفرضية H_0 التالية :

$$H_0 : \boldsymbol{\tau_1} = \boldsymbol{\tau_2} = \boldsymbol{\tau_3} = \cdots = \boldsymbol{\tau_g} = 0$$

$H_1 : \boldsymbol{\tau_k} \neq 0$ من أجل حالة واحدة على الأقل :

يمكننا أن نستخدم قيمة المؤشر النسبي F، الذي يساوي نسبة حجم مصفوفة التأثيرات البينية SSPB على حجم مصفوفة تأثيرات البواقي SSPW، وحتى نعبر عن حجم كل من هاتين المصفوفتين، نحسب القيمة العددية لمحدد كل منهما $|SSPB|$ و $|SSPW|$ ثم نقسمهما على درجة الحرية لكل منهما، ثم نقوم بحساب المؤشر النسبي F من العلاقة التالية :

$$F = \frac{\frac{|SSPB|}{g-1}}{\frac{|SSPW|}{n-g}} = \frac{(n-g) * |SSPB|}{(g-1) * |SSPW|} \qquad (34-7)$$

وبعد حساب قيمة F نقارنها مع القيمة الحرجة لمتحول F عند مستوى الدلالة ∝ ومقابل درجتي الحرية $v_1 = g-1$ و $v_2 = n-g$ والتي نرمز لها بـ $F_{(g-1),(n-g)}(\propto)$ ونتخذ القرار كما يلي:

$$(35-7) \qquad \text{إذا كانت} \quad F \leq F_{(g-1)(n-g)}(\propto) \quad \text{نقبل فرضية العدم } H_0$$

$$(36-7) \qquad \text{أما إذا كانت} \quad F > F_{(g-1)(n-g)}(\propto) \quad \text{نرفض } H_0 \text{ ونقبل } H_1$$

وعندما نرفض H_0 نبحث عن مصدر الاختلاف بين هذه المجتمعات، وذلك بإنشاء مجالات الثقة المتزامنة لها، حسب طريقة (بونفيروني) أو غيرها .

وهنا نشير إلى أنه يتم رفض الفرضية H_0 عندما تكون قيمة F كبيرة بقدر كاف يتجاوز القيمة الحرجة $F(\propto)$.

وهناك مؤشر آخر لاختبار الفرضية H_0 ، أكثر دقة ولكنه أصعب حساباً ، وهو نسبة محدد مصفوفة البواقي $|SSPW|$ على محدد المصفوفة الاجمالية $|SSPT|$ (بدون تقسيم أي منهما على درجة الحرية) فنحصل على مؤشر جديد يسمى بمؤشر Wilk's Lambda المعمم، ويعرف بالعلاقة التالية :

$$\Lambda^* = \frac{|SSPW|}{|SSPT|} = \frac{\left|\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})'\right|}{\left|\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{ki}} - \bar{\boldsymbol{X}})'\right|}$$

وبالاستفادة من العلاقة (7−31) نجد أن Λ^* تساوي :

$$\Lambda^* = \frac{|SSPW|}{|SSPB + SSPW|} \qquad (37-7)$$

إن مؤشر Λ^* يكافئ تماماً المؤشر النسبي F المعرف في (7−34) ، ولكنه يتميز عنه بأنه يرتبط مع مؤشر نسبة الامكانية العظمى Λ .

ولكن التوزيع الدقيق لـ Λ^* يأخذ عدة أشكال معقدة، وذلك حسب عدد المتحولات p وعدد المجتمعاتg .

وإن إجراء بعض التحويلات عليه مقابل بعض الحالات الخاصة لـ p و g ، أدت إلى استخراج العلاقات

التي تربط توزيعه بالتوزيع F المعروف ، وإن بعض هذه الحالات والتحويلات المقابلة معروضة في الجدول التالي، أما بالنسبة للحالات الأخرى وعندما تكون حجوم العينات كبيرة . فإننا نستخدم تعديلاً لـ Λ^* يسمى مؤشر Bartlitt لاختبار H_0 .

جدول (7-3): توزيعات Λ^* حسب بعض الحالات p و g .

توزيع F المقارب	التحويل المستخدم للحصول على توزيع المعاينة لبيانات طبيعية	عدد المجتمعات g	عدد المتحولات p
$F_{(g-1),(n-g)}$	$\left(\frac{n-g}{g-1}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right)$	$g \geq 2$	$p = 1$ ANOVA
$F_{2(g-1),(2n-g-1)}$	$\left(\frac{n-g-1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right)$	$g \geq 2$	$p = 2$
$F_{p,(n-p-1)}$	$\left(\frac{n-p-1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right)$	$g = 2$	$p \geq 1$
$F_{2p,2(n-p-2)}$	$\left(\frac{n-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\Lambda^*}\right)$	$g = 3$	$p \geq 1$

المصدر : (Johnson, Wichern P. 238) حيث $n = \sum_{k=1}^{g} n_k$

أما عندما يكون مجموع حجوم العينات كبيراً $\left(n = \sum_{k=1}^{g} n_k\right)$ فإننا نستخدم مؤشر Bartlitt، الذي يستند على أن المقدار المحول التالي :

$$Bart = -\left(n-1-\frac{p+g}{2}\right) ln\, \Lambda^* = -\left(n-1-\frac{p+g}{2}\right) ln\left(\frac{|SSPW|}{|SSPB+SSPW|}\right) \qquad \mathbf{(38-7)}$$

يخضع تقاربياً لتوزيع χ^2 بدرجة حرية $p(g-1)$. لذلك فإن Bartlitt اقترح أن نستخدم هذا المقدار في اختبار الفرضية H_0 ، وأن نقارنه مع القيمة الحرجة $\chi^2_{p(g-1)}(\propto)$ ، فإذا كانت القيمة المحسوبة:

$$Bart = -\left(n-1-\frac{p+g}{2}\right) ln\left(\frac{|SSPW|}{|SSPB+SSPW|}\right) \leq \chi^2_{p(g-1)}(\propto) \qquad \mathbf{(39-7)}$$

فإننا نقبل الفرضية H_0 باحتمال ثقة $(1-\propto)$ ونرفضها في حالة العكس بمستوى دلالة $\propto$.

ملاحظة: هناك اختبارات إحصائية أخرى تستخدم في تحليل MANOVA باتجاه واحد وهي:

1- **اختبار أثر (Hotelling – Lowley trace)**: وهو يعرف بالعلاقة التالية :

$$trace\, [SSPB(SSPW)^{-1}] = \sum_{k=1}^{g} \lambda_k \qquad (40-7)$$

حيث أن λ_k هي القيم الذاتية (الجذور الخاصة) للمصفوفة A المعرفة بالعلاقة التالية :

$$A = SSPB(SSPW)^{-1} \qquad (41-7)$$

2-**اختبار الجذر الأكبر لـ Roy's (Largest Root):** ويعرف بالعلاقة التالية :

$$RLr = max(\lambda_k) \qquad (42-7)$$

حيث λ_k هي القيم الذاتية أو الجذور الخاصة للمصفوفة A المعرفة في (7-41) .

3-**اختبار أثر بللاي (Pallai's trace):** ويعرف بالعلاقة :

$$trace\,[SSPB(SSPB+SSPW)^{-1}] = \sum_{k=1}^{g} \frac{\lambda_k}{1+\lambda_k} \qquad (43-7)$$

حيث أن λ_k هي القيم الذاتية للمصفوفة A المعرفة في (7-41) .

7-2-1 : إنشاء مجالات الثقة المتزامنة لتأثيرات المجتمعات $\boldsymbol{\tau_k}$:

يتم إنشاء تلك المجالات عندما يتم رفض فرضية العدم، التي تنص على أن جميع تأثيرات تلك المجتمعات متساوية وتساوي الصفر .

لذلك نقوم بمقارنة الأزواج الممكنة $(\boldsymbol{\tau_k - \tau_\ell})$ ونطبق عليها أسلوب (بونفيروني). لأنه يعطينا مجالات أقصر من المجالات التي نحصل عليها بواسطة المصفوفات المتعارضة . ولأنه يحتاج فقط إلى حساب القيم الحرجة للمتحول t مع تعديل بسيط على ∝ ليصبح $\left(\frac{\propto}{2m}\right)$.

ولإنشاء تلك المجالات المتزامنة بأسلوب (بونفيروني) نفترض أن : τ_{kj} هي المركبة j للشعاع $\boldsymbol{\tau_k}$ ، وعندها فإننا نجد أن تقدير تلك التأثيرات يساوي :

$$\boldsymbol{\tilde{\tau}_k} = (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}}) \Longrightarrow \begin{bmatrix} \tilde{\tau}_{k1} \\ \tilde{\tau}_{k2} \\ \vdots \\ \tilde{\tau}_{kp} \end{bmatrix} = \begin{bmatrix} X_{k1} - \bar{X} \\ X_{k2} - \bar{X} \\ \vdots \\ X_{kp} - \bar{X} \end{bmatrix} \qquad (44-7)$$

وبذلك نجد أن المركبة $\tilde{\tau}_{kj}$ المقابلة للمتحول j تقدر بما يلي :

$$\tilde{\tau}_{kj} = \left(\bar{X}_{kj} - \bar{X}\right) \qquad j{:}\, 1\ 2\ 3\ \dots\ p$$

ولدراسة الاختلافات في تأثير المجتمعات على المتحول X_j نحسب الفرق بين تأثيري المجتمعين k و ℓ على المتحول X_j كما يلي :

$$\left(\tilde{\tau}_{kj} - \tilde{\tau}_{\ell j}\right) = \left(\bar{X}_{kj} - \bar{X}\right) - \left(\bar{X}_{\mathrm{k}j} - \bar{X}\right) = \left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) \qquad (45-7)$$

حيث أن: $j{:}\, 1\ 2\ 3\ \dots\ p$

وبما أن عدد المجتمعات يساويg ، فإن عدد الأزواج الممكنة (غير المكررة)، التي يمكن تشكيلها لهذه الفروقات يساوي $\frac{g(g-1)}{2}$ ، وبما أن ذلك يحدث من أجل كل متحول j وأن عدد المتحولات يساوي p ، فإننا نجد أن عدد الأزواج الممكنة لمقارنة هذه الفروقات يساوي :

$$m = P * \frac{g(g-1)}{2} \qquad \left(\text{زوجاً}\right) \qquad (46-7)$$

ومن العلاقة (7-45) نجد أن الفرق $\left(\tilde{\tau}_{kj} - \tilde{\tau}_{\ell j}\right)$ يساوي :

$$\left(\tilde{\tau}_{kj} - \tilde{\tau}_{\ell j}\right) = \left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) \qquad (47-7)$$

وهذا يعني أنه يساوي الفرق بين متوسطي العينتين المستقلتين k و ℓ . لذلك فإن الأسلوب المناسب لإنشاء مجالات الثقة المتزامنة لهذه الفروقات هو أسلوب t (ستودينت) المعدل، الذي يعتمد على أسلوب (بونفيروني) وتعديل $\propto$ إلى $\left(\frac{\propto}{2m}\right)$. لذلك نحسب تباين الفرق $\left(\tilde{\tau}_{kj} - \tilde{\tau}_{\ell j}\right)$ فنجد (لأن العينتان مستقلتان) أن:

$$Var\left(\tilde{\tau}_{kj} - \tilde{\tau}_{\ell j}\right) = Var\left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) = Var\left(\bar{X}_{kj}\right) + V\left(\bar{X}_{\ell j}\right) = \frac{\sigma_{jj}^2}{n_k} + \frac{\sigma_{jj}^2}{n_\ell} = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\sigma_{jj}^2 \quad \mathbf{(48-7)}$$

حيث أن σ_{jj}^2 هو العنصر القطري j في مصفوفة التباينات المشتركة V للمتحولات المدروسة X .

وحسب العلاقة (7–33) التي تشير إلى أن تقدير **V** يعطى بما يلي :

$$\widetilde{\boldsymbol{V}} = \frac{SSPW}{n-g} \qquad (49-7)$$

أي أن أي عنصر في **V** يقدر بواسطة تقسيم العنصر المقابل له في SSPW على العدد $(n-g)$، أي أن $\sigma_{jj}^2 = \frac{W_{jj}}{n-g}$ ، وبذلك نجد أن تباين الفرق $\left(\bar{X}_{kj} - \bar{X}_{\ell j}\right)$ يساوي :

$$Var\left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right) * \frac{w_{jj}}{n-g} \qquad j : 1\ 2\ 3 \ldots P \qquad (50-7)$$

حيث أن w_{jj} هو العنصر القطري j في المصفوفة SSPW .

وحسب أسلوب t لإنشاء مجالات الثقة للفروقات الفعلية $\left(\tau_{kj} - \tau_{\ell j}\right)$ نستخدم العلاقة المعروفة والمرتبطة بأسلوب (بونفيروني)، والتي تأخذ الشكل التالي :

$$\left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) - t_{n-g}\left(\frac{\propto}{2m}\right) * \sqrt{Var\left(X_{kj} - \bar{X}_{\ell j}\right)} \le \left(\tau_{kj} - \tau_{\ell j}\right) \le$$
$$\le \left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) + t_{n-g}\left(\frac{\propto}{2m}\right) * \sqrt{Var\left(\bar{X}_{kj} - \bar{X}_{\ell j}\right)} \qquad (51-7)$$

وبتعويض ( ) Var و m من العلاقتين (7–50) و(7–46) نحصل على مجالات الثقة المتزامنة حسب أسلوب (بونفيروني) كما يلي:

$$\left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) - t_{n-g}\left(\frac{\propto}{pg(g-1)}\right) * \sqrt{\left(\frac{1}{n_k} + \frac{1}{n_\ell}\right) * \frac{w_{jj}}{n-g}} \le \left(\tau_{kj} - \tau_{\ell j}\right) \le$$
$$\le \left(\bar{X}_{kj} - \bar{X}_{\ell j}\right) + t_{n-g}\left(\frac{\propto}{pg(g-1)}\right) * \sqrt{\left(\frac{1}{n_k} + \frac{1}{n_\ell}\right) * \frac{w_{jj}}{n-g}} \qquad \mathbf{(52-7)}$$

حيث أن: w_{jj} هي العنصر القطري j في المصفوفة SSCW المقابل للمتحول X_j، وأن $n = \sum_{k=1}^{g} n_k$.

إن العلاقة (7–52) تعطينا مجالات الثقة المتزامنة للفروقات $\left(\tau_{kj} - \tau_{\ell j}\right)$ لكل متحول j وباحتمال ثقة قدره $(1-\propto)$، علماً بأن هذه المجالات تحسب لجميع المتحولات X_j حيث أن $j : 1\ 2\ 3 \ldots p$ ولجميع أزواج المجتمعات (k و ℓ) حيث أن : $k < l : 1\ 2\ 3 \ldots . g$، وبذلك يكون لدينا $P\left(\frac{g(g-1)}{2}\right)$ مجالاً متزامناً لهذه الفروقات .

مثال (7−1): سنقوم بتطوير بيانات المثال (6−2) السابق . ونضيف إلى المتحول X_1 متحولاً آخر X_2، وسنعمل على دراسة تغيرات هذين المتحولين في ثلاثة مجتمعات، سحبنا منها (3) عينات بحجومها السابقة: $n_1 = 3$, $n_2 = 2$, $n_3 = 3$ ، وحصلنا منها على القياسات والمتوسطات التالية :

$$\begin{array}{ccc} & & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ P_1: & \begin{array}{c} X_1 \\ X_2 \end{array} & \\ P_2: & \begin{array}{c} X_1 \\ X_2 \end{array} & \\ P_3: & \begin{array}{c} X_1 \\ X_2 \end{array} & \end{array} \begin{pmatrix} \begin{bmatrix} 9 \\ 3 \end{bmatrix} & \begin{bmatrix} 6 \\ 2 \end{bmatrix} & \begin{bmatrix} 9 \\ 7 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 \\ 0 \end{bmatrix} & - \\ \begin{bmatrix} 3 \\ 8 \end{bmatrix} & \begin{bmatrix} 1 \\ 9 \end{bmatrix} & \begin{bmatrix} 2 \\ 7 \end{bmatrix} \end{pmatrix} \qquad \begin{array}{l} \bar{\boldsymbol{X}}_{\boldsymbol{1}} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \\ \bar{\boldsymbol{X}}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \bar{X} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \\ \bar{\boldsymbol{X}}_3 = \begin{bmatrix} 2 \\ 8 \end{bmatrix} \end{array}$$

حيث حسبنا المتوسط الكلي لها من العلاقة :

$$\bar{\boldsymbol{X}} = \frac{n_1\bar{\boldsymbol{X}}_{\boldsymbol{1}} + n_2\bar{\boldsymbol{X}}_2 + n_3\bar{\boldsymbol{X}}_3}{n_1 + n_2 + n_3} = \frac{3\begin{bmatrix} 8 \\ 4 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 2 \\ 8 \end{bmatrix}}{3 + 2 + 3} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

ولحساب المصفوفة SSPB نحسب مجموع المربعات والجداءات المتقاطعة فيها من العلاقة :

$$SSPB = \sum_{k=1}^{3} n_k(\bar{\boldsymbol{X}}_{\boldsymbol{k}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{k}} - \bar{\boldsymbol{X}})' =$$

$$= n_1(\bar{\boldsymbol{X}}_{\boldsymbol{1}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{1}} - \bar{\boldsymbol{X}})' + n_2(\bar{\boldsymbol{X}}_{\boldsymbol{2}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{2}} - \bar{\boldsymbol{X}})' + n_3(\bar{\boldsymbol{X}}_{\boldsymbol{3}} - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_{\boldsymbol{3}} - \bar{\boldsymbol{X}})'$$

$$= 3\begin{bmatrix} 4 \\ -1 \end{bmatrix}[4, -1] + 2\begin{bmatrix} -3 \\ -3 \end{bmatrix}[-3, -3] + 3\begin{bmatrix} -2 \\ 3 \end{bmatrix}[-2, 3] =$$

$$= 3\begin{bmatrix} 16 & -4 \\ -4 & 1 \end{bmatrix} + 2\begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix} + 3\begin{bmatrix} 4 & -6 \\ -6 & 9 \end{bmatrix} =$$

$$SSPB = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$$

ولحساب المصفوفة SSPW نقوم بحساب عناصر مصفوفات التباين المشترك لهذين المتحولين في كل من المجتمعات الثلاثة فنجد أن :

$$\boldsymbol{S}_{\boldsymbol{1}} = \frac{1}{n_1 - 1}\sum_{i=1}^{3}(\boldsymbol{X}_{\boldsymbol{1i}} - \bar{\boldsymbol{X}}_{\boldsymbol{1}})(\boldsymbol{X}_{\boldsymbol{1i}} - \bar{\boldsymbol{X}}_{\boldsymbol{1}})'$$

$$= \frac{1}{3-1}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}[1, -1] + \begin{bmatrix} -2 \\ -2 \end{bmatrix}[-2, -2] + \begin{bmatrix} 1 \\ 3 \end{bmatrix}[1, 3]\right\}$$

$$= \frac{1}{2}\left\{\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}\right\}$$

$$\boldsymbol{S}_{\boldsymbol{1}} = \frac{1}{2}\begin{bmatrix} 6 & 6 \\ 6 & 14 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 7 \end{bmatrix}$$

$$\boldsymbol{S_2} = \frac{1}{n_2 - 1}\sum_{i=1}^{2}(\boldsymbol{X_{2i}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})(\boldsymbol{X_{2i}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})'$$

$$= \frac{1}{2-1}\left\{\begin{bmatrix}-1\\2\end{bmatrix}[-1,2] + \begin{bmatrix}1\\-2\end{bmatrix}[1,-2]\right\}$$

$$\boldsymbol{S_2} = \begin{bmatrix}1 & -2\\-2 & 4\end{bmatrix} + \begin{bmatrix}1 & -2\\-2 & 4\end{bmatrix} = \begin{bmatrix}2 & -4\\-4 & 8\end{bmatrix}$$

$$\boldsymbol{S_3} = \frac{1}{n_3 - 1}\sum_{i=1}^{3}(\boldsymbol{X_{3i}} - \overline{\boldsymbol{X}}_{\boldsymbol{3}})(\boldsymbol{X_{3i}} - \overline{\boldsymbol{X}}_{\boldsymbol{3}})'$$

$$= \frac{1}{3-1}\left\{\begin{bmatrix}1\\0\end{bmatrix}[1,0] + \begin{bmatrix}-1\\1\end{bmatrix}[-1,1] + \begin{bmatrix}0\\-1\end{bmatrix}[0,-1]\right\}$$

$$= \frac{1}{2}\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} + \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\}$$

$$\boldsymbol{S_3} = \frac{1}{2}\begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix} = \begin{bmatrix}1 & -\frac{1}{2}\\-\frac{1}{2} & 1\end{bmatrix}$$

ومنها نجد أن المصفوفة $SSPW$ تساوي :

$$SSPW = (n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S_2} + (n_3 - 1)\boldsymbol{S_3}$$

$$= 2\begin{bmatrix}3 & 3\\3 & 7\end{bmatrix} + 1\begin{bmatrix}2 & -4\\-4 & 8\end{bmatrix} + 2\begin{bmatrix}1 & -\frac{1}{2}\\-\frac{1}{2} & 1\end{bmatrix}$$

$$SSPW = \begin{bmatrix}10 & 1\\1 & 24\end{bmatrix}$$

وللحصول على المصفوفة SSPT يمكننا الاستفادة من العلاقة (7−31) فنجد أن:

$$SSPT = SSPB + SSPW = \begin{bmatrix}78 & -12\\-12 & 48\end{bmatrix} + \begin{bmatrix}10 & 1\\1 & 24\end{bmatrix}$$

$$SSPT = \begin{bmatrix}88 & -11\\-11 & 72\end{bmatrix}$$

ثم نقوم بحساب قيمة Λ^* من العلاقة (7−37) فنجد أن:

$$\Lambda^* = \frac{|SSPW|}{|SSPB + SSPW|} = \frac{|SSPW|}{|SSPT|} = \frac{\begin{vmatrix}10 & 1\\1 & 24\end{vmatrix}}{\begin{vmatrix}88 & -11\\-11 & 72\end{vmatrix}} = 0.0385$$

وبناء على ذلك ننظم جدول تحليل MANOVA كما يلي :

جدول (7–4): جدول MANOVA

مصدر التباين	مصفوفات المربعات والجداءات المتقاطعة	درجة الحرية	قيمة Λ^*
ما بين المجتمعات أو المجتمعات	$SSPB = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$	$g-1=3-1=2$	
داخل العينات أو البواقي (والخطأ العشوائي)	$SSPW = \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$	$n-g=8-3=5$	
المجموع الاجمالي المصحح	$SSPT = \begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$	$n-1=8-1=7$	$\Lambda^* = \frac{\lvert SSPW \rvert}{\lvert SSPT \rvert} = 0.0385$

ومن معطيات الجدول السابق نحسب قيمة Λ^* ثم نقارنها مع القيمة الحرجة لها . ولذلك علينا أن نختار من الجدول (7–3) التحويلة المقابلة للحالة $P=2$ و $g=3$ ، فنجد أن التحويلة المناسبة لاختبار فرضية العدم: $H_0 : \boldsymbol{\tau_1} = \boldsymbol{\tau_2} = \cdots = \boldsymbol{\tau_K} = 0$ (ضمن تحقق شروط الطبيعية وتساوي مصفوفات التباين المشترك) . هي قيمة المقدار التالي :

$$\left(\frac{n-g-1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = \frac{(8-3-1)}{3-1}\left(\frac{1-\sqrt{0.0385}}{\sqrt{0.0385}}\right) = 8.19$$

وبمقارنة هذه القيمة المحولة مع القيمة الحرجة لتوزيع F المقابلة لدرجتي الحرية $v_1 = 2(g-1)$ و $v_2 = (2n-g-1)$ وهي $F_{2(g-1),(2n-g-1)}(\propto)$ ، فنجد من جداول F الاحتمالية عند مستوى دلالة $\propto = 0,01$ ، أن :

$$F_{2(g-1),(2n-g-1)}(0\,,01) = F_{4\,,8}(0,01) = 7,01$$

وبمقارنة القيمة المحولة (8,19) مع القيمة الحرجة (7,01) أن: $8{,}19 > 7{,}01$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تنص على أن واحد على الأقل من التأثيرات $\boldsymbol{\tau_k} \neq 0$.

ملاحظة: عندما يكون عدد المتحولات p كبيراً . فإننا لا نقوم بحساب جداول MANOVA كاملة، بل نطلب من الحواسيب أن تقوم بإجراء الحسابات وطباعة المصفوفتين SSPB وSSPW فقط، وبذلك يمكننا الاستغناء عن كمية هائلة من المعلومات غير اللازمة .

ولكن يجب اختبار تحقق شرط الطبيعية لأشعة البواقي : $\boldsymbol{e_{ki}} = \boldsymbol{X_{ki}} - \boldsymbol{\bar{X}_k}$ والتأكد من أنها تخضع للتوزيع الطبيعي وإن توقعها يساوي الصفر .

ولنبحث الآن عن التأثيرات غير المعدومة التي أدت إلى رفض الفرضية H_0، ولذلك نقوم بحساب تقديرات الفروقات $\boldsymbol{\tau_K}$ من العلاقة $\boldsymbol{\tilde{\tau}_K} = (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}})$ ، فنجد أن:

$$\boldsymbol{\tilde{\tau}_1} = \begin{bmatrix} 4 \\ -1 \end{bmatrix} \neq 0\,, \boldsymbol{\tilde{\tau}_2} = \begin{bmatrix} -3 \\ -3 \end{bmatrix} \neq 0\,, \boldsymbol{\tilde{\tau}_3} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \neq 0$$

ولتحديد أي من هذه التأثيرات كان السبب في رفض H_0 ، نقوم بحساب الفروقات الزوجية بين المجتمعات فنجد أن :

$$\tilde{\boldsymbol{\tau}}_1 - \tilde{\boldsymbol{\tau}}_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix} \quad \tilde{\boldsymbol{\tau}}_1 - \tilde{\boldsymbol{\tau}}_3 = \begin{bmatrix} 6 \\ -4 \end{bmatrix} \quad \tilde{\boldsymbol{\tau}}_2 - \tilde{\boldsymbol{\tau}}_3 = \begin{bmatrix} -1 \\ -6 \end{bmatrix}$$

ثم نقوم بإنشاء مجالات الثقة المتزامنة (بأسلوب بونفيروني) لمركبات هذه الفروقات، فنجد أن مجال الثقة للمركبة الفعلية الأولى من $(\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2)$ ، وهي $(\tau_{11} - \tau_{21})$ والمقدرة بـ : $\tilde{\tau}_{11} - \tilde{\tau}_{21} = 7$ ، يعطى بالعلاقة (7−52) وهي :

$$(\tau_{11} - \tau_{21}) = (\tilde{\tau}_{11} - \tilde{\tau}_{21}) \pm t_{n-g}\left(\frac{\propto}{pg(g-1)}\right)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{w_{11}}{n-g}}$$

ولذلك نقوم أولاً بإيجاد قيمة t الحرجة عند مستوى دلالة 0.05 =∝ فنجد أن :

$$t_{n-g}\left(\frac{\propto}{pg(g-1)}\right) = t_{8-3}\left(\frac{0.05}{2*3*2}\right) = t_5(0,00417) = 4,22$$

ومن المصفوفة SSPW نلاحظ أن العنصر $w_{11} = 10$، ثم نعوض ذلك في العلاقة السابقة، فنجد أن مجال الثقة للمركبة الفعلية $(\tau_{11} - \tau_{21})$ يساوي:

$$(\tau_{11} - \tau_{21}) = 7 \pm (4,22) * \sqrt{\left(\frac{1}{3} + \frac{1}{2}\right)\frac{10}{8-3}} = 7 \pm 5,448$$

أي أن: $1,552 \le (\tau_{11} - \tau_{21}) \le 12,448$

وهنا نلاحظ أن المجال لا يتضمن نقطة الصفر وهذا يعني أن هذا الفرق معنوي

وكذلك نجد أن مجال الثقة للمركبة الفعلية الثانية من $(\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2)$ وهي$(\tau_{12} - \tau_{22})$ والمقدرة بـ $\tilde{\tau}_{12} - \tilde{\tau}_{22} = 2$ ، فنجد أن مجال الثقة لها يساوي:

$$(\tau_{12} - \tau_{22}) = +2 \pm (4,22)\sqrt{\left(\frac{1}{3} + \frac{1}{2}\right)\frac{24}{8-5}} = 2 \pm 8,44$$

أي أن $-6,44 < (\tau_{12} - \tau_{22}) < 10,44$ (لا يوجد فرق معنوي)

وكذلك نجد أن مجال الثقة للمركبة الفعلية الأولى من $(\boldsymbol{\tau}_1 - \boldsymbol{\tau}_3)$ وهي: $(\tau_{11} - \tau_{31})$ والمقدرة بـ $\tilde{\tau}_{11} - \tilde{\tau}_{31} = 6$ فنجد أنه يساوي :

$$(\tau_{11} - \tau_{31}) = 6 \pm (4,22)\sqrt{\left(\frac{1}{3} + \frac{1}{3}\right)\frac{10}{8-3}} = 6 \pm 4,873$$

أي أن: $1,127 \le (\tau_{11} - \tau_{31}) \le 12,873$ (يوجد فرق معنوي)

وكذلك نجد أن مجال الثقة للمركبة الفعلية الثانية من $(\boldsymbol{\tau_1}-\boldsymbol{\tau_3})$ وهي $(\tau_{12}-\tau_{32})$ المقدرة بـ $\tilde{\tau}_{12}-\tilde{\tau}_{32}=-4$ فنجد أنه يساوي :

$$(\tau_{12}-\tau_{32})=-4\pm(4{,}22)\sqrt{\left(\frac{1}{3}+\frac{1}{3}\right)\frac{24}{8-3}}=-4\pm7{,}549$$

أي أن: $-11{,}549\leq(\tau_{12}-\tau_{32})\leq3{,}549$ (لا يوجد فرق معنوي)

ونفعل الأمر ذاته مع المركبة الفعلية الأولى من $(\boldsymbol{\tau_2}-\boldsymbol{\tau_3})$ وهي: $(\tau_{21}-\tau_{31})$ والمقدرة بـ $\tilde{\tau}_{21}-\tilde{\tau}_{31}=-1$ فنجد أنه يساوي :

$$(\tau_{21}-\tau_{31})=-1\pm(4{,}22)\sqrt{\left(\frac{1}{2}+\frac{1}{3}\right)\frac{10}{8-5}}=-1\pm5{,}448$$

أي أن: $-6{,}548\leq(\tau_{21}-\tau_{31})\leq4{,}448$ (لا يوجد فرق معنوي)

وأن مجال الثقة للمركبة الفعلية الثانية من$(\boldsymbol{\tau_2}-\boldsymbol{\tau_3})$ وهي:$(\tau_{22}-\tau_{32})$ والمقدرة بـ $\tilde{\tau}_{22}-\tilde{\tau}_{32}=-6$ فنجد أنه يساوي :

$$(\tau_{22}-\tau_{32})=-6\pm(4{,}22)\sqrt{\left(\frac{1}{2}+\frac{1}{3}\right)\frac{24}{8-5}}=-6\pm8{,}44$$

أي أن: $-14{,}44\leq(\tau_{22}-\tau_{32})\leq2{,}44$ (لا يوجد فرق معنوي)

7-2-2: تحليل الصورة الجانبية (Profile Analysis) بواسطة المصفوفات المتعارضة (Contrast):

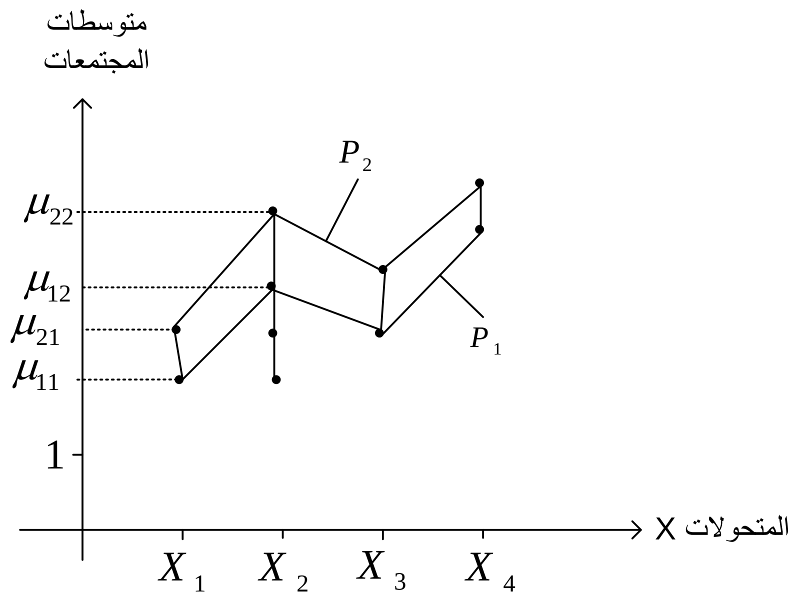
يستخدم تحليل الصورة الجانبية عندما يكون عدد المتحولات p كبيراً . وذلك لاختصار العمليات الحسابية عند اختبار الفروقات $(\tau_{kj}-\tau_{\ell j})$، أو عند إنشاء مجالات الثقة المتزامنة لها . وهو يشترط أن تكون استجابات المجتمعات المختلفة مستقلة عن بعضها البعض، وأن يكون جميع الاستجابات مقاسة بوحدة قياس موحدة .

وهنا يطرحون السؤال التالي: هل أشعة المتوسطات للمجتمعات متساوية ؟

ولتسهيل تحليل الصورة الجانبية يتم تجزئة هذا السؤال إلى عدة صور جانبية، بحيث تعكس كل صورة حالة شعاع المتوسطات في أحد المجتمعات، ولتسهيل فهم هذا التحليل نقتصر على مجتمعين فقط P_1 و P_2 يؤثران على أربعة متحولات $X_1\ X_2\ X_3\ X_4$، ولنفترض أن شعاع توقعاتها في المجتمع الأول هو $\boldsymbol{\mu_1}$، وفي المجتمع الثاني $\boldsymbol{\mu_2}$ ونكتب ذلك كما يلي :

$$\boldsymbol{\mu_1}=\begin{bmatrix}\mu_{11}\\\mu_{12}\\\mu_{13}\\\mu_{14}\end{bmatrix}\qquad\boldsymbol{\mu_2}=\begin{bmatrix}\mu_{21}\\\mu_{22}\\\mu_{23}\\\mu_{24}\end{bmatrix}\qquad(53-7)$$

وعند رسم الصورة الجانبية لمركبات هذه المتوسطات بدلالة المتحولات الأربعة نحصل على الشكل التالي:



الشكل (7-2):الصورة الجانبية لمركبات التوقعين

وعندها فإن فرضية العدم لشعاعي توقعات هذين المجتمعين تأخذ الشكل التالي :

$$H_0: \ \boldsymbol{\mu_1} = \boldsymbol{\mu_2} \qquad (54-7)$$

وهذا يعني أننا نفترض أن متوسطات تأثير هذين المجتمعين على المتحولات متساوية . وفي التحليل الجانبي يمكننا صياغة هذه الفرضية على شكل ثلاثة أسئلة مرحلية . كما يلي :

س1: هل الصور الجانبية متوازية ؟ وهذا يكافئ القول أن المسافات بين نقاط الخطين متساوية، ونكتب ذلك كما يلي :

$$H_{01}: \ \mu_{1j} - \mu_{2j} = \mu_{1(j-1)} - \mu_{2(j-1)} \qquad : j = 1\ 2\ 3 \ldots P$$

والتي يمكن كتابتها على الشكل التالي :

$$H_{01}: \ \mu_{1j} - \mu_{1(j-1)} = \mu_{2j} - \mu_{2(j-1)} \qquad : j = 1\ 2\ 3 \ldots P \qquad (55-7)$$

ولصياغة هذه الفرضية بدلالة المصفوفات نستخدم المصفوفة المتعارضة C المعرفة كما يلي:

$$\boldsymbol{C_{(P-1)P}} = \boldsymbol{C_{3*4}} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \qquad (56-7)$$

وعندها نجد أن جداء **C** في $\boldsymbol{\mu_1}$ يعطينا أن:

$$\boldsymbol{C\mu_1} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \end{bmatrix} = \begin{bmatrix} \mu_{12} - \mu_{11} \\ \mu_{13} - \mu_{12} \\ \mu_{14} - \mu_{13} \end{bmatrix} \qquad (57-7)$$

وإن جداءها في $\boldsymbol{\mu_2}$ يعطينا أن:

$$\boldsymbol{C\mu_2} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \end{bmatrix} = \begin{bmatrix} \mu_{22} - \mu_{21} \\ \mu_{23} - \mu_{22} \\ \mu_{24} - \mu_{23} \end{bmatrix} \qquad (58-7)$$

وعندها فإن فرضية العدم الأولى (7–55) والتي نرمز لها بـ H_{01} وتأخذ الشكل التالي:

$$H_{01}:\ \boldsymbol{C\mu_1} = \boldsymbol{C\mu_2} \qquad \text{(في المجتمعين)} \qquad (59-7)$$

وعندما تكون العينتان المسحوبتان من المجتمعين بحجمين n_1 و n_2 مستقلتين، فإنه يمكننا تقدير $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ بواسطة شعاعي متوسطيهما. $\bar{\boldsymbol{X}}_1$ و $\bar{\boldsymbol{X}}_2$ في هذين المجتمعين . وعندها نجد أن جداء هذين التقديرين بالمصفوفة **C** يأخذان الشكل التالي :

$$\begin{aligned} \boldsymbol{C} * \tilde{\boldsymbol{\mu}}_1 &= \boldsymbol{C} * \bar{\boldsymbol{X}}_1 \\ \boldsymbol{C} * \tilde{\boldsymbol{\mu}}_2 &= \boldsymbol{C} * \bar{\boldsymbol{X}}_2 \end{aligned} \qquad (60-7)$$

ويمكن البرهان على أن المصفوفة المدمجة للتباينات المشتركة لهما تساوي ما يلي :

$$\boldsymbol{C} * \boldsymbol{S_p} * \boldsymbol{C'} \qquad (61-7)$$

وكذلك على أن الشعاعين $\boldsymbol{C\bar{X}_1}$ و $\boldsymbol{C\bar{X}_2}$ يخضعان لتوزيعين طبيعيين، $N_{p-1}(\boldsymbol{C\mu_1}, \boldsymbol{CS_pC'})$ و $N_{p-1}(\boldsymbol{C\mu_2}, \boldsymbol{CS_pC'})$ في الفضاء R^{p-1} على الترتيب، وإن اختبار توازي الصورتين الجانبيتين يصبح كما يلي :

$$H_{01}:\ \boldsymbol{C_1\mu_1} = \mathbf{C_2}\boldsymbol{\mu_2} \qquad \text{(فرضية العدم)} \qquad (62-7)$$

وإن مؤشر الاختبار المناسب لها هو T^2 ويأخذ الشكل التالي :

$$T^2 = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{C'} * \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{CS_pC'}\right]^{-1} * \boldsymbol{C}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \qquad (63-7)$$

وهذا المؤشر مرتبط تقارباً بمتحول التوزيع : F_{p-1, n_1+n_2-p} وذلك من خلال العلاقة التالية :

$$T^2(\propto) = \frac{(n_1+n_2-2)(p-1)}{n_1+n_2-p}\ , F_{p-1, n_1+n_2-p}\,(\propto) \qquad (64-7)$$

وعندها فإننا نقبل فرضية العدم H_0 باحتمال ثقة $(1-\propto)$، إذا كانت :

$$T^2 \le T^2(\propto) = \frac{(n_1+n_2-2)(p-1)}{n_1+n_2-p}\ , F_{p-1, n_1+n_2-p}\,(\propto) \qquad (65-7)$$

ونرفض H_0 في حالة العكس بمستوى دلالة $\propto$.

س2: إذا كانت الصور الجانبية للمجتمعين متوازية فهل هي متوافقة (coincident) مع بعضها البعض، أي هل هي متساوية تقريباً ؟ إن مبرر هذا السؤال هو أنه إذا كانت الصور الجانبية متوازية فقد تكون مختلفة عن بعضها وتكون مقادير الاختلاف بين عناصرها متساوية تقريباً، وقد تكون عناصر المجتمع الأول أكبر من الثاني أو العكس، وحتى نبرهن على أن عناصر الصور الجانبية متوافقة نلاحظ أن الصور الجانبية تكون متوافقة فقط إذا كان مجموع متوسطات عناصر المجتمع الأول يساوي مجموع متوسطات عناصر المجتمع الثاني . أي عندما يتحقق الشرط التالي :

$$\mu_{11} + \mu_{12} + \mu_{13} + \cdots + \mu_{1p} = \mu_{21} + \mu_{22} + \mu_{23} + \cdots + \mu_{2p} \qquad (66-7)$$

ويمكن كتابة هذا الشرط على شكل جداء شعاعين كما يلي :

$$[1\,,1\;1\,,\ldots 1]*\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{13}\\ \vdots\\ \mu_{1p}\end{bmatrix}=[1\,,1\,,1\ldots 1]*\begin{bmatrix}\mu_{21}\\ \mu_{22}\\ \mu_{23}\\ \vdots\\ \mu_{2p}\end{bmatrix} \qquad (67-7)$$

وإذا رمزنا للشعاع الواحدي بالرمز **1** فإن الشرط السابق يكتب كما يلي :

$$\mathbf{1}'*\boldsymbol{\mu_1}=\mathbf{1}'*\boldsymbol{\mu_2} \qquad (68-7)$$

وبناء على ذلك نضع فرضية العدم الثانية على الشكل التالي :

$$H_{02}:\ \mathbf{1}'*\boldsymbol{\mu_1}=\mathbf{1}'*\boldsymbol{\mu_2} \qquad (69-7)$$

ويمكننا اختبار هذه الفرضية بواسطة اختبار t للفرق بين متوسطي قيم متحول طبيعي واحد X في عينتين مستقلتين والذي يأخذ الشكل التالي :

$$T^2=\mathbf{1}'(\overline{\boldsymbol{X}}_\mathbf{1}-\overline{\boldsymbol{X}}_\mathbf{2})\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{1}'\boldsymbol{S_p}\mathbf{1}\right]^{-1}\mathbf{1}'(\overline{\boldsymbol{X}}_\mathbf{1}-\overline{\boldsymbol{X}}_\mathbf{2}) \qquad (86-6)$$

والذي يتحول بعد الاصلاح إلى الشكل التالي (لأنها أعداد) :

$$T^2=\left[\frac{\mathbf{1}'(\overline{\boldsymbol{X}}_\mathbf{1}-\overline{\boldsymbol{X}}_\mathbf{2})}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{1}'\boldsymbol{S_p}\mathbf{1}}}\right]^2=t^2_{n_1+n_2-2}\left(\frac{\propto}{2}\right)=F_{1\ n_1+n_2-2}\ (\propto) \qquad (\mathbf{69-7})$$

وبمقارنة قيمة T^2 المحسوبة مع قيمتها الحرجة $t^2_{n_1+n_2-2}\left(\frac{\propto}{2}\right)$ أو مع $F_{1\ n_1+n_2-2}(\propto)$ ، نتخذ القرار حول H_{02} كما يلي :

نقبل الفرضية H_{02} وباحتمال قدره $(1-\propto)$ إذا كانت :

$$T^2\le F_{1\ n_1+n_2-2}(\propto) \qquad (70-9)$$

ونرفض H_{02} إذا كانت $T^2>F_{1\ n_1+n_2-2}(\propto)$ بمستوى دلالة $\propto$.

وهنا نذكّر أنه في حالة اختبار فرضية توافق عناصر الصورتين الجانبيتين يجب أن تكون القياسات المشاهدة للمتحول X مسحوبة من مجتمعين خاضعين لتوزيع طبيعي موحد $N(\mu\,,\sigma^2)$.

س3: إذا كانت متوسطات الصورة الجانبية متوافقة (متساوية تقريباً) فهل هي ثابتة لجميع المتحولات؟

(وهذا يشترط أن تكون هذه المتحولات مقاسة بوحدة قياس موحدة) ، وحتى نستطيع مناقشة هذه القضية ، ننطلق من أن الفرضيتين H_{01} و H_{02} محققتين، وعندها فإن شعاع المتوسط المشترك $\boldsymbol{\mu}$ لهما يقدر بواسطة (n_1+n_2) مشاهدة وباستخدام العلاقة التالية :

$$\widetilde{\boldsymbol{\mu}}=\overline{\boldsymbol{X}}=\frac{\sum_{i=1}^{n_1}\boldsymbol{X_{1i}}+\sum_{i=1}^{n_2}\boldsymbol{X_{2i}}}{n_1+n_2}=\frac{n_1}{n_1+n_2}\overline{\boldsymbol{X}}_\mathbf{1}+\frac{n_2}{n_1+n_2}\overline{\boldsymbol{X}}_\mathbf{2} \qquad (71-7)$$

وإذا كانت متوسطات الصور الجانبية ثابتة، فهذا يعني أن :

$$\begin{aligned}\boldsymbol{\mu_1}&=\boldsymbol{\mu_2}\\ \boldsymbol{\mu_1}-\boldsymbol{\mu_2}&=\mathbf{0}\end{aligned} \qquad (\mathbf{90-6})$$

وإن فرضية العدم الثالثة H_{03} حول ذلك تأخذ الشكل التالي :

$$H_{03} : \boldsymbol{C} * \boldsymbol{\mu} = 0 \qquad (72-7)$$

حيث **C** هي المصفوفة المتعارضة المعطية في (7–56) .

وبطريقة مشابهة لما عرضناه في حالة H_{01} نجد أن مؤشر الاختبار المناسب لهذه الفرضية هو T^2 المعرف كما يلي:

$$T^2 = (n_1 + n_2)\overline{\boldsymbol{X}}'\boldsymbol{C}' * \left[\boldsymbol{C} * \boldsymbol{S_p} * \boldsymbol{C}'\right]^{-\mathbf{1}} * \boldsymbol{C}\overline{\boldsymbol{X}} \qquad (73-7)$$

وهو مقدار يخضع تقاربياً للتوزيع $F_{p-1\,,n_1+n_2-p}$.

ولذلك نقارن T^2 المحسوبة مع $T^2(\propto)$ الحرجة أو مع $F_{p-1\,,n_1+n_2-p}(\propto)$ ونتخذ القرار كما يلي :

نقبل الفرضية H_{03} إذا كانت :

$$T^2 \le F_{p-1\,,n_1+n_2-p}(\propto) \qquad (74-7)$$

ونرفض H_{03} إذا كان العكس .

ملاحظة: لا يمكن إجراء اختبار H_{03} إلا إذا كانت جميع المتحولات مقاسة بواحدة قياس موحدة .

مثال (7–2): [مأخوذ من Johnston, Wichem ص 246 بتصرف وإضافة]

في دراسة عن الحب والزواج، قام فريق من الباحثين الاجتماعيين بسحب عينة من الذكور المتزوجين بحجم $n_1 = 30$ وعينة من النساء المتزوجات بحجم $n_2 = 30$. وتوجهوا إليهم بسؤالين عن تقييمهم لعملية الإقدام على الزواج ولحصيلة ذلك الزواج كما يلي :

1–كيف تقيم عملية إقدامك على الزواج ؟ ورمزوا له بـ X_1 .

2–كيف تقيم حصيلة الزواج التي قمت به ؟ ورمزوا له بـ X_2 .

وطلبوا منهم أن تكون أجوبتهم بوضع إشارة على أحد (8) خيارات كما يلي :

جدول (7–5): خيارات الإجابة على السؤالين (1) و(2)

8	7	6	5	4	3	2	1
إيجابي لأقصى درجة	إيجابي جداً	إيجابي وسط	إيجابي ضعيف	سلبي ضعيف	سلبي وسط	سلبي جداً	سلبي لأقصى درجة

كما توجهوا إليهم بسؤالين آخرين عن تبادل العواطف والمجاملات كما يلي :

3– ماهي درجة تبادل عواطف المحبة التي تشعر بها مع شريكك ؟ ورمزوا له بـ X_3 .

4–ماهي درجة تبادل المجاملات الودية التي تشعر بها مع شريكك ؟ ورمزوا له بـ X_4 .

وطلبوا من أفراد العينتين الإجابة على هذين السؤالين بوضع إشارة على أحد الخيارات الخمسة التالية :

جدول (7–6): خيارات الإجابة على السؤالين (3) و(4)

5	4	3	2	1
كثيرة جداً	كثيرة	أحياناً	قليلة	قليلة جداً

وبعد جمع الإجابات من أفراد هاتين العينتين تم حساب شعاعي المتوسطات في كل منهما حسب التثقيلات المرافقة، فكانا كما يلي :

$$\bar{\boldsymbol{X}}_{\mathbf{1}} = \begin{bmatrix} 6{,}833 \\ 7{,}033 \\ 3{,}967 \\ 4{,}700 \end{bmatrix} \qquad \bar{\boldsymbol{X}}_{\mathbf{2}} = \begin{bmatrix} 6{,}633 \\ 7{,}000 \\ 4{,}000 \\ 4{,}533 \end{bmatrix}$$

علماً بأن المتحولين X_1 و X_2 يأخذان قيمهما من (1 حتى 8) حسب السؤالين (1) و(2)

وأن المتحولين X_3 و X_4 يأخذان قيمهما من (1 حتى 5) حسب السؤالين (3) و(4)

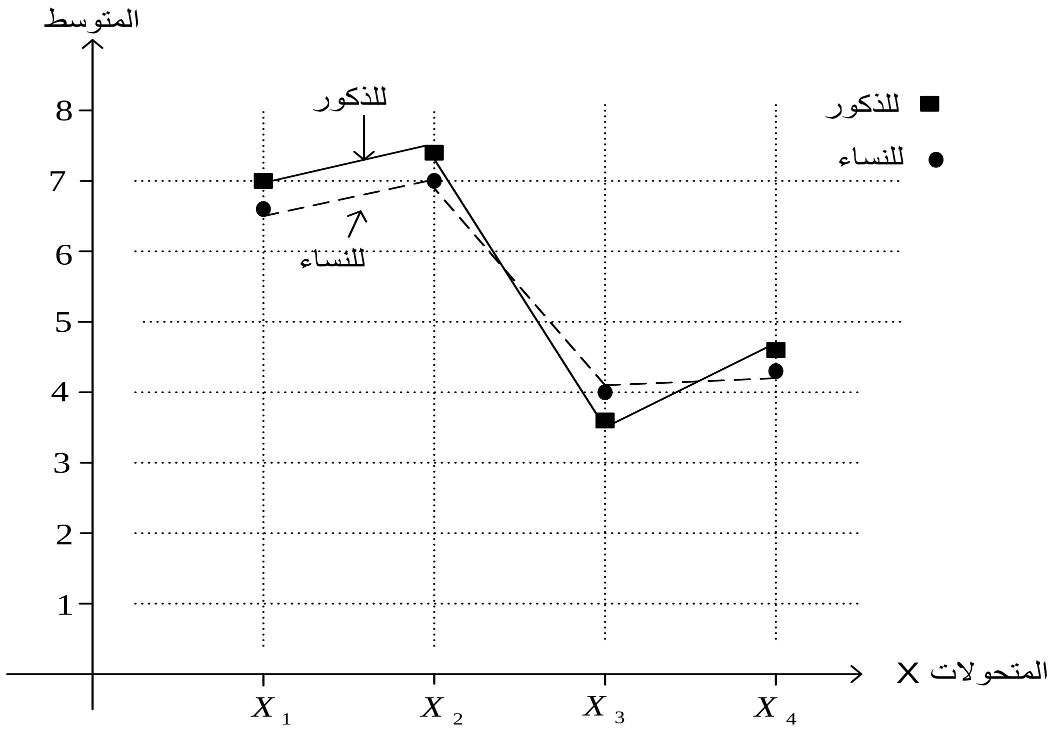
وبفرض أن مصفوفة التباين المشترك موحدة وتساوي$\mathbf{V}$، ولتقدير $\mathbf{V}$ قاموا بحساب مصفوفات التباين المشترك لهذه المتحولات في كلتا العينتين فحصلوا على$\boldsymbol{S}_{\mathbf{1}}$ و $\boldsymbol{S}_{\mathbf{2}}$. ثم قاموا بحساب المصفوفة المدمجة للتباين المشترك$\boldsymbol{S}_{\boldsymbol{p}}$ من العلاقة:

$$\boldsymbol{S}_p = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2}$$

وبعد إجراء الحسابات اللازمة حصلوا على أن:

$$\boldsymbol{S}_{\boldsymbol{p}} = \begin{bmatrix} 0.606 & 0.262 & 0.066 & 0.161 \\ 0.262 & 0.637 & 0.173 & 0.143 \\ 0.066 & 0.173 & 0.810 & 0.029 \\ 0.161 & 0.143 & 0.029 & 0.306 \end{bmatrix}$$

ثم قاموا برسم الصورة الجانبية لشعاعي المتوسطات في المجتمعين فكانا كما يلي :



الشكل (7–3): الصورة الجانبية للمتوسطات

وبما أن حجما العينتين كبيران، تم الاستناد إلى نظرية النهاية المركزية واعتبار هذه المتحولات طبيعية (رغم إنها تأخذ قيماً صحيحة) . ولإجراء الاختبارات اللازمة على هذه البيانات بدأوا باختبار التوازي كما يلي :

1– **اختبار التوازي :** وهو اختبار صحة الفرضية : $H_{01} : \boldsymbol{C\mu_1} = \boldsymbol{C\mu_2}$

ولاختبار هذه الفرضية كان عليهم تطبيق الاختبار T^2 المعرف في (7–63) ولهذا قاموا بحساب الجداء التالي :

$$\boldsymbol{C} * \boldsymbol{S_p} * \boldsymbol{C'} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \boldsymbol{S_p} \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$\boldsymbol{C} * \boldsymbol{S_p} * \boldsymbol{C'} = \begin{bmatrix} 0.719 & -0.268 & -0.125 \\ -0.268 & 1.101 & -0.751 \\ -0.125 & -0.251 & 1.058 \end{bmatrix}$$

ثم قاموا بحساب الجداء التالي :

$$\boldsymbol{C} * (\overline{\boldsymbol{X}}_1 - \overline{\mathbf{X}}_2) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0.200 \\ 0.033 \\ -0.033 \\ 0.167 \end{bmatrix} = \begin{bmatrix} -0.167 \\ -0.066 \\ 0.200 \end{bmatrix}$$

ثم قاموا بحساب T^2 من العلاقة (7–63) كما يلي:

$$T^2 = [-0.167\,, -0.066\,, 0.200] * \left(\frac{1}{30} + \frac{1}{30}\right)^{-1} S_p^{-1} * \begin{bmatrix} -0.167 \\ -0.066 \\ 0.200 \end{bmatrix}$$

$$T^2 = 15(0.67) = 1.005$$

وبفرض أن: $\propto = 0.05$ قاموا بإيجاد القيمة الحرجة $T^2(\propto)$ من العلاقة (7–64) فكانت كما يلي :

$$T^2(\propto) = \left[\frac{(30+30-2)(4-1)}{30+30-4}\right] * F_{1\ 56}(0.05) = 3.11 * (2{,}8) = 8{,}7$$

وبمقارنة T^2 المحسوبة مع القيمة الحرجة وجدوا أن :

$$T^2 = 1.005 < 8{,}7 = T^2(\propto)$$

لذلك تم قبول الفرضية H_{01} ، التي تنص على توازي الصورة الجانبية لمتوسطات للرجال والنساء واعتبروها ممكنة، وذلك كما هو موضح على الشكل (7–3) .

2– **ولاختبار فيما إذا كانت تلك المتوسطات متوافقة** (متساوية تقريباً)، انطلقوا من نتيجة الاختبار الأول الذي يعتبر أن الصورة الجانبية للمتوسطات متوازية، وعملوا على اختبار الفرضية الثانية التالية :

$$H_{02} : \mathbf{1}' * \boldsymbol{\mu_1} = \mathbf{1}' * \boldsymbol{\mu_2}$$

ولاختبار هذه الفرضية تم تطبيق المؤشر T^2 المعرف في (7–69) لذلك قاموا بحساب عناصره كما يلي:

$$\mathbf{1}'(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) = [1\,, 1\,, 1\,, 1] * \begin{bmatrix} 0.200 \\ 0.033 \\ -0.033 \\ 0.167 \end{bmatrix} = 0.367$$

وهو مجموع عناصر $(\bar{X}_1 - \bar{X}_2)$:

$$\mathbf{1}'\boldsymbol{S_p}\mathbf{1} = [1\,,1\,,1\,,1]\boldsymbol{S}_p\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = 4.204$$

وهوعبارة عن مجموع عناصر المصفوفة $\boldsymbol{S_p}$.

ثم قاموا بحساب قيمة المؤشر T^2 من العلاقة (7–69) كما يلي:

$$T^2 = \left(\frac{0.367}{\sqrt{\left(\frac{1}{30}+\frac{1}{30}\right)4.027}}\right)^2 = 0.708$$

وبفرض أن $\propto = 0.05$ تم إيجاد القيمة الحرجة $T^2(\propto)$ من العلاقة:

$$T^2(0.05) = F_{1\ 58}(0.05) = 4.0$$

وبمقارنة قيمة T^2 مع $T^2(\propto)$ نجد أن :

$$T^2 = 0.708 < 4.0$$

وبذلك تم قبول الفرضية H_{02} والتي تنص على أن متوسطات الإجابات عند الرجال متوافقة مع متوسطات الإجابات عند النساء .

3– **ثم انتقلوا لاختبار الفرضية الثالثة** H_{03} التي تنص على أن قيم الصورة الجانبية ثابتة (تساوي مقداراً ثابتاً) . فلاحظوا مباشرة أنه لا يمكن إجراء هذا الاختبار على المتحولات المذكورة، لأنها لا تقاس بواحدة قياس موحدة . فالمتحولان X_1 وX_2 يقاسان بواحدات قياس من (1 حتى 8)، أما المتحولان X_3 وX_4 فيقاسان بواحدات قياس من (1 حتى 5) . وهذا الأمر يجعل تطبيق الاختبار الثالث دون معنى ما لم يتم توحيد واحدات القياس لهذه المتحولات.

وأخيراً نشير إلى أنه إذا كانت حجوم العينات صغيرة فإن تحليل الصورة الجانبية سيعتمد على فرضية المتحولات الطبيعية. ويمكن التأكد من تحقق هذه الفرضية بإجراء أحد اختبارات الطبيعية المذكورة في الفصل الثالث .

ولإجراء تحليل الصورة الجانبية لعدة مجتمعات (أكثر من 2) نتبع نفس الإجراءات المعروضة لمجتمعين وإن المقاييس العامة للمقارنة مشابهة لتلك المقاييس المستخدمة في هذه الفقرة .

7–3: تحليل التباين المتعدد باتجاهين (n مشاهدة) (MANOVA tow way)

إن تحليل التباين المتعدد باتجاهين هو تعميم لتحليل التباين البسيط باتجاهين، ويختلف جوهرياً عن تحليل التباين المتعدد باتجاه واحد . وهو يُستخدم لدراسة تغيرات عدة متحولات (p متحولاً) هي: $X_1\,,X_2\,,\dots,X_p$ ، الناتجة عن تأثير عاملين F_A وF_B (وليس عن مجتمعات محددة) . وإن الباحث يمكن أن يتحكم بهذين العاملين خلال مجريات التجارب، ثم يقوم بدراسة تأثيرهما على تلك المتحولات .

فمثلاً: يمكننا دراسة تغيرات أسعار الفساتين X وكميات المبيعات منها Y بتأثير عاملين هما نوعية القماش F_A وله (3) أنواع أو حالات، وشكل الزخرفة F_B وله شكلان فقط، وعندها يجب علينا تأمين قياس واحد على الأقل (أو n قياساً) في كل حجرة مقابلة لتقاطعات حالات هذين العاملين ولكل من المتحولين X و Y ووضعها في جدول مناسب كما يلي :

جدول (7-7): بيانات MANOVA باتجاهين tow way .

حالات F_A / حالات F_B	A_1		A_2		A_3	
B_1	X	Y	X	Y	X	Y
	X_{111} X_{112} $\vdots$ X_{11n}	Y_{111} Y_{112} $\vdots$ Y_{11n}	X_{121} X_{122} $\vdots$ X_{12n}	Y_{121} Y_{122} $\vdots$ Y_{12n}	X_{131} X_{132} $\vdots$ X_{13n}	Y_{131} Y_{132} $\vdots$ Y_{13n}
B_2	X	Y	X	Y	X	Y
	X_{211} X_{212} $\vdots$ X_{21n}	Y_{211} Y_{212} $\vdots$ Y_{21n}	X_{221} X_{222} $\vdots$ X_{22n}	Y_{221} Y_{222} $\vdots$ Y_{22n}	X_{231} X_{232} $\vdots$ X_{23n}	Y_{231} Y_{232} $\vdots$ Y_{23n}

وبطريقة مشابهة لما فعلناه في حالة التحليل البسيط ذي الاتجاهين، وباستخدام نفس الرموز السابقة، يمكننا التعبير عن مركبات شعاع المتحولات $X_1, X_2, \dots, X_p$ ومتوسطاتها وتوقعاتها كما يلي :

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \qquad \bar{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \qquad (76-7)$$

ولصياغة النموذج الرياضي نقوم باستبدال الرموز العددية في العلاقات السابقة بالأشعة المقابلة ، فنجد أن العلاقة (6–100) نأخذ الشكل الشعاعي التالي :

$$\boldsymbol{X_{k\ell i}} = \boldsymbol{\mu} + \propto_{\boldsymbol{k}} + \boldsymbol{\beta_\ell} + \boldsymbol{\gamma_{k\ell}} + \boldsymbol{e_{k\ell i}} \qquad (77-7)$$

حيث أن: $k = 1\ 2\ 3 \dots g$: دليل عدد حالات العامل الأول F_A

وأن: $\ell = 1\ 2\ 3 \dots q$: دليل عدد حالات العامل الثاني F_B

وأن: $i = 1\ 2\ 3 \dots n$: دليل عناصر العينة في كل حجرة **(بفرض أن حجوم العينات متساوية** $n =$)

وهنا نشير إلى أن الأشعة $\propto_{\boldsymbol{k}}$ و $\boldsymbol{B_\ell}$ و $\boldsymbol{\gamma_{k\ell}}$ هي أعمدة من المرتبة $(P * 1)$ ويجب أن تحقق حسب (6–98) الشروط التالية :

$$\sum_{k=1}^{g} \propto_{k} = 0 \quad \sum_{\ell=1}^{q} \boldsymbol{B}_{\ell} = 0$$
$$\sum_{k=1}^{g} \boldsymbol{\gamma}_{k\ell} = 0 \quad \sum_{\ell=1}^{q} \boldsymbol{\gamma}_{k\ell} = 0 \qquad (78-7)$$

وإن كل من الأشعة $\boldsymbol{e}_{k\ell i}$ هي أشعة عشوائية في كل حجرة وتخضع للتوزيع الطبيعي $N(0\ ,\boldsymbol{V})$، الذي توقعه صفر وله مصفوفة موحدة $\mathbf{V}$ للتباين المشترك . وكل من هذه الأشعة يتضمن n قياساً لـ p متحولاً في كل تقاطع ممكن لحالات العاملين F_A و F_B .

وبناءً على العلاقة (6-101) يمكننا تقدير شعاع المشاهدات $X_{k\ell i}$ وكتابة العلاقة (6-111) كما يلي:

$$\boldsymbol{X}_{k\ell i} = \overline{\boldsymbol{X}} + (\overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}}) + (\overline{\boldsymbol{X}}_{\ell} - \overline{\boldsymbol{X}}) + (\overline{\boldsymbol{X}}_{k\ell} - \overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}}_{\ell} + \overline{\boldsymbol{X}}) + (\boldsymbol{X}_{k\ell i} - \overline{\boldsymbol{X}}_{k\ell}) \quad (79-7)$$

حيث أن $\overline{\boldsymbol{X}}$ هو شعاع المتوسط الكلي لأشعة المشاهدات التجريبية .

وأن $\overline{\boldsymbol{X}}_{k}$ هو شعاع متوسط أشعة المشاهدات في الحالة k للشعاع F_A .

وأن $\overline{\boldsymbol{X}}_{\ell}$ هو شعاع متوسط أشعة المشاهدات في الحالة ℓ للشعاع F_B .

وأن $\overline{\boldsymbol{X}}_{k\ell}$ هو شعاع متوسط أشعة المشاهدات المتقاطعة للحالة k للشعاع F_A مع الحالة ℓ للشعاع F_B.

ثم نقوم بطرح $\overline{\boldsymbol{X}}$ من طرفي العلاقة (7-79) وتربيع حدود الطرفين، ثم نأخذ المجاميع الثلاثة لهذه المربعات وتقاطعاتها، فنحصل على العلاقة التالية (بدلالة الأشعة) :

$$\begin{aligned}\sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(\boldsymbol{X}_{k\ell i} - \overline{\boldsymbol{X}})(\boldsymbol{X}_{k\ell i} - \overline{\boldsymbol{X}})' &= qn\sum(\overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}})' \\ &+gn\sum(\overline{\boldsymbol{X}}_{\ell} - \overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}_{\ell} - \overline{\boldsymbol{X}})' \\ &+n\sum_{k=1}^{g}\sum_{\ell=1}^{q}(\overline{\boldsymbol{X}}_{k\ell} - \overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}}_{\ell} + \overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}_{k\ell} - \overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}}_{\ell} + \overline{\boldsymbol{X}})' \\ &+\sum_{\boldsymbol{k=1}}^{\boldsymbol{g}}\sum_{\boldsymbol{\ell=1}}^{\boldsymbol{q}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\boldsymbol{X}_{k\ell i} - \overline{\boldsymbol{X}}_{k\ell})\,(\boldsymbol{X}_{k\ell i} - \overline{\boldsymbol{X}})'\end{aligned} \qquad (80-7)$$

واختصاراً نرمز لهذه الحدود بالرموز التالية :

$$SSPT = SSPA + SSPB + SSPAB + SSPE \qquad (81-7)$$

علماً بأن درجات الحرية المقابلة لهذه الحدود تساوي :

$$g*q*n-1 = (g-1) + (q-1) + (g-1)(q-1) + g*q(n-1) \qquad (82-7)$$

وهنا نلاحظ أن حدود العلاقة (7-80) مستخرجة من حدود العلاقة (6-104) في (ANOVA tow way)، وذلك بتعميم حدود المتحول الواحد إلى التحليل المتعدد وبدلالة الأشعة واستبدال كل مربع عددي مثل: $(\bar{X}_k - \bar{X})^2$ بالمصفوفة الناتجة عن جداء الشعاع بمنقوله $(\overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}_{k} - \overline{\boldsymbol{X}})'$..وهكذا دواليك.

وأخيراً نضع نتائج الحسابات في جدول MANOVA ذي الاتجاهين والذي يأخذ الشكل التالي :

جدول (7−8): جدول MANOVA باتجاهين :

مصدر التباين	مصفوفة مجموع مربعات الانحرافات وجداء تقاطعاتها SSP	درجة الحرية	Λ^*
العامل الأول Factor A	$SSPA = qn\sum_{k=1}^{g}(\bar{X}_k - \bar{X})(\bar{X}_k - \bar{X})'$	$(g-1)$	
العامل الثاني Factor B	$SSPB = gn\sum_{\ell=1}^{q}(\bar{X}_\ell - \bar{X})(\bar{X}_\ell - \bar{X})'$	$(q-1)$	
تداخل أو تفاعل العاملين A و B	$SSPAB = n\sum_{k=1}^{g}\sum_{\ell=1}^{q}(\bar{X}_{k\ell} - \bar{X}_k - \bar{X}_\ell + \bar{X}) * * (\bar{X}_{k\ell} - \bar{X}_k - \bar{X}_\ell + \bar{X})'$	$(g-1)(q-1)$	$\Lambda^*_{AB} = \frac{\lvert SSPE\rvert}{\lvert SSPAB + SSPE\rvert}$
الخطأ العشوائي أو البواقي	$SSPE = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(X_{k\ell i} - \bar{X}_{k\ell})\,(X_{k\ell i} - \bar{X}_{k\ell})'$	$gq(n-1)$	
الاجمالي (المصحح أو المميز)	$SSPT = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(X_{k\ell i} - \bar{X})\,(X_{k\ell i} - \bar{X})'$	$g*q*n-1$	

وبعد أعداد هذا الجدول يكون لدينا ثلاث فرضيات مختلفة هي:

الفرضية الإولى: فرضية عدم التداخل : وهي فرضية تنص على عدم وجود تأثير لتداخل العاملين F_A و F_B على المتحولات التابعة $X_1, X_2, \dots, X_p$ ، ونكتب ذلك كما يلي :

$$H_{01}: \gamma_{11} = \gamma_{12} = \cdots = \gamma_{gq} = 0 \qquad (83-7)$$

مقابل الفرضية البديلة: من أجل زوج واحد على الأقل $(k \neq \ell)$ $\quad H_{11}: \gamma_{k\ell} \neq 0$

أما مؤشر الاختبار المناسب لذلك فهو المؤشر Λ^* (المرتبط بنسبة الإمكانية العظمى Λ) والمعرف بالعلاقة :

$$\Lambda^*_{AB} = \frac{|SSPE|}{|SSPAB + SSPE|} \qquad (wilk's\ lambda) \qquad (84-7)$$

ويتم رفض H_0 إذا كانت قيمة Λ^*_{AB} صغيرة بقدر كافٍ لتحقيق مستوى الدلالة $\propto$ وهنا يشترط أن يكون : $p \leq gq(n-1)$ وذلك حتى تكون المصفوفة SSPE محددة إيجابياً .

أما عندما تكون حجوم العينات n كبيرة . فإنه يمكن تحويل المؤشر Λ^* ليصبح خاضعاً تقاربياً للتوزيع χ^2 ، ثم تطبيق أسلوب اختبار χ^2 عليه ، وذلك باستخدام تحويل Bartlitt المعرف في (7−38) لتحويل توزيع Λ^* تقاربياً إلى التوزيع χ^2 مع الاحتفاظ بشكل فرضية العدم السابقة :

$H_{01}: \boldsymbol{\gamma_{11}} = \boldsymbol{\gamma_{12}} = \cdots = \boldsymbol{\gamma_{gq}} = 0$. وعندها فإننا نقبل فرضية العدم H_0 باحتمال ثقة $(1-\propto)$ إذا تحقق الشرط التالي:

$$-\left[gq(n-1) - \frac{(p+1)-(g-1)(q-1)}{2}\right] ln\,\Lambda^*_{AB} \leq \chi^2_{(g-1)(q-1)*p}(\propto) \qquad \mathbf{(85-7)}$$

حيث أن Λ^* هي القيمة المحسوبة من العلاقة (7−84)، وأن $\chi^2_{(g-1)(q-1)*p}(\propto)$ هي القيمة الحرجة لمتحول χ^2 المقابل لمستوى الدلالة $\propto$ ولدرجة الحرية $(g-1)(q-1)*p$

ونرفض الفرضية H_0 بمستوى دلالة $\propto$ إذا كان الشرط (7−85) غير محقق .

ملاحظة هامة :

إن منطق الأفضليات يقتضي أن يتم تنفيذ اختبار تأثير التداخل قبل إجراء اختباري تأثير العاملين F_A و F_B . فإذا كان تأثير التداخل موجوداً، فلن يكون لتأثير العاملين F_A و F_B تفسيراً واضحاً . وعندها ومن وجهة نظر عملية ليس من الصواب متابعة إجراء الاختبارات المتعددة حول تأثير العاملين F_A و F_B . وبدلاً من ذلك يجب إجراء تحليل التباين البسيط ذي الاتجاهين (لكل متحول X_j بمفرده) لتحري ظهور التداخل في بعض هذه المتحولات وعدم ظهوره في بعضها الآخر، وإن المتحولات التي لا تظهر أي تداخل لتأثير للعاملين F_A و F_B فيها، تمكننا من تفسير تأثير كل منهما (إن وجد) من خلال مصطلحات التأثير الخاص لهذين العاملين، ولكن بشرط أن تكون التأثيرات الأخيرة موجودة .

:والآن لنفترض أنه لا يوجد تداخل لتأثير العاملين F_A و F_B على المتحولات $X_1, X_2, \dots, X_p$. وعندها فقط ننطلق لاختبار التأثيرات الرئيسية لكل من F_A و F_B .

الفرضية الثانية: عدم تأثير العامل الأول: لاختبار تأثير F_A ، نضع فرضية العدم الثانية H_{02} مقابل الفرضية البديلة H_{12} بالنسبة لـ F_A كما يلي :

$$H_{02}:\ \propto_{\mathbf{1}}=\ \propto_{\mathbf{2}}=\cdots=\ \propto_{\boldsymbol{g}}=0$$

$$H_{12}:\ \propto_{\boldsymbol{k}}\neq \mathbf{0} \qquad \text{وذلك من أجل حالة واحدة على الأقل} \qquad (80-7)$$

وإن مؤشر الاختبار المناسب لهذه الفرضية هو المؤشر Λ_A^* المعرف بالعلاقة التالية :

$$\Lambda_A^*=\frac{|SSPE|}{|SSPA+SSPE|} \qquad (88-7)$$

ويتم رفض H_0 إذا كانت قيمة Λ_A^* صغيرة بقدر كاف لتحقيق مستوى الدلالة $\propto$.

وعندما تكون حجوم العينات n كبيرة فإننا نطبق تحويل Bartlitt على Λ_A^* ، حتى يصبح خاضعاً تقاربياً لتوزيع χ^2 . وعندها نقبل فرضية العدم باحتمال ثقة $(1-\propto)$ إذا تحقق الشرط التالي :

$$-\left[gq(n-1)-\frac{(p+1)-(g-1)}{2}\right] ln\,\Lambda_A^* \leq \chi^2_{(g-1)*p}(\propto) \qquad (89-7)$$

حيث أن Λ_A^* هي قيمة مؤشر Wilk's Lambda المحسوبة من العلاقة (7−88) .

أما $\chi^2_{(g-1)*p}(\propto)$ فهي القيمة الحرجة لمتحول توزيع χ^2 المقابلة لمستوى الدلالة $\propto$ ولدرجة الحرية $(g-1)p$.

الفرضية الثانية: عدم تأثير العامل الثاني: وبطريقة مشابهة نقوم باختبار تأثير العامل F_B ونضع فرضية العدم H_{03} مقابل الفرضية البديلة H_{13} كما يلي :

$$H_{03} : \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \cdots = \boldsymbol{\beta_q} = 0$$
$$H_{13} : \boldsymbol{\beta_\ell} \neq 0 \quad \text{وذلك من أجل حالة واحدة على الأقل} \qquad (90-7)$$

وإن مؤشر الاختبار المناسب لذلك F_B هو المؤشر Λ_B^* المعرف بالعلاقة التالية :

$$\Lambda_B^* = \frac{|SSPE|}{|SSPB + SSPE|} \qquad (91-7)$$

وإننا نرفض H_0 إذا كانت قيمة Λ_B^* صغيرة بقدر كافٍ لتحقيق مستوى الدلالة $\propto$.

أما إذا كانت حجوم العينات n كبيرة فإننا نطبق تحويل Bartlitt على Λ_B^* لنجعله خاضعاً تقاربياً للتوزيع χ^2 . وعندها نقبل فرضية العدم H_{03} باحتمال ثقة $(1-\propto)$ إذا تحقق الشرط التالي :

$$-\left[gq(n-1) - \frac{(p+1)-(q-1)}{2}\right] ln\, \Lambda_B^* \leq \chi^2_{(q-1)*p}(\propto) \qquad (92-7)$$

حيث أن Λ_B^* هي قيمة مؤشر Wilk's Lambda المحسوبة من العلاقة (7-91) .

أما $\chi^2_{(q-1)p}(\propto)$ فهي القيمة الحرجة لمتحول χ^2 المقابلة لمستوى الدلالة $\propto$ ولدرجة الحرية $(q-1)p$.

أما بالنسبة لإنشاء مجالات الثقة المتزامنة أولمقارنة الفروقات بين معالم النموذج، يمكننا الاستناد إلى طبيعة تأثيرات هذه العوامل، والاعتماد على العلاقات التي استخرجناها في تحليل التباين البسيط ذي الاتجاهين.

وعندما وفقط عندما يكون تأثير التداخل مهملاً، فإنه يمكننا التركيز على مقارنة الفروقات بين التأثيرات الرئيسية للعاملين F_A و F_B واختبار تأثير كل منهما.

كما يمكننا تطبيق أسلوب (بونفيروني) لإنشاء المجالات المتزامنة للفروقات بين المركبات $(\propto_{\boldsymbol{k}} - \propto_{\boldsymbol{m}})$ لتأثيرات العامل F_A ، وللفروقات $(\beta_\ell - \beta_h)$ لتأثيرات العامل F_B .

وبذلك نجد أن المجال المتزامن حسب أسلوب (بونفيروني)، الذي تنتمي إليه مركبات الفروقات $(\propto_{kj} - \propto_{mj})$ باحتمال ثقة $(1-\propto)$ هو المجال التالي :

$$(\propto_{kj} - \propto_{mj}) = (\bar{x}_{kj} - \bar{x}_{mj}) \pm t\left(\frac{\propto}{pg(g-1)}\right) * \sqrt{\frac{E_{jj}}{v} * \frac{2}{qn}} \qquad \mathbf{(93-7)}$$

حيث أن: $v = g * q(g-1)$ وهي درجة الحرية لـ SSPT

وأن: E_{jj} هي قيمة العنصر القطري j في المصفوفة SSPE

وأن: الفرق $(\bar{x}_{kj} - \bar{x}_{mj})$ هي المركبة j في الشعاع $(\bar{X}_k - \bar{X}_m)$.

وبطريقة مشابهة نجد أن مجال الثقة المتزامن، الذي تنتمي إليه الفروقات $(\beta_{\ell j} - \beta_{hj})$ باحتمال $(1-\propto)$ هو المجال التالي :

$$(\beta_{\ell j} - \beta_{hj}) = (\bar{x}_{\ell j} - \bar{x}_{hj}) \pm t\left(\frac{\propto}{pq(q-1)}\right) * \sqrt{\frac{E_{jj}}{v} * \frac{2}{qn}} \qquad \mathbf{(94-7)}$$

حيث أن: $v = g * q(g-1)$ وإن: E_{jj} هو العنصر القطري j في SSPE

وإن: الفرق $(\bar{x}_{\ell j} - \bar{x}_{hj})$ هو المركبة j في الشعاع $(\bar{X}_\ell - \bar{X}_h)$.

الحالة الخاصة ($\boldsymbol{n=1}$): لقد اعتبرنا عند إجراء تحليل التباين المتعدد باتجاهين أن المشاهدات تتكرر n مرة في كل حجرة مقابلة لتقاطعات حالات العاملين F_A و F_B (n موحدة ويمثل حجم العينة في كل حجرة) . وإن هذه الخاصية (التكرار) تستخدم لفحص وجود تأثيرات التداخل بين العاملين F_A و F_B .

ولكن يمكننا في بعض الحالات إجراء ذلك التحليل بالاعتماد على مشاهدة واحدة في كل حجرة ($n=1$) ، وعندها فإن النموذج الرياضي ذي الاتجاهين لا يعطينا الحد الخاص بتأثير التداخل $\gamma_{k\ell}$. وبذلك فإن جدول MANOVA في هذه الحالة يكون مختصراً ويتضمن فقط تأثيرات العامل F_A وتأثيرات العامل F_B وحد الخطأ العشوائي (البواقي) . وهذه التأثيرات هي التي تشكل مركبات التباين الكلي SSPT .

وإن النموذج لـ MANOVA في هذه الحالة ($n=1$) يأخذ الشكل التالي :

$$\boldsymbol{x_{k\ell}} = \boldsymbol{\bar{X}} + \boldsymbol{\propto_k} + \boldsymbol{\beta_\ell} + \boldsymbol{e_{k\ell}} \qquad (95-7)$$

حيث أن: $\propto_k$ في B_ℓ تحقق الشرطين التاليين :

$$\sum_{k=1}^{g} \boldsymbol{\propto_k} = 0 \qquad \sum_{\ell=1}^{g} \boldsymbol{\beta_\ell} = 0 \qquad (96-7)$$

وأن $\boldsymbol{e_{k\ell}}$: هي أشعة مستقلة وتخضع للتوزيع الطبيعي المتعدد $N(0\ ,\boldsymbol{V})$ ، الذي توقعه صفر ومصفوفة تباينه $\mathbf{V}$.

وبطريقة مشابهة للمعالجة التي أجريناها في الفقرة السابقة نقوم بالإجراءات التالية :

1- نكتب مشاهدات المتحولات $X_1 , X_2 , \ldots , X_p$ على شكل تصفيفات كما فعلنا في المثال (6-1) كما يلي:

$$\boldsymbol{x_{k\ell}} = \boldsymbol{\bar{X}} + (\boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}}) + (\boldsymbol{\bar{X}_\ell} - \boldsymbol{\bar{X}}) + (\boldsymbol{x_{k\ell}} - \boldsymbol{\bar{X}_k} - \boldsymbol{\bar{X}_\ell} + \boldsymbol{\bar{X}}) \qquad (98-7)$$

حيث أن $\bar{X}$: هو المتوسط الكلي . وأن $\bar{X}_k$ هو متوسط الحالة k للعامل F_A . وأن $\bar{X}_\ell$ هو متوسط الحالة ℓ للعامل F_B .

2- نعيد تنظيم أسطر التصفيفات التي حصلنا عليها في البند (1) ونشكل منها شعاعاً واحداً (طويلاً) ونحسب مجاميع مربعات الانحرافات ومجاميع الجداءات المتقاطعة لها فنحصل على ما يلي :

$$SSPT = SSPA + SSPB + SSPE \qquad (99-7)$$

والتي تقابلها درجات الحرية التالية :

$$gq - 1 = (g-1) + (q-1) + (g-1)(q-1) \qquad (100-7)$$

3- نلخص حسابات البند (2) في جدول MANOVA المختصر الذي يتألف من تأثيرات العاملين F_A و F_B والبواقي عنهما عندما يكون ($n=1$) . وهنا يجب أن نلاحظ أنه عندما يكون $n=1$ ، فإن SSPE في جدول الاتجاهين يصبح مساوياً لمصفوفة صفرية، وبدرجة حرية تساوي الصفر . وإن

مصفوفة التداخل بين العاملين (مجموع المربعات وجداءات التقاطعات) تصبح مساوية لمصفوفة الأخطاء (مجموع مربعات البواقي وجداءات تقاطعاتها والتي تقابل $(g-1)(q-1)$ درجة حرية) .

4-نستفيد من الخلاصة الحسابية التي حصلنا عليها في البند (3) لاختبار تأثيرات العاملين F_A و F_B وبمستوى دلالة $\propto = 0.05$.

ولإجراء هذا الاختبار نستخدم العلاقتين (6-130) و(6-131) ولكن بعد استبدال درجة الحرية $gq(n-1)$ بدرجة الحرية $(g-1)(q-1)$.

وهنا نؤكد على وجوب تحقق الشرط التالي: $p \le (g-1)(q-1)$ وذلك حتى تكون المصفوفة SSPE محددة إيجابياً باحتمال يساوي الواحد .

مثال (7-3): [مأخوذ من Jonhson, Wichern P. 255 بتصرف وإضافة]

لدراسة تغيرات (3) متحولات X_1, X_2, X_3، الناتجة عن تأثيرات عاملين F_A و F_B ، لكل منهما حالتين فقط (منخفضة ومرتفعة)، قمنا بأخذ عينة بحجم $(n=5)$ قياسات لهذه المتحولات في كل حجرة مقابلة لتقاطعات حالات العاملين F_A و F_B فحصلنا على البيانات الواردة في الجدول التالي :

جدول (7-9): بيانات المثال

العوامل	الحالات	العامل الثاني F_B الحالة المنخفضة Low X_1	X_2	X_3	العامل الثاني F_B الحالة المرتفعة High X_1	X_2	X_3
العامل الأول F_A	الحالة المنخفضة Low	[6,5	9,5	4,4]	[6,9	9,1	5,7]
		[6,2	9,9	6,4]	[7,2	10,0	2,0]
		[5,8	9,6	3,0]	[6,9	9,9	3,9]
		[6,5	9,6	4,1]	[6,1	9,5	1,9]
		[6,5	9,2	0,8]	[6,3	9,4	5,7]
	الحالة المرتفعة High	[6,7	9,1	2,8]	[7,1	9,2	8,4]
		[6,6	9,3	4,1]	[7,0	8,8	5,2]
		[7,2	8,3	3,8]	[7,2	9,7	6,9]
		[7,1	8,4	1,6]	[7,5	10,1	2,7]
		[6,8	8,5	3,4]	[7,6	9,2	1,9]

ولمتابعة الدراسة تم حساب مصفوفات مجموع المربعات والجداءات المتقاطعة للحالات الممكنة من العلاقات (7-80) وتم وضعها مباشرة في جدول MANOVA ذي الاتجاهين، والذي يأخذ الشكل التالي :

جدول (7–10): جدول MANOVA باتجاهين

مصدر التباين	رمز المصفوفة	المصفوفات المختصرة	درجة الحرية	قيمة المحدد D
العامل الأول F_A	SSPA	$\begin{matrix} 1.7505 & -1.5045 & 0.8555 \\ \dots\dots\dots & 1.3005 & -0.7.95 \\ \dots\dots\dots & \dots\dots\dots & 0.4205 \end{matrix}$	$g-1=1$	44.3114
العامل الثاني F_B	SSPB	$\begin{matrix} 0.7605 & 0.6825 & 1.9305 \\ \dots\dots\dots & 0.6125 & 1.7325 \\ \dots\dots\dots & \dots\dots\dots & 4.9005 \end{matrix}$	$q-1=1$	251.4289
التداخل لـF_A و F_B Interaction	SSPAB	$\begin{matrix} 0.0005 & 0.0165 & 0.0554 \\ \dots\dots\dots & 0.5445 & 1.4685 \\ \dots\dots\dots & \dots\dots\dots & 3.9605 \end{matrix}$	$(g-1)(q-1)=1$	79.0808
الخطأ العشوائي أو البواقي Residual	SSPE	$\begin{matrix} 1.7640 & 0.0200 & -3.0700 \\ \dots\dots\dots & 2.6280 & -0.5520 \\ \dots\dots\dots & \dots\dots\dots & 64.9240 \end{matrix}$	$gq(n-1)=16$	275.7098
الاجمالي (المصحح أو المعير بـ $\bar{X}$)	SSPT	$\begin{matrix} 4.2655 & -0.7855 & -0.2395 \\ \dots\dots\dots & 5.0855 & 1.9095 \\ \dots\dots\dots & \dots\dots\dots & 74.2055 \end{matrix}$	$gq\mathrm{n}-1=19$	1052.53

ولإجراء الاختبارات اللازمة، نبدأ باختبار التداخل (Interaction) لأنه يوضح لنا النتيجة مباشرة، فإذا كان تأثيره معنوياً، فإنه لا يمكننا تفسير تأثير العاملين F_A و F_B بشكل واضح، وعندها لا توجد جدوى من اختبار تأثير F_A و F_B .

أما إذا كان تأثير التداخل غير معنوي، فيمكن إهماله والانتقال إلى اختبار تأثير F_A ثم تأثير F_B .

ولإجراء اختبار التداخل نقوم بحساب قيمة المؤشر Λ^* من العلاقة (أنظر الجدول السابق):

$$\Lambda_{AB}^* = \frac{|SSPE|}{|SSPAB + SSPE|} = \frac{275.7098}{354.7906} = 0.7771$$

ثم نقوم بحساب قيمة المتحول F من العلاقة المناسبة لهذه الحالة $[(g-1)(q-1)=1]$ وهي:

$$F = \left(\frac{1-\Lambda_{AB}^*}{\Lambda_{AB}^*}\right) * \frac{\frac{gq(n-1)-p+1}{2}}{\frac{[|(g-1)(q-1)-p|+1]}{2}} = \left(\frac{1-\Lambda_{AB}^*}{\Lambda_{AB}^*}\right) * \frac{v_2/2}{v_1/2}$$

علماً بأن F يخضع تماماً للتوزيع $F(X)$ ذي درجتي الحرية $v_1 = |(g-1)(q-1)-p|+1$ و $v_2 = gq(n-1)-p+1$ وبالتعويض نجد أن :

$$F = \left(\frac{1-0.7771}{0.7771}\right) \frac{(2)(2)(4)-3+1}{\frac{|1*1-3|+1}{2}} = 1.34$$

ثم نبحث في جداول F وعند مستوى الدلالة $\propto = 0.05$ عن القيمة الحرجة $F_{v_1,v_2}(\propto)$ المقابلة لدرجتي الحرية

$$v_1 = |1*1-3|+1 = 3 \qquad , \quad v_2 = (2)(2)(4)-3+1 = 14$$

فنجد أن القيمة الحرجة $F_{v_1,v_2}(\propto)$ تساوي :

$$F_{v_1,v_2}(\propto) = F_{3\,,14}(0.05) = 3.34$$

وبمقارنة القيمة المحسوبة $F = 1.34$ مع القيمة الحرجة $F(\propto)$ نجد أنه : $1.34 < 3.34$ ، لذلك نقبل فرضية العدم H_0 التي تقول أن: $H_{01} : \boldsymbol{\gamma_{11}} = \boldsymbol{\gamma_{12}} = \boldsymbol{\gamma_{21}} = \boldsymbol{\gamma_{22}} = 0$. ونستنتج أنه لا يوجد تأثير معنوي لتداخل العاملين F_A و F_B .

ملاحظة: يمكن إجراء هذا الاختبار باستخدام طريقة χ^2 التقريبية ، ولذلك نقوم بحساب قيمة χ^2 من العلاقة (7–85) التالية:

$$\chi^2 = -\left[(2)(2)(5-1) - \frac{3+1-(1)(1)}{2}\right] ln(0.7771) = 3.66$$

وبمقارنة هذه القيمة المحسوبة لـχ^2 مع القيمة الحرجة $\chi^2_{v_1}(\propto)$ عند مستوى دلالة $\propto = 0.05$ والتي تساوي :

$$\chi^2_{v_1}(\propto) = \chi^2_3(0.05) = 7.81$$

نجد أن : $\chi^2 = 3.66 < 7.81$ ، لذلك نقبل فرضية العدم H_0 السابقة . ونحصل على نفس النتيجة التي حصلنا عليها من التوزيع F. وبما أن تأثير التداخل غير معنوي، فإننا نهمله وننتقل لاختبار تأثير العامل الأول F_A، ونقوم بحساب Λ^*_A من العلاقة المناسبة التالية :

$$\Lambda^*_A = \frac{|SSPE|}{|SSPA + SSPE|} = \frac{275.7098}{722.0121} = 0.3819$$

ثم نقوم بحساب قيمة المتحول F_A المناسبة لهذه الحالة $(g-1=1\ ,\ q-1=1)$ من العلاقة التالية:

$$F_A = \left(\frac{1-\Lambda^*_A}{\Lambda^*_A}\right) * \frac{(gq(n-1)-p+1)/2}{(|(g-1)-p|+1)/2}$$

$$F_A = \left(\frac{1-0.3819}{0.3819}\right)\frac{(16-3+1)/2}{(|1-3|+1)/2} = 7.55$$

ولمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $F_{v_1,v_2}(\propto)$ نفترض أن $\propto = 0.05$ ثم نحسب درجتي الحرية كما يلي :

$$v_1 = |1-3| + 1 = 3 \qquad , \quad v_2 = (16-3+1) = 14$$

ومن جداول F نجد أن :

$$F_{v_1,v_2}(\propto) = F_{3\,,14}(0.05) = 3.34$$

وبمقارنة هذه القيمة المحسوبة $F_A = 7.55$ مع القيمة الحرجة $F(\propto)$ نجد أن : $7.55 > 3.34$ ، لذلك نرفض فرضية العدم H_{02} التي تقول أن: $H_{02}: \propto_{\mathbf{1}} = \propto_{\mathbf{2}} = 0$ ، وذلك بمستوى دلالة $\propto = 0.05$.ونستنتج أنه يوجد تأثير معنوي للعامل F_A على تغيرات المتحولات المذكورة $X_1\ , X_2\ , X_3$.

وبطريقة مشابهة يمكننا اختبار تأثير العامل الثاني F_B، لذلك نقوم بحساب Λ_B^* من العلاقة المناسبة التالية:

$$\Lambda_B^* = \frac{|SSPE|}{|SSPB + SSPE|} = \frac{275.7098}{527.1347} = 0.5230$$

ثم نقوم بحساب قيمة المتحول F_B المناسبة لهذه الحالة $(g-1=1 \ , \ q-1=1)$ من العلاقة التالية :

$$F_B = \left(\frac{1-\Lambda^*}{\Lambda^*}\right) * \frac{(gq(n-1)-p+1)/2}{(|(q-1)-p|+1)/2} = \left(\frac{1-0.5230}{0.5230}\right) * \frac{(16-3+1)/2}{(|1-3|+1)/2} = 4.26$$

وبمقارنة هذه القيمة المحسوبة F_B مع القيمة الحرجة $F_{v_1,v_2}(\propto)$ والتي تساوي نفس القيمة السابقة .

$$F_{3\ ,14}(\propto) = F_{3\ ,14}(0.05) = 3.34$$

نجد أن $F_B = 4.26 > 3.34$، لذلك نرفض فرضية العدم H_{03} التي تقول: $H_{03}\text{:}\ \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = 0$, وذلك بمستوى دلالة $\propto = 0.05$. ومن ذلك نستنتج أنه يوجد تأثير معنوي للعامل الثاني F_B على المتحولات الثلاثة $X_1\ ,X_2\ ,X_3$.

وأخيراً يمكننا حساب النسبة المئوية لتأثير كل من F_A و F_B وكذلك النسبة المئوية لتأثير كل من التداخل والأخطاء من خلال تقسيم محددات المصفوفات المقابلة لها على محدد المصفوفة الاجمالية والذي يساوي $|SSPT| = 1052.52$ فنجد أن النسب المئوية للتأثير تساوي ما يلي :

جدول (7−11): النسب المئوية لتأثير العوامل والتداخل والبواقي :

	الخطأ العشوائي	Interaction	F_B	F_A	مصدر التباين
%100	26.19	7.51	23.89	42.40	النسبة المئوية %

مراجع الجزء الثاني:

1- Anderson, T. W.: An International to Multivariate Statistical Analysis (2nd Ed.) New york. John willey 1984 .

2- Bartllitt M.S. : Multivariate Analysis, Journal of Royal Statictics Society, Sapploment (B) 9 1947 .

3- Chrestmann, S, Van, Aelest: Robust Estimation of cronbache Alpha, JMA 97 2006 P. 1660 .

4- Daniel, C. Wood, F. C. : Fitting Equation to Data (2nd.ed) New york J.W. 1980 .

5- Daniel. D.: Statisique Theorique Et App lique, paris, masson Et. C. Editears, 1958 .

6- Fielibin j. j. The Probabality Plat correlation coeffeiciens Test For Normality Technomitrics J. 17, No 1 ,1975 ,P. 111-117 .

7- Johhson, R. Wichern, D: App Lied Multivariate Statistical ANALYSIS (2nd ed.),Pre Hall, 1988 .

8- Gopal, k. kauji, 100 statistical tests 3rd ed. SG pwal. London (2006) .

9- Kendall. M. Staley A. The Advanced Theory of statistics. London ch. Co. 1978 (Russion) .

10- Lvovsky. E. statistal Methods of Fittiug emperetical Form Moscow, High School, 1982 .

11- Metropolslcy A. K. Techneca of statistical colculetions, Moscow Nawoka, 1971 .

12- Morrison D.F. Maltivariate Metheds (2nd ed), New york, Macgraw Hill, 1976 .

13- Scheffe, H. The Anelysis of Variance, New york J.W. 1959 .

14- Triol, M. Triol, F.: Biostatistics For Biologicalin Heal Seinces. Edic person, New york, London 2006 .

15- Wilks S.S.: Cartion Generolizition in Analysis of Varionce Biometrika, 24, 1932 .

16- www/: wykepedia . www/: Raarech Gate .

17- الإمام، محمد الطاهر: تصميم وتحليل التجارب- جامعة الملك سعود- الرياض 1994 .

18- العلي، إبراهيم محمد، وعكروش، محمد: الإحصاء التطبيقي، كلية الاقتصاد- جامعة تشرين- سورية **2005** .

19- العلي،إبراهيم محمد، عكروش، محمد: نظرية الاحتمالات،ك.الاقتصاد- جامعة تشرين- سورية 2007.

20- العلي، إبراهيم محمد، كابوس أمل: الإحصاء الرياضي، كلية الاقتصاد- جامعة حلب- سورية 1986.

أسس
التحليل الإحصائي متعدد المتغيرات

الجزء الثالث
الانحدار الخطي والارتباط القانوني

الفصل الثالث: الارتباط القانوني :

3−1: تمهيد .

3−2: المفاهيم العامة للارتباط القانوني .

3−3: الشروط المفروضة على المتحولات X وY .

3−4: التحليل الرياضي للارتباط القانوني .

3−5: اختبار معنوية القيم الذاتية λ_k^2 أو معاملات الارتباط ρ_k .

3−6: خلاصة التحليل الرياضي .

3−7: الارتباط القانوني المعياري .

3− 8: حساب التحميلات القانونية (Loading) .

3− 9: تقييم كفاءة النماذج القانونية $(U_k\ , V_k)$ من التحميلات المباشرة .

3− 10: تقييم كفاءة النماذج القانونية $(V_k\ , U_k)$ من التحميلات العابرة .

3−11: مثال عام .

المراجع المعتمدة:

الفصل الأول

الانحدار الخطي البسيط Simple Linear Regression

1-1: تمهيد:

إن الانحدار الخطي البسيط يدرس العلاقة بين متحولين كميين فقط ، متحول كمي تابع ونرمز له بـ Y، ومتحول كمي مستقل ومؤثر في Y نرمز له بـ X، كالعلاقة بين وزن الطفل Y وعمره X ، أو كالعلاقة بين إنفاق الأسرة Y وعدد أفرادها X، أو كالعلاقة بين الناتج المحلي للدولة Y وعدد السكان فيها X ...الخ .

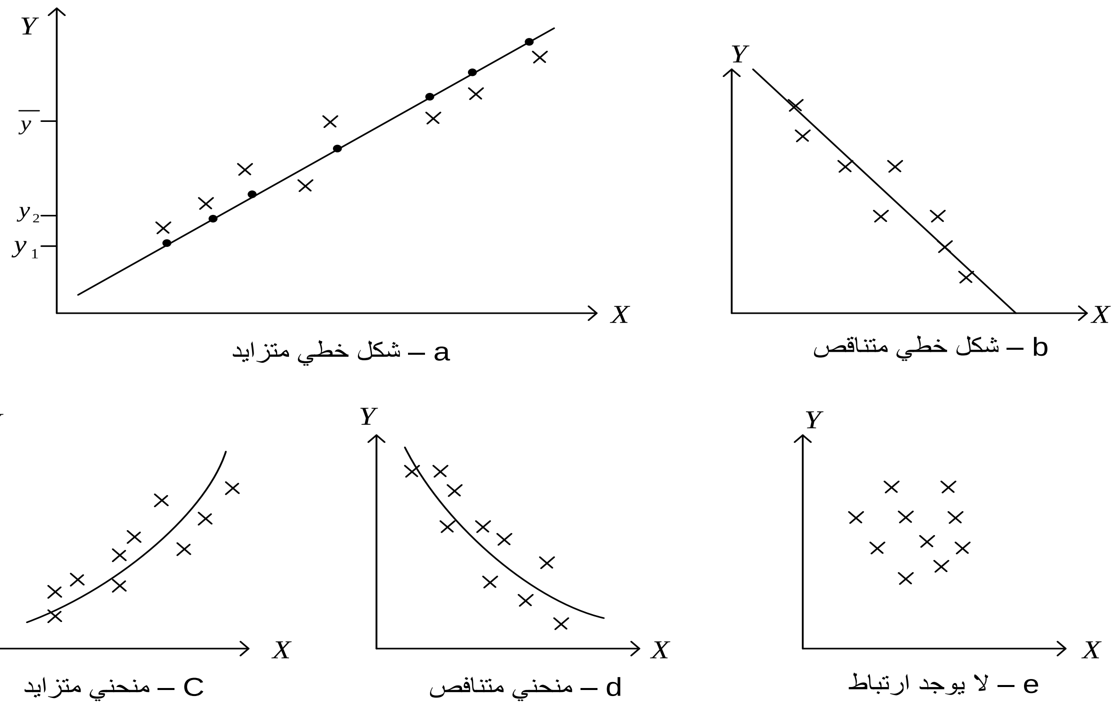
ولدراسة مثل هذه العلاقات في المجتمع المفروض نسحب عينة عشوائية من عناصره بحجم n عنصراً، ونأخذ من كل عنصر فيها القياسات المتقابلة (x_i , y_i) لكل من Yو X. ونضعها في جدول منظم كالتالي :

جدول (1-1): بيانات العينة لـ X وY:

رقم عنصر العينة i	1	2	3	....	i	....	n	المتوسط	الانحراف المعياري
x_i	x_1	x_2	x_3	....	x_i	....	x_n	$\bar{x}$	σ_x
y_i	y_1	y_2	y_3	....	y_i	....	y_n	$\bar{y}$	σ_y

حيث يتم حساب $\bar{x}$ و$\bar{y}$ و σ_x و σ_y من العلاقات المعروفة في الإحصاء .

وعند رسم هذه النقاط المتقابلة (x_i , y_i) على المستوى XoY نحصل على ما يسمى بشكل الانتشار، والذي يأخذ أحد الأشكال التالية :



الشكل (1-1): أشكال الانتشار

ومن هذه الأشكال نلاحظ أنه يوجد لدينا عدة أشكال ممكنة للعلاقة بين Y وX هي:

- الارتباط أو الانحدار الخطي المتزايد والمتناقص كما على الرسمين a وb من الشكل (1−1) .
- الارتباط أو الانحدار المنحني المتزايد والمتناقص كما في الرسمين c وd من الشكل (1−1) .
- عدم وجود ارتباط بين المتحولين Y وX كما في الرسم e من الشكل (1−1) .

1−2: صياغة النموذج:

وسنركز اهتمامنا هنا على الانحدار الخطي المبين في الرسمين a وb من الشكل (1−1) . وسنفترض أن الطبيعة العامة للعلاقة بين Y وX هي علاقة سببية ، وإن الصيغة الرياضية لها في المجتمع هي من الشكل الخطي التالي :

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{(النموذج في المجتمع)} \qquad (1-1)$$

حيث أن ε هو حد الخطأ العشوائي . وهو متحول عشوائي مؤلف من حدود مستقلة عن بعضها البعض ومستقلة عن المتحول المستقل X ، وإن توقعها يساوي الصفر $E(\varepsilon) = 0$ ، وإن تباينها ثابت ويساوي σ^2 ، وهي تخضع للتوزيع الطبيعي $N(0\ ,\sigma^2)$.

أما الأمثال β_0 و β_1 فهي أعداد حقيقية، تعمل على تحديد وضعية المستقيم (1−1) في المستوى XoY وبحيث يكون ميله مساوياً لـ β_1 وقاطعه مع oY مساوياً لـ β_0 . ولكن بما أن الأمثال β_0 و β_1 مازالت مجهولة فإن وضع المستقيم (1−1) في المستوى يبقى غير محدد .

ولتحديد الوضع المناسب لذلك المستقيم في المستوى نعتمد على بيانات العينة المسحوبة من ذلك المجتمع، ونرسم شكل الانتشار للنقاط الهندسية $(x_i\ , y_i)$، فإذا كان شكلها يوحي لنا باتجاه خط مستقيم نفترض أن العلاقة بين Y وX في العينة هي علاقة خطية أيضاً وتأخذ الشكل التقديري التالي :

$$Y = b_0 + b_1 X + e \qquad \text{(نموذج العينة)} \qquad (2-1)$$

حيث أن e هو حد خطأ التقدير، وهو متحول عشوائي آخر (يختلف عن ε) ومؤلف من حدود مستقلة وخاضعة للتوزيع الطبيعي $N(0\ ,\sigma^2)$. لأن وضعية المستقيم تختلف من عينة لأخرى .

وإذا أخذنا قيمة محددة لـ X مثل x_i، فإننا نحصل على قيمة عددية لـ Y نرمز لها بـ y_i وتساوي :

$$y_i = b_0 + b_1 x_i + e_i \qquad (3-1)$$

وإذا اعتمدنا على معادلة مستقيم العينة $Y = b_0 + b_1 X$ وأخذنا قيمة محددة لـ X مثل x_i فإننا سنحصل مقابلها على قيمة تقديرية لـ Y. نرمز لها بـ $\tilde{y}_i$ ونكتبها كما يلي :

$$\tilde{y}_i = b_0 + b_1 x_i \qquad \text{(قيمة } y_i \text{ التقديرية أو النظرية)} \qquad (4-1)$$

وبتعويض ذلك في (1−3) نجد أن بيانات العينة تعطينا أن:

$$y_i = \tilde{y}_i + e_i \qquad (i : 1\ 2\ 3 \ldots n) \qquad (5-1)$$

حيث أن $\tilde{y}_i = b_0 + b_1 x_i$ وهو تقدير y_i . وأن : e_i هو مقدار الخطأ في تقدير y_i . وهو يساوي :

$$e_i = y_i - \tilde{y}_i \qquad (i : 1\ 2\ 3 \ldots n) \qquad (6-1)$$

1-3: الافتراضات الموضوعة على النموذج الخطي هي:

1- أن تكون قياسات المتحول المستقل X دقيقة ومحددة (غير عشوائية) .

2- أن يكون المتحول X مؤثراً على Y ويساهم في تفسير تغيراته(أي أن العلاقة بينهما سببية) .

3- أن لا يقل عدد المشاهدات الزوجية عن (3) مشاهدات .

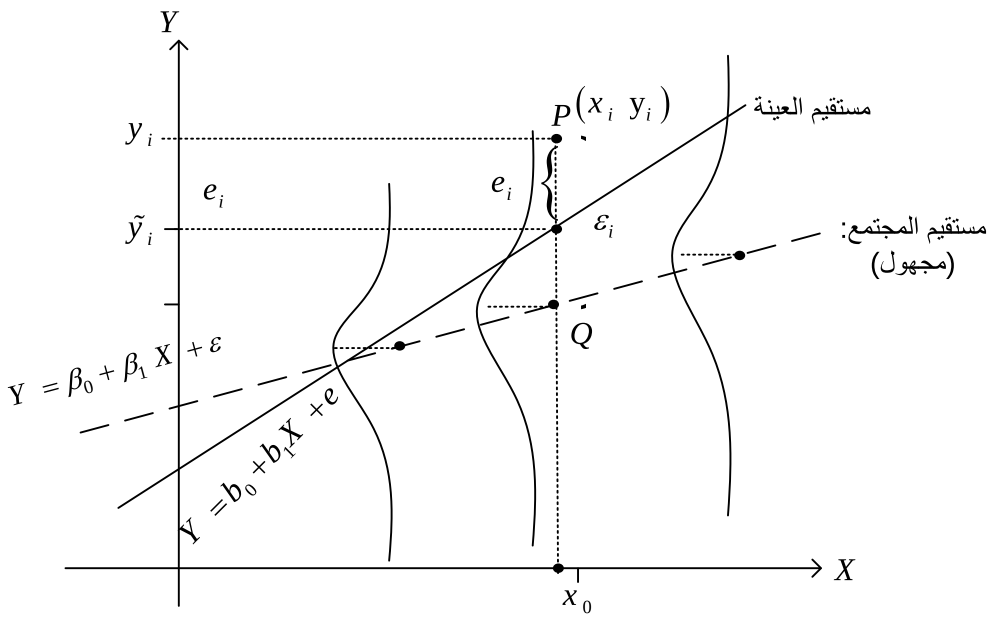
4- أن تكون القيمة المتوقعة لحدود الخطأ العشوائي ε مساوية للصفر $E(\varepsilon) = 0$.

5- أن يكون قيمة تباين حدود الخطأ العشوائي ε ثابتة عند كل قيم X وتساوي σ^2 .

6- أن تكون حدود الخطأ العشوائي ε مستقلة عن بعضها البعض وتخضع للتوزيع الطبيعي $N(0\ ,\sigma^2)$.

7- أن تكون حدود الخطأ العشوائي ε مستقلة عن المتحول المستقل X .

ولتوضيح هذه الأمور نعرضها على الشكل البياني التالي :



الشكل (1-2): مستقيما الانحدار في المجتمع والعينة

1-4: حساب معالم النموذج بطريقة المربعات الصغرى :

حتى يكون مستقيم الانحدار في العينة ممثلاً لشكل الانتشار يجب أن يأخذ أفضل وضعية ممكنة بين نقاط الانتشار، وهذه الوضعية هي التي تجعل مجموع مربعات أخطاء التقدير $\left(\sum_{i=1}^{n} e_i^2\right)$ اصغر ما يمكن، وهذا هو مبدأ طريقة المربعات الصغرى ونكتبه كما يلي :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = f(b_0 b_1) \rightarrow min \quad (7-1)$$

ولجعل المقدار $f\left(\tilde{b}_0 \tilde{b}_1\right)$ أصغر ما يمكن نشتقه بالنسبة لـ b_0 ثم بالنسبة لـ b_1 ونضع هذين المشتقين مساويين للصفر فنحصل على ما يلي:

$$\frac{\partial f}{\partial b_0} = 2(y_i - b_0 - b_1 x_i)(-1) = 0$$
$$\frac{\partial f}{\partial b_1} = 2(y_i - b_0 - b_1 x_i)(-x_i) = 0 \qquad (8-1)$$

وبعد إصلاح هذين المشتقين نحصل على المعادلتين العاديتين التاليتين :

$$nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
$$b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (9-1)$$

ثم نحل هاتين المعادلتين الخطيتين اللتين تتضمنان المجهولين b_0 و b_1 ، فنحصل على قيمتين محددتين لهما ، كما يمكن حسابهما مباشرة من العلاقتين المكافئتين لـ (1-9) التاليتين :

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \qquad (10-1)$$

$$b_0 = \bar{y} - b_1 \bar{x} \qquad (11-1)$$

حيث أن $\bar{x}$ و$\bar{y}$ هما متوسطا قيم X وY على الترتيب .

1-5: حساب معامل الارتباط (البيرسوني) :

لحساب معامل الارتباط (البيروسوني) r_{xy} بينY وX نستخدم العلاقة المعرفة كما يلي :

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n * \sigma_{\bar{x}} * \sigma_y} = \frac{\sum x_i\, y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} \qquad (12-1)$$

ومن خواص r_{xy} أن: $-1 \le r_{xy} \le +1$ ، وكلما كانت قيمته قريبة من (1+) كان الارتباط قوياً وطردياً، وكلما كانت قيمته قريبة من (1−) كان الارتباط قوياً وعكسياً . كما أن قيمة r_{xy} لا تتأثر بواحدات القياس المستخدمة في قياس قيم X أو Y .

1-6: حساب القيم النظرية لـ Y .

بعد حصولنا على b_0 و b_1 من المعادلتين (1-9) نعوضهما في معادلة النموذج (1-4) ، فنحصل على المعادلة التقديرية التالية:

$$\tilde{Y} = b_0 + b_1 X \qquad (13-1)$$

وبتعويض قيم X فيها نحصل مقابل كل قيمة x_i على قيمة تقديرية نظرية $\tilde{y}_i$ تساوي:

$$\tilde{y}_i = b_0 + b_1 x_i \qquad (14-1)$$

وبعد أن نحصل على القيم النظرية $\tilde{y}_i$ ، نقوم بحساب الأخطاء $e_i = (y_i - \tilde{y})$ ثم نربعها ونضعها في جدول مناسب كما يلي :

جدول (1−2): جدول مساعد لإجراء الحسابات اللازمة

الانحراف المعياري	المتوسط	n	…	I	…	3	2	1	رقم العنصر i
σ_x	$\bar{x}$	x_n	…	x_i	…	x_3	x_2	x_1	x_i
σ_y	$\bar{y}$	y_n	…	y_i	…	y_3	y_2	y_1	y_i
——	$\bar{\tilde{y}} = \bar{y}$	$\tilde{y}_n$	…	$\tilde{y}_i$	…	$\tilde{y}_3$	$\tilde{y}_2$	$\tilde{y}_1$	$\tilde{y}_i$
——	0	$(y_n - \tilde{y}_n)$	…	$(y_i - \tilde{y}_i)$	…	$(y_3 - \tilde{y}_3)$	$(y_2 - \tilde{y}_2)$	$(y_1 - \tilde{y}_1)$	$e_i = (y_i - \tilde{y}_i)$
——	——	$(y_n - \tilde{y}_n)^2$		$(y_i - \tilde{y}_i)^2$		$(y_3 - \tilde{y}_3)^2$		$(y_1 - \tilde{y}_1)^2$	$(y_i - \tilde{y}_i)^2$

ومن الجدول (1−2) نحسب الخطأ المعياري للتقديرات من العلاقة التالية :

$$S_{y/x} = \sqrt{\frac{\sum(y_i - \tilde{y}_i)^2}{n-2}} = \sqrt{\frac{\sum e_i^2}{n-2}} \qquad \mathbf{(15-1)}$$

ولتقدير جودة التمثيل نقوم بحساب معامل التحديد R^2 العادي من العلاقة :

$$R^2 = 1 - \frac{\sum e_i^2}{\sum(y_i - \bar{y})^2} \qquad (16-1)$$

كما يمكن حساب معامل التحديد المعدل من العلاقة :

$$Ad_jR^2 = 1 - \frac{S_{y/x}^2}{\sigma_y^2} \qquad (17-1)$$

وكلما كانت قيمة R^2 قريبة من الواحد كان التقدير جيداً .

كما يمكننا اختبار معنوية التقديرات b_0 و b_1 و r_{xy} و R^2 حسب القواعد المعروفة في الاقتصاد القياسي.

وسنتعرض لهما في الانحدار المتعدد .

1−7: اختبار جودة التمثيل وتحليل التباين:

لاختبار صلاحية النموذج نستخدم تحليل التباين ANOVA لدراسة تغيرات Y بدلالة X . ولذلك نقوم بتحليل مجموع مربعات انحرافات Y عن متوسطها الحسابي $\bar{y}$ كما يلي:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \tilde{y}_i + \tilde{y}_i - \bar{y})^2$$

$$\sum(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 + \sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 + 2(0) \qquad \mathbf{(18-1)}$$

ومن جهة أخرى نلاحظ أن الحد الأول يساوي $\sum_{i=1}^{n} e_i^2$، أما الحد الثاني فنعالجه بتعويض $\tilde{y}_i$ و $\bar{y}$ كما يلي :

$$\sum_{i=1}^{n}(\tilde{y}_i - \bar{y}) = \sum_{i=1}^{n}(\,b_0 + b_1x_i - \,b_0 - b_1\bar{x})^2 = b_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2$$

وبذلك نجد (بعد تبديل موقعي الحدين) أن (1−18) تأخذ الشكل التالي:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n} e_i^2 \qquad \mathbf{(19-1)}$$

وباستخدام الرموز المعروفة لهذه المجاميع نجد أن:

$$SST = SSX + SSE \qquad (20-1)$$

وإن درجات الحرية المقابلة لها هي كما يلي:

$$n-1 = 1 + (n-2) \qquad (21-1)$$

ثم نقوم بوضع فرضتي الاختبار كما يلي:

$$H_0 : b_1 = 0 \quad \text{فرضية العدم (النموذج غير صالح)}$$
$$H_1 : b_1 \neq 0 \quad \text{الفرضية البديلة} \qquad (22-1)$$

وبعد إجراء الحسابات اللازمة نضع النتائج في جدول ANOVA كما يلي:

جدول (1−3): جدول ANOVA **لاختبار علاقة** Y بـ X:

مصدر التباين	درجة الحرية v	مجموع المربعات	متوسط المربعات	قيمة F
المتحول X	1	$SSX = \sum^{n}(\tilde{y}_i - \bar{y})^2 = b_1^2 \sum (x_i - \bar{x})^2$	$MSX = \frac{SSX}{1} = b_1^2 \sum (x_i - \bar{x})^2$	$F = \frac{MSX}{MSE}$
الأخطاء أو البواقي	$n-2$	$SSE = \sum^{n}(y_i - \tilde{y}_i)^2 = \sum^{n} e_i^2$	$MSE = \frac{SSE}{n-2}$	
الاجمالي	$n-1$	$SST = \sum (y_i - \tilde{y})^2$		

واعتماداً على بيانات الجدول (1−3) نقوم باختبار صلاحية التمثيل باستخدام المؤشر F المعرف بالعلاقة :

$$F = \frac{MSX}{MSE} = \frac{b_1^2 \sum (x_i - \bar{x})^2}{\sum e_i^2/(n-2)} \qquad (23-1)$$

وهو متحول عشوائي يخضع لتوزيع F بدرجتي حرية $v_1 = 1$ و $v_2 = n-2$، لذلك نقوم بمقارنة قيمة F المحسوبة من (1-23) مع القيمة الحرجة لمتحول $F_{v_1,v_2}(\propto)$ عند مستوى الدلالة $\propto$ ودرجتي الحرية v_1 و v_2، ونتخذ القرار كما يلي :

إذا كانت $F \leq F_{1,n-2}(\propto)$ نقبل فرضية العدم H_0 . التي تقول أن أمثال X في النموذج معدومة $b_1 = 0$ ، ونستنتج أن النموذج لا يمثل العلاقة المدروسة بين Y و X .

أما إذا كانت $F > F_{1,n-2}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية التي تقول أن $b_1 \neq 0$ ونقول أن النموذج يمثل العلاقة بين Y و X باحتمال $(1 - \propto)$.

8-1: الاستدلال من تابع الانحدار المقدر:

وهو يستخدم لمعالجة مسائل التنبؤ بقيمة Y المقابلة لقيمة معينة لـ X مثل x_0 . علماً بأن عملية التنبؤ تعني إيجاد تقدير لقيمة Y المجهولة ثم إيجاد مجال الثقة الذي يحتوي هذه القيمة باحتمال ثقة$(1-\propto)$.

لذلك نفترض إننا اخترنا قيمة لـ X مثل x_0 . ونريد أن نتنبأ بقيمة Y المقابلة لها، وإذا ذهبنا إلى الشكل القادم (1−3) نلاحظ أنه مقابل أية قيمة x_0 يوجد لدينا (3) قيم لـ Y هي:

1− القيمة الفعلية أو الحقيقية لـ Y ونرمز لها بـ y_0 ، وهي قيمة مجهولة وتقابل النقطة $p(x_0, y_0)$ على الشكل (1−3)، وهي تساوي حسب النموذج العام ما يلي:

$$Y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0 \qquad (24-1)$$

2−القيمة المتوقعة لـ y_0 : لأن y_0 هي عبارة عن متحول عشوائي يخضع في المجتمع لتوزيع احتمالي معين، ويمكن أن يأخذ أية قيمة ممكنة له (حسب العينات المختلفة). وتتمركز هذه القيم حول توقعها الرياضي $E(y_0)$، وهو يقع على مستقيم الانحدار في المجتمع ويقابل القيمة x_0، وإن هذا التوقع مجهول أيضاً . ولكنه يساوي حسب النموذج العام ما يلي :

$$E(y_0) = E(\beta_0 + \beta_1 x_0 + \varepsilon_0) = \beta_0 + \beta_1 x_0 = \quad \left(\text{مستقيم المجتمع}\right) \qquad (25-1)$$

وهي نقطة تقع على مستقيم المجتمع (المجهول أيضاً) وتقابل القيمة x_0 ، ولقد رمزنا لها على الشكل (1−3) بالرمز Q ، وهي تمثل مركز التوزيع الاحتمالي الشرطي لـ y_0 المقابلة لـ x_0 .

3− القيمة التقديرية لـ y_0: ونرمز لها بـ $\widetilde{y_0}$ وهي قيمة تقديرية لـ y_0وتحسب من معادلة مستقيم الانحدار للعينة بعد حسابه أصولاً وتساوي ما يلي:

$$\widetilde{y}_0 = b_0 + b_1 x_0 = \quad \left(\text{مستقيم العينة}\right) \qquad (26-1)$$

وهي قيمة معلومة وتقع على مستقيم العينة ، وهي تعتبر تقديراً غير متحيز للقيمة الفعلية y_0 (حسب خواص تقديرات المربعات الصغرى) .

وإذا أخذنا توقع هذه القيمة التقديرية (حسب العينات المختلفة) نجد أن:

$$E(\widetilde{y}_0) = E(b_0 + b_1 x_0) = E(b_0) + E(b_1)x_0 = \beta_0 + \beta_1 x_0 \qquad (27-1)$$

وذلك لأن $E(b_0) = \beta_0$ و $E(b_1) = \beta_1$ (حسب تقديرات المربعات الصغرى) وبمقارنة العلاقتين (1−25) و(1−27) نحصل على العلاقة التالية:

$$E(\widetilde{y}_0) = \beta_0 + \beta_1 x_0 = E(y_0) \qquad (28-1)$$

أي أن التقدير $\widetilde{y}_0$ هو تقدير غير متحيز للتوقع $E(y_0)$ أيضاً، لأن توقع $\widetilde{y_0}$ يساوي $E(y_0)$.

وهكذا نستنتج مما سبق إنه مقابل كل قيمة x_0 يكون لدينا قيمتان مجهولتان هما: القيمة الفعلية y_0، والقيمة المتوقعة لها $E(y_0)$، ويكون لدينا قيمة معلومة هي القيمة التقديرية $\widetilde{y_0}$ المحسوبة من معادلة الانحدار(1−26) من بيانات العينة المدروسة، وسنستفيد من هذه القيمة المعلومة $\widetilde{y_0}$ في عمليات التنبؤ بقيمة y_0 الفعلية وبقيمة القيمة المتوقعة لها $E(y_0)$. وذلك لأنه قد وجدنا من العلاقتين (1−26)

و(28-1) أن $\widetilde{y_0}$ هي تقدير غير متحيز لكل من y_0 الفعلية وللقيمة المتوقعة لها $E(y_0)$. لذلك سنعتمد عليها في عملية التنبؤ بقيمة y_0وبقيمة $E(y_0)$كما يلي:

أ- التنبؤ بالقيمة المتوقعة لـ y_0 وهي $E(y_0)$ المقابلة لـ x_0 :

نقوم أولاً بحساب القيمة التقديرية لها $\widetilde{y_0}$ من النموذج المقدر (1-4) فنجد أن:

$$\widetilde{y_0} = b_0 + b_1 x_0 \qquad (29-1)$$

ثم نقوم بحساب الانحراف المعياري لها من العلاقة التالية (انظر البرهان عند العشعوش، والعريبد. ص113) .

$$Sd(\widetilde{y_0}) = S_{y/x} * \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \qquad (30-1)$$

حيث أن: $S_{y/x}$ هو الخطأ المعياري المعرف بالعلاقة (1-15)، ثم نقوم بإنشاء مجال الثقة ذي الاحتمال $(1-\propto)$ لـ $E(y_0)$ من العلاقة التالية :

$$\widetilde{y_0} - t_{n-2}\left(\frac{\propto}{2}\right) * Sd(\widetilde{y_0}) \le E(y_0) \le \widetilde{y_0} + t_{n-2}\left(\frac{\propto}{2}\right) * Sd(\widetilde{y_0}) \qquad (31-1)$$

وهو مجال يحتوي على القيمة المتوقعة $E(y_0)$ باحتمال $(1-\propto)$. ولكن هذا المجال يزداد اتساعاً كلما ابتعدنا عن المتوسط $\bar{x}$، وذلك بسبب الحد $(x_0 - \bar{x})^2$.

ب- التنبؤ بالقيمة الفعلية y_0المقابلة لـ x_0 :

نقوم أولاً بإيجاد تقدير القيمة الفعلية y_0المقابلة لـ x_0 من النموذج المقدر (1-4) فنجد أيضاً أن:

$$\widetilde{y_0} = b_0 + b_1 x_0 \qquad (32-1)$$

ثم نقوم بحساب الانحراف المعياري للتقدير $\widetilde{y_0}$ الذي يتضمن في هذه الحالة مصدرين للخطأ هما:

- خطأ تقدير القيمة المتوقعة $E(y_0)$ بواسطة $\widetilde{y_0}$.
- خطأ القياس الفردي لـ y_0 والذي تباينه σ_e^2 والذي يقدر بتباين الأخطاء $S_{y/x}^2$.

وبالتالي نستنتج أن مجال الثقة للقيمة الفعلية y_0 يجب أن يكون أوسع من مجال الثقة للقيمة المتوقعة $E(y_0)$. ولهذا فإن الانحراف المعياري للتقدير $\widetilde{y_0}$ –في هذه الحالة– يحسب من العلاقة التالية: (انظر البرهان عند العشعوش، والعريبد ص116) .

$$SSd(\widetilde{y_0}) = S_{y/x} * \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2}} \qquad (33-1)$$

حيث أن: $S_{y/x}$ هو الخطأ المعياري المعرف بالعلاقة (1-15)

ثم نقوم بإنشاء مجال الثقة للقيمة الفعلية y_0 ذي الاحتمال $(1-\propto)$ من العلاقة التالية:

$$\widetilde{y_0} - t_{n-2}\left(\frac{\propto}{2}\right) * SSd(\widetilde{y_0}) \le y_0 \le \widetilde{y_0} + t_{n-2}\left(\frac{\propto}{2}\right) * SSd(\widetilde{y_0}) \qquad (34-1)$$

وهو مجال يحتوي على القيمة الفعلية y_0 باحتمال $(1-\propto)$. وإنه يزداد اتساعاً كلما ابتعدنا عن المتوسط $\bar{x}$ بسبب الحد $(x_0-\bar{x})^2$.

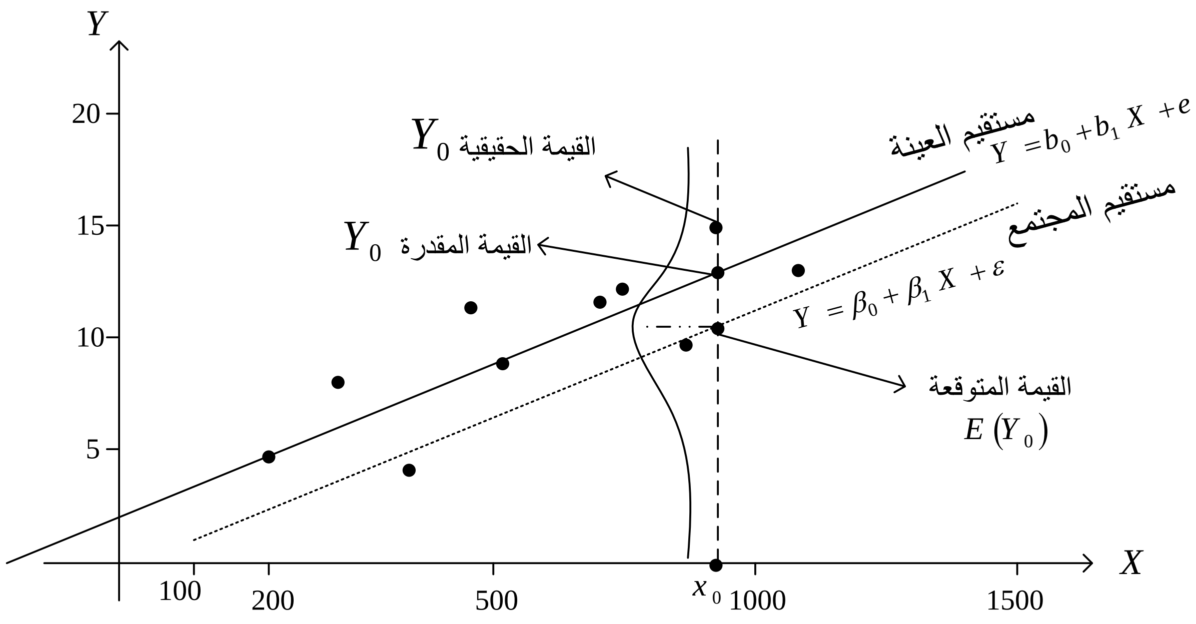
ملاحظة: عند إجراء عمليات التنبؤ الخارجي يجب أن لا نبتعد كثيراً عن حدود نطاق بيانات العينة . لأن ذلك يزيد الخطأ المرتكب ويقلل من دقة النتائج .

مثال (1−1): أدرس العلاقة بين مقدار دخل الأسرة X (بالدولار) وكمية استهلاكها من اللحم Y (كغ/شهريا) ثم تنبأ بكمية الاستهلاك عندما $X=800$ و$X=1200$ ، وذلك اعتماداً على بيانات عينة مؤلفة من $(n=10)$ أسر مختلفة والمبينة في الجدول التالي :

جدول (1−4): بيانات العينة (المصدر فرضي)

رقم الاسرة	1	2	3	4	5	6	7	8	9	10	المتوسط	الانحراف المعياري
دخل الاسرة X$	210	290	350	480	490	730	780	850	920	1010	611	282.51
كمية الاستهلاك Y kg	5	7	6	11	8	11	12	8	15	12	9.5	3.171

الحل: قبل كل شيء نرسم شكل الانتشار فنجد أنه كما يلي:



الشكل (1−3): شكل الانتشار

ومن الجدول (1−4) نجد أن $\bar{x}=611$ و $\bar{y}=9.5$ و $\sigma_x=282.51$ و $\sigma_y=3.171$

ونفترض أن العلاقة بين Y وX في العينة هي علاقة خطية من الشكل :

$$Y=b_0+b_1X+e$$

وحتى نستطيع حساب المعلمتين b_0 و b_1 من المعادلتين (1−9) أو من المعادلتين (1−10) و(1−11) ، علينا أن نقوم بإعداد الحسابات الواردة في الجدول التالي :

جدول (1−5): الحسابات المساعدة

رقم الأسرة i	1	2	3	4	5	6	7	8	9	10	المجموع
x_i^2	44100	48100	122500	250400	240100	532900	608400	7225000	896400	1020100	4451500
y_i^2	25	49	36	121	64	121	144	64	225	144	993
$x_i y_i$	1050	2050	2100	5280	5920	8030	9360	6800	13800	12120	64490
القيم النظرية $\tilde{y}_i = b_0 + b_1 x_i$	5.91	6.62	7.16	8.33	8.41	10.57	11.02	11.64	12.27	13.08	95
$e_i^2 = (y_i - \tilde{y}_i)^2$	0.8281	0.1444	1.3456	7.1209	0.1681	0.1849	0.9604	13.2496	6.4529	1.1664	32.6432

نعوض نتائج هذه الحسابات في المعادلتين (1−10) و(1−11)، فنجد أن :

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{64490 - 10(611)(9.5)}{4451500 - 10(611)^2} = \frac{6445}{718290} = 0.00897$$

$$b_0 = \bar{y} - b_1 \bar{x} = 9.5 - (0.00897)(611) = 4.019$$

وبذلك نحصل على أن معادلة العلاقة الانحدارية تأخذ الشكل التالي :

$$\tilde{y}_i = 4.019 + (0.00897)x_i$$

ومن هذه العلاقة حسبنا القيم النظرية $\tilde{y}_i$ ووضعناها في السطر الخامس من الجدول (1−5) السابق، ثم قمنا بحساب مربعات الفروقات $(y_i - \tilde{y}_i)^2$ ووضعناها في السطر الأخير فوجدنا أن :

$$\sum (y_i - \tilde{y}_i)^2 = \sum e_i^2 = 32.6432 \quad \Rightarrow \quad S_{y/x}^2 = \frac{32.6432}{8} = 4.0804$$

ومنها نحسب معاملي التحديد العادي والمعدل بعد حساب المقام لهما كما يلي:

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = \sum_{i}^{10} y_i^2 - n\bar{y} = 90.5$$

فيكون لدينا:

$$\sigma^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{90.5}{10} = 9.05$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{32.6432}{90.5} = 0.639 \approx 0.64$$

ثم نحسب معامل التحديد المعدل من العلاقة:

$$Ad_j R^2 = 1 - \frac{S_{y/x}^2}{\sigma_y^2} = 1 - \frac{4.0804}{9.05} = 0.594$$

ومنهما نستنتج أن جودة التمثيل جيدة (لأن $R^2 > 0.50$). كما يمكننا حساب معامل الارتباط r_{yx} من العلاقة :

$$r_{yx} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1) * \sigma_x \sigma_y} = \frac{64490 - 10(611)(9.5)}{9 * (282.51)(9.17)} \approx 0.80$$

وهنا نلاحظ أن مربع معامل الارتباط يساوي معامل التحديد أي أن ($r^2 = R^2$) .

وأخيراً نقوم بإجراء تحليل التباين فنجد أن المجاميع المطلوبة تساوي:

$$SST = \sum_{i=1}^{10} (y_i - \bar{y})^2 = (n-1)\sigma_y^2 = 9(3.171)^2 = 90.5$$

$$SSE = \sum (y_i - \tilde{y}_i)^2 = \sum e_i^2 = 32.6432$$

$$SSX = SST - SSE = 90.5 - 32.6432 = 57.86$$

وبذلك نجد أن جدول تحليل التباين ANOVA لهذا النموذج يأخذ الشكل التالي :

جدول (1-6): تحليل التباين ANOVA

مصدر التباين	الرمز	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F
المتحول X	SSX	57.85	1	57.80	$F = \frac{57.86}{4.0804} = 14.18$
الأخطاء (البواقي)	SSE	32.64	8	4.0804	
الاجمالي	SST	90.50	9	—	

ثم نقوم بإيجاد قيمة $F(\propto)$ الحرجة المقابلة لمستوى الدلالة $\propto= 0.05$ ولدرجتي الحرية $v_1 = 1$ و$v_2 = 8$ فنجد أنها تساوي :

$$F_{v_1, v_2}(\propto) = F_{1, 8}(0.05) = 5.32$$

وبمقارنة قيمة F المحسوبة مع قيمتها الحرجة $F(\propto)$ نجد أن: $F = 14.18 > 5.32$، لذلك نرفض فرضية العدم ($H_0: b_1 = 0$) ونقبل ($H_1: b_1 \neq 0$) ونعتبر النموذج صالحاً لتمثيل العلاقة بين Y وX.

ولإجراء عملية التنبؤ للقيمة المتوقعة لـ Y المقابلة لقيمة X مثل: $x_0 = 800$، نعوضها في معادلة النموذج فنحصل على قيمة Y المقدرة :

$$\tilde{y}_{800} = 4.019 + (0.00897)(800) = 11.195 \quad \left(\text{كغ / شهريا}\right)$$

ولكن بما أن هذه القيمة المتوقعة لـ y_{800} مجهولة. فإننا نقوم بإنشاء مجال الثقة لها ذي الاحتمال (0.95) من العلاقة (1-31) فنجد أن القيمة المتوقعة لـ y_{800} تتراوح في المجال التالي :

$$\tilde{y}_{800} - t_{n-2}\left(\frac{\propto}{2}\right) * Sd(\tilde{y}_{800}) \leq E(y_{800}) \leq \tilde{y}_{800} + t_{n-2}\left(\frac{\propto}{2}\right) * Sd(\tilde{y}_{800})$$

ولذلك نحسب $Sd(\tilde{y}_{800})$ من العلاقة (1-25) التالية:

$$Sd(\tilde{y}_{800}) = S_{y/x} * \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sqrt{4.0804}\sqrt{\frac{1}{10} + \frac{(800 - 611)^2}{718290}} = 0.7816$$

وذلك لأن:

$$\sum_{i=1}^{10}(x_i-\bar{x})^2=\sum x_i^2-n(\bar{x})^2=4451500-10(611)^2=718290$$

ومن جداول(ستودينت) نجد أن القيمة الحرجة $t_{n-2}\left(\frac{\propto}{2}\right)$ تساوي: $t_{n-2}\left(\frac{\propto}{2}\right)=t_8(0.025)=2.306$ ، وبالتعويض في مجال الثقة نجد أن:

$$11.195-(2.306)(0.7816)\leq E(y_{800})\leq 11.195+(2.306)(0.7816)$$
$$9.39\leq \mathrm{E}(y_{800})\leq 12.997\approx 13$$

ولإجراء عملية تنبؤ للقيمة الفعلية لـ (y_{800}) المقابلة لـ$x_0=800$، نعوضها في معادلة النموذج فنجد أن:

$$\tilde{y}_{800}=4.019+(0.00897)(800)=11.195 \quad (\text{كغ / شهريا})$$

ولإنشاء مجال الثقة للقيمة الحقيقية لـ y_{800} نطبق العلاقة (34−1) فنجد أن:

$$11.195-(2.306)\sqrt{4.0804}\sqrt{1+\frac{1}{10}+\frac{(800-611)^2}{718290}}\leq y_{800}\leq 11.195+(2.306)\sqrt{4.0804}\sqrt{1+\frac{1}{10}+\frac{(800-611)^2}{718290}}$$

$$11.195-4.995\leq y_{800}\leq 11.195+4.995$$
$$6.200\leq y_{800}\leq 16.195$$

وهو مجال يتضمن القيمة الفعلية y_{800} المقابلة لـ $x=800$ باحتمال قدره $0.95=\propto-1$.

وهنا نلاحظ أن مجال الثقة لـ y_{800} الفعلية أوسع من مجال الثقة للقيمة المتوقعة لها $E(y_{800})$. لأن القيمة الفعلية y_0 أكثر تشتتاً من القيمة المتوقعة $E(y_{800})$. وللتنبؤ بقيمة الاستهلاك Y عندما يكون $X=1200$ ، ندعو القارئ أن يقوم بها على سبيل التمرين، وعليه أن يتبع نفس الخطوات السابقة .

1−9: حساب تقدير معالم نموذج الانحدار تحت قيود خطية عليها :

لنفترض أن نموذج الانحدار الخطي العام (في المجتمع) كان على الشكل التالي:

$$Y=\beta_0+\beta_1X+\varepsilon \qquad (35-1)$$

وإذا سحبنا عينة عشوائية من عناصر المجتمع وأخذنا قياسات X منها وقياسات Y المقابلة لها، فإننا سنحصل من كل عنصر i منها على قيمة x_i وعلى قيمة y_i . وإذا عوضنا هذه القيم في معادلة النموذج (1−35) نحصل على n معادلة خطية كما يلي:

$$\begin{aligned} y_1&=\beta_0+\beta_1x_1+\varepsilon_1\\ y_2&=\beta_0+\beta_1x_2+\varepsilon_2\\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ y_n&=\beta_0+\beta_1x_i+\varepsilon_n \end{aligned} \qquad (36-1)$$

حيث : ε_i هو حد الخطأ العشوائي المقابلة للعنصر i، ولتسهيل المعالجات الرياضية نكتب هذه المعادلات بدلالة الأشعة والمصفوفات على الشكل التالي :

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}1&x_1\\1&x_2\\\vdots&\vdots\\1&x_n\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix} \qquad (37-1)$$

وبذلك يمكننا كتابة النموذج (1−35) مصفوفياً على الشكل التالي:

$$Y_{n*1} = X_{n*2} * \beta_{2*1} + \varepsilon_{n*1} \qquad (38-1)$$

حيث أن: X هي المصفوفة الموسعة لـ X وتتضمن n سطراً وعمودين. وإن عمودها الأول يتألف من واحدات حتى يتوافق مع المعلمة الثابتة β_0 : أي أن :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (39-1)$$

وعندما يتم تقدير المعلمتين β_0 و β_1 من بيانات العينة من إحدى العلاقتين (1−9) أو (1−10)، فإننا سنحصل على تقديرهما b_0 و b_1، وعندها فإن صيغة النموذج تأخذ الشكل التالي :

$$Y = b_0 + b_1X + e = X * b + e \qquad (40-1)$$

حيث أن: $b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$، وأن e هو شعاع الأخطاء في العينة وهو يختلف عن ε، ولكنه يخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$.

والآن نفترض أن هناك مجموعة من القيود أو الشروط مفروضة على معالم النموذج المجهولة β_0 و β_1 مثل:

$$\beta_0 = 10 \quad , \quad \beta_0 + \beta_1 = 5 \quad , \quad \beta_0 - \beta_1 = 0$$

وهنا نلاحظ أنه يمكننا كتابة هذه القيود مصفوفياً على الترتيب على الشكل التالي :

$$[1\ \ ,0] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 10 \qquad [1\ \ ,1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 5 \qquad [1\ \ ,-1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 0 \qquad (41-1)$$

وبصورة عامة يمكننا كتابة هذه القيود على شكل معادلة مصفوفية واحدة على الشكل التالي :

$$R_{g*2} * \beta_{2*1} = C_{g*1} \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \\ 0 \end{bmatrix} \qquad (42-1)$$

حيث أن R: هي مصفوفة معاملات القيود في معادلاتها الخطية وهي مؤلفة من g سطراً وعمودين . حيث g عدد القيود المفروضة . وأن C: هو عمود الثوابت التي في الطرف الثاني لمعادلات القيود ، أي أنه يمكننا كتابة القيود المفروضة على معالم النموذج β بشكل منفرد أو بشكل جماعي ومشترك كما في العلاقتين (1−41) و(1−42) على الترتيب . وهكذا نجد أن المصفوفة R والعمود C يساويان (في هذه الحالة) ما يلي:

$$R = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad C = \begin{bmatrix} 10 \\ 5 \\ 0 \end{bmatrix}$$

كما يمكننا كتابة أي شرط من القيود السابقة بمفرده على الشكل التالي :

$$[1\ \ ,0] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 10 \quad : \quad R_1\,\beta = 10$$

$$[1\ \ ,1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 5 \quad : \quad R_2\,\beta = 5$$

$$[1\ \ ,-1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 0 \quad : \quad R_3\,\beta = 0$$

والآن علينا أن نقوم بتقدير المعالم β_1 و β_0 تحت القيود R المذكورة في (1−42)، علماً بأن النموذج المقدر بدون قيود يأخذ الشكل التالي :

$$Y = b_0 + b_1X + e \qquad (43-1)$$

حيث يتم حساب التقديرين b_1 و b_0 بطريقة المربعات الصغرى من بيانات العينة قبل الأخذ بعين الاعتبار الشروط المفروضة .

ولتقدير المعلمتين β_1 و β_0 تحت القيود R، نرمز لتقديرهما المقيد بالرمزين: $\tilde{\beta}_{1R}$ و $\tilde{\beta}_{0R}$ ، وعندها يمكننا كتابة نموذج الانحدار المقيد والمقدر من العينة على الشكل التالي :

$$Y_R = \tilde{\beta}_{0R} + \tilde{\beta}_{1R}X + e_R = X * \tilde{\beta}_R + e_R \qquad (44-1)$$

حيث أن: $\tilde{\beta}_R$ هو عمود التقديرات المقيدة بالقيود المفروضة .

وأن: e_R هو عمود الأخطاء المقيدة بالقيود المفروضة .

ويمكن حساب الشعاع المقيد $\tilde{\beta}_R$ بعد حساب الشعاع غير المقيد b من بيانات العينة ، وذلك باستخدام العلاقة التالية [انظر البرهان عند عناني ص 703] .

$$\tilde{\beta}_R = b - (X'X)^{-1} * R'[R(X'X)^{-1}R']\,[R\ \ b - C] \qquad (45-1)$$

ومنها نحصل على عمود التقديرات المقيدة $\tilde{\beta}_R$ بالاعتماد على طريقة المربعات الصغرى وبدلالة التقديرات غير المقيدة b .

علماً بأن التقديرات المقيدة تكون تقديرات غير متحيزة للمعالم β المجهولة، إذا كانت القيود $(R\beta = C)$ صحيحة، وتكون متحيزة إذا كانت تلك القيود غير صحيحة . كما أنها تكون أكثر كفاءة من التقديرات b إذا كانت القيود صحيحة، لأن مصفوفة التباينات المشتركة للشعاع $\tilde{\beta}_R$ في هذه الحالة تكون **أقل أو تساوي من مصفوفة التباينات المشتركة للتقديرات** b، أي أن:

$$COV(\tilde{\beta}_R) \leq COV(b) \qquad \left(\text{إذا كان القيود صحيحة}\right) \qquad (46-1)$$

وهذا يعني إن فرض بعض القيود الخطية الصحيحة على معالم النموذج β يجعل مقدرات تلك المعالم المقيدة $\tilde{\beta}_R$ أكثر كفاءة لأنها ستكون أقل تبايناً .

ولاختبار معنوية القيود الخطية المفروضة في (1−42) نضع الفرضيتين كما يلي:

$$\begin{aligned} H_0 &: R\beta = C \qquad \left(\text{القيود صحيحة}\right) \\ H_1 &: R\beta \neq C \qquad \left(\text{القيود غير صحيحة}\right) \end{aligned} \qquad \mathbf{(47-1)}$$

ثم نحسب مؤشر الاختبار F من العلاقة التالية :

$$F = \frac{\dfrac{[\boldsymbol{e'_R} * \boldsymbol{e_R} - \boldsymbol{e'} * \boldsymbol{e}]}{g}}{\dfrac{\boldsymbol{e'} * \boldsymbol{e}}{n-k}} \sim F_{g\,,n-k}(x) \qquad \mathbf{(48-1)}$$

وهو يخضع للتوزيع $F(x)$ وبدرجتي حرية: $v_1 = g$ و $v_2 = n - k$

حيث أن: k هو عدد المعالم β في النموذج .

وأن: $e'_R * e_R$ هو يمثل مجموع مربعات الأخطاء في النموذج المقيد e_R .

وأن: $e' * e$ هو مجموع مربعات الأخطاء في النموذج غير المقيد e .

وأن g عدد الشروط المفروضة على المعالم β .

علماً بأنه يمكن استبدال البسط في العلاقة (1−48) بما يساويه كما يلي:

$$(e'_R * e_R - e' * e) = [Rb - C]'[R(X'X)^{-1}R']^{-1} * [Rb - C] \qquad (49-1)$$

وعندها تأخذ العلاقة (1−48) لمؤشر الاختبار F الشكل التالي :

$$F = \frac{\dfrac{[Rb - C]'[R(X'X)^{-1}R']^{-1} * [Rb - C]}{g}}{\dfrac{e' * e}{n-k}} \qquad (50-1)$$

وهي علاقة تعطينا قيمة F بدلالة بيانات العينة X والتقديرات b ومصفوفة القيود R وعددها g .

وبعد حساب F من العلاقة (1−48) أو(1−50) نقوم بمقارنة قيمتها المحسوبة مع القيمة الحرجة $F_{g,n-k}(\propto)$ وعند مستوى الدلالة $\propto$. ونتخذ القرار حول معنوية التقديرات المقيدة $\tilde{\beta}_R$ كما يلي:

إذا كانت $F \le F_{g,n-k}(\propto)$ نقبل فرضية العدم H_0 (القيود صحيحة) .

وإذا كانت $F > F_{g,n-k}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تعني أن القيود غير صحيحة .

كما يمكن إنشاء مجال ثقة لكل من التقديرات المقيدة $\tilde{\beta}_R$. حسب القواعد المذكورة سابقاً للتقديرات غير المقيدة b .

مثال (1−2): [مأخوذ من عنانيَ ص. 713 بتصرف]

ادرس العلاقة بين الإنفاق على الاتصالات Y وعدد أفراد الأسرة X. وذلك بناء على بيانات عينة مؤلفة من $(n = 7)$ أسر والمبينة في الجدول التالي :

جدول (1−7): بيانات المسألة

المتوسط	7	6	5	4	3	2	1	رقم الأسرة
$\bar{X} = 5$	10	2	3	6	5	4	5	عدد الأفراد x_i
$\bar{Y} = 14$	17	11	13	14	12	16	15	مقدار الإنفاق بالآلاف Y

وذلك تحت القيود الخطية التالية (كل على حدة) .

$$\beta_0 + \beta_1 = 5 \qquad \text{ثم} \qquad \beta_0 = 10$$

الحل: نقوم أولاً بحساب التقديرات غير المقيدة b من العلاقة (1−9) أو من العلاقة التالية:

$$b = (X'X)^{-1} * X'Y$$

لذلك نشكل المصفوفة X الموسعة كما يلي:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

وبذلك نجد أن:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} * \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n\,, & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

وبعد الحساب نجد أن:

$$X'X = \begin{bmatrix} 7 & 35 \\ 35 & 215 \end{bmatrix}$$

ومنه نجد أن:

$$(X'X)^{-1} = \frac{1}{280} \begin{bmatrix} 215 & -35 \\ -35 & 7 \end{bmatrix}$$

وكذلك نجد أن:

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_0 \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} 98 \\ 514 \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \frac{1}{280} \begin{bmatrix} 215 & -35 \\ -35 & 7 \end{bmatrix} \begin{bmatrix} 98 \\ 514 \end{bmatrix} = \begin{bmatrix} 11 \\ 0.6 \end{bmatrix}$$

ومنها نجد أن نموذج الانحدار المقدر غير المقيد يأخذ الشكل التالي :

$$\tilde{Y} = 11 + 0.6X$$

ولاختبار معنوية β_0 و β_1 نقوم بحساب المجاميع التالية:

$$SST = Y'Y - n\bar{Y}^2 = 1400 - 7(14)^2 = 28$$
$$SSX = b'X'Y - n\bar{Y}^2 = 1386.4 - 1372 = 14.4$$
$$SSE = SST - SSX = 28 - 14.4 = 13.6$$
$$\tilde{\sigma}_e^2 = \frac{SSE}{n-2} = \frac{13.6}{5} = 2.72$$
$$Var(b) = \tilde{\sigma}_e^2 (X'X)^{-1} = \frac{2.72}{280} \begin{bmatrix} 215 & -35 \\ -35 & 7 \end{bmatrix} = \begin{bmatrix} 2.089 & -0.345 \\ -0.345 & 0.068 \end{bmatrix}$$

أي أن:

$$Var(b_0) = 2.089 \qquad Var(b_1) = 0.068$$
$$\tilde{\sigma}_{b_0} = 1.445 \qquad \tilde{\sigma}_{b_1} = 0.261$$

وبناءً على هذه الحسابات نقوم باختبار معنوية β_0 و β_1 فنجد أن:

أولاً: لاختبار معنوية $\boldsymbol{\beta_0}$: نضع الفرضيتين كما يلي:

$$H_0 : \beta_0 = 0 \qquad , \qquad H_1 : \beta_0 \neq 0$$

ثم نقوم بحساب قيمة المؤشر t من العلاقة :

$$t = \frac{b_0 - (\beta_0)_0}{\tilde{\sigma}_{b_0}} = \frac{11 - 0}{1.445} = 7.612$$

ومن جداول t نجد أن القيمة الحرجة لـ t عندما $\propto = 0.5$ و $v = n - 2$ يساوي:

$$t_{n-2}\left(\frac{\propto}{2}\right) = t_5(0.025) = 2.571$$

وعند المقارنة نجد أن $t > t_5(\propto)$ ، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تفيد أن $H_1\ :\ \beta_1 \neq 0$ أي أن قيمة β_0 معنوية .

ثانياً: لاختبار معنوية $\boldsymbol{\beta_1}$: نضع الفرضيتين كما يلي:

$$H_0\ :\ \beta_1 = 0 \qquad , \qquad H_1\ :\ \beta_1 \neq 0$$

ثم نقوم بحساب مؤشر الاختبار t من العلاقة :

$$t = \frac{b_1 - (\beta_1)_0}{\tilde{\sigma}_{b_1}} = \frac{0.6}{0.261} = 2.299$$

وبالمقارنة بالقيمة الحرجة السابقة: $t_{n-2}\left(\frac{\propto}{2}\right)$ التي تساوي أيضاً 2.571 نجد أن:

$t < t_5(\propto)$ ، لذلك نقبل فرضية العدم H_0 والتي تفيد أن قيمة β_1 غير معنوية لأن $H_0\ :\ \beta_1 = 0$.

والآن نقوم بإعادة تقدير معالم النموذج المقيد β_0 و β_1، تحت القيد الأول فقط وهو $\beta_0 = 10$، والذي يمكن كتابته كما يلي:

$$R_1\beta = C$$

$$[1\ \ 0]\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = 10 \ \Rightarrow\ \beta_0 - 0 = 10$$

ولحساب عناصر العلاقة (1−50) نقوم بحساب ما يلي:

$$(X'X)^{-1} * R_1' = \frac{1}{280}\begin{bmatrix}215 & -35\\ -35 & 7\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix} = \begin{bmatrix}\frac{215}{280}\\ \frac{-35}{280}\end{bmatrix}$$

ثم نقوم بحساب الجداء التالي :

$$R_1(X'X)^{-1}R_1' = [1\ \ 0]\begin{bmatrix}\frac{215}{280}\\ \frac{-35}{280}\end{bmatrix} = \frac{215}{280} \qquad \text{(عدد)}$$

ثم نحسب المقلوب التالي :

$$[R_1(X'X)^{-1}R_1']^{-1} = \frac{280}{215}$$

وكذلك نحسب الحد التالي :

$$R_1b - C = [1\ \ 0]\begin{bmatrix}11\\ 0.6\end{bmatrix} - 10 = 11 - 10 = 1$$

وبذلك نجد أن التقديرات المقيدة لـ β_R تساوي حسب (1−45) تساوي ما يلي:

$$\tilde{\beta}_{R_1} = \begin{bmatrix}11\\ 0.6\end{bmatrix} - \begin{bmatrix}\frac{215}{280}\\ \frac{-35}{280}\end{bmatrix}\left(\frac{280}{215}\right)(1) = \begin{bmatrix}11\\ 0.6\end{bmatrix} - \begin{bmatrix}1\\ -0.1627\end{bmatrix} = \begin{bmatrix}10\\ 0.7627\end{bmatrix}$$

وهنا نلاحظ أن القيد المفروض قد تحقق وأعطانا أن: $\tilde{\beta}_{oR_1} = 10$

ولكن قيمة β_1 أصبحت تساوي: 0.7267

وإن النموذج المقيد يأخذ الشكل التالي : $\tilde{X}_R = 10 + 0.7627X$

ولاختبار صحة القيد الأول نضع الفرضيتين كما يلي:

$$H_0 : R_1\beta = C \iff H_0 : \beta_0 = 10$$
$$H_1 : R_1\beta = C \iff H_1 : \beta_0 \neq 10$$

ثم نقوم بحساب قيمة مؤشر الاختبار F من العلاقة التالية:

$$F = \frac{(\boldsymbol{R_1 b} - \boldsymbol{C})' \left[\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{\boldsymbol{-1}}\boldsymbol{R}_1'\right]^{\boldsymbol{-1}} (\boldsymbol{R_1 b} - \boldsymbol{C})/g}{\boldsymbol{ee}'/(n-2)} =$$

$$F = \frac{\left(1 * \frac{280}{215} * 1\right)/1}{2.72} = 0.4788$$

ثم نقوم بحساب القيمة الحرجة لـ F عندما $\propto = 0.05$ وعند درجتي الحرية $v_1 = g = 1$

و $v_2 = n - k = 5$ فنجد أن:

$$F_{g\ n-k}(\propto) = F_{1,5}(0.05) = 6.61$$

وعند المقارنة نجد أن: $F < F_{1,5}(\propto)$ ، لذلك نقبل فرضية العدم H_0 التي تقول أن القيمة $\beta_0 = 10$ هي قيمة معنوية وأن القيد الأول عليها هو قيد صحيح ومقبول .

والآن نقوم بإعادة تقدير معالم للنموذج المقيد بالشرط التالي:

$$\beta_0 + \beta_1 = 5$$

والذي يمكن كتابته على الشكل التالي :

$$R_2\beta = C \implies [1\ \ 1]\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = 5$$

ثم نقوم بحساب العناصر التالية:

$$(X'X)^{-1}R_2' = \frac{1}{280}\begin{bmatrix}215 & -35\\ -35 & 7\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \begin{bmatrix}\frac{180}{280}\\ \frac{-28}{280}\end{bmatrix}$$

$$R_2(X'X)^{-1}R_2' = [1\,,1]\begin{bmatrix}\frac{180}{280}\\ \frac{-28}{280}\end{bmatrix} = \frac{152}{280} \qquad (\text{عدد})$$

$$[R_2(X'X)^{-1}R_2']^{-1} = \frac{280}{152}$$

$$R_2 b - C = [1\,,1]\begin{bmatrix}11\\ 0.6\end{bmatrix} - 5 = 11.6 - 5 = 6.6$$

وبذلك نجد أن التقديرات المقيدة بالشرط الثاني $\left(\beta_0 + \beta_1 = 5\right)$ تحسب من العلاقة(1−50) كما يلي:

$$\tilde{\beta}_{oR_2} = \begin{bmatrix} 11 \\ 0.6 \end{bmatrix} - \begin{bmatrix} \frac{180}{280} \\ \frac{-28}{280} \end{bmatrix} \left(\frac{280}{152}\right)(6.6) = \begin{bmatrix} 11 \\ 0.6 \end{bmatrix} - \begin{bmatrix} 7.815 \\ -1.215 \end{bmatrix} = \begin{bmatrix} 3.185 \\ 1.815 \end{bmatrix}$$

وهنا نلاحظ أن القيد الثاني قد تحقق وأصبح المجموع $(\beta_0 + \beta_1 = 5)$، ولاختبار صحة القيد الثاني نضع الفرضيتين على الشكل التالي :

$$H_0 : R_2\beta = C \iff H_0 : \beta_0 + \beta_1 = 5$$
$$H_1 : R_2\beta \neq C \iff H_1 : \beta_0 + \beta_1 \neq 5$$

ثم نقوم بحساب قيمة مؤشر الاختبار من العلاقة (1−50) فنجد أن:

$$F = \frac{(6.6)\left(\frac{280}{152}\right)(6.6)/1}{2.72} = \frac{80.242}{2.72} = 29.500$$

وبمقارنة هذه القيمة لـ F مع القيمة الحرجة (6.61) نجد أن: $F > F_{1,5}(\propto)$ لذلك نرفض فرضية العدم H_0 ونقبل الفرضية H_1 التي تفيد أن القيد الثاني المفروض غير صحيح بالنسبة للمثال المدروس .

الفصل الثاني

الانحدار الخطي المتعدد والمضاعف

Multivariate & Multiple Linear Regression

2-1: تمهيد :

إن الانحدار المتعدد هو تعميم للانحدار البسيط ، وهو يتناول دراسة العلاقة بين متحول كمي تابع Y مع عدة متحولات كمية مستقلة $X_1\ X_2\ ...\ X_p$. حيث p هو عدد المتحولات المستقلة .

وذلك مثل: دراسة علاقة إنفاق الأسرة Y مع دخلها X_1 وعدد أفرادها X_2 . ودراسة علاقة الناتج المحلي الاجمالي Y مع عدد السكان X_1 وكمية الصادرات X_2 والواردات X_3 ...الخ .

ولدراسة مثل هذه العلاقات في المجتمع، نفترض أولاً أن العلاقة بين التابع Y والمتحولات المستقلة $X_1\ X_2\ ...\ X_p$ في المجتمع هي علاقة خطية من الشكل التالي (بدلالة أسماء المتحولات) :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \qquad (1-2)$$

حيث أن $X_1\ X_2\ ...\ X_p$ هي متحولات مستقلة ومرتبطة خطياً مع Y

وحيث أن $\beta_0\ \beta_1\ \beta_2\ ...\ \beta_p$ هي معالم عددية مجهولة ، وأن β_j تشير إلى مقدار تأثير المتحول X_j في التابع Y عندما يزداد X_j بمقدار وحدة واحدة، وأما β_0 فهي القيمة الثابتة لـ Y في حال انعدام وجود أوتأثير المتحولات X_j .

وحيث أن الحد ε يمثل الخطأ العشوائي ، وهو متحول عشوائي يشترط فيه أن يكون مستقلاً عن المتحولات X ، وأن يكون توقعه مساوياً للصفر، وأن يكون تباينه مساوياً لمقدار ثابت σ_ε^2 ، وأن يكون خاضعاً للتوزيع الطبيعي المتعدد $N(0\ ,\sigma^2 I)$ الذي توقعه الشعاع (0) وتباينه المصفوفة $V = \sigma_\varepsilon^2 I$.

وهنا نشير إلى أن β_j وε وσ_ε^2 هي مقادير مجهولة يجب العمل على تقديرها من بيانات العينة المسحوبة من المجتمع . وإذا سحبنا عينة من عناصر المجتمع بحجم n عنصراً وأخذنا قياسات المتحولات $X_1\ X_2\ ...\ X_p$ منها وقياسات Y المقابلة لها ، ووضعناها في جدول كالتالي :

جدول (2-1): بيانات العينة المسحوبة من المجتمع بحجم n عنصراً.

رقم العنصر في العينة	Y	X_1	X_2	......	X_p
1	y_1	x_{11}	x_{12}	......	x_{1p}
2	y_2	x_{21}	x_{22}	......	x_{2p}
3	y_3	x_{31}	x_{32}	......	x_{3p}
⋮	⋮	⋮	⋮	......	⋮
n	y_n	x_{n1}	x_{n2}	......	x_{np}

وإذا عوضنا هذه القيم في النموذج (2−1) نحصل على جملة من المعادلات الخطية (n معادلة) تأخذ الشكل التالي :

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ y_3 &= \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \cdots + \beta_p x_{3p} + \varepsilon_3 \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n \end{aligned} \qquad (2-2)$$

حيث أن ε_i هي حدود الخطأ العشوائي ε المقابلة لعنصر العينة i .

ولتسهيل المعالجات الرياضية المختلفة، سنقوم باستخدام المصفوفات، ونكتب المعادلات (2−2) على الشكل التالي :

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ i & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (3-2)$$

وبذلك يمكننا كتابة النموذج (2−1) على الشكل التالي :

$$Y_{n*1} = X_{n*(p+1)} * B_{(p+1)*1} + \varepsilon_{n*1} \qquad (4-2)$$

وهنا يجب أن نشير إلى أن تشكيل المصفوفة الموسعة X يتضمن $(p+1)$ عموداً، وإن عمودها الأول يتألف من الواحدات، وذلك حتى يتوافق مع المعلمة الثابتة β_0، كما إن مرتبة X تساوي $n*(p+1)$، حيث $(p+1)$ يساوي عدد المعلمات $\beta_0\ \beta_1\ \beta_2 \dots \beta_p$ الداخلة في النموذج الخطي (2−1). وللاستفادة من هذا النموذج في عمليات التحليل والتنبؤ، علينا تقدير المعلمات المجهولة $\beta_0\ \beta_1\ \beta_2 \dots \beta_p$ وتقدير حدود الخطأ ε، ولكننا قبل أن نبحث في ذلك نستعرض الافتراضات أو الشروط التي يجب أن تحققها عناصر هذا النموذج .

2−2: الافتراضات الموضوعة على نموذج الانحدار المتعدد:

تصنف هذه الافتراضات إلى نوعين: هيكلية وعشوائية:

- **الافتراضات الهيكلية:** وهي الشروط التي يجب أن تتوفر في المتحولات المستقلة والمتحول التابع وهي غير مرتبطة بالمتحول العشوائي ε وهذه الافتراضات هي:

1− أن يكون حجم العينة n أكبر من $(P+1)$ عدد المعلمات β_j، أي أن يكون $n > P+1$ وذلك لضرورات رياضية سنراها لاحقاً .

2− أن تكون المتحولات $X_1\ X_2 \dots X_p$ مستقلة عن بعضها البعض، وهذا يعني أن تكون معاملات الارتباط بينها ضعيفة ويمكن التأكد من ذلك من خلال المصفوفة الإرتباطية المشتركة لها ، وإن هذا

الاستقلال يعني أن تكون رتبة المصفوفة X كاملة وتساوي $P+1$، وهذا يكافئ أن تكون أعمدة المصفوفة X مستقلة خطياً. أي لا يمكن لأي منها أن ينشأ من تركيب خطي لعمود آخر (أو أكثر).

3- أن تكون العلاقة بين Y وكل من المتحولات المستقلة $X_1 \ X_2 \ \dots \ X_p$ علاقة خطية . ويمكن التأكد من ذلك من خلال حساب معاملات الارتباط الثنائية بين Y وتلك المتحولات .

وفي التطبيقات العملية يمكننا التحقق من الافتراضين (2) و(3) معاً من خلال المصفوفة الارتباطية المشتركة لـ X و Y .

4- أن تكون القياسات المشاهدة x_{ij} محددة ودقيقة، أي غير متضمنة لأخطاء القياس (غير عشوائية).

- **الافتراضات العشوائية:** وهي افتراضات أو شروط موضوعة على المتحول العشوائي ε وهي:

1- أن يكون التوقع الرياضي لكل ε_i معدوماً . أي أن يكون :

$$E(\varepsilon_i) = 0 \qquad i : 1\ 2\ 3 \dots n \qquad (5-2)$$

2- أن يكون تباين ε_i مساوياً لمقدار ثابت σ^2 ، أي أن :

$$Var(\varepsilon_i) - E(\varepsilon_i - 0)^2 = E\left(\varepsilon_i^2\right) = \sigma^2 \qquad \text{عدد ثابت :} \qquad (6-2)$$

3- أن يكون التباين المشترك لأي حدين مختلفين ε_k و ε_ℓ معدوماً . أي أن يكون :

$$cov(\varepsilon_k , \varepsilon_\ell) = E(\varepsilon_k , \varepsilon_\ell) = 0 \qquad : k \neq \ell \qquad (7-2)$$

وهذا يعني أن حدود الخطأ العشوائي المقابلة لعناصر مختلفة من العينة مستقلة عن بعضها البعض .

4- يمكن دمج الشرطين (2) و(3) في شرط واحد وكتابتهما ضمن مصفوفة التباين المشترك لـ ε كما يلي:

$$V = cov(\varepsilon\ ) = E[(\varepsilon - 0)(\varepsilon - 0)'] = E(\varepsilon * \varepsilon') = \sigma^2\ * I \qquad (8-2)$$

وذلك لأن :

$$V =\ cov(\varepsilon\ ) = E(\varepsilon * \varepsilon') = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} * [\varepsilon_1\ \varepsilon_2\ \varepsilon_3\ \dots\ \varepsilon_n] = \begin{bmatrix} E(\varepsilon_1\ \varepsilon_1) & E(\varepsilon_1\ \varepsilon_2) \dots E(\varepsilon_1\ \varepsilon_n) \\ E(\varepsilon_2\ \varepsilon_1) & E(\varepsilon_2\ \varepsilon_2) \dots E(\varepsilon_2\ \varepsilon_n) \\ & \vdots \\ E(\varepsilon_n\ \varepsilon_1) & E(\varepsilon_n\ \varepsilon_2) \dots E(\varepsilon_n\ \varepsilon_n) \end{bmatrix}$$

وبناء على الافتراضات السابقة نجد أن:

$$V =\ cov(\varepsilon\ ) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2\ * I \qquad (9-2)$$

أي أن مصفوفة التباينات المشتركة لـ ε هي مصفوفة قطرية وعناصر قطرها الرئيسي متساوية وتساوي σ^2 .

وباختصار يمكننا دمج الافتراضات (1) و(2) و(3) في عبارتين رياضيتين كما يلي :

$$E(\varepsilon) = E\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \vdots\\ \varepsilon_n\end{bmatrix} = \begin{bmatrix}E(\varepsilon_1)\\ E(\varepsilon_2)\\ E(\varepsilon_3)\\ \vdots\\ E(\varepsilon_n)\end{bmatrix} = \begin{bmatrix}0\\ 0\\ 0\\ \vdots\\ 0\end{bmatrix} \qquad (10-2)$$

$$V = cov(\varepsilon) = E(\varepsilon * \varepsilon') = \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \vdots\\ \varepsilon_n\end{bmatrix} * [\varepsilon_1\ \varepsilon_2\ \varepsilon_3 \dots \varepsilon_\mathrm{n}] = \sigma^2 * I \qquad (11-2)$$

5- ومن أجل إجراء اختبارات المعنوية اللاحقة يشترط أن يكون ε خاضعاً للتوزيع الطبيعي المتعدد الذي توقعه الشعاع (0) وله مصفوفة التباين المشترك $V = \sigma^2 I$، أي ان يكون خاضعاً لـ $N(0\,,\sigma^2 I)$.

6- أن يكون ε مستقلاً عن المتحولات X أي أن يكون:

$$cov(X\,,\varepsilon\,) = 0 \qquad (12-2)$$

وذلك حتى نتمكن من قياس وتفسير أثر المتحولات X على التابعY دون أي التباس بسبب تقاطعها مع ε. وسيتم الاعتماد على هذه الافتراضات في البراهين اللاحقة .

2-3 : تقدير معلمات النموذج بطريقة المربعات الصغرى :

لتقدير معلمات النموذج نسحب عينة من عناصر المجتمع بحجم n ونأخذ منها القياسات المتقابلة لكل من Y والمتحولات $X_1\ X_2 \dots X_p$، ثم نفترض أن الأعداد $b_0\ b_1\ b_2 \dots b_p$ هي تقديرات معينة للمعلمات $\beta_0\ \beta_1\ \beta_2 \dots \beta_p$ في المجتمع . وبذلك نجد أنه يمكننا حساب قيمة المشاهدة y_i من النموذج (2-1) السابق، ولكن بخطأ آخر e_i يسمى خطأ التقدير، ونكتب ذلك كما يلي:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} \dots b_p x_{ip} + e_i \qquad i : 1\ 2\ 3 \dots n \qquad (13-2)$$

وإذا أخذنا جميع المشاهدات فإنه يمكننا كتابتها مصفوفياً كما يلي:

$$\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_i\\ \vdots\\ y_n\end{bmatrix} = \begin{bmatrix}1 & x_{11} & x_{12} & \dots & x_{1p}\\ 1 & x_{21} & x_{22} & \dots & x_{2p}\\ \vdots & \vdots & \vdots & \dots & \vdots\\ i & x_{i1} & x_{i2} & \dots & x_{ip}\\ \vdots & \vdots & \vdots & \dots & \vdots\\ i & x_{n1} & x_{n2} & \dots & x_{np}\end{bmatrix}\begin{bmatrix}b_0\\ b_1\\ \vdots\\ b_i\\ \vdots\\ b_p\end{bmatrix} + \begin{bmatrix}e_1\\ e_2\\ \vdots\\ e_i\\ \vdots\\ e_n\end{bmatrix} \qquad (14-2)$$

وباستخدام الرموز المصفوفية يمكننا كتابة النموذج السابق كما يلي:

$$Y = X * b + e \qquad (15-2)$$

حيث أن e هو شعاع خطأ التقدير (البواقي) وهو يختلف عن شعاع الخطأ العشوائي ε ، أما b فهو شعاع تقدير المعلمات β . وهنا نلاحظ أنه يمكننا التغاضي عن الخطأ e وتقديرY من العلاقة (2-15)، وذلك بوضع ذلك التقدير مساوياً لـ $X * b$ ونكتب ذلك كما يلي:

$$\tilde{Y} = X * b \qquad (16-2)$$

ومن العلاقتين (2–15)(2–16) نحصل على أن Y الفعلية تساوي :

$$Y = \tilde{Y} + e \qquad (17-2)$$

وبذلك نجد أن شعاع البواقي e يساوي :

$$e = Y - \tilde{Y} = Y - X * b \qquad (18-2)$$

وأنه مقابل أي مشاهدة i يكون لدينا خطأ e_i يساوي ما يلي:

$$e_i = y_i - \tilde{y}_i \qquad i : 1\ 2\ 3 \dots n \qquad (19-2)$$

حيث أن $\tilde{y}_i$ هي ناتج جداء السطر i من المصفوفة X في العمود b ، وهو يساوي:

$$\tilde{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} \dots b_p x_{ip} \qquad (20-2)$$

وعلينا الآن إيجاد قيم المعلمات $b_0\ b_1\ \ b_2 \dots b_p$ بحيث يكون مجموع مربعات الأخطاء التقديرية في العينة أصغر ما يمكن . أي بحيث يكون المقدار:

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = Min \qquad (21-2)$$

والذي يمكن كتابته مصفوفياً كما يلي :

$$Q = e' * e = [(Y - Xb)'(Y - Xb)] = Min \qquad (22-2)$$

ولحساب عناصر الشعاع b نشتق Q بالنسبة لـ b (حسب قواعد الاشتقاق في المصفوفات) ونضعه مساوياً للصفر فنجد أن:

$$\frac{\partial Q}{\partial b} = \left[\frac{\partial}{\partial b}(Y - Xb)'\right](Y - Xb) + (Y - Xb)'\left[\frac{\partial}{\partial b}(Y - Xb)\right]$$

$$\frac{\partial Q}{\partial b} = [-X'](Y - Xb) + (Y - Xb)'[-X] = 0$$

$$\frac{\partial Q}{\partial b} = -X'(Y - Xb) - (Y - Xb)'\,(X) = 0 \qquad (23-2)$$

ولكن بما أن الجداء الأخير $(Y - Xb)'X_{n\,p+1}$ هو شعاع من المرتبة $(p+1)$ ، لذلك فإنه يساوي منقوله، أي يكون لدينا ما يلي :

$$(Y - Xb)'X = [(Y - Xb)'X]' = X'(Y - Xb) \qquad (24-2)$$

وبالتعويض في المشتق الأخير (2–23) نحصل على ما يلي:

$$\frac{\partial Q}{\partial b} = -X'(Y - Xb) - X'(Y - Xb) = -2X'(Y - Xb) = 0 \qquad (25-2)$$

وبعد الإصلاح نحصل منها على المعادلة التالية .

$$X'Y = X'Xb$$

أو على المعادلة

$$X'Xb = X'Y \qquad (26-2)$$

فإذا كانت المصفوفة $X'X$ نظامية (محددها غير معدوم $|X'X| \neq 0$) فإن مقلوبها يكون موجوداً .

وعندها نضرب من اليسار طرفي المعادلة (2–26) بـ $(X'X)^{-1}$ ، فنحصل على أن:

$$(X'X)^{-1} * X'X * b = (X'X)^{-1} * X'Y$$

$$I * b = (X'X)^{-1} * X'Y$$

ومنها نجد أن b تساوي :

$$b = (X'X)^{-1} * X'Y \qquad (27-2)$$

ومنها يمكننا حساب عناصر الشعاع b بإجراء مطابقة بين عناصر الطرفين، ثم نعوض b في العلاقة (26–2) ، فنحصل على الصيغة المقدرة لتابع الانحدار التالية :

$$\tilde{Y} = X * b \qquad : b = \tilde{\beta} \qquad (28-2)$$

ولقد اعتمدنا الرمز b بدلاً عن الرمز $\tilde{\beta}$ لتسهيل التعامل معه في المعالجات الرياضية.

ملاحظة: يسمى شعاع التقديرات b بشعاع تقديرات المربعات الصغرى للمعلماتβ ، وهنا نشير إلى أنه يمكننا حساب عناصر الشعاع b من منقوله b' ، الذي يساوي :

$$b' = [(X'X)^{-1} * X'Y]' = Y'X * (X'X)^{-1} \qquad (29-2)$$

وذلك لأن المصفوفة $(X'X)^{-1}$ متناظرة ومنقولها يساويها .

ولدراسة جودة تمثيل النموذج نحلل مجموع مربعات انحرافات التابع Y كما يلي :

$$\sum^{n}(y_i - \bar{y})^2 = \sum(y_i - \tilde{y}_i + \tilde{y}_i - \bar{y})^2 = \sum^{n}(y_i - \tilde{y})^2 + \sum^{n}(\tilde{y}_i - \bar{y})^2 + 0 \qquad (\mathbf{30-2})$$

ونرمز لهذه المجاميع بالرموز المعروفة التالية:

$$SST = SSE + SSX \qquad (31-2)$$

ثم نحسب معامل التحديد R^2 من إحدى العلاقات التالية :

$$R^2 = \frac{SSX}{SST} = \frac{\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum^{n}(y_i - \tilde{y})^2}{\sum_{i=1}(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST} \qquad (\mathbf{32-2})$$

وهو يعبر عن النسبة المئوية التي يفسرها النموذج من التباين الكلي .

وأخيراً يمكننا حساب معامل الارتباط المتعدد من العلاقة :

$$R = +\sqrt{R^2} \qquad (33-2)$$

كما يمكننا حساب معامل التحديد المعدل من العلاقة :

$$Ad_j.R^2 = 1 - (\frac{n-1}{n-p-1})(1-R^2) \qquad (34-2)$$

وعندما يكون حجم العينة n كبيراً ، فإن : $Ad_j.R^2 \simeq R^2$

حالات خاصة :

يمكننا أن نصيغ العلاقة (27–2) لحساب شعاع المعلمات b على شكل معادلات خطية بدلالة بيانات المصفوفة X والشعاع Y. ولكننا هنا سنستخدم العلاقة (26–2) لأنها تعطينا جملة من المعادلات الخطية للمعلمات المجهولة $b' = (b_0\ b_1\ \ b_2 \dots b_p)$ بدلالة بيانات العينة . وسنكتفي باستخراج هذه المعادلات للحالتين الخاصتين التاليتين :

- **الحالة الأولى:** حالة الانحدار الخطي البسيط . وعندها يكون شكل النموذج كما يلي:

$$Y = b_0 + b_1 X_1 + e = bX + e \qquad (35-2)$$

ولنفترض أننا سحبنا عينة حجمها n عنصراً وأخذنا منها القياسات المطلوبة لـ X و Y، فعندها نجد أن أشعة نموذج الانحدار البسيط تأخذ الأشكال التالية :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix} \quad b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \qquad (36-2)$$

وبتطبيق العلاقة (2−26) نجد أن:

$$X' * X * b = X * Y$$

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} * \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} * \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

وبإجراء عمليات الضرب في الطرفين نحصل على أن:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} * \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \qquad \mathbf{(37-2)}$$

وبإجراء الضرب في الطرف الأيسر والقيام بمطابقة العناصر المتقابلة في الشعاعين، نحصل على المعادلتين الخطيتين التاليتين:

$$\begin{aligned} nb_0 + b_1 \sum x_i &= \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 &= \sum x_i y_i \end{aligned} \qquad \mathbf{(38-2)}$$

وهما نفس المعادلتين (1−9) اللتين حصلنا عليهما في الانحدار البسيط . ومنهما نحسب كل b_0 و b_1 . وكان يمكن تطبيق المعادلة (2−27) فنحصل منها مباشرة على العلاقتين (1−10) و(1−11) ونترك ذلك للقارئ .

- **الحالة الثانية:** حالة الانحدار المتعدد الخطي بمتحولين X_1 و X_2 فقط، وعندها يكون شكل النموذج الخطي كما يلي:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + e = bX + e \qquad \mathbf{(39-2)}$$

وبذلك تكون المصفوفات المقابلة لعناصر العينة ذات الحجم n كما يلي :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ 1 & x_{13} & x_{23} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix} \quad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \qquad \mathbf{(40-2)}$$

وبتطبيق العلاقة (2−26) نجد أنها تساوي : $X'Xb = X' * Y$

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ 1 & x_{13} & x_{23} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

وبإجراء الضرب في الطرفين والقيام بمطابقة العناصر المتقابلة ، نحصل على ثلاث معادلات خطية هي:

$$\begin{aligned} nb_0 + b_1 \sum^n x_i + b_2 \sum^n x_i^2 &= \sum^n y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 &= \sum x_i y \\ b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 &= \sum x_i^2 * y_i \end{aligned} \qquad \mathbf{(41-2)}$$

ومنها نحسب المعلمات المجهولة : $b_0 \ , \ b_1 \ , \ b_2$

مثال تمهيدي (2−1): أوجد النموذج الرياضي للعلاقة الخطية البسيطة بين Yو X من البيانات التالية :

جدول (2−2):

i	1	2	3	4	5	المتوسط
x_i	0	1	2	3	4	2
y_i	1	4	3	8	9	5

الحل: نفترض أن النموذج الرياضي هو من الشكل التالي :

$$Y = b_0 + b_1 X_1 + e = Xb + e$$

ولحساب المعلمتين b_1 و b_0 نشكل المصفوفات والأشعة كما يلي:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \ , X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \ , b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{5*30 - 10*10} \begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 25 \\ 70 \end{bmatrix}$$

ومما سبق نجد أن الشعاع b يساوي :

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (X'X)^{-1} * X'Y = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix} * \begin{bmatrix} 25 \\ 70 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

أي أن معادلة النموذج المطلوب هي: $\tilde{Y} = 1 + 2X$

ولحساب القيم النظرية $\tilde{Y}$ نطبق العلاقة :

$$\tilde{Y} = X * b = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}$$

ولحساب الفروقات نحسب شعاع البواقي e فنجد أن :

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} = Y - \tilde{Y} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

ومنها نلاحظ أن مجموع البواقي يساوي :

$$\sum e_i = 0 + 1 - 2 + 1 + 0 = 0$$

ولكن مجموع مربعاتها يساوي :

$$\sum e_i^2 = e'e = [0 \quad 1 \quad -2 \quad 1 \quad 0] \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} = 0^2 + 1^2 + (-2)^2 + 1^2 + 0^2 = 6$$

ولحساب معامل التحديد R^2 نقوم بحساب مجموع مربعات انحرافات المتحول التابع Y عن المتوسط$(\bar{y} = 5)$, فنجد أن:

$$\sum (y_i - \bar{y})^2 = (1-5)^2 + (4-5)^2 + (3-5)^2 + (8-5)^2 + (9-5)^2 = 46$$

وبذلك نجد أن:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{6}{46} = \frac{40}{46} = 0.8696$$

ومنه نجد أن معامل الارتباط المتعدد يساوي :

$$R_{yx} = +\sqrt{R^2} = 0.9325$$

وهي مؤشرات جيدة وتدل على جودة تمثيل النموذج للعلاقة بين Y وX المدروسين .

4-2 : حساب تباينات شعاع التقديرات b :

لحساب هذه التباينات نقوم بإيجاد مصفوفة التباينات والتباينات المشتركة للشعاع b عن الشعاع β ، فنجد أنه يمكننا الاستفادة من العلاقة (2-4) وكتابة العلاقة (2-27) كما يلي:

$$b = (X'X)^{-1} * X'Y = (X'X)^{-1}X'(X\beta + \varepsilon)$$

$$b = (X'X)^{-1} * X'X\beta + (X'X)^{-1}X' * \varepsilon = I_{p+1} * \beta + (X'X)^{-1}X' * \varepsilon$$

$$b = \beta + (X'X)^{-1}X' * \varepsilon \qquad (42-2)$$

ومنها نجد أن الفرق $(b-\beta)$ يساوي :

$$(b-\beta) = (X'X)^{-1}X' * \varepsilon \qquad (43-2)$$

وبناء على ذلك يمكننا حساب مصفوفة التباينات المشتركة لـ b بأخذ التوقع الرياضي على جميع العينات الممكنة لجداءات تلك الفروقات كما يلي :

$$COV(b) = E[(b-\beta)(b-\beta)'] \qquad (44-2)$$

وبتعويض $(b-\beta)$ من العلاقة (4-43) نجد أن:

$$COV(b) = E\{[(X'X)^{-1}X' * \varepsilon] * [(X'X)^{-1}X' * \varepsilon]'\}$$

$$COV(b) = \mathrm{E}[(X'X)^{-1}X'\varepsilon * \varepsilon'X(X'X)^{-1}] \qquad \text{(لأن } (X'X)^{-1} \text{ متناظرة)}$$

$$COV(b) = (X'X)^{-1}X'E(\varepsilon * \varepsilon') * X(X'X)^{-1}]$$

وبالاستفادة من (2-11) نجد أن:

$$COV(b) = (X'X)^{-1}X' * \sigma_\varepsilon^2 I * X(X'X)^{-1}]$$

$$COV(b) = \sigma_\varepsilon^2(X'X)^{-1} * X'X * (X'X)^{-1}] = \sigma_\varepsilon^2(X'X)^{-1} * I$$

وأخيراً نجد أن:

$$COV(b) = \sigma_\varepsilon^2(X'X)^{-1} \qquad (45-2)$$

ولحساب عناصر المصفوفة $COV(b)$ علينا أن نجد تقديراً لتباين الخطأ العشوائي σ_ε^2 المجهول، ولذلك نلجأ إلى خطأ التقدير e ، ونرمز لمجموع مربعاتها بالرمز Q فيكون لدينا:

$$Q = \sum_{i=1}^{n} e_i^2 = e' * e \qquad (46-2)$$

ثم نأخذ التوقع الرياضي على جميع العينات الممكنة لـ Q فنجد من العلاقة (2-18) أن:

$$E(Q) = E(e' * e) = E[(Y-Xb)'(Y-Xb)] \qquad (47-2)$$

ومن جهة أخرى نجد من العلاقتين (2-4) و(2-27) أن:

$$Y - Xb = (X\beta + \varepsilon) - X[(X'X)^{-1}X' * (X\beta + \varepsilon)]$$

$$Y - Xb = X\beta + \varepsilon - X(X'X)^{-1}X'\mathrm{X}\beta - X(X'X)^{-1}X' * \varepsilon$$

$$Y - Xb = \varepsilon - X(X'X)^{-1}X' * \varepsilon = [I_n - X(X'X)^{-1}X'] * \varepsilon \qquad (48-2)$$

والآن لنرمز للمصفوفة ضمن القوس الأخير بالرمز Γ وهي من المرتبة $n * n$ ونكتبها كما يلي:

$$\Gamma = I_n - X(X'X)^{-1}X' \qquad (49-2)$$

علماً بأن المصفوفة Γ هي مصفوفة مشهورة وتتمتع بخاصتين هامتين هما :

1- إن المصفوفة Γ متناظرة، وذلك لأن منقولها يساويها، وللبرهان على ذلك نجد أن:

$$\Gamma' = [I_n - X(X'X)^{-1}X']' = [I_n' - X''(X'X)^{-1}X'] = [I_n - X(X'X)^{-1}X'] = \Gamma \qquad \mathbf{(50-2)}$$

2- إنها عديمة القوة (أو جامدة) لأن جداءها بنفسها يساويها، وللبرهان على ذلك نجد أن :

$$\begin{aligned}\Gamma * \Gamma &= [I - X(X'X)^{-1}X'] * [I - X(X'X)^{-1}X'] \\ &= I^2 - 2IX(X'X)^{-1}\mathrm{X}' + [X(X'X)^{-1}X'][X(X'X)^{-1}X'] \\ &= I - 2X(X'X)^{-1}X' + [X(X'X)^{-1}X' * X(X'X)^{-1}X'] \\ &= I - 2X(X'X)^{-1}X' + X(X'X)^{-1}IX'\end{aligned}$$

$$\Gamma * \Gamma = I - X(X'X)^{-1}X' = \Gamma \qquad (51-2)$$

والآن نعود إلى معالجة العلاقة (2–47) ونعوض $(Y - Xb)$ من العلاقة (2–48) فيها فنجد أن :

$$E(Q) = E\{[I_n - X(X'X)^{-1}X'] * \varepsilon\}]'[I_n - X(X'X)^{-1}X']\varepsilon \}$$

$$E(Q) = E[\varepsilon'(I_n - X(X'X)^{-1}X')][(I_n - X(X'X)^{-1}X') * \varepsilon] = E\left[\varepsilon' * \Gamma * \Gamma * \varepsilon\right]$$

وبناء على العلاقة (2–51) نجد أن:

$$E(Q) = E\left[\varepsilon' * \Gamma * \varepsilon\right] \qquad (52-2)$$

وإذا رمزنا لعناصر المصفوفة Γ ذات المرتبة $n * n$ بالرمز γ_{ij} ، فإننا نجد أنه يمكننا كتابة الجداء $\left(\varepsilon' * \Gamma * \varepsilon\right)$ كما يلي :

$$\varepsilon' * \Gamma * \varepsilon = \underset{(1*n)}{[\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]} * \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \cdots & \gamma_{2n} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \cdots & \gamma_{3n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \gamma_{nn} \end{bmatrix} * \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{(n\ 1)} = \text{عدد} \qquad \mathbf{(53-2)}$$

وبعد إجراء عمليات الضرب نحصل على عدد سلمي ينتج عن حصيلة الجداءات المختلفة . وبصورة عامة يمكننا أن نرتب ونكتب تلك الجداءات حسب مجاميعها القطرية وغير القطرية كما يلي :

$$\varepsilon' * \Gamma * \varepsilon = \sum_{i=1}^{n} \gamma_{ii}\varepsilon_i^2 + \sum_{i=1}^{n}\sum_{j\neq \mathrm{i}}^{n} \gamma_{ij} * \varepsilon_i * \varepsilon_j \qquad (54-2)$$

وبذلك نجد أن :

$$E(Q) = E\left[\varepsilon' * \Gamma * \varepsilon\right] = \sum_{i=1}^{n} \gamma_{ii}E(\varepsilon_i^2) + \sum_{i=1}^{n}\sum_{j\neq 1}^{n} \gamma_{ij} * E\left(\varepsilon_i * \varepsilon_j\right) \qquad (55-2)$$

وبما أن الافتراضات الأساسية في(2–7) تشترط أن تكون $E(\varepsilon_k * \varepsilon_\ell)$ مساوية للصفر ، فإن $E(Q)$ تأخذ الشكل التالي:

$$E(Q) = \sum_{i=1}^{n} \gamma_{ii}E(\varepsilon_i^2) = \sum_{i=1}^{n} \gamma_{ii} * \sigma_\varepsilon^2$$

$$E(Q) = \sigma_\varepsilon^2 \sum_{i=1}^{n} \gamma_{ii} = \sigma_\varepsilon^2 * \mathrm{tr}\,\Gamma \qquad (56-2)$$

وذلك لأن المجموع الأخير ما هو إلا مجموع العناصر القطرية للمصفوفة Γ . وهو ما يسمى بأثر المصفوفة Γ ويرمز له بالرمز $\mathrm{tr}\,\Gamma$ .. وهكذا نجد أن :

$$E(Q) = \sigma_\varepsilon^2 * \mathrm{tr}\,\Gamma = \sigma_\varepsilon^2 * \mathrm{tr}[I_n - X(X'X)^{-1}X'] \qquad (57-2)$$

ولحساب أثر المصفوفة الأخيرة نشير إلى أن الأثر tr يتمتع بالخاصتين التاليتين :

$$\mathrm{tr}(A \pm B) = \mathrm{tr}A \pm \mathrm{tr}B \qquad \text{(تجميعي)} \qquad (58-2)$$

$$\mathrm{tr}(A * B) = tr(B * A) \qquad \text{(تبديلي)} \qquad (59-2)$$

وبناء على ذلك نجد أن $E(Q)$ يساوي :

$$E(Q) = \sigma_\varepsilon^2[\mathrm{tr}I_n - \mathrm{tr}(X(X'X)^{-1} * X']$$

$$E(Q) = \sigma_\varepsilon^2[\mathrm{tr}I_n - \mathrm{tr}(X' * X(X'X)^{-1}]$$ (تبديل)

$$E(Q) = \sigma_\varepsilon^2[\mathrm{tr}I_n - \mathrm{tr}I_{p+1}]$$

$$E(Q) = \sigma_\varepsilon^2[n - (p+1)] \qquad (60-2)$$

وذلك لأن I_n مصفوفة واحدية من المرتبة n وإن مجموع عناصرها القطرية $\mathrm{tr}I_n = n$. ولأن I_{p+1} مصفوفة واحدية من المرتبة$(p+1)$ وان مجموع عناصرها القطرية $\mathrm{tr}\, I_{p+1} = p+1$.

وهكذا نجد أن التوقع $E(Q)$ يساوي التباين المجهول σ_ε^2 مضروباً بعدد معلوم هو $[n-(p+1)]$ ، وهذا يعني أنه يمكننا أن نستخرج من Q تقديراً غير متحيز لـ σ_ε^2 وذلك بتقسيم Q على العدد $[n-(p+1)]$، فنحصل على المقدار $\frac{Q}{n-(p+1)}$ ، وعندما نحسب توقعه نجد أن :

$$E\left(\frac{Q}{n-(p+1)}\right) = \frac{1}{n-(p+1)}E(Q) = \frac{1}{n-(p+1)} * \sigma_\varepsilon^2[n-(p+1)]$$

$$E\left(\frac{Q}{n-(p+1)}\right) = \sigma_\varepsilon^2 \qquad (61-2)$$

أي أن المقدار $\frac{Q}{n-(p+1)}$ هو تقدير غير متحيز لـ σ_ε^2 ، لذلك نرمز له بالرمز S_e^2 ونكتبه كما يلي :

$$\tilde{\sigma}_\varepsilon^2 = S_e^2 = \frac{Q}{n-(p+1)} = \frac{\sum_{i=1}^n e_i^2}{n-(p+1)} \qquad (62-2)$$

ويسمى المقدار S_e^2 بتقدير تباين الأخطاء العشوائية σ_ε^2 .

والآن نعود إلى قضية تقدير مصفوفة تباينات شعاع التقديرات b ، فنجد أنه يمكننا الآن تقدير $COV(b)$ في (2–45) ، وذلك بالاعتماد على التقدير $\tilde{\sigma}_\varepsilon^2$ الوارد في (2–62) ، فنحصل على أن :

$$\widehat{COV}(b) = S_e^2(X'X)^{-1} \qquad (63-2)$$

وهي مصفوفة مربعة من المرتبة $(p+1)(p+1)$. ومنها نجد أن تقديرات تباينات عناصر شعاع التقديرات b تساوي عناصر القطر الرئيسي للمصفوفة $(X'X)^{-1}$ مضروبة بـ S_e^2 ، فتباين العنصر b_j يساوي العنصر الواقع على القطر الرئيسي وعلى السطر j مضروباً بالتباين S_e^2 .

أما تقدير التباين المشترك لأي زوج من الأزواج الممكنة (مثل b_j, b_k) فيساوي عنصر المصفوفة $(X'X)^{-1}$ الواقع على تقاطع السطر j مع العمود k مضروباً بالتباين S_e^2 . ويمكننا أن نوضح ذلك كما يلي :

$$\widehat{COV}_{(p+1)(p+1)}(b) = S_e^2(X'X)^{-1} = S_e^2 * \begin{bmatrix} \sigma_{b_0}^2 & COV(b_0b_1) & \dots & COV(b_0b_p) \\ COV(b_1b_0) & \sigma_{b_1}^2 & \dots & COV(b_0b_p) \\ \dots & \dots & \dots & \dots \\ COV(b_pb_0) & COV(b_pb_1) & \dots & \sigma_{b_p}^2 \end{bmatrix} \qquad (\mathbf{64-2})$$

ويستفاد من هذه التباينات في عمليات اختبار معنوية المعلمات β ، ومعرفة أثر كل من المتحولات X على التابع Y .

- **خواص شعاع التقديرات b (المحسوبة بطريقة المربعات الصغرى)**

1- إن الشعاع b مرتبط مع Y بالعلاقة الخطية التالية :

$$b = (X'X)^{-1}X' * Y = C_{n*n}Y \qquad (65-2)$$

2- إن الشعاع b مرتبط مع الخطأ العشوائي ε بالعلاقة التالية :

$$b = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon \qquad (66-2)$$

ومنها نستنتج أن:

$$b - \beta = (X'X)^{-1}X'\varepsilon \qquad (67-2)$$

3- إن التوقع الرياضي للشعاع b يساوي شعاع المعلمات β، وللبرهان على ذلك نجد مباشرة من العلاقتين (2-66) و(2-5) أن :

$$E(b) = E[\beta + (X'X)^{-1}X'\varepsilon] = \beta + (X'X)^{-1}X'E(\varepsilon) = \beta \qquad (68-2)$$

أي أن الشعاع b هو تقدير غير متحيز لشعاع المعلمات β .

4- أن الشعاع b هو تقدير فعّال للشعاع β، لأن تباينه أصغر من تباين أي تقدير آخر (**انظر البرهان في** [Johnson, wichern P.236]).

- **خواص شعاع البواقي e:**

1- إن شعاع البواقي e يرتبط مع Yبالعلاقة التالية :

$$e = Y - \tilde{Y} = Y - Xb = Y - X(X'X)^{-1}X'Y = [I - X(X'X)^{-1}X']Y = \Gamma * Y \qquad (\mathbf{69-2})$$

2- إن شعاع البواقي e يرتبط مع ε بالعلاقة التالية :

$$e = \Gamma * Y = \Gamma(X\beta + \varepsilon) = \Gamma X\beta + \Gamma * \varepsilon == \Gamma * \varepsilon \qquad (\mathbf{70-2})$$

وذلك لأن الجداء $\Gamma X = 0$ لأن: $\Gamma * X = [I - X(X'X)^{-1}X']X = X - X = 0$

3- إن التوقع الرياضي للشعاع e يساوي الصفر، لأن :

$$E(e) = E\big(\Gamma * \varepsilon\big) = \Gamma * E(\varepsilon) = 0 \qquad (71-2)$$

وذلك لأن $E(\varepsilon) = 0$ حسب الافتراضات الأساسية (2-10) .

4- إن مصفوفة التباينات المشتركة لعناصر الشعاع e يساوي حسب العلاقتين (2-70) و(2-11) مايلي:

$$COV(e) = COV\big(\Gamma * \varepsilon\big) = \Gamma' * COV(\varepsilon) * \Gamma = \sigma_\varepsilon^2 * \Gamma' * \Gamma = \sigma_\varepsilon^2 * \Gamma \qquad (\mathbf{72-2})$$

أي أن تباين أي عنصر من e مثل e_i يساوي العنصر القطري في السطر i من المصفوفة Γ مضروباً بالتباين المجهول σ_ε^2 أي أن :

$$Vor(e_i) = \sigma_\varepsilon^2 * (\gamma_{ii}) \qquad (73-2)$$

وبما أن σ_ε^2 يقدر من العلاقة (2-62) بواسطة S_e^2 فإننا نجد أن:

$$\widetilde{Vor}(e_i) = S_e^2 * (\gamma_{ii}) \qquad (74-2)$$

وإن انحرافه المعياري يقدر كما يلي:

$$\widetilde{Sd}(e_i) = \sqrt{S_e^2 * (\gamma_{ii})} = S_e * \sqrt{\gamma_{ii}} \qquad (75-2)$$

5-إن شعاع البواقي e مستقل عن شعاع التقديرات b لأن :

$$COV(b\,,e) = E[(b-\beta)(e-0)']$$

ومن (2-67) نجد أن: $(b-\beta) = (X'X)^{-1}X'\varepsilon$ ومن (2-70) نجد أن: $e = \Gamma * \varepsilon$، وبذلك نجد أن:

$$COV(b\,,e) = E\left[(X'X)^{-1}X' * \varepsilon\varepsilon' * \Gamma\right] = (X'X)^{-1}X'E(\varepsilon\varepsilon') * [I - X(X'X)^{-1}X'] =$$
$$= \sigma_\varepsilon^2[(X'X)^{-1}X'(I - X(X'X)^{-1}X')] = 0 \qquad (76-2)$$

وذلك لأن الجداء :

$$X'(I - X(X'X)^{-1}X') = X' - X' = 0 \qquad (77-2)$$

6-إن شعاع البواقي e مستقل عن المتحولات X . لأنه اعتماداً على (2-77) نجد أن :

$$X' * e = X' * \Gamma Y = X'[I - X(X'X)^{-1}X'] * Y = 0 * Y = 0 \qquad (78-2)$$

7-إن شعاع البواقي e مستقل عن شعاع القيم المقدرة $\tilde{Y}$ لأنه اعتماداً على (2-72) نجد أن :

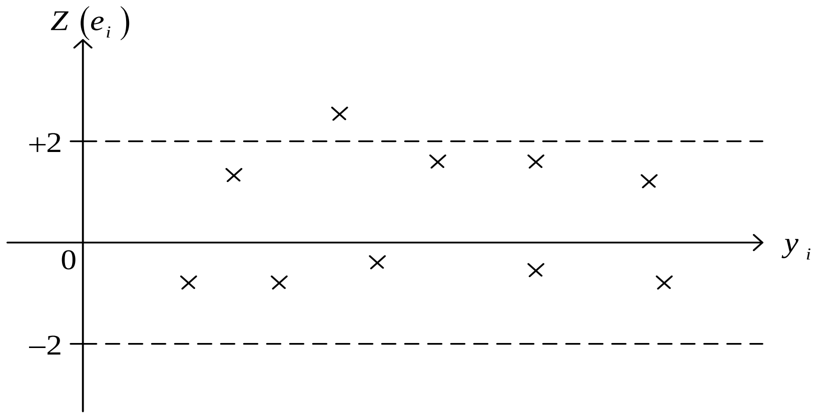
$$\tilde{Y}' * e = (Xb)' * e = b'X' * e = b' * 0 = 0 \qquad (79-2)$$

8-لتحليل ورسم شعاع البواقي نقوم بتحويله إلى شعاع معياري وحساب القيم المعيارية له $Z(e_i)$ من العلاقة التالية:

$$Z(e_i) = \frac{e_i - \bar{e}}{Sd(e_i)} = \frac{e_i - 0}{S_e\sqrt{\gamma_{ii}}} \qquad (80-2)$$

ثم نقوم برسم شكل الانتشار لهذه النقاط المعيارية بدلالة علاقتها مع $\tilde{Y}$ (لأنه متعامد معها)، أي نقوم بحساب ورسم النقاط الهندسية $\left(\tilde{y}_i\,, Z(e_i)\right)$ على المستوى الذي محوره الأفقي $\tilde{y}_i$ ومحوره العمودي $Z(e_i)$. كما يمكننا رسم شكل الانتشار لـ $Z(e_i)$، بدلالة علاقتها مع X (لأنه متعامد معها أيضاً) .

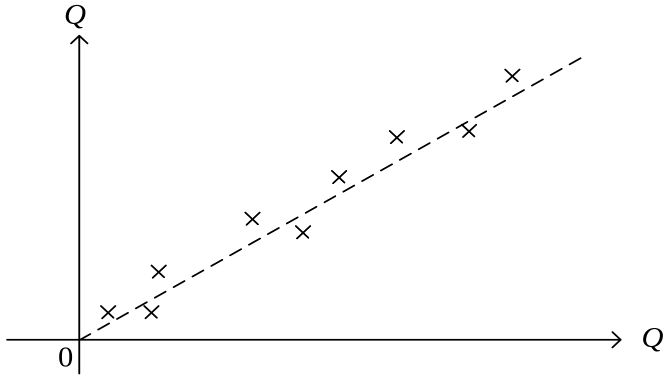
وبما أن $Z(e_i)$ قيم معيارية فإنها تخضع للتوزيع الطبيعي المعياري $N(0\,,1)$ ، ويجب أن تكون 95% من قيمتها محصورة في المجال $[-2\,,+2]$، وبالتالي نحصل على شكل لتوزع هذه النقاط على المستوى كما يلي:



الشكل (2-1): شريط توزع الأخطاء المعيارية $Z(e_i)$ بدلالة y_i

فإذا كان النموذج صالحاً لتمثيل العلاقة المطلوبة، فإن معظم نقاط$Z(e_i)$ ستقع ضمن الخطين المرسومين $[-2\,,+2]$ وإذا كانت مبعثرة خارجه فإن التمثيل يكون غير جيد .

وأخيراً يمكننا أن نحسب ونرسم مخطط $Q-Q$، فإذا كان النموذج صالحاً لتمثيل العلاقة فإن مخطط $Q-Q$ سيأخذ شكل خط مستقيم كما يلي:



الشكل (2-2): مخطط $Q-0$ للأخطاء المعيارية

مثال (2-2): ادرس العلاقة الانحدارية الخطية بين الانفاق الشهري للأسرة Y مع دخلها X_1 وعدد أفرادها X_2 ، وذلك من بيانات عينة عشوائية مؤلفة من15أسرة ، المبينة في الجدول التالي:

جدول (2-3): بيانات العينة (الانفاق والدخل بآلاف الليرات السورية)

رقم الأسرة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∑
الانفاق y_i	70	62	65	70	75	80	85	75	68	80	88	90	94	98	100	1200
الدخل X_{1i}	100	90	95	110	115	120	135	140	115	115	120	130	125	140	150	1800
X_{2i} عدد الأفراد	3	4	4	5	6	7	8	5	6	8	8	7	7	8	9	95
$\tilde{y}_i$	62.35	62.7	63.63	72.35	77.35	82.35	90.45	81.65	77.35	84.25	85.80	85.46	83.90	92	98.55	1200
$(y-\tilde{y})^2$	58.52	0.49	0.56	5.52	5.52	5.52	29.70	44.22	87.42	18.06	4.84	20.70	102.01	36	2.10	420.66

الحل: نفترض أن نموذج الانحدار يأخذ الشكل المتعدد التالي :

$$Y = b_0 + b_1X_1 + b_2X_2 + e$$
$$Y = Xb + e$$

حيث أن:

$$Y=\begin{bmatrix}70\\62\\65\\\vdots\\68\\\vdots\\100\end{bmatrix}\quad X=\begin{bmatrix}1 & 100 & 3\\1 & 90 & 4\\\cdots & \cdots & \cdots\\1 & 115 & 6\\1 & 115 & 8\\\cdots & \cdots & \cdots\\1 & 140 & 8\\1 & 150 & 9\end{bmatrix}\quad e=\begin{bmatrix}e_1\\e_2\\\vdots\\e\\e\\\vdots\\e_n\end{bmatrix}\quad b=\begin{bmatrix}b_0\\b_1\\b_2\end{bmatrix}$$

ولحساب تقديرات المعاملات b نبدأ بحساب المصفوفة $(X'X)$ فنجد أن:

$$X'X = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 100 & 90 & 95 & \dots & 125 & 140 & 150 \\ 3 & 4 & 4 & \dots & 7 & 8 & 9 \end{bmatrix} * \begin{bmatrix} 1 & 100 & 3 \\ 1 & 90 & 4 \\ 1 & 95 & 4 \\ \vdots & \vdots & \vdots \\ 1 & 125 & 7 \\ 1 & 140 & 8 \\ 1 & 150 & 9 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 15 & 1800 & 95 \\ 1800 & 220150 & 11725 \\ 95 & 11725 & 647 \end{bmatrix} = \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2, & \sum x_{1i}x_{2i} \\ \sum x_{2i} + & \sum x_{2i}x_{1i} & \sum x_{2i}^2 \end{bmatrix}$$

ثم ننتقل إلى حساب مقلوبها $(X'X)^{-1}$، لذلك يجب علينا أن نتأكد من أنها نظامية ، ولهذا نقوم بحساب محددها D فنجد أنه يساوي:

$$D = |X'X| = 1237625 \neq 0 \qquad \left(X'X \text{ نظامية}\right)$$

ثم نقوم بحساب المصفوفة المساعدة لها، المؤلفة من المصغرات المقابلة لعناصرها . ونرمز لها $M(X'X)$ فنجد أنها تساوي :

$$M(X'X) = \begin{bmatrix} 4961425 & -50725 & 190750 \\ -50725 & 680 & -4875 \\ 190750 & -4875 & 62250 \end{bmatrix}$$

وبما هذه المصفوفة متناظرة فإن منقولها يساويها نفسها و وبذلك نجد أن المقلوب $(X'X)^{-1}$ يساوي :

$$(X'X)^{-1} = \frac{1}{D}M = \frac{1}{1237625}\begin{bmatrix} 4961425 & -50725 & 190750 \\ -50725 & 680 & -4875 \\ 190750 & -4875 & 62250 \end{bmatrix} =$$

ثم نقوم بحساب الجداء $(X'Y)$ فنجد أنه يساوي :

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 100 & 90 & 95 & \dots & 125 & 140 & 150 \\ 3 & 4 & 4 & \dots & 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 70 \\ 62 \\ 65 \\ \vdots \\ 68 \\ \vdots \\ 100 \end{bmatrix} = \begin{bmatrix} 1200 \\ 146405 \\ 7857 \end{bmatrix}$$

ومما سبق يمكننا حساب شعاع التقديرات b من العلاقة :

$$b = (X'X)^{-1} * X'Y = \frac{1}{1237625}\begin{bmatrix} 4961425 & -50725 & 190750 \\ -50725 & 680 & -4875 \\ 190750 & -4875 & 62250 \end{bmatrix} * \begin{bmatrix} 1200 \\ 146405 \\ 7857 \end{bmatrix}$$

وبعد إجراء الحسابات نجد أن:

$$b = \begin{bmatrix} 21 \\ 0.31 \\ 3.45 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

وبذلك نحصل على النموذج المقدر التالي :

$$\tilde{Y} = 21 + 0.31X_1 + 3.45X_2$$

ولدراسة جودة تمثيل هذا النموذج نقوم بحساب القيم النظرية $\tilde{y}_i$ ، ثم نقوم بحساب مجموع مربعات البواقي فنجد من الجدول (2-3) الأساسي أن:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = 420.66$$

ثم نقوم بحساب مجموع مربعات انحرافات Y عن متوسطها $(\bar{y} = 80)$ فنجد أنه يساوي :

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 23337$$

ثم نقوم بحساب معامل التحديد R^2 من العلاقة :

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{420.66}{23337} = 0.82$$

ومنه نجد أن معامل الارتباط المتعدد $R = +\sqrt{0.82} = 0.91$

ولتقدير التباينات والتباينات المشتركة للتقديرات b علينا أن نحسب التباين S_e^2 من العلاقة :

$$S_e^2 = \frac{SSE}{n - (p + 1)} = \frac{420.66}{15 - (2 + 1)} = 35.055$$

ثم نحسب مصفوفة تلك التباينات المشتركة من العلاقة :

$$\widetilde{COV}(b) = S_e^2 (X'X)^{-1} = \frac{35.055}{1237625} \begin{bmatrix} 4961425 & -50725 & 190750 \\ -50725 & 680 & -4875 \\ 190750 & -4875 & 62250 \end{bmatrix}$$

ثم نقوم بحساب تباينات التقديرات وانحرافاتها المعيارية من العناصر القطرية للمصفوفة الناتجة فنجد أن:

$$S_{b_0}^2 = 35.055 \frac{4961425}{1237625} = 140.53$$

$$S_{b_0} = \sqrt{140.53} = 11.85$$

$$S_{b_1}^2 = 35.055 \frac{680}{1237625} = 0.0193$$

$$S_{b_1} = \sqrt{0.0193} = 0.139$$

$$S_{b_2}^2 = 35.055 \frac{62250}{1237625} = 1.763$$

$$S_{b_2} = \sqrt{1.763} = 1.33$$

ولحساب التباينات المشتركة (إذا لزم ذلك) نقوم بأخذ العناصر غير القطرية فمثلاً نجد أن :

$$COV(b_0, b_1) = 35.055 \frac{(-50725)}{1237625} = -1.4368$$

وهكذا دواليك ....

2-4-1: اختبار معنوية المعلمات β :

بعد أن قمنا بتقدير شعاع المعلمات β للنموذج (1−30) بواسطة الشعاع b المعرف بالعلاقة (1−55) والتي نكتبها كما يلي :

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = (X'X)^{-1} * X'Y \qquad (\mathbf{81-2})$$

وقمنا بتقدير تباين كل من هذه التقديرات من خلال مصفوفة التباين والتباين المشترك المعرفة بالعلاقة التالية :

$$\widehat{COV}(b) = S_e^2 (X'X)^{-1} \qquad (82-2)$$

والتي نستنتج منها أن تباين كل تقدير b_j يساوي العنصر القطري المقابل له في المصفوفة $(X'X)^{-1}$ مضروباً بـ S_e^2 ونرمز له كما يلي :

$$\sigma_{b_j}^2 = S_e^2 (X'X)_{jj}^{-1} \qquad (83-2)$$

وبذلك نجد أن الانحراف المعياري للتقدير b_j يساوي :

$$\sigma_{\mathrm{b}_j} = S_e \sqrt{(X'X)_{jj}^{-1}} \qquad (84-2)$$

ولاختبار معنوية أي تقدير b_j نستخدم اختبار (ستودينت) t ، ونضع الفرضيتين على المعالم الأساسية β_j في المجتمع كما يلي :

$$H_{0j} : \beta_j = 0 \qquad (85-2)$$

$$H_{1j} : \beta_j \neq 0 \qquad j : 1\ 2\ 3\ ..\ p \qquad (\text{بدون } \beta_0)$$

ثم نقوم بحساب مؤشر الاختبار t من العلاقة:

$$t_j = \frac{b_j - \beta_{0j}}{\sigma_{b_j}} = \frac{b_j - 0}{S_e \sqrt{(X'X)_{jj}^{-1}}} \qquad j = 0 \quad 1\ 2\ 3 \ldots p \qquad (\mathbf{86-2})$$

وهو متحول يخضع لتوزيع (ستودينت) بدرجة حرية $v = n - (p + 1)$. لذلك نقارن قيمة t_j المحسوبة مع القيمة الحرجة $t_v(\frac{\propto}{2})$ المقابلة لـ $v = n - p - 1$ درجة الحرية . ونتخذ القرار حول معنوية β_j أو عدمها كما يلي:

إذا كانت $|t_j| \leq t_v(\frac{\propto}{2})$ نقبل H_0 وتكون المعلمة β_j غير معنوية (لأن $H_0 : \beta_j = 0$) .

أما إذا كانت $|t_j| > t_v(\frac{\propto}{2})$ فإننا نرفض H_0 ونقبل H_1 التي تقول أن $\beta_j \neq 0$، ونستنتج أن تأثير المتحول المرافق له يقدر بالأمثال b_j .

كما يمكننا إنشاء مجال الثقة ذي الاحتمال $(1-\propto)$ للمعلمة β_j باستخدام العلاقة التالية :

$$b_j - t_{n-p-1}\left(\frac{\propto}{2}\right) * S_e \sqrt{(X'X)_{jj}^{-1}} \leq \beta_j \leq b_j + t_{n-p-1}\left(\frac{\propto}{2}\right) * S_e \sqrt{(X'X)_{jj}^{-1}} \qquad (\mathbf{87-2})$$

حيث أن : $j = 0\ 1\ 2\ 3 \ldots p$ **وأن:** $(X'X)_{jj}^{-1}$ **هو العنصر القطري j من المصفوفة** $(X'X)^{-1}$.

5-2 : تحليل التباين لدراسة صلاحية النموذج ككل :

لإجراء تحليل التباين للمتحول التابع Y نقوم بتحليل مجموع مربعات انحرافات Y عن المتوسط $\bar{Y}$ إلى مركباته الأساسية (كما فعلنا في الانحدار البسيط) كما يلي:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \tilde{y}_i + \tilde{y}_i - \bar{y})^2 \qquad (88-2)$$

وبعد التربيع والإصلاح والترتيب نحصل على أن:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 + 0 \qquad (89-2)$$

والتي نرمز لها كالعادة بالرموز التالية :

$$SST = SSX\left(\text{لمتحولات النموذج}\right) + SSE\left(\text{البواقي}\right) \qquad (90-2)$$

والتي يكون لها درجات الحرية التالية:

$$n - 1 = p + (n - p - 1) \qquad (91-2)$$

ولحساب هذه المجاميع باستخدام المصفوفات يمكننا تحويلها وحسابها من المعادلات التالية :

$$\begin{aligned} SST &= Y'Y - n\bar{Y}^2 \\ SSX &= b'X'Y - n\bar{Y}^2 \\ SSE &= Y'Y - b'X'Y \end{aligned} \qquad (92-2)$$

ولاختبار صلاحية النموذج ككل نضع الفرضيتين كما يلي (بدون β_0) .

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0 \qquad (93-2)$$

$H_1 : \beta_j \neq 0$ من أجل j واحدة على الأقل

ثم نقوم بتنظيم جدول تحليل التباين كما يلي :

جدول (2-3): تحليل التباين ANOVA

قيمة F	متوسط المربعات	درجة الحرية	الرموز	مصدر التباين
$F = \dfrac{MSSX}{MSE}$	$MSSX = \dfrac{SSX}{P}$	P	SSX	المتحولات $X_1\, X_2 \ldots X_p$
	$MSE = \dfrac{SSE}{n-p-1}$	$n-p-1$	SSE	البواقي (الأخطاء)
		$n-1$	SST	الإجمالي

ومنه نحسب قيمة مؤشر الاختبار F المعرف بالعلاقة التالية:

$$F = \frac{MSSX}{MSE} = \frac{\frac{SSX}{P}}{\frac{SSE}{n-P-1}} = \frac{(b'X'Y - n\bar{Y}^2)/P}{\left(Y'Y - b'X'Y\right)/n - P - 1} \qquad \mathbf{(94-2)}$$

وهو متحول عشوائي يخضع للتوزيع F بدرجتي حرية $v_1 = p$ و $v_2 = \mathrm{n} - p - 1$ لذلك نقوم بمقارنة القيمة المحسوبة لـ F مع القيمة الحرجة $F_{v_1,v_2}(\propto)$ المقابلة لدرجتي الحرية v_1 و v_2، ونتخذ القرار حول صلاحية النموذج لتمثيل العلاقة بين التابع Y والمتحولات $X_1\ X_2\ ...\ \mathrm{X}_p$ كما يلي :

إذا كانت $F \le F_{v_1,v_2}(\propto)$، فإننا نقبل الفرضية H_0 ، التي تقول أن جميع β_j معدومة، أي أن المتحولات $X_1\ X_2\ ...\ X_p$ لا تؤثر على Y ولا ترتبط به ، أي أن النموذج غير صالح لتمثيل العلاقة بين Y والمتحولات $X_1\ X_2\ ...\ X_p$ (وعندها تكون $P \ge \propto$) .

أما إذا كانت $F > F_{v_1,v_2}(\propto)$، فإننا نرفض الفرضية H_0 ونقبل H_1 ، التي تقول أن واحدة على الأقل من المعلمات β_j لا تساوي الصفر. أي أن النموذج يمثل العلاقة بين Y و X (وعندها تكون $P < \propto$) .

2-6: الاستدلال من تابع الانحدار المقدر:

بعد أن يحصل الباحث على النموذج الانحداري المقدر يمكن أن يستخدمه في حل مسائل التقدير والتنبؤ بقيم Y مقابل أية نقطة محددة مثل x_0 . وبما أن احداثيات النقطة x_0 تأخذ شكل سطر مشابه لأسطر المصفوفة X لذلك سنرمز لها على شكل سطر x'_0 ونكتبه كما يلي:

$$x'_0 = [1\ \ , x_{01}\ , x_{02}, \dots , x_{0p}] \qquad (94\ a - 2)$$

وبالعودة إلى الشكل (1-2) والفقرة (1-8) في الانحدار البسيط نلاحظ أنه يوجد لدينا (3) قيم لـ Y مقابل النقطة x_0 ، وهي:

- القيمة الفعلية لـ Y المقابلة لـ x_0، ونرمز لها بـ y_0 وهي قيمة مجهولة .
- القيمة المتوقعة لـ y_0، ونرمز لها بـ $E(y_0)$. وهي أيضاً قيمة مجهولة ولكنها تقع على مستقيم المجتمع .
- القيمة التقديرية لـ y_0، ونرمز لها بـ $\tilde{y}_0$. وهي قيمة معلومة وتحسب من نموذج الانحدار (2-26) وتقع على مستقيم العينة .

لذلك سنعتمد على هذه القيمة التقديرية $\tilde{y}_0$ للتنبؤ بالقيمتين المجهولتين $E(y_0)$ و y_0 الفعلية .

1- التنبؤ بالقيمة المتوقعة لـ y_0 وهي $E(y_0)$ المقابلة للنقطة x_0 .

بما أن عملية التنبؤ تتألف من خطوتين هما:

- إيجاد تقدير غير متحيز للقيمة المتوقعة $E(y_0)$.
- إنشاء مجال ثقة لـ $E(y_0)$ باحتمال ثقة $(1 - \propto)$.

ولذلك نفترض أن y_0 هي القيمة الفعلية للتابع Y في المجتمع المقابلة لقيم المتحولات المستقلة في النقطة x_0 ، أي المقابلة للسطر $x'_0 = [1\ , x_{01}\ , x_{02}, \dots , x_{0p}]$ ، وبالتعويض في العلاقة (2-4) نجد أن القيمة الفعلية لـ Y في المجتمع تساوي :

$$y_0 = x'_0\beta + \varepsilon_0 \qquad \text{(وهي مجهولة)} \qquad (95 - 2)$$

حيث أن ε_0 هو مقدار الخطأ العشوائي في x_0 .

وإذا أخذنا توقع الطرفين على كل العينات الممكنة وملاحظة أن $(x_0'\beta)$ في المجتمع هو عدد سلمي نحصل على أن :

$$E(y_0) = E(x_0'\beta + \varepsilon_0) = x_0'\beta + E(\varepsilon_0) = x_0'\beta + 0 \qquad (96-2)$$

أي أن:

$$E(y_0) = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \cdots + \beta_{0p} x_{0p} = x_0'\beta \qquad (97-2)$$

ومن جهة أخرى نجد أن القيمة التقديرية $\tilde{y}_0$ تحسب من النموذج المقدر في (2–6) كما يلي:

$$\tilde{y}_0 = b_0 + b_1 x_{01} + b_2 x_{02} + \cdots + b_{0p} x_{0p} = x_0'b \qquad (98-2)$$

وإذا أخذنا التوقع الرياضي لـ $\tilde{y}_0$ على جميع العينات وبملاحظة أن $E(b) = \beta$ نجد أن:

$$E(\tilde{y}_0) = E(x_0'b) = x_0'E(b) = x_0'\beta \qquad (99-2)$$

وبمقارنة العلاقة (2–96) مع العلاقة (2–99) نجد أن:

$$E(\tilde{y}_0) = E(y_0) = x'\beta \qquad (100-2)$$

وبذلك نجد أن النموذج الخطي (2–4) والمقدر بالعلاقة (2–15) يعطينا أن المقدار: $\tilde{y}_0 = x_0'b$ ، وحسب (2–100)هو تقدير غير متحيز للقيمة المتوقعة : $E(y_0) = x_0'\beta$ المجهولة، وهو تقدير فعال لأن تباينه يأخذ أصغر قيمة ممكنة (حسب المربعات الصغرى) ويساوي :

$$\begin{aligned} Var(x_0'b) &= x_0'Cov(b)x_0 = x_0'[X'X)^{-1}\sigma_\varepsilon^2]x_0 \\ Var(x_0'b) &= \sigma_\varepsilon^2 * x_0'(X'X)^{-1}x_0 \end{aligned} \qquad (101-2)$$

وذلك لأن $COV(b) = (X'X)^{-1}\sigma_\varepsilon^2$ حسب العلاقة (2–45) .

وبما أننا نفترض أن متحول الخطأ العشوائي ε يخضع للتوزيع الطبيعي $N(0\,,\sigma_\varepsilon^2)$. فإن المقدار $\tilde{y}_0 = x_0'b$ يكون خاضعاً للتوزيع الطبيعي $N[x_0'\beta,\ \sigma_\varepsilon^2 x_0'(X'X)^{-1}x_0]$، الذي توقعه حسب ماسبق $E(\tilde{y}_0) = x_0'\beta$ وتباينه : $\sigma_\varepsilon^2 * x_0'(X'X)^{-1}x_0$ ، وبذلك فإن التحويل المعياري Z_0 التالي :

$$Z_0 = \frac{\tilde{y}_0 - E(yY_0)}{\sqrt{\sigma_\varepsilon^2 * x_0'(X'X)^{-1}x_0}} \sim N(0\ 1) \qquad (102-2)$$

يكون خاضعاً للتوزيع الطبيعي المعياري$N(0\ 1)$.

وبما أن التقدير غير المتحيز لـ σ_ε^2 هو S_e^2 المعرف بالعلاقة (2–62) التالية :

$$\tilde{\sigma}_\varepsilon^2 = S_e^2 = \frac{Q}{n-p-1} = \frac{\sum_{i=1}^n e_i^2}{n-p-1} \qquad (103-2)$$

فإن النسبة التالية: $\chi^2 = \frac{(n-p-1)S_e^2}{\sigma_\varepsilon^2}$ تخضع لتوزيع χ^2 بـ $v = (n-p-1)$ درجة حرية .

وإذا شكلنا من χ^2 وZ النسبة التالية: $\frac{Z}{\sqrt{\frac{\chi^2}{v}}}$ ، فإننا سنحصل على متحول (ستودينت) t التالي :

$$t = \frac{Z_0}{\sqrt{\frac{\chi^2}{n-p-1}}} = \frac{\tilde{y}_0 - E(y_0)}{\sqrt{\sigma_\varepsilon^2 * x_0'(X'X)^{-1}x_0}} * \sqrt{\frac{\sigma_\varepsilon^2}{S_e^2}} \qquad (104-2)$$

وبالاختصار على σ_ε نجد أن t يساوي:

$$t = \frac{\tilde{y}_0 - E(y_0)}{S_e\sqrt{x_0'(X'X)^{-1}x_0}} \sim t_{n-p-1} \qquad (105-2)$$

ويكون خاضعاً لتوزيع (ستودينت) بـ$(n-p-1)$ درجة حرية، واعتماداً على العلاقة الأخيرة (2-105) يمكننا إنشاء مجال الثقة ذي الاحتمال $(1-\propto)$ للقيمة المتوقعة المجهولة $E(y_0)$ من العلاقة التالية :

$$\tilde{y}_0 - t_{n-p-1}\left(\frac{\propto}{2}\right) * S_e\sqrt{x_0'(X'X)^{-1}x_0} \le E(y_0) \le \tilde{y}_0 + t_{n-p-1}\left(\frac{\propto}{2}\right) * S_e\sqrt{x_0'(X'X)^{-1}x_0} \quad \mathbf{(106-2)}$$

مع الانتباه إلى أن x_0' هو سطر مشابه لأسطر المصفوفة X ويكتب في النقطة x_0 كما يلي:

$$x_0' = [1\ , x_{01}\ , x_{02}, \dots, x_{0p}]$$

2- التنبؤ بالقيمة الفعلية y_0 المقابلة للنقطة $\mathbf{x_0}$:

إن عملية التنبؤ بالقيمةy_0 المقابلة للنقطة x_0 هي أكثر تعقيداً وغموضاً من تقدير القيمة المتوقعة $E(y_0)$ لـy_0 ، لذلك نرمز لإحداثيات النقطة x_0 على شكل سطر مشابه لأسطر المصفوفة X ونكتبها كما يلي:

$$x_0' = [1\ , x_{01}\ , x_{02}, \dots, x_{0p}] \qquad (107-2)$$

وعندما نعوض إحداثيات x_0 في النماذج (2-99) و(2-15) و(1-16) نحصل على أن:

$$y_0 = x_0'\beta + \varepsilon_0 \quad , \quad y_0 = x_0'b + e_0 \quad , \quad \tilde{y}_0 = x_0'b \qquad (108-2)$$

ومنها نستخلص أن:

$$y_0 = x_0'\beta + \varepsilon_0 = x_0'b + e_0 = \tilde{y}_0 + e_0 \qquad (109-2)$$

$$\left(\text{القيمة الفعلية } y_0\right) = \left[\text{القيمة المقدرة لـ } y_0\right] + \left[\text{خطأ جديد } e_0\right]$$

حيث أن الأخطاء e_0 خاضعة للتوزيع الطبيعي$N(0\ ,\sigma_\varepsilon^2)$، وهي مستقلة عن$\varepsilon$ ومستقلة عن التقديرات b وعن التباينS_e^2 ، بينما الأخطاء ε تؤثر على b و S_e^2 من خلال التابع Y، ولكن e_0 لا يفعل ذلك .

وباتباع نفس المعالجة السابقة التي اتبعناها في مسألة التنبؤ بـ$E(y_0)$ ، واعتماداً على الافتراضات الابتدائية نجد أن التوقع الرياضي للقيمة الفعلية المتنبأ بها y_0 يساوي :

$$E(y_0) = E(x_0'b + e_0) = x_0'E(b) + E(e_0) = x_0'\beta + 0 \qquad (110-2)$$

وأن تباين القيمة الفعلية المتنبأ بها يساوي :

$$\begin{aligned} Var(y_0) &= Var(x_0'b + e_0) = x'Cov(b)x_0 + Var(e_0) + 2(0) \\ &= x_0'[(X'X)^{-1}\sigma_\varepsilon^2]x_0 + \sigma_\varepsilon^2 = \sigma_\varepsilon^2[1 + x_0'(X'X)^{-1}x_0] \end{aligned}$$

وإن الانحراف المعياري يقدر بالعلاقة (لأن $\tilde{\sigma}_\varepsilon^2 = S_e^2$):

$$S\widetilde{Sd}(y_0) = S_e\sqrt{[1 + x_0'(X'X)^{-1}x_0]} \qquad (111-2)$$

ومما سبق وبناء على الافتراضات الابتدائية وبإجراء معالجة مشابهة للمعالجة المتبعة في الحالة السابقة، نجد أنه يمكننا أن ننشأ مجال الثقة ذي الاحتمال $(1-\propto)$ للقيمة الحقيقية y_0 من العلاقة التالية :

$$y_0 = \tilde{y}_0 \pm t_{n-p-1}\left(\frac{\propto}{2}\right) * S_e\sqrt{[1 + x_0'(X'X)^{-1}x_0]} \qquad (112-2)$$

مع الانتباه إلى أن x_0' هو سطر إحداثيات النقطة X_0 ويكتب كما يلي:

$$x_0' = [1\ , x_{01}\ , x_{02}, \dots, x_{0p}]$$

مثال (2–3): لنتابع المثال (2–2) ونقوم باختبار معنوية شعاع المعلمات β والذي كان تقديره مساوياً لـ:

$$\tilde{\beta} = b = \begin{bmatrix} 21 \\ 0.31 \\ 3.45 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

ونضع فرضيات الاختبار على المعلمتين β_1 و β_2 فقط كما يلي: (لأن b_0 هو حد ثابت لا يدخل في التأثير على تغيرات Y) .

$$H_{01} : \beta_1 = 0 \qquad \text{مقابل}\ H_{11} : \beta_1 \neq 0$$

$$H_{02} : \beta_2 = 0 \qquad \text{مقابل}\ H_{12} : \beta_2 \neq 0$$

ثم نقوم باختبار معنوية كل من β_1 و β_2 على حدة كما يلي:

- **لاختبار معنوية المعلمة β_1** : نقوم بحساب مؤشر الاختبار t من العلاقة (2–86) ، واعتماداً على الحسابات السابقة في المثال (2–1) ، نجد أن:

$$t_1 = \frac{b_1 - \beta_{01}}{Sd(b_1)} = \frac{b_1 - 0}{Sd(b_1)} = \frac{0.31 - 0}{0.139} = 2.23$$

ثم نقوم بمقارنة هذه القيمة المحسوبة t_1 مع القيمة الحرجة $t_{n-p-1}\left(\frac{\propto}{2}\right)$ والتي تساوي عندما ($\propto = 0.05$) ما يلي:

$$t_{n-p-1}\left(\frac{\propto}{2}\right) = t_{15-2-1}(0.025) = t_{12}(0.025) = 2.179$$

وعند المقارنة نجد أن $|t_1| > 2.179$، لذلك نرفض فرضية العدم H_{01} (حول أن $\beta_1 = 0$) ونعتبر المعلمة β_1 معنوية ونستنتج أن المتحول X_1 يؤثر على تغيرات Y .

ولإنشاء مجال الثقة ذي الاحتمال 95% لـ β_1 نطبق العلاقة (2–87) فنجد أن:

$$0.31 - 2.179(0.139) \leq \beta_1 \leq 0.31 + 2.179(0.139)$$
$$0.0071 \leq \beta_1 \leq 0.61$$

وهو مجال لا يتضمن الصفر. وهذا يؤكد لنا معنوية المعلمة β_1.

- **ولاختبار معنوية المعلمة β_2:** نقوم بحساب مؤشر الاختبار t_2 اعتماداً على الحسابات السابقة في المثال (2–1) من العلاقة التالية :

$$t_2 = \frac{b_2 - \beta_{02}}{Sd(b_2)} = \frac{3.45 - 0}{1.33} = 2.59$$

وبمقارنة هذه القيمة المحسوبة t_2 مع القيمة الحرجة $t_{n-p-1}\left(\frac{\propto}{2}\right)$ نفسها والتي تساوي $t_{12}(0.025) =$ 2.179 ، فنجد أن: $|t_2| > 2.179$، لذلك نرفض فرضية العدم H_{02} (حول أن $\beta_2 = 0$) ونقبل

الفرضية H_{12} ، أي نقبل أن تكون $\beta_2 \neq 0$ ، ومنها نستنتج أن المعلمة β_2 معنوية، وإن المتحول X_2 يؤثر على تغيرات Y، وبشكل أكثر من X_1 .

ولإنشاء مجال الثقة لـ β_2 نطبق العلاقة (2−87) فنجد أن:

$$3.45 - 2.179(1.33) \leq \beta_2 \leq 3.45 + 2.179(1.33)$$
$$0.552 \leq \beta_2 \leq 6.348$$

وهو مجال لا يتضمن الصفر. وهذا يؤكد لنا معنوية المعلمة β_2 .

ولاختبار صلاحية النموذج ككل نقوم بإجراء تحليل التباين لتغيرات Y ولذلك نحسب مجاميع المربعات التالية : [ارجع إلى حسابات المثال (2−2)] .

$$SST = Y'Y - n(\bar{Y})^2 = \sum_{i=1}^{n} y_i^2 - n(\bar{Y})^2 = 89052 - 15\left(\frac{1200}{15}\right)^2 = 2052$$

$$SSX = b'X'Y - n\bar{Y}^2 = [21\ , 0.31\ , 3.45]\begin{bmatrix} 1200 \\ 145405 \\ 7857 \end{bmatrix} - 15\left(\frac{1200}{15}\right)^2 = 1692.2$$

وبذلك نجد أن: $SSE = SST - SSX = 2052 - 1692.2 = 359.8$

ثم نضع فرضية العدم (للمعلمتين معاً) $H_0\ :\ \beta_1 = \beta_2 = 0$

مقابل الفرضية البديلة $H_1\ :\ \beta_j \neq 0$ من أجل j واحدة على الأقل

ثم نضع هذه النتائج في جدول كالتالي:

جدول (4-2): جدول ANOVA

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F
المتحولات $X_1X_2 \dots X_p$	$SSX = 1692.2$	2	846.1	$F = \frac{846.1}{29.98} = 28.22$
البواقي أو الأخطاء	$SSE = 359.8$	12	29.98	
الاجمالي	$SST = 2052$	14	——	

وبمقارنة هذه القيمة المحسوبة لـ F مع القيمة الحرجة $F_{v_1,v_2}(\propto)$ والتي تساوي :

$$F_{v_1,v_2}(\propto) = F_{2\ ,12}(0.05) = 3.888$$

نجد أن $F > 3.88$ ، لذلك نرفض فرضية العدم H_0 ونقبل H_1، التي تقول أنه توجد علاقة معنوية بين المتحولات $X_1X_2 \dots X_p$ والتابع Y، ونعتبر أن النموذج المقدر صالحاً لتمثيل هذه العلاقة .

ولإجراء عملية التنبؤ بالقيمة المتوقعة $E(y_0)$ ، المقابلة لنقطة اختيارية لأسرة دخلها $X_{01} = 120$ وعدد أفرادها $X_{02} = 5$ ، وعندها يكون لدينا $x'_0 = [1\ , 120.5]$. وبتعويض ذلك في النموذج نجد أن:

$$\tilde{y}_0 = x'_0 * b = (1\ \ ,120\ \ ,5)\begin{bmatrix} 21 \\ 0.31 \\ 3.45 \end{bmatrix} = 75.45$$

ولإنشاء مجال الثقة للقيمة المتوقعة $E(y_0)$ لإنفاق هذه الأسرة وباحتمال ثقة 0.95، نقوم بتطبيق العلاقة (2–106) التالية:

$$\tilde{y}_0 - t_{12}\left(\frac{\alpha}{2}\right) S_e\sqrt{x_0'(X'X)^{-1}x_0} \le \mathrm{E}(y_0) \le \tilde{y}_0 + t_{12}\left(\frac{\alpha}{2}\right) S_e\sqrt{x_0'(X'X)^{-1}x_0}$$

لذلك نقوم أولاً بحساب الجداء $x_0'(X'X)^{-1}x_0$ كما يلي:

$$x_0'(X'X)^{-1}\mathrm{x}_0 = \frac{(1\,,120\,,5)}{1237625}\begin{bmatrix} 4961425 & -50725 & 190750 \\ -50725 & 680 & -4875 \\ 190750 & -4875 & 62250 \end{bmatrix} * \begin{bmatrix} 1 \\ 120 \\ 5 \end{bmatrix}$$

$$x_0'(X'X)^{-1}x_0 = \frac{(1\,,120\,,5)}{1237625}\begin{bmatrix} -171825 \\ 6500 \\ -83000 \end{bmatrix}$$

$$x_0'(X'X)^{-1}x_0 = 0.156085$$

ومن نتائج حسابات المثال (2–2) نجد أن: $t_{12}(0.025) = 2.179$ وأن: $S_e^2 = 35.005$

وبالتعويض في العلاقة (2–106) نحصل على أن:

$$75.15 - 2.179\sqrt{35.005}\sqrt{0.156085} \le \mathrm{E}(y_0) \le 75.15 + 2.179\sqrt{35.005}\sqrt{0.156085}$$

$$70.35 \le \mathrm{E}(y_0) \le 80.55$$

ولإجراء عملية التنبؤ للقيمة الفعليةy_0 المقابلة لنفس الأسرة التي دخلها $X_{01} = 120$ وعدد أفرادها $X_{02} = 5$ اشخاص، نجد أن $x_0' = (1\,,120\,,5)$ وأن :

$$\tilde{y}_0 = x_0' * b = (1\,,120\,,5)\begin{bmatrix} 21 \\ 0.31 \\ 3.45 \end{bmatrix} = 75.45$$

ولإنشاء مجال الثقة للقيمة الفعلية y_0 نطبق العلاقة (2–112) فنجد أن:

$$75.45 - 2.179\sqrt{35.005}\sqrt{1 + 0.156085} \le Y_0 \le 75.45 + 2.179\sqrt{35.005}\sqrt{1 + 0.156085}$$

$$61.588 \le Y_0 \le 89.312$$

وهو مجال أوسع من المجال السابق لـ $\mathrm{E}(y_0)$.

2–7: الانحدار المتعدد الخطي المعياري:

يستخدم هذا الانحدار في كثير من الأبحاث العلمية، وذلك لتسهيل الحسابات ولتخفيض مراتب المصفوفات ولتصغير الأرقام ولتوحيد واحدات القياس لـ Y وللمتحولات المستقلة $X_1\ X\ X_2 \ldots X_n$، وهو يعتمد على تحويل جميع المتحولات X والتابع Y إلى متحولات معيارية Z . وذلك باستخدام العلاقات التالية :

$$Z_j = \frac{X_j - \bar{X}_j}{\sigma_j} \qquad j : 1\ 2\ 3 \ldots p \qquad (113-2)$$

$$Z_y = \frac{Y - \bar{Y}}{\sigma_y} \qquad (114-2)$$

وبعد التحويل نحصل على بيانات معيارية للمتحولات Z_j وللتابع Y، وتكون مجردة من واحدات القياس، وتكون قيمتها صغيرة (نسبياً) . ويكون متوسط كل Z_j مساوياً للصفر $\left(\bar{Z}_j = 0\right)$ وتباينه $\sigma^2_{Z_j} = 1$.

ثم نضع نموذج الانحدار المعياري بدون حد ثابت (لأنه منسوب إلى المركز $(\bar{x}\,,\bar{y})$) ونكتبه كما يلي:

$$Z_y = \tilde{\beta}_1 Z_1 + \tilde{\beta}_2 Z_2 + \tilde{\beta}_3 Z_3 + \cdots + \tilde{\beta}_p Z_p + e^* \qquad (115-2)$$

وهو يتضمن P معلمة فقط مرافقة لـ P متحولاً معيارياً ، إضافة إلى خطأ التقدير المعياري e^* .

ولإيجاد العلاقة بين هذا النموذج والنموذج العام نعيده إلى المتحولات الأصلية . ونكتب ذلك كما يلي:

$$\frac{Y-\bar{Y}}{\sigma_y} = +\tilde{\beta}_1\left(\frac{X_1-\bar{X}_1}{\sigma_1}\right) + \tilde{\beta}_2\left(\frac{X_2-\bar{X}_2}{\sigma_2}\right) + \cdots \tilde{\beta}_P\left(\frac{X_P-\bar{X}_P}{\sigma_P}\right) + e^* \qquad (\mathbf{116-2})$$

وبعد الإصلاح والمعالجة نجد أن:

$$\begin{aligned} Y = &\left(\bar{Y} - \tilde{\beta}_1\frac{\sigma_y}{\sigma_1}\bar{X}_1 - \tilde{\beta}_2\frac{\sigma_y}{\sigma_2}\bar{X}_2 + \cdots - \tilde{\beta}_P\frac{\sigma_y}{\sigma_P}\bar{X}_P\right) + \\ &\tilde{\beta}_1\frac{\sigma_y}{\sigma_1}X_1 + \tilde{\beta}_2\frac{\sigma_y}{\sigma_2}X_2 + \cdots + \tilde{\beta}_P\frac{\sigma_y}{\sigma_P}X_P + e^* \end{aligned} \qquad (117-2)$$

وبمقارنة هذا الشكل مع الشكل العام لنموذج الانحدار الخطي المتعدد، والذي كان يساوي:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_PX_P + e \qquad (118-2)$$

نجد أن أمثال النموذج العام ترتبط مع أمثال النموذج المعياري بالعلاقات التالية:

$$\begin{aligned} &b_0 = \left(\bar{Y} - \tilde{\beta}_1\frac{\sigma_y}{\sigma_1}\bar{X}_1 - \tilde{\beta}_2\frac{\sigma_y}{\sigma_2}\bar{X}_2 + \cdots - \tilde{\beta}_P\frac{\sigma_y}{\sigma_P}\bar{X}_P\right) \\ &b_1 = \tilde{\beta}_1\frac{\sigma_y}{\sigma_1} \quad \Leftrightarrow \quad \tilde{\beta}_1 = b_1\frac{\sigma_1}{\sigma_y} \\ &b_2 = \tilde{\beta}_2\frac{\sigma_y}{\sigma_2} \quad \Leftrightarrow \quad \tilde{\beta}_2 = b_2\frac{\sigma_2}{\sigma_y} \\ &\dots\dots\dots\dots\dots\dots \\ &b_P = \tilde{\beta}_P\frac{\sigma_y}{\sigma_P} \quad \Leftrightarrow \quad \tilde{\beta}_P = b_P\frac{\sigma_P}{\sigma_y} \end{aligned} \qquad (119-2)$$

ولكننا بعد أن نقوم بتحويل المتحولات المستقلة X والتابع Y إلى متحولات معيارية، لسنا بحاجة إلى إجراء كل هذه الحسابات لإيجاد تقديرات المعلمات المعيارية $\tilde{\beta}_1\ \tilde{\beta}_2\ \dots\ \tilde{\beta}_P$، ويمكننا اتباع نفس الخطوات التي اتبعناها في معالجة النموذج العام للحصول على تقديرات غير متحيزة لـ $\tilde{\beta}_1\ \tilde{\beta}_2\ \dots\ \tilde{\beta}_P$ وعلى تقدير لمصفوفة تبايناتها $COV(\tilde{\beta})$.

لذلك نعوض قيم المشاهدات المتقابلة في معادلة النموذج المعياري (2−115) ، فنحصل على n معادلة كما يلي:

$$\begin{aligned} Z_{y1} &= \tilde{\beta}_1 Z_{11} + \tilde{\beta}_2 Z_{12} + \tilde{\beta}_3 Z_{13} + \cdots + \tilde{\beta}_P Z_{1\mathrm{P}} + \mathrm{e}_1^* \\ Z_{y2} &= \tilde{\beta}_1 Z_{21} + \tilde{\beta}_2 Z_{22} + \tilde{\beta}_3 Z_{23} + \cdots + \tilde{\beta}_P Z_{2P} + \mathrm{e}_2^* \\ Z_{y3} &= \tilde{\beta}_1 Z_{31} + \tilde{\beta}_2 Z_{32} + \tilde{\beta}_3 Z_{33} + \cdots + \tilde{\beta}_P Z_{3P} + \mathrm{e}_3^* \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ Z_{yn} &= \tilde{\beta}_1 Z_{n1} + \tilde{\beta}_2 Z_{n2} + \tilde{\beta}_3 Z_{n3} + \cdots + \tilde{\beta}_P Z_{nP} + \mathrm{e}_p^* \end{aligned} \qquad (120-2)$$

نقوم بكتابة هذه المعادلات بشكل مصفوفي كما يلي:

$$\underset{Z_y}{\begin{bmatrix} Z_{y1} \\ Z_{y2} \\ Z_{y3} \\ \vdots \\ Z_{yn} \end{bmatrix}} = \underset{Z}{\begin{bmatrix} Z_{11}\ Z_{12}\ \ Z_{13} + \cdots + Z_{1P} \\ Z_{21}\ Z_{22}\ Z_{23} + \cdots + Z_{2P} \\ Z_{31}\ Z_{32}\ Z_{33} + \cdots + Z_{3P} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Z_{n1}\ Z_{n2}\ Z_{n3} + \cdots + Z_{nP} \end{bmatrix}} * \underset{\tilde{B}}{\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \\ \vdots \\ \tilde{\beta}_P \end{bmatrix}} + \underset{e^*}{\begin{bmatrix} e_1^* \\ e_2^* \\ e_3^* \\ \vdots \\ e_p^* \end{bmatrix}} \qquad \mathbf{(121-2)}$$

وباختصار يمكننا كتابة النموذج مصفوفياً على الشكل التالي :

$$Z_y = Z * \tilde{\beta} + e^* \qquad (122-2)$$

حيث أن e^* هو خطأ التقدير المعياري ويتميز بأن متوسطه يساوي الصفر، ولكن تباينه يختلف عن الواحد .

ولحساب التقديرات $\tilde{\beta}_1\, \tilde{\beta}_2\, ...\, \beta_P$ من النموذج (2−121) بطريقة المربعات الصغرى نتبع نفس الخطوات السابقة ونرمز للجزء النظري في النموذج (2−122) بالرمز $\tilde{Z}_y$ ونكتبه كما يلي :

$$\tilde{Z}_y = Z * \tilde{\beta} \qquad (123-2)$$

وبذلك نجد أن:

$$Z_y = \tilde{Z}_y + e^* \qquad (124-2)$$

ومنها نجد أن شعاع خطأ التقدير e^* يساوي :

$$e^* = Z_y - \tilde{Z}_y = \left(Z_y - Z\tilde{\beta}\right) \qquad (125-2)$$

وعلينا أن نحسب قيم $\tilde{\beta}_1\, \tilde{\beta}_2\, ...\, \tilde{\beta}_P$ التي تجعل مجموع مربعات الأخطاء e_i^* أصغر ما يمكن . أي التي تجعل المقدار Q^* التالي أصغر ما يمكن :

$$Q^* = \left[\sum_{i=1}^{n} e_i^{*2}\right] = e^{*\prime} * e^* = \left(Z_y - Z\tilde{\beta}\right)'\left(Z_y - Z\tilde{\beta}\right) \Longrightarrow min \qquad (126-2)$$

لذلك نقوم باشتقاق Q^* بالنسبة للشعاع $\tilde{\beta}$ (حسب قواعد اشتقاق المصفوفات) ونضع ذلك المشتق مساوياً للصفر فنجد أن:

$$\frac{\partial Q^*}{\partial \tilde{\beta}} = -Z'\left(Z_y - Z\tilde{\beta}\right) + \left(Z_y - Z\tilde{\beta}\right)'(-Z) = 0 \qquad (127-2)$$

وبما أن الجداء $\left(Z_y - Z\tilde{\beta}\right)'_{(1,n)} * Z_{n*P}$ هو شعاع من المرتبة $(1 * P)$ فإن منقوله يساوي نفسه أي أن:

$$\left(Z_y - Z\tilde{\beta}\right)' * Z = \left[\left(Z_y - Z\tilde{\beta}\right)' * Z\right]' = Z'\left(Z_y - Z\tilde{\beta}\right) \qquad (128-2)$$

وبالتعويض في (2−127) نحصل على أن:

$$\frac{\partial Q^*}{\partial \tilde{\beta}} = -2Z'\left(Z_y - Z\tilde{\beta}\right) = 0$$

ومنها نحصل على المعادلة التالية :

$$Z'Z_y = Z'Z\tilde{\beta} \qquad (129-2)$$

وبذلك نجد أنه إذا كانت مصفوفة الجداء $(Z'Z)$ نظامية $(|Z'Z| \neq 0)$، فإنه يمكننا حساب شعاع التقديرات المعيارية من العلاقة التالية :

$$\tilde{\beta} = (Z'Z)^{-1} * Z'Z_y \qquad (130-2)$$

وهي علاقة شبيهة بالعلاقة (2–27) لحساب شعاع المعلمات العادية b، ولكن المصفوفة Z المستخدمة هنا مختلفة عن المصفوفة X، حيث أنها لا تتضمن عمود الواحدات.

وكذلك يمكننا أن نحسب تقدير مصفوفة التباينات والتباينات المشتركة لهذه المعلمات المعيارية من العلاقة:

$$\widehat{COV}(\tilde{\beta}) = S_e^2(Z'Z)^{-1} \qquad (131-2)$$

حيث أن S_e^2 هو تقدير تباين الأخطاء e^* ويحسب من العلاقة التالية :

$$S_e^2 = \frac{\sum e^{*}{}_i^2}{n-P} = \frac{\sum_{i=1}^{n}\left(Z_y - \tilde{Z}_y\right)^2}{n-P} \qquad (132-2)$$

وذلك لأن النموذج (2–115) يتضمن P معلمة فقط هي $\tilde{\beta}_1\ \tilde{\beta}_2\ \dots\ \tilde{\beta}_P$.

ملاحظة: إن جميع البرامج الإحصائية تعطينا شعاع التقديرات العادية b وإلى جانبه شعاع التقديرات المعيارية $\tilde{\beta}$ (ولكن بدون الحد الثابت)، وتسمى المعلمات المعيارية في هذه البرامج بالمصطلح (Beta coefficients) ويمكن حساب حدود الثقة لكل منها من علاقات مشابهة للعلاقة (2–87) .

ورغم سهولة التعامل مع الانحدار المعياري من حيث الشكل والحسابات . ولكن عملية تفسير معلماته $\tilde{\beta}_j$ تختلف عن تفسيرات المعلمات العادية b_j في النموذج العادي ، وقد تكون غير مفهومة في بعض النماذج المتعددة .

لتوضيح ذلك نأخذ المثال التالي :

مثال (2–4): لنفترض أن معادلة الانحدار البسيط بين إنفاق الأسرةY ودخلها الشهري X كانت كما يلي:

$$Y = 100 + 0.5X = b_0 + b_1X \qquad (133-2)$$

وهنا يمكننا تفسير b_0 بأنه الانفاق الثابت في حال انعدام الدخل $(X = 0)$، أما تفسير العامل b_1 فيعني أن مقدار زيادة الانفاق Y عندما يزداد الدخل X بمقدار وحدة واحدة، لأنه عندما نعوض $(X + 1)$ في المعادلة نجد أن:

$$Y' = 100 + 0.5(X + 1) = 100 + 0.5 + 0.5 = Y + 0.5 \qquad (134-2)$$

وهكذا نجد أن $b_1 = 0.5$ هي مقدار زيادة الانفاق Y المقابل لزيادة X بمقدار الواحد ليصبح $(x + 1)$.

وعندما نقوم بإيجاد معادلة الانحدار المعياري لهذه العلاقة نعتمد على قيمة الانحراف المعياري لكل من X وY ولنفترض أن $\sigma_x = 6$ و$\sigma_y = 4$، وعندها نجد أن معادلة الانحدار المعياري تحسب من(2−123) و(2−119) كما يلي:

$$\tilde{Z}_y = Z_x * \tilde{\beta}_1$$
$$\tilde{Z}_y = b_1 \frac{\sigma_x}{\sigma_y} Z_x = 0.5 \frac{6}{4} * Z = 0.75\, Z \qquad (135-2)$$

حيث أن Z تقاس بواحدات الانحراف المعياري لـ X والذي $\sigma_x = 6$ ، وإن Z_y تقاس بواحدات الانحراف المعياري لـ Y وهو $\sigma_y = 4$. وهذا يعني أنه إذا ازداد المتحول المعياري Z بمقدار واحد انحراف معياري σ_x ، فإن Z_y ستزداد بمقدار 0.75 من واحدة الانحراف معياري σ_y لأن:

$$Z'_y = 0.75(Z+1) = 0.75Z + 0.75 = Z_y + 0.75 \qquad (136-2)$$

ولكننا من جهة أخرى نجد أن زيادة Z بمقدار واحد انحراف تكافئ زيادة X بمقدار 6 وحدات، وإن زيادة Z_y بمقدار 0.75 يعني زيادة Y بمقدار $4 * 0.75 = 3$ واحدات، فهل هذا صحيح ؟ للتأكد من ذلك نعوض في المعادلة الأولى فنجد أن:

$$Y' = 100 + 0.5(X+6) = 100 + 0.5X + 3 = Y + 3 \qquad (137-2)$$

وهذا يتطابق مع التحليل المعياري، ولكن مسألة اعتبار الانحرافات المعيارية كواحدات لقياس تغيرات X وY يعتبر عملاً غير مألوف، وعلى الباحثين التمعن في ذلك عند إجراء التحليل والاستدلال .

ملاحظة: إن قضايا الانحدار المتعدد كثيرة ومتشعبة كالارتباط الجزئي والخطية المتعددة والارتباط الذاتي واستخدام المتحولات الصورية وقضايا الإبطاء الزمني ...الخ، وهي معروضة في معظم مراجع الاقتصاد القياسي . لذلك رأينا عدم التعرض إليها في هذا المنشور حتى نبقى ضمن الإطار المحدد له كتحليل متعدد المتغيرات .

2−8: الانحدار المتعدد المضاعف (متعدد من الطرفين):

يتناول هذا الانحدار دراسة العلاقة بين عدة متحولات تابعة $Y_1\ Y_2\ Y_3\ \dots\ Y_r$، ومجموعة من المتحولات المستقلة المؤثرة عليها أو المفسرة لها $X_1\ X_2\ X_3\ \dots\ X_P$، حيث أن r عدد المتحولات التابعة، وp عدد المتحولات المستقلة المؤثرة .

ويفترض أن يكون كل تابع Y_k مرتبطاً مع المتحولات $X_1\ X_2\ \dots\ X_P$ في المجتمع وفق نموذج انحدار خطي خاص كما يلي:

$$\begin{aligned}
Y_1 &= \beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \cdots + \beta_{P1}X_P + \varepsilon_1 \\
Y_2 &= \beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \cdots + \beta_{P2}X_P + \varepsilon_2 \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
Y_r &= \beta_{0r} + \beta_{1r}X_1 + \beta_{2r}X_2 + \cdots + \beta_{Pr}X_P + \varepsilon_r
\end{aligned} \qquad (138-2)$$

ويشترط في شعاع الخطأ ε أن يحقق الشرطين التاليين :

$$E(\varepsilon) = 0 \qquad Var(\varepsilon) = V$$

علماً بأن حدود الأخطاء ε_k المتقابلة مع التوابع المختلفة Y_k يمكن أن تكون مترابطة .

ولمعالجة قضايا هذا الانحدار علينا أن نتبع نفس الخطوات التي اتبعناها في الانحدار المتعدد العادي، والعمل على تقدير جميع المعلمات β_{jk} الواردة في العلاقات (2−138)، لذلك نرمز للمصفوفة الموسعة لقيم المتحولات $X_1\ X_2\ X_3\ \dots\ X_P$ والمأخوذة من عينة بحجم n عنصراً كما يلي:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \dots x_{1P} \\ 1 & x_{21} & x_{22} \dots x_{2P} \\ 1 & x_{31} & x_{32} \dots x_{3P} \\ \vdots & \dots\dots & \dots\dots\dots \\ 1 & x_{n1} & x_{n2} \dots x_{np} \end{bmatrix} \qquad (139-2)$$

ونرمز لقيم التوابع Y_k المأخوذة من عناصر نفس العينة بالرمز الخاص التالي (نرمز لكل عمود في المصفوفة Y بعنصر في سطر)

$$\begin{array}{cc} & \begin{array}{cccc} Y_1 & Y_2 & \dots & Y_r \end{array} \\ Y_{n*r} = & \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1r} \\ y_{21} & y_{22} & \cdots & y_{2r} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nr} \end{bmatrix} = [Y(1) \vdots Y(2) \vdots \cdots \vdots Y(r)] \end{array} \qquad (140-2)$$

كما نرمز للمعلمات β_{jk} في المجتمع بالرمز الخاص التالي:

$$\beta = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0r} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{P1} & \beta_{P2} & \cdots & \beta_{Pr} \end{bmatrix} = [\beta(1) \vdots \beta(2) \vdots \cdots \vdots \beta(r)] \qquad (141-2)$$

كما نرمز لمصفوفة الأخطاء العشوائية ε بالرمز الخاص التالي:

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1r} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2r} \\ \vdots & \vdots & \cdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nr} \end{bmatrix} = [\varepsilon(1) \vdots \varepsilon(2) \vdots \cdots \vdots \varepsilon(r)] = \begin{bmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \vdots \\ \varepsilon_n' \end{bmatrix} \qquad (142-2)$$

وعندها يمكننا أن نكتب النموذج الخطي للانحدار المتعدد المضاعف على الشكل التالي:

$$Y_{n*r} = X_{n(P+1)} * \beta_{(P+1)r} + \varepsilon_{n*r} \qquad (143-2)$$

حيث أن:

$$\begin{array}{ll} E[\varepsilon(k)] = 0 & k = 1\ 2\ 3 \dots r \\ COV[\varepsilon(k)\,, \varepsilon(h)] = \sigma_{kh} * I & k\,, h\!: 1\ 2\ 3\ \dots r \end{array} \qquad (144-2)$$

وحيث أن الـ r مشاهدة في التجربة i (أو من العنصر i) يكون لها مصفوفة تباين مشترك تساوي :

$$V = COV(i\,, r) = [\sigma_{ir}] \qquad (145-2)$$

ولكن المشاهدات المأخوذة من تجارب أو عناصر مختلفة تكون غير مرتبطة .

علماً بأن عناصر المصفوفة β وعناصر المصفوفة V تشكل المعلمات المجهولة في هذه القضية والمطلوب تقديرها، ولتبسيط الأمور نأخذ أحد التوابع $Y(k)$ ونكتب معادلته الخطية كما يلي :

$$Y(k) = X * \beta(k) + \varepsilon(k) \qquad k; 1\ 2\ 3 \dots r \qquad (146-2)$$

ورغم أن: $COV(\varepsilon(k)) = \sigma_{kk} * I$، نجد أن أخطاء التوابع المختلفة في نفس التجربة قد تكون مترابطة. ولإيجاد تقدير غير متحيز لعناصر المصفوفة β نتبع طريقة المربعات الصغرى، ونأخذ بيانات المتحولات التابعة Y وبيانات المتحولات X من عينة بحجم n عنصراً من المجتمع المدروس .

وبطريقة مشابهة لما فعلناه في الانحدار المتعدد العام . نتوصل إلى أنه يمكننا تقدير عناصر المعلمة $\beta(k)$ للنموذج التابع $Y(k)$ من علاقة مشابهة للعلاقة (2–27) ونكتبها كما يلي:

$$\tilde{\beta}(k) = (X' * X)^{-1} * X' * Y(k) \qquad k: 1\ 2\ 3 \dots r \qquad (147-2)$$

وبتجميع هذه التقديرات المختلفة لـ $\beta(k)$ في مصفوفة واحدة β يمكننا أن نكتب تقدير عناصر المصفوفة β في (2–147) كما يلي:

$$\tilde{\beta} = [\beta(1) \vdots \beta(2) \vdots \beta(3) \vdots \cdots \vdots \beta(r)] = (X' * X)^{-1} * X'[Y(1) \vdots Y(2) \vdots \cdots \vdots Y(r)] \qquad (\mathbf{148-2})$$

أو على الشكل العام التالي :

$$\tilde{\beta} = (X' * X)^{-1} * X' * Y \qquad (149-2)$$

وإذا اعتمدنا القيمة التقديرية للمعلمات β وكتبناها كما يلي $\tilde{\beta} = [\tilde{\beta}(1) \vdots \tilde{\beta}(2) \vdots \cdots \vdots \tilde{\beta}(r)]$ فإنه يمكننا أن نحصل من النموذج (2–143) على تقدير للمصفوفة Y (مصفوفة القيم النظرية) كما يلي:

$$\tilde{Y} = X * \tilde{\beta} \qquad (150-2)$$

ومنه نجد أن :

$$Y = \tilde{Y} + \tilde{\varepsilon} \qquad (151-2)$$

وبذلك نجد أن مصفوفة الأخطاء ε تقدر كما يلي:

$$\tilde{\varepsilon} = Y - \tilde{Y} = Y - X\tilde{\beta} \qquad (152-2)$$

وأن مصفوفة مجموع مربعات هذه الأخطاء وجداءاتها المتقابلة تساوي:

$$\tilde{\varepsilon}' * \tilde{\varepsilon} = (Y - X\tilde{\beta})'(Y - X\tilde{\beta}) = \qquad (153-2)$$

$$\tilde{\varepsilon}'_{r*r} * \tilde{\varepsilon} = \begin{bmatrix} \left(Y(1) - X\tilde{\beta}(1)\right)' (Y(1) - X\tilde{\beta}(1)) \dots \left(Y(1) - X\tilde{\beta}(1)\right)' (Y(r) - X\tilde{\beta}(r)) \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \left(Y(k) - X\tilde{\beta}(k)\right)' (Y(1) - X\tilde{\beta}(1)) \dots \left(Y(k) - X\tilde{\beta}(k)\right)' (Y(r) - X\tilde{\beta}(r)) \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \left(Y(r) - X\tilde{\beta}(r)\right)' (Y(1) - X\tilde{\beta}(1)) \dots \left(Y(r) - X\tilde{\beta}(r)\right)' (Y(r) - X\tilde{\beta}(r)) \end{bmatrix} \qquad (\mathbf{154-2})$$

وهي مصفوفة من المرتبة $r * r$ وكل حد قطري فيها يمثل مجموع مربعات الأخطاء للتابع المقابل له $Y(k)$، والحدود الأخرى تمثل الجداءات المختلفة، فإذا كانت التقديرات $\tilde{\beta}(k)$ مأخوذة من طريقة المربعات الصغرى والمحسوبة من العلاقة (2–149) فإنها تجعل مجموع المربعات في كل حد قطري k أصغر ما يمكن .

وهذا يعني أن التقدير العام $(\tilde{\beta})$ المأخوذ من العلاقة (2–149)، يجعل مجموع العناصر القطرية في المصفوفة الكلية $\varepsilon' * \varepsilon$ أصغر ما يمكن، أي أنه يجعل أثر المصفوفة الكلية والمساوي لـ $tr[(Y - 1Xb)'(Y - Xb)]$ أصغر ما يمكن . وهذا يعني انه يجعل التباين الكلي في المصفوفة التالية $[(Y - 1Xb)'(Y - Xb)]$ أصغر ما يمكن .

2-8-1: خواص تقديرات المربعات الصغرى $\tilde{\beta}$ في الانحدار المضاعف :

عند استخدام تقديرات طريقة المربعات الصغرى $\tilde{\beta}$ فإنه يمكننا صياغة المصفوفات السابقة كما يلي:

$$\tilde{Y} = X\tilde{\beta} = X * (X' * X)^{-1} * X' * Y \quad \text{(مصفوفة القيم النظرية)} \quad (155-2)$$

$$\tilde{\varepsilon} = Y - \tilde{Y} = Y - X(X' * X)^{-1}X' * Y = [I - X(X' * X)^{-1}X']Y = \Gamma * Y \quad \text{(مصفوفة الأخطاء)} \quad (156-2)$$

وكذلك نجد أن:

$$\tilde{\varepsilon} = \Gamma * Y = [I - X(X'X)^{-1}X][X\beta + \varepsilon) = [X\beta + \varepsilon - X\beta - X(X'X)^{-1}X * \varepsilon] = \Gamma * \varepsilon$$

وتتمتع هذه التقديرات والمصفوفات بنفس شروط التعامد المذكورة في الانحدار المتعدد العام . وهي:

1- إن: $X' * \Gamma = 0$ أي أن X متعامدة دائماً مع Γ وذلك لأن :

$$X' * \Gamma = X'[I - X(X' * X)^{-1}X'] = X' - IX' = 0 \quad (157-2)$$

2- إن: $X'\tilde{\varepsilon} = 0$. أي أن X متعامدة مع البواقي $\tilde{\varepsilon}$ لأن :

$$X'\tilde{\varepsilon} = X' * \Gamma Y = X'[I - X(X' * X)^{-1}X']Y = [X' - IX'] * Y = 0 \quad (158-2)$$

3- إن: $\tilde{Y}'\tilde{\varepsilon} = 0$ أي أن مصفوفة القيم النظرية $\tilde{Y}$ متعامدة مع البواقي $\tilde{\varepsilon}$. لأن:

$$\tilde{Y}'\tilde{\varepsilon} = (X\beta)'[I - X(X' * X)^{-1}X']Y = \beta'X'[I - X(X' * X)^{-1}X']Y = 0 \quad (159-2)$$

4- إن: $Y'Y = \tilde{Y}'\tilde{Y} + \tilde{\varepsilon}' * \tilde{\varepsilon}$ وذلك لأن $\left(Y = \tilde{Y} + \tilde{\varepsilon}\right)$ ولأن:

$$Y'Y = \left(\tilde{Y} + \tilde{\varepsilon}\right)'\left(\tilde{Y} + \tilde{\varepsilon}\right) = \tilde{Y}' * \tilde{Y} + \tilde{Y}' * \tilde{\varepsilon} + \tilde{\varepsilon}' * \tilde{Y} + \tilde{\varepsilon}' * \tilde{\varepsilon}$$

وبما أن: $\tilde{Y}'\tilde{\varepsilon} = \tilde{\varepsilon}'\tilde{Y} = 0$ فإننا نحصل على العلاقة الهامة التالية :

$$Y'Y = \tilde{Y}'\tilde{Y} + \tilde{\varepsilon}' * \tilde{\varepsilon} \quad (160-2)$$

$$\begin{bmatrix}\text{المجموع الاجمالي لمربعات} \\ \text{الانحرافات وجداءاتها}\end{bmatrix} = \begin{bmatrix}\text{المجموع المقدر لمربعات} \\ \text{الانحرافات وجداءاتها}\end{bmatrix} + \begin{bmatrix}\text{مجموع مربعات البواقي} \\ \text{(الأخطاء) وجداءاتها}\end{bmatrix} \quad (161-2)$$

ومما سبق نستخلص أن مصفوفة مجموع مربعات البواقي وجداءاتها تساوي :

$$\tilde{\varepsilon}' * \tilde{\varepsilon} = Y'Y - \tilde{Y}'\tilde{Y} = Y'Y - \tilde{\beta}'X'X\tilde{\beta} \quad (162-2)$$

5- إن التوقع الرياضي للتقديرات $\tilde{\beta}(k)$ يساوي $\beta(k)$ لأن:

$$\tilde{\beta}(k) = (X'X)^{-1}X'Y = (X'X)^{-1}X'\left(X\beta(k) + \varepsilon\right) = \beta(k) + (X'X)^{-1}X' * \varepsilon \quad (163-2)$$

$$E\left(\tilde{\beta}(k)\right) = \beta(k) + (X'X)^{-1}X'\, E(\varepsilon) = \beta(k) \quad (163a-2)$$

وبشكل عام يكون لدينا:

$$E\left(\tilde{\beta}\right) = \beta \quad (163\, b-2)$$

6- إن التباين المشترك لكل زوج $\tilde{\beta}(k)$ و $\tilde{\beta}(h)$ يساوي :

$$COV\left(\tilde{\beta}(k), \tilde{\beta}(h)\right) = \sigma_{kh}(X' * X)^{-1} \quad : k, h = 1\ 2\ 3 \ldots . r + 1 \quad (164-2)$$

7- إن مصفوفة البواقي $\tilde{\varepsilon}$ والتي تساوي :

$$\tilde{\varepsilon} = [\tilde{\varepsilon}(1) \vdots \tilde{\varepsilon}(2) \vdots \cdots \vdots \tilde{\varepsilon}(r)] = Y - X\tilde{\beta}$$

تحقق العلاقات التالية :

$$E\left(\tilde{\varepsilon}(k)\right) = 0$$
$$E[\tilde{\varepsilon}'(k), \tilde{\varepsilon}(h)] = (n - P - 1) * \sigma_{kh} \qquad (165-2)$$

وبصورة عامة إن $\tilde{\varepsilon}$ تحقق العلاقتين التاليتين :

$$E(\tilde{\varepsilon}) = 0$$
$$E\left(\frac{\tilde{\varepsilon}' * \tilde{\varepsilon}}{n-p-1}\right) = V(\tilde{\varepsilon}) \qquad (166-2)$$

أي أن المصفوفة $\frac{\tilde{\varepsilon}' * \tilde{\varepsilon}}{n-P-1}$ تصلح لأن تكون تقديراً غير متحيز لـ V ونكتب ذلك كما يلي :

$$\tilde{V}(\tilde{\varepsilon}) = \frac{\tilde{\varepsilon}' * \tilde{\varepsilon}}{n-p-1} \qquad (167-2)$$

8- إن التباين المشترك لكل $\tilde{\beta}(k), \tilde{\varepsilon}(h)$ يساوي الصفر أي أن:

$$COV\left(\tilde{\beta}(k), \tilde{\varepsilon}(h)\right) = 0 \qquad (168-2)$$

أي أن كل عنصر من $\tilde{\beta}$ مستقل عن أي عنصر من $\tilde{\varepsilon}(k)$، وللبرهان على ذلك نلاحظ أن:

$$COV\big(\tilde{\beta}(k), \tilde{\varepsilon}(h)\big) = E\left[\left(\tilde{\beta}(k) - \beta(k)\right)(\tilde{\varepsilon}(h) - 0)'\right]$$
$$= E[(X'X)^{-1}X'\ \varepsilon(k) * \varepsilon'(h) * \Gamma]$$
$$= [(X'X)^{-1}X']E[\varepsilon(k) * \varepsilon'(h)] * [I - X(X'X)^{-1}X']$$
$$= \sigma_{kh} * [(X'X)^{-1}X' - I(X'X)^{-1}X'] = 0$$

ملاحظة: إن البراهين على هذه الخواص موجودة في كثير من المراجع المختصة [انظر Johnston, Wichern P304+305] .

مثال (2-5): ادرس الانحدار المتعدد المضاعف للعلاقة بين التابعين Y_1 و Y_2 مع متحول مستقل واحد X (للسهولة) وذلك من خلال البيانات التالية :

جدول (2-5) بيانات المسألة:

رقم العنصر *i*	1	2	3	4	5
قيم X_1	0	1	2	3	4
قيم Y_1	1	4	3	8	9
قيم Y_2	−1	−1	2	3	2

الحل: نفترض أن العلاقة بين كل متحول تابع Y_1 و Y_2 مع X هي علاقة خطية من الشكل:

$$Y_1 = X * \beta_1 + \varepsilon_1 \quad \Longrightarrow \quad y_{i1} = \beta_{01} + \beta_{11}X_i + \varepsilon_{i1}$$
$$Y_2 = X * \beta_2 + \varepsilon_2 \quad \Longrightarrow \quad y_{i2} = \beta_{02} + \beta_{12}X_i + \varepsilon_{i2}$$

حيث أن: $i: 1\ 2\ 3\ 4\ 5$ لأن حجم العينة $n = 5$

ومن البيانات السابقة نجد أن المصفوفات المطلوبة تساوي :

$$X=\begin{bmatrix}1&0\\1&1\\1&2\\1&3\\1&4\end{bmatrix} \quad ,Y=\begin{bmatrix}1&-1\\4&-1\\3&2\\8&3\\9&2\end{bmatrix} \quad ,\beta=\begin{bmatrix}\beta_{01}&\beta_{02}\\\beta_{11}&\beta_{12}\end{bmatrix} \quad ,\varepsilon=\begin{bmatrix}\varepsilon_{11}&\varepsilon_{12}\\\varepsilon_{21}&\varepsilon_{22}\\\varepsilon_{31}&\varepsilon_{32}\\\varepsilon_{41}&\varepsilon_{42}\\\varepsilon_{51}&\varepsilon_{52}\end{bmatrix}$$

وهكذا نجد أن:

$$X'X=\begin{bmatrix}1&1&1&1&1\\0&1&2&3&4\end{bmatrix}*\begin{bmatrix}1&0\\1&1\\1&2\\1&3\\1&4\end{bmatrix}=\begin{bmatrix}5&10\\10&30\end{bmatrix}$$

$$(X'X)^{-1}=\frac{1}{D}\begin{bmatrix}30&-10\\-10&5\end{bmatrix}=\frac{1}{150-100}\begin{bmatrix}5&-10\\-10&30\end{bmatrix}=\begin{bmatrix}0.6&-0.2\\-0.2&0.1\end{bmatrix}$$

ولحساب $X'Y$ نكتبه كما يلي :

$$X'Y=X'[Y(1)\vdots Y(2)]$$

$$X'Y(1)=\begin{bmatrix}1&1&1&1&1\\0&1&2&3&4\end{bmatrix}*\begin{bmatrix}1\\4\\3\\8\\9\end{bmatrix}=\begin{bmatrix}25\\70\end{bmatrix}$$

$$X'Y(2)=\begin{bmatrix}1&1&1&1&1\\0&1&2&3&4\end{bmatrix}*\begin{bmatrix}-1\\-1\\2\\3\\2\end{bmatrix}=\begin{bmatrix}5\\20\end{bmatrix}$$

وبما أن المصفوفة $\tilde{\beta}$ تساوي :

$$\tilde{\beta}=(X'X)^{-1}X'*[Y(1),Y(2)]=\begin{bmatrix}\beta_{01}&\beta_{02}\\\beta_{11}&\beta_{12}\end{bmatrix}=[\beta(1)\vdots\beta(2)]$$

ومنها نجد أن:

$$\tilde{\beta}(1)=(X'X)^{-1}X'*Y(1)=\begin{bmatrix}0.6&-0.2\\-0.2&0.1\end{bmatrix}*\begin{bmatrix}25\\70\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}$$

$$\tilde{\beta}(2)=(X'X)^{-1}X'*Y(2)=\begin{bmatrix}0.6&-0.2\\-0.2&0.1\end{bmatrix}*\begin{bmatrix}5\\20\end{bmatrix}=\begin{bmatrix}-1\\+1\end{bmatrix}$$

وبذلك نجد أن المصفوفة $\tilde{\beta}$ تساوي:

$$\tilde{\beta}=[\tilde{\beta}(1)\vdots\tilde{\beta}(2)]=\begin{bmatrix}1&-1\\2&+1\end{bmatrix}$$

وكان يمكن حسابها مباشرة من العلاقة التالية :

$$\tilde{\beta}=(X'X)^{-1}X'*Y=\begin{bmatrix}0.6&-0.2\\-0.2&0.1\end{bmatrix}*\begin{bmatrix}1&1&1&1&1\\0&1&2&3&4\end{bmatrix}*\begin{bmatrix}1&-1\\4&-1\\3&2\\8&3\\9&2\end{bmatrix}=$$

$$\tilde{\beta} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix} * \begin{bmatrix} 25 & 5 \\ 70 & 20 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & +1 \end{bmatrix}$$

وبذلك نحصل على نموذجي الانحدار للتابعين Y_1 و Y_2 كما يلي:

$$\tilde{Y}_1 = X\tilde{\beta}(1) = [1\,, \mathrm{X}] \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 + 2X$$

$$\tilde{Y}_2 = X\tilde{\beta}(2) = [1\,, \mathrm{X}] \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 + X$$

ولإيجاد القيم النظرية $\tilde{Y}$ نستخدم العلاقة التالية:

$$\tilde{Y} = X * \tilde{\beta} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 2 & +1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix}$$

ومنها نحسب مصفوفة الأخطاء $\tilde{\varepsilon}$ كما يلي :

$$\tilde{\varepsilon} = Y - \tilde{Y} = \begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ -2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix}$$

ومنه نجد أن الخواص المذكورة محققة لأن:

$$X' * \tilde{\varepsilon} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ -2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\tilde{Y}' * \tilde{\varepsilon} = \begin{bmatrix} 1 & 3 & 5 & 5 & 9 \\ -1 & 0 & 1 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ -2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$Y' * Y = \begin{bmatrix} 1 & 4 & 3 & 8 & 9 \\ -1 & -1 & 2 & 3 & 2 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix} = \begin{bmatrix} 171 & 43 \\ 43 & 19 \end{bmatrix}$$

$$\tilde{Y}' * \tilde{Y} = \begin{bmatrix} 1 & 3 & 5 & 5 & 9 \\ -1 & 0 & 1 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 165 & 45 \\ 45 & 15 \end{bmatrix}$$

$$\tilde{\varepsilon}' * \tilde{\varepsilon} = \begin{bmatrix} 0 & 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ -2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}$$

منها نجد أن مجاميع الجداءات تحقق العلاقة التالية :

$$Y' * Y = \tilde{Y}' * \tilde{Y} + \tilde{\varepsilon}' * \tilde{\varepsilon}$$

2-8-2: التنبؤ بالقيمة المتوقعة للتابع المضاعف $Y(k)$:

لتقدير شعاع القيمة المتوقعة لتابع الانحدار المنفرد $Y(k)$ المقابل لنقطة معينة xX_0 من فضاء العينة ذي الـ $(p+1)$ بعداً . نرمز لإحداثيات النقطة x_0، التي تأخذ شكل سطر ونكتبها كما يلي:

$$x_0' = [1, x_{01}, x_{02}, \dots x_{0P}]$$

لتقدير القيمة المتوقعة لـ Y نقوم أولاً بحساب القيمة النظرية لـ $Y(k)$ المقابلة لـ x_0' من العلاقة السابقة (2-155) فنجد أن:

$$\tilde{y}_0(k) = x_0' \, \tilde{\beta}(k) \qquad : k: 1\ 2\ 3 \dots r \qquad (167-2)$$

ويمكننا أن نقوم بتركيب هذه القيم في مصفوفة واحدة . ونكتبها كما يلي:

$$\tilde{y}_0 = x_0' \, \tilde{\beta} = x_0'[\beta(1) \vdots \beta(2) \vdots \cdots \vdots \beta(r)] \qquad (168-2)$$

وهنا نلاحظ أن $\tilde{y}_0(k)$ هو تقدير غير متحيز لقيمة تابع الانحدار $y_0(k)$، وذلك لأنه من أجل أية مركبة K نجد أن:

$$E\big(\tilde{y}_0(k)\big) = E\left(x_0'\tilde{\beta}(k)\right) = x_0'E\big[\tilde{\beta}(k)\big] = x_0'\beta(k) = y_0(k) \qquad (169-2)$$

ولحساب مصفوفة التباين المشترك لأخطاء هذه التقديرات نأخذ فروقاتها عن توقعاتها ، والتي تساوي عند كل تابع $Y(k)$ ما يلي:$\tilde{y}_0(k)$

$$\varepsilon(k) = y(k) - \tilde{y}(k) = x_0'\beta(k) - x_0'\tilde{\beta}(k) = x_0'\big[\beta(k) - \tilde{\beta}(k)\big] \qquad (170-2)$$

وإن عناصر مصفوفة التباينات المشتركة لأخطاء التقدير تحسب من العلاقة :

$$E(\varepsilon(k) * \varepsilon'(h)) = E\left[\mathrm{x}_0' \left(\beta(k) - \tilde{\beta}(k)\right) * \left(\beta(h) - \tilde{\beta}(h)\right)' x_0\right] =$$

$$= x_0'E\left[\left(\beta(k) - \tilde{\beta}(k)\right)\left(\beta(h) - \tilde{\beta}(h)\right)'\right] x_0$$

$$= x_0' CoV\left(\tilde{\beta}(k), \tilde{\beta}(h)\right) x_0 = x_0'[\sigma_{kh}(X'X)^{-1}]x_0$$

$$E(\varepsilon(k) * \varepsilon'(h)) = \sigma_{kh} * x_0'(X'X)^{-1}x_0 \qquad (171-2)$$

حيث أن σ_{kh} هو عنصر المصفوفة V . وهو يقدر من العنصر المقابل له في المصفوفة $\tilde{V}$ المقدرة من العلاقة التالية :

$$\tilde{V}(\tilde{\varepsilon}) = \frac{\tilde{\varepsilon}' * \tilde{\varepsilon}}{n-p-1} \qquad \mathbf{(172-2)}$$

ويستفاد من هذه العلاقات في الاستدلال حول قطوع الثقة للقيم المتوقعة للتوابع $Y_1\ Y_2 \dots Y_r$، وذلك باستخدام التوزيعات الاحتمالية كما يلي :

إن القيمة المقدرة لـ Y في النقطة $x_0' = [1, x_{01}, x_{02}, \dots x_{0P}]$ تساوي حسب (2-155) :

$$\tilde{y}_0 = x_0'\tilde{\beta} \qquad (173-2)$$

ولكن القيمة المتوقعة والمجهولة لـ Y في تلك النقطة تساوي: $E(\ y_0) = x_0'\beta$. وبناء على الافتراضات الابتدائية يمكننا أن نبرهن على أن القيمة المقدرة $\tilde{y}_0 = x_0'\tilde{\beta}$، تخضع للتوزيع الطبيعي المتعدد الذي توقعه$E(\ y_0) = x_0'\beta$، ومصفوفة تبايناته المشتركة تساوي $[x_0'(X'X)^{-1}x_0 * V]$، علماً بأن المقدار $[n\,\tilde{V}]$ يخضع لتوزيع (ويشارت) التالي: $W_{n-p-1}(.\,|V)$، وبناء على ذلك وقياساً على المعالجات التي أجريناها في الفصل الخامس من الجزء الثاني حول المؤشر T^2 (مؤشر Hotelling)، يمكننا تعريف المؤشر T^2 الذي يعبر عن المسافة الإحصائية من القيمة المقدرة $x_0'\tilde{\beta}$ إلى القيمة المتوقعة $x_0'\beta$ بواسطة العلاقة التالية :

$$T^2 = \left(\frac{x_0'\tilde{\beta} - x_0'\beta}{\sqrt{x_0'(X'X)^{-1}x_0}}\right)' \left(\frac{n\,\tilde{V}}{n-p-1}\right)^{-1} \left(\frac{x_0'\tilde{\beta} - x_0'\beta}{\sqrt{x_0'(X'X)^{-1}x_0}}\right) \qquad (174-2)$$

وبذلك نجد أن قطع الثقة ذي الاحتمال $(1-\propto)$ للقيمة المتوقعة المجهولة$E(\ y_0) = x_0'\beta$. يعطى بواسطة متراجحة من الشكل التالي :

$$T^2 \le C^2(\propto) \qquad (175-2)$$

حيث $C^2(\propto)$ عدد موجب متعلق بـ $\propto$.

ولكن بما أن T^2 يخضع لتوزيع Hotelling، الذي يرتبط – في هذه الحالة– مع التوزيع F بالعلاقة التالية:

$$T^2 \sim \frac{r(n-p-1)}{n-p-r} F_{r\,,n-p-r} \qquad (176-2)$$

حيث $v_1 = r$ و $v_2 = n-p-r$.

لذلك فإننا نقارن قيمة T^2 مع القيمة الحرجة لها $T^2(\propto)$ والتي تحسب من العلاقة (2–176) ، ثم نشكل معادلة قطع الثقة ذي الاحتمال $(1-\propto)$ من خلال المتراجحة التالية :

$$T^2 \le \frac{r(n-p-1)}{n-p-r} F_{r\,,n-P-r}(\propto) = C^2(\propto) \qquad (177-2)$$

وبتبديل قيمة T^2 من (2–174) في (2–177) وإجراء بعض الإصلاحات نحصل على معادلة القطع المطلوب التالية :

$$\left(x_0'\beta - x_0'\tilde{\beta}\right)' \left(\frac{n\,\tilde{V}}{n-p-1}\right)^{-1} * \left(x_0'\beta - x_0'\tilde{\beta}\right) \le x_0'(X'X)^{-1}x_0 \left[\frac{r(n-p-1)}{n-p} * F_{r\,,n-p-r}(\propto)\right] \qquad \mathbf{(178-2)}$$

ولإنشاء مجالات الثقة المتزامنة ذات الاحتمال $(1-\propto)$ للقيمة المتوقعة للتابع $y_0(k)$، وهي $E(y_0(k\,)) = x_0'\beta(k)$. نطبق العلاقة التالية:

$$E\left(y_0(k\,)\right) = $$

$$= x_0'\tilde{\beta}(k) \pm \sqrt{\frac{r(n-p-1)}{n-p-r} * F_{r\,,n-P-r}(\propto)} * \sqrt{x_0'(X'X)^{-1}x_0\left(\frac{n}{n-p-1}\tilde{\sigma}_{kk}\right)} \qquad \mathbf{(179-2)}$$

حيث أن $k = 1\ 2\ 3 \ldots r$ ، وأن $\tilde{\beta}(k)$ هو العمود k من المصفوفة $\tilde{\beta}$، وأن $\tilde{\sigma}_{kk}$ هو العنصر القطري في المصفوفة $\tilde{V}$. وهكذا نحصل على المجالات المتزامنة المقابلة لجميع التوابع $y_0(k)$.

2-8-3 : التنبؤ بشعاع القيم الفعلية للتابع المضاعف لـ $Y(k)$ المقابل لنقطة x_0 :

لنأخذ نقطة من فضاء العينة مثل السطر $x_0' = [1, x_{01}, x_{02}, \ldots x_{0P}]$، ولنرمز للشعاع المطلوب في النقطة x_0 بالرمز :

$$Y_0 = [Y_{01} \vdots Y_{02} \vdots Y_{03}, \ldots \vdots Y_{0r}] \qquad (180-2)$$

وعندها نجد أن معادلة الانحدار العام تعطينا أن معادلة التابع $Y(k)$ في النقطة X_0 تعطى بالعلاقة :

$$y_0(k) = x_0' * \beta(k) + \varepsilon_0(k) \ \Rightarrow\ \varepsilon_0(k) = y_0(k) - x_0'\beta(k) \qquad (181-2)$$

حيث أن ε_0 هو شعاع الخطأ الجديد ونرمز له بـ :

$$\varepsilon_0 = [\varepsilon_0(1), \varepsilon_0(2), \varepsilon_0(3), \ldots, \varepsilon_0(r)]' \qquad (182-2)$$

علماً بإن الشعاع ε_0 الجديد مستقل عن الأخطاء ε ويحقق العلاقتين التاليتين :

$$E\big(\varepsilon_0(k)\big) = 0 \qquad COV\big(\varepsilon_0(k), \varepsilon_0(h)\big) = \sigma_{kh} \qquad (183-2)$$

وبذلك نجد أن خطأ التنبؤ من أجل المركبة k من Y_0 يساوي ما يلي (نضيف ونطرح $x_0'\beta(k)$) :

$$\begin{aligned} y_0(k) - x_0'\tilde{\beta}(k) &= y_0(k) - x_0'\tilde{\beta}(k) + x_0'\beta(k) - x_0'\beta(k) \\ &= \varepsilon_0(k) - x_0' * \big(\tilde{\beta}(k) - \beta(k)\big) \end{aligned} \qquad (184-2)$$

وهو يحقق الخاصة التالية :

$$E\big[y_0(k) - x_0'\tilde{\beta}(k)\big] = E[\varepsilon_0(k)] - x_0'E\big[\tilde{\beta}(k) - \beta(k)\big] = 0 - 0 = 0 \qquad (185-2)$$

وهذا يدل على أن المقدار $x_0'\tilde{\beta}(k)$ هو تقدير غير متحيز للقيمة الفعلية $Y_0(k)$. ولإيجاد عناصر مصفوفة التباينات المشتركة لأخطاء التنبؤ نجد أنها تحسب بناء على العلاقة (2-184) كما يلي :

$$\begin{aligned} &E\left[\left(y_0(k) - x_0'\tilde{\beta}(k)\right)\left(Y_0(h) - x_0'\tilde{\beta}(h)\right)'\right] = \\ &= E\left\{\left[\varepsilon_0(k) - x_0'\left(\tilde{\beta}(k) - \beta(k)\right)\right]\left[\varepsilon_0(h) - x_0'\left(\tilde{\beta}(h) - \beta(h)\right)\right]'\right\} \\ &= E\big(\varepsilon_0(k) * \varepsilon_0'(h)\big) + x_0'\mathrm{E}\left[\left(\tilde{\beta}(k) - \beta(k)\right)\left(\tilde{\beta}(h) - \beta(h)\right)' x_0\right] - \\ &\quad - x_0'E\left[\left(\tilde{\beta}(k) - \beta(k)\right)\varepsilon_0'(h)\right] - E\left[\varepsilon_0(k)\left(\tilde{\beta}(h) - \beta(h)\right)' x_0\right] \\ &= \sigma_{kh} + x_0'(\sigma_{kh}(X'X)^{-1})x_0 + 0 + 0 \\ &= \sigma_{kh}(1 + x_0'(X'X)^{-1}x_0) \end{aligned} \qquad (\mathbf{186-2})$$

ذلك لأن الحدين الأخيرين معدومين أي أن:

$$E\big[(\tilde{\beta}(k) - \beta(k)\varepsilon_0'(h)\big] = 0 \ , \quad E\left[\varepsilon_0(k)\left(\tilde{\beta}(h) - \beta(h)\right)\right] = 0 \qquad (\mathbf{187-2})$$

وذلك لأن $\tilde{\beta}(k)$ مستقلة عن ε_0 حسب الخاصة 8 **والعلاقة** (2-168) **ولأنها حسب** (2-147) **تساوي :**

$$\tilde{\beta}(k) = (X'X)^{-1}X'\,Y(k) = (X'X)^{-1}X'[X * \beta(x) + \varepsilon(k)] = \beta(k) + (X'X)^{-1}X' * \varepsilon(k) \quad (\mathbf{188-2})$$

وكذلك الأمر بالنسبة للحد الثاني حيث نجد أيضاً أن:

$$\tilde{\beta}(h) = (X'X)^{-1}X'\,Y(h) = (X'X)^{-1}X'[X * \beta(h) + \varepsilon(h)] = \beta(h) + (X'X)^{-1}X' * \varepsilon(h) \quad (\mathbf{189-2})$$

وبناء على ذلك نجد أن النموذج العام يعطينا بعد تجميع مركباته أن قيمة Y_0 الفعلية والمجهولة تساوي :

$$y_0 = x_0'\beta + \varepsilon \qquad (190-2)$$

ومن جهة أخرى نجد أن $\tilde{y}_0$ المتنبأ بها من النموذج (1−150) تساوي :

$$\tilde{y}_0 = x_0'\tilde{\beta} \qquad (191-2)$$

وبذلك نجد أن أخطاء التنبؤ تساوي :

$$\begin{aligned} y_0 - \tilde{y}_0 &= y_0 - x_0'\tilde{\beta} = x_0'\beta + \varepsilon_0 - x_0'\tilde{\beta} \\ y_0 - \tilde{y}_0 &= x_0'(\beta - \tilde{\beta}) + \varepsilon_0 \end{aligned} \qquad (192-2)$$

وإن هذه الأخطاء خاضعة للتوزيع الطبيعي المتعدد $N_r[0\,,(1 + x_0'(X'X)^{-1}x_0) * V]$ وهي مستقلة عن التقدير $(n\,\tilde{V})$.

كما نجد أن قطع الثقة التنبؤي ذي الاحتمال $(1 - \propto)$، لقيمة y_0 الفعلية والمجهولة (حسب المعالجة السابقة) يأخذ شكل المتراجحة التالية :

$$\left(y_0 - x_0'\tilde{\beta}\right)'\left(\frac{n\,\tilde{V}}{n-p-1}\right)^{-1}\left(y_0 - x_0'\tilde{\beta}\right) \le (1 + x_0'(X'X)^{-1}x_0)\left[\frac{r(n-p-1)}{n-p-r}F_{r\,,n-P-r}(\propto)\right] \qquad \mathbf{(193-2)}$$

كما أن مجالات الثقة التنبؤية المتزامنة ذات الاحتمال$(1 - \propto)$ للقيم الحقيقية للتوابع المنفردة $y_0(k)$ تأخذ الشكل التالي:

$$y_0(k) = \tilde{y}_0(k) \pm \sqrt{\frac{r(n-P-1)}{n-P-r} * F_{r\,,n-P-r}(\propto)} * \sqrt{(1 + x_0'(X'X)^{-1}x_0)\left(\frac{n}{n-P-1}\tilde{\sigma}_{kk}\right)} \qquad \mathbf{(194-2)}$$

حيث أن $k\ :\ 1\ 2\ 3 \ldots r$ وأن: $\tilde{y}_0\ \ (k) = x_0'\tilde{\beta}(k)$.

حيث أن$\tilde{\beta}(k)$ و $\tilde{\sigma}_{kk}$ و$(\propto)$ $F_{r\,,n-P-r}$هي نفس الرموز المستخدمة في العلاقتين (2−178) و(2−179) وهنا نلاحظ أن المجالات التنبؤية للقيم الفعلية للتوابع $Y(k)$ هي أوسع من المجالات المقابلة لها للقيمة المتوقعة ، علماً بأن التوسع الزائد يؤثر على سلوك الخطأ العشوائي $\varepsilon_0(k)$ ويزيد من قيمته .

مثال (2−6): [مأخوذ من Johnson- Wichern P.293+313 **بتصرف وإضافة**]

في دراسة لمتطلبات الشركات من أدوات الحاسوب اللازمة لتطوير إدارة المخزون في الشركات المتشابهة، قام الخبراء بتحديد متحولين مستقلين X_1 و X_2 كما يلي:

- X_1 – عدد طلبات الزبائن (بالآلاف) .
- X_2 – عدد الطلبات المحذوفة (بالآلاف) .

كما حددوا متحولين تابعين Y_1 و Y_2 كما يلي:

- Y_1– مدة تشغيل وحدة المعالجة المركزية Cpu (بالساعات) .
- Y_2– السعة التخزينية لقرص المدخلات والمخرجات (كيلوبايت) .

ثم قاموا بجمع البيانات اللازمة عن هذه المتحولات من (7) شركات متشابهة فكانت كما في الجدول التالي :

جدول (2-6) بيانات المثال:

Y_2- السعة التخزينية (كيلوبايت)	Y_1- مدة التشغيل (ساعة)	X_2 – الطلبات المحذوفة (ألف)	X_1 – الطلبات (ألف)	رقم الشركة
301.8	141.5	2.108	123.8	1
396.1	168.9	9.213	146.1	2
328.2	154.8	1.905	133.9	3
307.4	146.5	0.815	128.5	4
362.4	172.8	1.061	151.5	5
369.5	160.1	8.603	136.2	6
229.1	108.5	1.125	92.0	7
?	?	7.5	130	نقطة التنبؤ x_0

المصدر: [Johnson , Wichen P.293]

والمطلوب إيجاد معادلات الانحدار المتعدد المضاعف للتابعين Y_1 و Y_2 بدلالة المتحولين المستقلين X_1 و X_2، ثم القيام بتقدير هذين التابعين عند النقطة $X'_0 = [1\,,130\,,7.5]$ ثم إيجاد مجالات الثقة المتزامنة لهما . ثم إيجاد معادلة الناقص الذي يشملها باحتمال 0.95 .

الحل: نقوم أولاً بتحديد وحساب المصفوفات اللازمة وهي:

$$X = \begin{array}{c} \quad\quad X_1 \quad\quad X_2 \\ \begin{bmatrix} 1 & 123.4 & 2.108 \\ 1 & 146.1 & 9.213 \\ 1 & 133.9 & 1.905 \\ 1 & 128.5 & 0.813 \\ 1 & 151.5 & 1.061 \\ 1 & 136.2 & 8.603 \\ 1 & 0.92 & 1.125 \end{bmatrix} \end{array} ,Y = \begin{array}{c} Y(1) \quad Y(2) \\ \begin{bmatrix} 141.5 & 301.8 \\ 168.9 & 396.1 \\ 154.8 & 328.2 \\ 146.5 & 307.4 \\ 172.5 & 362.4 \\ 160.1 & 369.5 \\ 108.5 & 229.1 \end{bmatrix} \end{array} = [Y(1)Y(2)]$$

$$(X'X)^{-1} = \begin{bmatrix} 8.17969 & & \\ -0.06411 & 0.00052 & \\ 0.08831 & -0.00107 & 0.0140 \end{bmatrix}$$

ثم نفترض أن معادلتي انحدار Y_1 و Y_2 على X_1 و X_2 كما يلي:

$$Y(1) = \beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \varepsilon_1$$
$$Y(2) = \beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \varepsilon_2$$

ونرمز لها مصفوفياً كما يلي:

$$Y = X * \beta + \varepsilon$$

حيث أن:

$$\beta = \begin{array}{c} \beta(1) \quad \beta(2) \\ \begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \end{array}$$

وباستخدام الحاسوب نقوم بتقدير عناصر المصفوفة β من العلاقة (2−149) فنجد أن:

$$\tilde{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 8.42 & 14.14 \\ 1.080 & 2.25 \\ 0.42 & 5.67 \end{bmatrix}$$

وبذلك نحصل على معادلتي الانحدار المتعدد المضاعف التاليتين :

$$\tilde{Y}\ (1) = 8.42 + 1.080X_1 + 0.42X_2$$
$$Y\ (2) = 14.14 + 2.25X_1 + 5.67X_2$$

كما نجد أن قيمة الانحراف المعياري للأخطاء في كل منهما تساوي:

$$S_2 = 1.812 \qquad\qquad S_1 = 1.204$$

ولإيجاد القيمة التقديرية لـ $y_0(1)$ و $y_0(2)$ عند النقطة السطرية $x_0' = [1\,, 130\,, 7.5]$ ، نعوض ذلك في المعادلتين السابقتين، فنجد أن :

$$\tilde{y}_0(1) = 8.42 + 1.080(130) + 0.42(7.5) = 151.97$$
$$\tilde{y}_0(2) = 14.14 + 2.25(130)_1 + 5.67(7.5) = 349.17$$

وكان يمكن حسابها من العلاقة المصفوفية التالية :

$$\tilde{Y}_0 = x_0' * \beta = [1\,, 130\,, 7.5] * \begin{bmatrix} 8.42 & 14.14 \\ 1.080 & 2.25 \\ 0.42 & 5.67 \end{bmatrix} = \begin{bmatrix} 151.97 \\ 349.17 \end{bmatrix}$$

ولإنشاء قطع ثقة لكل من هاتين القيمتين نطبق العلاقة (2−193)، لذلك نقوم أولاً بحساب الجداء $x_0'(X'X)^{-1}x_0$ ثم حساب المصفوفة $n\tilde{V}$ فنجد أن:

$$x_0'(X'X)^{-1}x_0 = [1\,, 130\,, 7.5] * \begin{bmatrix} 8.17969 & -0.06411 & 0.08831 \\ -0.06411 & 0.00052 & -0.00107 \\ 0.08831 & -0.00107 & 0.01440 \end{bmatrix} \begin{bmatrix} 1 \\ 130 \\ 7.5 \end{bmatrix} = 0.34725$$

$$n\tilde{V} = \begin{bmatrix} \left(Y(1) - X\tilde{\beta}(1)\right)'\left(Y(1) - X\tilde{\beta}(1)\right), \left(Y(1) - X\tilde{\beta}(1)\right)'\left(Y(2) - X\tilde{\beta}(2)\right) \\ \left(Y(2) - X\tilde{\beta}(2)\right)'\left(Y(1) - X\tilde{\beta}(1)\right), \left(Y(2) - X\tilde{\beta}(2)\right)'\left(Y(2) - X\tilde{\beta}(2)\right) \end{bmatrix}$$

وبعد إجراء الحسابات اللازمة على الحاسوب نجد أن:

$$n\tilde{V} = \begin{bmatrix} 5.80 & 5.30 \\ 5.30 & 13.13 \end{bmatrix} \implies \tilde{V} = \begin{bmatrix} \frac{5.80}{7} & \frac{5.30}{7} \\ \frac{5.30}{7} & \frac{13.13}{7} \end{bmatrix}$$

ثم نقوم بتشكيل معادلة القطع الناقص للثقة للقيمة المتوقعة $E\left[Y_0\left(Y(1), Y(2)\right)\right]$ وذي الاحتمال 0.95 من العلاقة (2−188) كما يلي:

$$[(y(1) - 151.97), (y(2) - 349.17) * \frac{1}{4}\begin{bmatrix} 5.80 & 5.30 \\ 5.30 & 13.13 \end{bmatrix}^{-1} * \begin{bmatrix} Y(1) - 151.97 \\ Y(2) - 349.17 \end{bmatrix} \leq (0.34725)\left[\frac{2(4)}{3} * F_{2\,3}(0.05)\right]$$

وبعد تعويض $F_{2\,3}(0.05) = 9.55$ وإجراء بعض الإصلاحات حصلوا على قطع ناقص بدلالة $Y(1)$ و $Y(2)$ ، مركزه (151.97 , 349,17)، ويمكن تحديد اتجاه وطول كل من محوريه (الكبير والصغير) بواسطة القيم الذاتية والأشعة الذاتية للمصفوفة $(n\tilde{V})$ كما فعلنا سابقاً في الفصل الرابع .

ولإنشاء مجالات الثقة المتزامنة للقيمة المتوقعة لـ $y_0(1)$ نطبق العلاقة (2−179)، ولذلك نقوم أولاً بحساب ما يلي: [انظر المصفوفة $\tilde{V}$]:

$$\sqrt{x_0'(X'X)^{-1}x_0\left(\frac{n}{n-P-1}\tilde{\sigma}_{11}\right)} = \sqrt{0.34725\,,\frac{7}{4}\left(\frac{5.80}{7}\right)} = 0.71$$

ثم نقوم بإيجاد المقدار: $\frac{r(n-P-1)}{n-P-r} * F_{r\,,n-P-n}(\propto) = \frac{2(4)}{3} * F_{2\,3}(0.05) = 25.466$

وهكذا نجد أن مجال الثقة ذي الاحتمال 0.95 لـ $E\big(Y_0(1)\big)$ يساوي :

$$E\big(y_0(1)\big) = \tilde{Y}_0(1) \pm \sqrt{\frac{r(n-p-1)}{n-p-n} * F_{r\,,n-p-r}(\propto)}\sqrt{x_0'(X'\mathrm{X})^{-1}x_0\left(\frac{n}{n-p-1}\tilde{\sigma}_{kk}\right)} =$$

$$E\big(y_0(1)\big) = 151.97 \pm \sqrt{25.466}\sqrt{0.71} = 151.97 \pm 4.25$$

وهذا يعطينا المجال التالي: $146.75 \leq E\big[y_0(1)\big] \leq 156.22$

وهكذا نحسب مجالات الثقة المتزامنة لـ $E[y_0(2)]$، ونترك ذلك للقارئ على سبيل التمرين .

ولإيجاد معادلة القطع الناقص التنبؤي للقيمة الفعلية لـ $Y_0\big(y_0(1)\,,y_0(2)\big)$ نطبق العلاقة (2−178)، لذلك نحسب أولاً قيمة المقدار :

$$1 + x_0'(X'X)^{-1}x_0 = 1 + 0.34725 = 1.34725$$

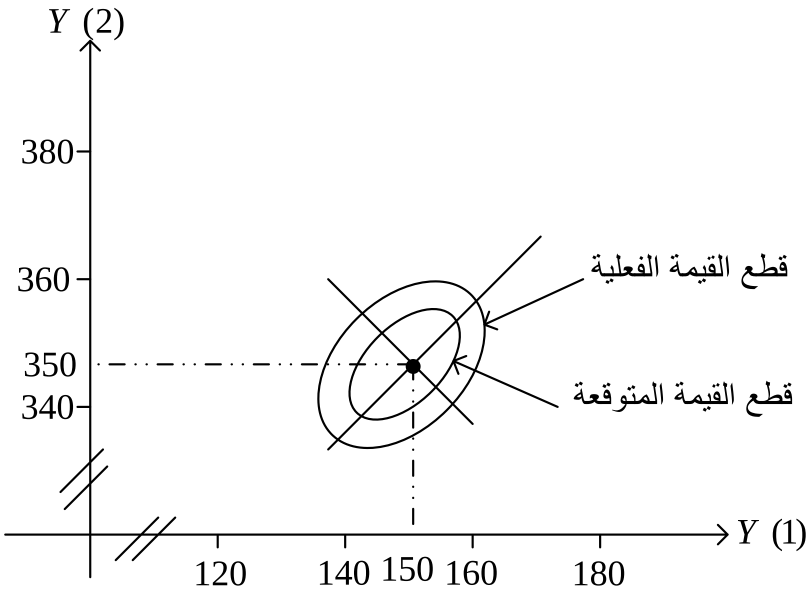
ثم نعوض ذلك في العلاقة (2−178) فنجد أن:

$$[(y_0(1) - 151.97), (y_0(2) - 349.17)]\frac{1}{4}\begin{bmatrix}5.80 & 5.30\\ 5.30 & 13.13\end{bmatrix}^{-1} * \begin{bmatrix}Y_0(1) - 151.97\\ Y_0(2) - 349.17\end{bmatrix}$$
$$\leq (1.34725)\left[\frac{2(4)}{3} * F_{2\,3}(0.05)\right]$$

وبتعويض $F_{2\,3}(0.05) = 9.55$ وإجراء بعض الإصلاحات نحصل على قطع ناقص جديد بدلالة $y_0(1)$ و$y_0(2)$ ، وهو يتمركز أيضاً في النقطة $(151.97\,, 349{,}17)$ ويكون لمحاوره نفس اتجاه محاور قطع الثقة السابق، ولكن أطوال محوري تكون أكبر من أطوال محوري قطع الثقة السابق . وذلك لأننا استبدلنا المقدار $x_0'(X'X)^{-1}x_0 = 0.34725$ بالمقدار:

$$1 + x_0'(X'X)^{-1}x_0 = 1 + 0.34725 = 1.34725$$

وبذلك يكون القطع الثاني أكبر أو أوسع من القطع الأول. وهما يرسمان على المحورين $OY(1)$ و $OY(2)$ الشكل التالي:



الشكل (2-3)

ولإيجاد مجالات الثقة المتزامنة للقيمة الفعلية للتابع الأول $Y(1)$ نطبق العلاقة (2-194) . لذلك نحسب أولاً المقدارين :

$$\sqrt{(1+x_0'(X'X)^{-1}x_0)\left(\frac{n}{n-p-1}\sigma_{11}\right)} = \sqrt{1.34725 * \frac{7}{4} * \frac{5.80}{7}} = 1.40$$

$$\sqrt{\frac{r(n-P-1)}{n-P-r} * F_{2\ 3}(\propto)} = \sqrt{\frac{2(4)}{3}9.55} = \sqrt{25.466} = 5.05$$

وهكذا نجد أن مجال الثقة للقيمة الفعلية لـ $y_0(1)$ يساوي :

$$y_0(1) = \tilde{y}_0(1) \pm \sqrt{\frac{r(n-p-1)}{n-p-r} * F_{P,n-P-r}(\propto)}\sqrt{(1+x_0'(X'X)^{-1}x_0)\left(\frac{n}{n-p-1}\sigma_{11}\right)} = 151.97 \pm (5.05)(1.40)$$

وهو يعطينا المجال التالي لـ $Y_0(1)$ الفعلية :

$$144.9 \leq y_0(1) \leq 159.04$$

ولإيجاد مجال الثقة المتزامن للقيمة الفعلية $y_0(2)$ نطبق نفس العلاقة (2-194) لذلك نحسب المقدار :

$$\sqrt{1+X_0'(X'X)^{-1}X_0\left(\frac{n}{n-p-1}\sigma_{22}\right)} = \sqrt{(1.34725)\left(\frac{7}{4} * \frac{13.13}{7}\right)} = 2.10$$

ولقد وجدنا سابقاً أن:

$$\sqrt{\frac{r(n-p-1)}{n-p-r} * F_2(0.05)} = 5.05$$

نعوض ذلك في العلاقة (2–194) فنجد أن:

$$y_0(2) = \tilde{y}_0(2) \pm \sqrt{\frac{r(n-p-1)}{n-p-r} * F_2(0.05)}\sqrt{(1 + x_0'(X'X)^{-1}x_0)\left(\frac{n}{n-p-1}\sigma_{22}\right)} =$$
$$= 349.17 \pm (5.05)(2.10) = 349.17 \pm 10.62$$

وهو يعطينا المجال التالي لقيمة $y_0(2)$ الفعلية :

$$338.55 \leq y_0(2) \leq 362.79$$

الفصل الثالث

الارتباط القانوني Canonical Correlation

3-1 تمهيد:

إن دراسة العلاقات بين المتحولات الكمية تعد من أولى مهام علم الاحصاء، ومنه ظهرت نظرية الارتباط والانحدار ثم تطورت وتفرعت إلى عدة فروع واختصاصات هي:

1- الارتباط البسيط (الانحدار البسيط): وهو يختص بدراسة العلاقة بين متحولين كميين فقط X , Y . وحساب معامل الارتباط بينهما ثم استخراج النموذج الرياضي الذي يمثل تلك العلاقة أو يعبر عنها أفضل تعبير .

2- الارتباط المتعدد (الانحدار المتعدد): وهو يختص بدراسة العلاقة بين متحول تابع Y وعدة متحولات مؤثرة عليه نرمز لها بالرموز: $X_1, X_2, X_3, \dots\dots, X_p$.

3- الارتباط القانوني: وهو يختص بدراسة العلاقة بين عدة متحولات تابعة نرمز لها بـ $Y_1, Y_2, Y_3, \dots\dots, Y_q$ وعدة أخرى من المتحولات مؤثرة فيها نرمز لها بالرموز: $X_1, X_2, X_3, \dots\dots, X_p$.

4- وهناك فروع أخرى تتناول هذا الموضوع مثل: التحليل العاملي والتحليل التمييزي.

وإننا سنعتبر أن القارئ قد اطلع على الفرعين الأول والثاني (الانحدار البسيط والمتعدد)، وإنه على دراية كافيةٍ بجبر المصفوفات وجداءاتها وجذورها الكامنة، وغير ذلك من الأمور الرياضية التي سنعتمد عليها عند معالجة مسائل الارتباط القانوني (انظر الجزء الأول حول المصفوفات) . وسنستعرض المفاهيم العامة للارتباط القانوني كما يلي :

3-2 المفاهيم العامة للارتباط القانوني:

لنفترض أننا نريد دراسة العلاقة الارتباطية بين مجموعتين من المتحولات الكمية هما:

1- مجموعة المتحولات المستقلة : وتتألف من p متحولاً مستقلاً (مؤثراً) ونرمز لها بالرموز $X_1, X_2, X_3, \dots, X_p$

2- مجموعة المتحولات التابعة: وتتألف من q متحولاً تابعاً (متأثراً) ونرمز لها بالرموز $Y_1, Y_2, Y_3, \dots, Y_q$

كما نفترض أننا جمعنا عن كل من هذه المتحولات المستقلة والتابعة البيانات اللازمة أو المشاهدات الميدانية وحصلنا على n مشاهدة متقابلة عن كل منها. أي حصلنا عن كل من المتحولات المستقلة X والتابعة Y على n من القيم العددية المتقابلة من حيث الزمان أو المكان أو كلاهما، ووضعنا النتائج في جدول مناسب كما يلي :

الجدول (3-1): البيانات الخاصة والمشاهدات الميدانية للمتحولات المستقلة X والتابعة Y

المتحولات	قيم المشاهدات للمتحولات المستقلة X					قيم المشاهدات للمتحولات التابعة Y				
رقم المشاهدة	X_1	X_2	X_3	....	X_p	Y_1	Y_2	Y_3	....	Y_q
1	x_{11}	x_{12}	x_{13}	....	x_{1p}	y_{11}	y_{12}	y_{13}	....	y_{1q}
2	x_{21}	x_{22}	x_{23}	....	x_{2p}	y_{21}	y_{22}	y_{23}	....	y_{2q}
3	x_{31}	x_{32}	x_{33}	....	x_{3p}	y_{31}	y_{32}	y_{33}	....	y_{3q}
4	x_{41}	x_{42}	x_{43}	....	x_{4p}	y_{41}	y_{42}	y_{43}	....	y_{4q}
5	x_{51}	x_{52}	x_{53}	....	x_{5p}	y_{51}	y_{52}	y_{53}	....	y_{5q}
6	x_{61}	x_{62}	x_{63}	....	x_{6p}	y_{61}	y_{62}	y_{63}	....	y_{6q}
⋮	⋮	⋮	⋮	....	⋮	⋮	⋮	⋮	....	⋮
n	x_{n1}	x_{n2}	x_{n3}	....	x_{np}	y_{n1}	y_{n2}	y_{n3}	....	y_{nq}
المتوسط أو التوقع	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	....	$\bar{X}_p$	$\bar{Y}_1$	$\bar{Y}_2$	$\bar{Y}_3$	....	$\bar{Y}_q$
التباين	s_1^2	s_2^2	s_3^2	....	s_p^2	σ_1^2	σ_2^2	σ_3^2	....	σ_q^2

وبصورة عامة يمكننا أن نضع جملة المتحولات المستقلة على شكل شعاع عمود ونرمز له بـ X وأن نضع جملة المتحولات التابعة على شكل شعاع عمود آخر ونرمز له بـ Y كما يلي :

$$X_{p*1} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{bmatrix} \qquad Y_{q*1} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_q \end{bmatrix}$$

وسيكون هدفنا دراسة العلاقة الارتباطية بين متحولات المجموعة Y مع متحولات المجموعة X

وسنرمز لمتوسطات X ولمتوسطات Y بالعمودين التاليين:

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \\ \vdots \\ \bar{X}_p \end{bmatrix} \qquad \bar{Y} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \\ \vdots \\ \bar{Y}_q \end{bmatrix}$$

كما سنرمز لتباينات X ولتباينات Y بالعمودين التاليين :

$$S^2 = \begin{bmatrix} s_1^2 \\ s_2^2 \\ s_3^2 \\ \vdots \\ s_p^2 \end{bmatrix} \qquad \sigma^2 = \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \\ \vdots \\ \sigma_q^2 \end{bmatrix}$$

علماً بأن هذه التباينات تحسب من العلاقة المصححة كما يلي :

$$S_i^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ki} - \bar{X}_i)^2 \quad , \quad \sigma_j^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(y_{kj} - \bar{Y}_j\right)^2 \quad (1-3)$$

*** مصفوفة التباينات والتباينات المشتركة :**

نقوم بحساب مصفوفة التباينات والتباينات المشتركة $cov(X_i , X_j)$ لكل المتحولات في المجموعة X .
وبحساب مصفوفة التباينات والتباينات المشتركة $cov(Y_k , Y_h)$ لكل المتحولات في المجموعة Y .
وبحساب مصفوفة التباينات والتباينات المتبادلة $cov(X_i , Y_k)$ لكل المتحولات المتقابلة في المجموعتين X , Y ونضعها في جدول منظم كما يلي :

الجدول (3-2): مصفوفات التباينات والتباينات المشتركة لـــ X , Y

المتحولات	X_1	X_2	X_3	....	X_p	Y_1	Y_2	Y_3	....	Y_q
X_1	s_{11}	s_{12}	s_{13}	....	s_{1p}	c_{11}	c_{12}	c_{13}	....	c_{1q}
X_2	s_{21}	s_{22}	s_{23}	....	s_{2p}	c_{21}	c_{22}	c_{23}	....	c_{2q}
X_3	s_{31}	s_{32}	s_{33}	....	s_{3p}	c_{31}	c_{32}	c_{33}	....	c_{3q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$
X_p	s_{p1}	s_{p2}	s_{p3}	....	s_{pp}	σ_{p1}	σ_{p2}	σ_{p3}	....	c_{pq}
Y_1	c_{11}	c_{12}	c_{13}	....	c_{1p}	σ_{11}	σ_{12}	σ_{13}	....	σ_{1q}
Y_2	c_{21}	c_{22}	c_{23}	....	c_{2p}	σ_{21}	σ_{22}	σ_{23}	....	σ_{2q}
Y_3	c_{31}	c_{32}	c_{33}	....	c_{3p}	σ_{31}	σ_{32}	σ_{33}	....	σ_{3q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$
Y_q	c_{q1}	c_{q2}	c_{q3}	....	c_{qp}	σ_{q1}	σ_{q2}	σ_{q3}	....	σ_{qq}

وهو يتضمن أربع مصفوفات مختلفة. علماً بأن التباينات المشتركة cov لمتحولات المجموعة X ولمتحولات المجموعة Y تحسب من العلاقتين التاليتين:

$$S_{ij} = cov(X_i , X_j) = \frac{1}{n-1} \sum_{k=1}^{n}(x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j) \qquad (2-3)$$

$$\sigma_{ij} = cov(Y_i , Y_j) = \frac{1}{n-1} \sum_{k=1}^{n}(y_{ki} - \bar{Y}_i)(y_{kj} - \bar{Y}_j) \qquad (3-3)$$

ومما سبق نلاحظ أن: $S_{ij} = S_{ji}$ أي أن المصفوفة S_{ij} متناظرة، وأن: $\sigma_{ij} = \sigma_{ji}$ أي أن المصفوفة σ_{ij} متناظرة .

وكذلك نلاحظ أن: $S_{ii} = S_i^2$ وأن: $\sigma_{ii} = \sigma_i^2$ ، أي أن العناصر القطرية هي التباينات S_i^2 و σ_i^2 ، للمتحولات X وللمتحولات Y ، أما التباينات المشتركة المتبادلة لـ X , Y فتحسب من العلاقة :

$$C_{ij} = cov(X_i , Y_j) = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{X}_i)(y_{kj} - \bar{Y}_j) \qquad (4-3)$$

كما نلاحظ أن مصفوفة التباينات المشتركة المتبادلة متناظرة وتحقق العلاقة : $C_{ij} = C_{ji}$ ، أي أن المصفوفة C_{ji} تساوي منقول المصفوفة C_{ij}، أي أن $C_{ji} = C'_{ij}$.

وبذلك يمكننا كتابة الجدول (2) السابق على شكل مصفوفات برموز مختلفة مع مراتبها المناسبة كما يلي:

الجدول (3-3) : مصفوفات التباينات والتباينات المشتركة لــ X , Y :

X_1 X_2 X_3 $\vdots$ X_p	$C_{xx}(p*p)$	$C_{xy}(p*q)$
Y_1 Y_2 Y_3 $\vdots$ Y_q	$C_{yx}(q*p)$	$C_{yy}(q*q)$

وهنا نلاحظ أن: $C_{yx}(q*p) = C'_{xy}(p*q)$

* **المصفوفات الارتباطية للمجموعتين** X و Y :

إن المصفوفة الارتباطية مؤلفة من معاملات الارتباط الزوجية لكل متحولين ضمن المجموعتين X , Y أو بينهما، ونحسبها من معادلة (بيرسون) المعرفة بالعلاقة الآتية:

$$r_{ij} = \frac{cov(X_i , X_j)}{\sqrt{s_i^2} * \sqrt{s_j^2}} \quad , \quad r_{ij}^* = \frac{cov(X_i , Y_j)}{\sqrt{s_i^2} * \sqrt{\sigma_j^2}} \qquad (5-3)$$

فنحصل على الجدول التالي:

الجدول (3-4) : المصفوفات الارتباطية للمجموعتين X و Y .

المتحولات	X_1	X_2	X_3	....	X_p	Y_1	Y_2	Y_3	....	Y_q
X_1	1	r_{12}	r_{13}	....	r_{1p}	r_{11}^*	r_{12}^*	r_{13}^*	....	r_{1q}^*
X_2	r_{21}	1	r_{23}	....	r_{2p}	r_{21}^*	r_{22}^*	r_{23}^*	....	r_{2q}^*
X_3	r_{31}	r_{32}	1	....	r_{3p}	r_{31}^*	r_{32}^*	r_{33}^*	....	r_{3q}^*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$
X_p	r_{p1}	r_{p2}	r_{p3}	....	1	r_{p1}^*	r_{p2}^*	r_{p3}^*	....	r_{pq}^*

المتحولات	X_1	X_2	X_3	....	X_p	Y_1	Y_2	Y_3	....	Y_q
Y_1	r^*_{11}	r^*_{21}	r^*_{31}	....	r^*_{p1}	1	r'_{12}	r'_{13}	....	r'_{1q}
Y_2	r^*_{21}	r^*_{22}	r^*_{32}	....	r^*_{p2}	r'_{21}	1	r'_{23}	....	r'_{2q}
Y_3	r^*_{31}	r^*_{32}	r^*_{33}	....	r^*_{p3}	r'_{31}	r'_{32}	1	....	r'_{3q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$
Y_q	r^*_{1q}	r^*_{2q}	r^*_{3q}	....	r^*_{pq}	r'_{p1}	r'_{p2}	r'_{p3}	....	r'_{pq}

واختصارا لكل ذلك نرمز للمصفوفات الأربعة التي يحتويها الجدول السابق بالرموز التالية:

R_{yx} , R_{xy} , R_{yy} , R_{xx} ونضعها مع مراتبها في جدول خاص كما يلي :

الجدول (3-5): مصفوفات معاملات الارتباط الزوجية لــ Y , X :

المتحولات	$X_1 \ , X_2 \ , \ldots \ , X_p$	$Y_1 \ , Y_2 \ , \ldots \ , Y_q$
X_1 X_2 X_3 $\vdots$ X_p	$R_{xx} \ (p * p)$	$R_{xy} \ (p * q)$
Y_1 Y_2 Y_3 $\vdots$ Y_q	$R_{yx} \ (q * p)$	$R_{yy} \ (q * q)$

وهنا نلاحظ أن: $R_{yx} \ (q * p) = R'_{xy} \ (p * q)$

3-3 الشروط المفروضة على المتحولات Y , X :

إن التحليل القانوني مؤسس على توفر عدة شروط وفرضيات في المتحولات المدروسة، وقبل أن نباشر عمليات الحساب والتحليل لابد لنا من التأكد من توفر الشروط المفروضة على المتحولات المستخدمة في الارتباط القانوني، وإن أهم هذه الشروط هي:

1- وجود علاقة سببية واضحة بين مجموعة المتحولات المستقلة (المؤثرة) X ومجموعة المتحولات التابعة Y ، مثل علاقة مجموعة المؤشرات السكانية (عدد السكان + معدل الولادة + معدل الوفاة+ مدة الحياة) مع مجموعة المؤشرات الاقتصادية (الناتج+ الدخل الفردي+ معدل العمالة+ معدل البطالة) ...الخ .

2- أن يكون للمتحولات المدروسة في المجموعة X وفي المجموعة Y صفة عشوائية، أي أن تأخذ قيمتها بطريقة عشوائية غير معروفة مسبقاً. وأن لا تتضمن قيماً شاذة (متطرفة) لأن وجود القيم

الشاذة يؤثر كثيراً على نتائج التحليل، وإن وجدت يجب استبعادها مع القيم المقابلة لها في المتحولات الأخرى .

3- أن تكون المتحولات ضمن كل مجموعة مستقلة عن بعضها البعض. أي أن تكون تغيرات كل متحول في المجموعة مستقلة وغير مشروطة بتغيرات أي متحول آخر فيها. وذلك بغض النظر عن وجود ارتباط نظري غير متين بين بعض المتحولات ضمن كل مجموعة .

4- يفترض أن تكون العلاقات الثنائية بين أي متحولين ضمن X أو ضمن Y، أو بين أي متحول من X مع أي متحول من Y علاقة خطية غير تامة. وإذا كانت العلاقة بينهما تامة أو شبه تامة (بمعامل ارتباط أكبر من 95%) فإننا نقوم بحذف أحدهما (الأقل أهمية للدراسة). أما إذا كانت العلاقة بينهما غير خطية(وعندها تكون $p > \alpha$) فإننا نحذف أحدهما أيضاً أو نقوم بتحويله إلى متحول آخر (حسب هدف الدراسة)، ويتم التحقق من ذلك خلال دراسة مصفوفات الارتباط الزوجية .

5- أن يكون حجم العينة n لكل المتحولات المدروسة يتراوح من 10 إلى 20 حالة أو مشاهدة، وأن تكون قيمها خاضعة للتوزيع الطبيعي أو متقاربة منه . ويتم التحقق من ذلك من خلال إجراء اختبار الطبيعية على بيانات كل متحول على حدة .

6- أن تكون المتحولات في كل مجموعة ذات طبيعة واحدة ، وأن تكون بياناتها متجانسة أو متقاربة من القيمة المتوقعة. ويتم التحقق من ذلك بتطبيق اختبار F على متحولات كل مجموعة على حدة .

7- أن تكون المتحولات ضمن كل مجموعة قابلة للتركيب الخطي فيما بينها .

3-4 التحليل الرياضي للارتباط القانوني

يهدف الارتباط القانوني إلى دراسة العلاقة بين مجموعة من المتحولات المستقلة X ومجموعة من المتحولات التابعة Y حيث أن:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{bmatrix} \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_q \end{bmatrix}$$

حيث p هو عدد المتحولات في X وq هو عدد المتحولات في Y، ولدراسة هذه العلاقة نشكل من كل منهما تركيباً خطياً مثقلاً بأمثال مجهولة a_i و b_j كما يلي:

$$U = a_1X_1 + a_2X_2 + a_3X_3 + \cdots a_pX_p \qquad (6-3)$$

$$V = b_1Y_1 + b_2Y_2 + b_3Y_3 + \cdots b_qY_q \qquad (7-3)$$

وإذا رمزنا للأمثال a_i بالشعاع $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ ولمنقوله بالرمز $a'(a_1, a_2, \dots a_p)$

وإذا رمزنا للأمثال b_j بالشعاع $b=\begin{bmatrix} b_1\\ b_2\\ \vdots\\ b_q\end{bmatrix}$ ولمنقوله بالرمز $b'(b_1,b_2,\dots b_q)$

فإنه يمكننا كتابة التركيبين (3−6) و (3−7) على الشكل التالي :

$$U=\left(a_{\,1,}a_{\,2,}a_{\,3,}\cdots a_{\,p}\right)\begin{bmatrix} X_1\\ X_2\\ X_3\\ \vdots\\ X_p\end{bmatrix}=a'*X \qquad (8-3)$$

$$V=\left(b_{1,}b_{2,}b_{3,}\cdots b_q\right)\begin{bmatrix} Y_1\\ Y_2\\ Y_3\\ \vdots\\ Y_q\end{bmatrix}=b'*Y \qquad (9-3)$$

وهنا لابد من التمييز بين المتحولات المستخدمة في العلاقات (3−6) و (3−7) أو (3−8)و (3−9) وفق التالي :

إن متحولات المجموعة X تسمى بالمتحولات المستقلة (Independent variables) وهي متحولات معروفة ولدينا عن كل منها n قيمة عددية (مشاهدة) متقابلة كما في الجدول (1) .

وإن متحولات المجموعة Y تسمى بالمتحولات التابعة (Dependent variables) وهي أيضاً معروفة ولدينا عن كل منها n قيمة عددية (مشاهدة) متقابلة كما في الجدول (1).

أما U , V فهي توابع جديدة ومجهولة ويطلق على كل منهما اسم التابع القانوني أو المركب القانوني (Canonical varaite) . فيكون لدينا:

U – هو المركب القانوني للمجموعة X (التابع القانوني لـ X) وأرغب في تسميته بالمتابع القانوني لـ X.

V – هو المركب القانوني للمجموعة Y (التابع القانوني لـ Y) وأرغب في تسميته بالمتابع القانوني لـ Y.

ولكن الأمثال $a_{\,i}$ و b_j فهي أعداد مجهولة أيضاً ويجب علينا حسابها أو تقديرها حتى تصبح العلاقات (3-6),(3-7) أو (3-8),(3-9) علاقات محددة. وبما أنها يمكن أن تأخذ قيماً متعددة، فإن العلاقتين (3−6) و(3−7) يمكن أن تعطيانا عدداً من أزواج المركبات القانونية نرمز لها بـ (U_1,V_1) و(U_2,V_2) و (U_3,V_3) ...... كما يلي :

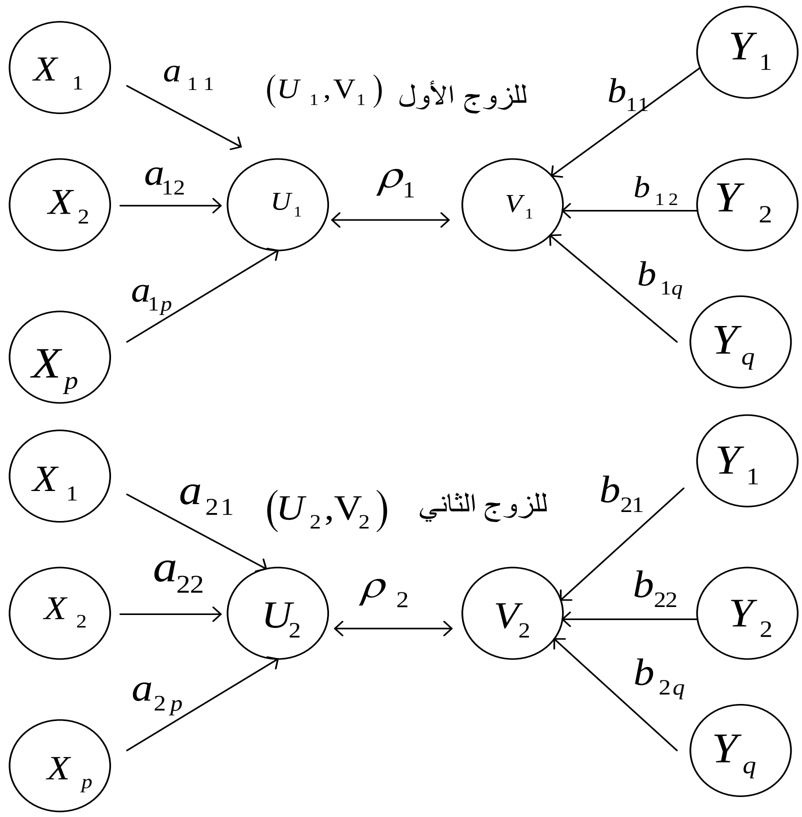
$$\begin{aligned} U_1&=a_{\,11}X_1+a_{\,12}X_2+a_{\,13}X+\cdots a_{\,1p}X_p\\ V_1&=b_{11}Y_1+b_{12}Y_2+b_{13}Y_3+\cdots b_{1q}Y_q \end{aligned} \qquad (10-3)$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\begin{aligned} U_2&=a_{\,21}X_1+a_{\,22}X_2+a_{\,23}X_3+\cdots a_{\,2p}X_p\\ V_2&=b_{21}Y_1+b_{22}Y_2+b_{23}Y_3+\cdots b_{2q}Y_q \end{aligned} \qquad (11-3)$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

ويمكن تمثيل العلاقة بين هذه المتحولات ومركباتها القانونية كما يلي :



الشكل (3-1) : تمثيل العلاقات القانونية

ولدراسة العلاقة بين المجموعتين X وY نقوم بحساب معامل الارتباط الزوجي ρ_k بين المركبين القانونيين في كل زوج من الأزواج الممكنة (U_1,V_1) ثم (U_2,V_2) ثم (U_3,V_3) ...الخ، وذلك باستخدام معادلة بيرسون كما يلي :

$$\rho_k = \frac{cov(U_k,V_k)}{\sqrt{var(U_k)} * \sqrt{var(V_k)}} \qquad (12-3)$$

ويسمى هذا المعامل ρ_k بمعامل الارتباط القانوني للزوج (U_k,V_k). وبالتالي سيكون لدينا عدداً من معاملات الارتباط القانونية ρ_k مساوياً لعدد الأزواج الممكنة لـ (U_k,V_k). وسنرى لاحقاً أن عدد هذه المعاملات الممكنة ρ_k يساوي s . حيث أن s هو أصغر العددين p أو q أي أن s=min (p,q)، كما إن بعضها يمكن أن يكون معدوماً أو ضعيفاً، وعندها سيتم إهماله خلال التحليل والدراسة بعد إجراء الاختبارات اللازمة لذلك.

ولنفرض الآن إننا حصلنا على جميع معاملات الارتباط القانونية ρ_k المقابلة للأزواج الممكنة وقمنا بتربيعها (لأنها قد تكون سالبة) ثم ترتيبها تنازلياً فإننا سنحصل على متتالية منها كما يلي :

$$\rho_1^2 \geq \rho_2^2 \geq \rho_3^2 \geq \cdots \rho_s^2 \geq 0 \qquad (13-3)$$

وسنصطلح هنا على أن نسمي الزوج (U ,V) الذي يقابل المعامل الأكبر ρ_1^2 بالزوج الأول ونرمز له بـ (U_1,V_1)، كما نسمي الزوج (U, V) الذي يقابل المعامل الثاني ρ_2^2 بالزوج الثاني ونرمز له بـ (U_2,V_2)، وهكذا نرمز للزوج الذي يقابل المعامل ρ_k^2 بالرمز (U_k,V_k) ونسميه بالزوج k ...الخ .

وبعد اختبار معنوية هذه المعاملات الارتباطية نستبعد المعاملات غير المعنوية ويقتصر الأمر على المعاملات الأولى فقط (الأول ثم الثاني ثم الثالث).

وبصورة عامة سنقوم بمعالجة هذه المسألة ودراسة الارتباط القانوني بين كل زوجين مركبين (U ,V) لا على التحديد من خلال حساب معامل الارتباط القانوني بينهما والمعرف بالعلاقة التالية:

$$\rho(U,V)=\frac{cov(U,V)}{\sqrt{var(U)}*\sqrt{var(V)}} \qquad (14-3)$$

لحساب قيمة هذا المعامل يلزمنا حساب التباين var (U) والتباين var (V) والتباين المشترك $cov(U,V)$ ، ويتم حسابهم من العلاقتين (3−6) و(3−7) باستخدام العلاقات التالية:

$$var(U)=\sum_{i=1}^{p}\sum_{j=1}^{p}a_ia_jcov(X_i,X_j) \qquad (15-3)$$

$$var(V)=\sum_{i=1}^{q}\sum_{j=1}^{q}b_ib_jcov(Y_i,Y_j) \qquad (16-3)$$

$$cov(U,V)=\sum_{i=1}^{p}\sum_{j=1}^{q}a_ib_jcov(X_i,Y_j) \qquad (17-3)$$

ويمكن تحويل هذه المجاميع إلى جداء مصفوفات وأشعة إذا علمنا أن:

$$var(U)=E(U-\bar{U})^2=E(a'X-a'\bar{X})^2=a'E(X-\bar{X})^2a \qquad (18-3)$$

وبما أن التوقع $E(X-\bar{X})^2$ مأخوذ على جميع الأشعة $(X-\bar{X})$ فإنه يساوي مصفوفة التباينات المشتركة لمتحولات المجموعة X وهي C_{xx} . ومنها نستنتج أن :

$$var(U)=a'*C_{xx}*a \qquad (19-3)$$

أي أن التباين $var(U)$ يحسب من العلاقة:

$$var(U)=\left(a_1a_2a_3\cdots a_p\right)*\begin{bmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1p}\\ s_{12} & s_{22} & s_{23} & \cdots & s_{2p}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ s_{p1} & s_{p2} & s_{p3} & \cdots & s_{pp}\end{bmatrix}*\begin{bmatrix} a_1\\ a_2\\ \cdots\\ a_p\end{bmatrix} \qquad (20-3)$$

وبطريقة مشابهة يمكننا أن نجد أن التباين $var(V)$ يحسب من العلاقة :

$$var(V)=b'*C_{yy}*b \qquad (21-3)$$

وكذلك يمكننا أن نجد أن التباين المشترك $cov(U,V)$ يحسب من العلاقة :

$$cov(U,V)=a'*C_{xy}*b=b'*C_{yx}*a \qquad (22-3)$$

حيث أن: C_{xx} و C_{yy} و C_{xy} و C_{yx} هي المصفوفات التباينية المعرفة سابقاً في الجدول (3) .

وبذلك تأخذ العلاقة (3-14) السابقة لمعامل الارتباط القانوني الشكل التالي:

$$\rho(U,V) = \frac{a' * C_{xy} * b}{\sqrt{a' * C_{xx} * a} * \sqrt{b' * C_{yy} * b}} \qquad (23-3)$$

وللحصول على معامل الارتباط القانوني للزوج الأول (U_1, V_1) من العلاقة (3-23) علينا أن نقوم بتقدير الأمثال $b_j\ , a_i$ الواردة في الشعاعين a و b . بحيث نحصل من (3–23) على أكبر قيمة ممكنة لمعامل الارتباط القانوني ρ(تعظيم قيمة ρ). وهي مسألة معقدة ولها لا نهاية من الحلول المقبولة. وحتى نحصل على حل محدد ومثالي لهذه المسألة لابد من وضع شروط إضافية على عناصر الشعاعين a و b .

ولتحقيق هذا الهدف نعلم قيمة معامل الارتباط ρ لا تتأثر بتغيير واحدات القياسي لـ U أوV. فمثلاً إذا ضربنا المركب U بعدد ثابت k ، فإن قيمة ρ تبقى ثابتة لأنه يكون لدينا مايلي:

$$\rho_{(kU,V)} = \frac{k * a' * C_{xy} * b}{\sqrt{k^2 a' * C_{xx} * a} * \sqrt{b' * C_{yy} * b}} = \frac{a' * C_{xx} * b}{\sqrt{a' * C_{xx} * a} * \sqrt{b' * C_{yy} * b}} = \rho(U,V) \quad (\mathbf{24-3})$$

وهذا يعني أنه إذا قمنا بتغيير واحدات القياس المركب U (أو كلا المركبين U و V)، فإن ذلك لا يؤثر على قيمة ρ. ولذلك نقوم باختيار واحدات قياس مناسبة للمركبين القانونيين U وV. وإن أبسط واحدات قياس لهما هي التي تجعل التباينين لهما $var\ (U)$ و$var\ (V)$ متساويين ومساويين للواحد الصحيح. لذلك نضع عليهما الشرطين التاليين :

$$var\ (U) = a' * C_{xx} * a = 1 \qquad (25-3)$$

$$var\ (V) = \grave{b} * C_{yy} * b = 1 \qquad (26-3)$$

وبذلك تتحول المعادلة (3-23) إلى الشكل البسيط التالي:

$$\rho = \frac{a' * C_{xy} * b}{1 * 1} = a' * C_{xy} * b \qquad (27-3)$$

وتتحول مسألة تعظيم قيمة المعامل ρ إلى مسألة تعظيم البسط $\left(\grave{a} * C_{xy} * b\right)$ ويصبح هدفنا إيجاد القيمة العظمى لـρ من العلاقة(3-27) ضمن الشرطين(3-25),(3-26) ولنرمز لهذه القيمة العظمى بـρ_1 فيكون لدينا:

$$\rho_1 = max\left(a' * C_{xy} * b\right) \qquad (28-3)$$

ويسمى الزوج U , V الذي يقابل هذه القيمة ρ_1 بالزوج الأول ونرمز له بـ (U_1, V_1)

ولإيجاد القيمة العظمى الشرطية لذلك البسط $\left(a' * C_{xy} * b\right)$ في(3-27) نشكل منه تابع لاغرانج مع الشرطين (3-25) , (3-26) كما يلي:

$$L = a' * C_{xy} * b - \frac{\lambda_x}{2}(a' * C_{xx} * a - 1) - \frac{\lambda_y}{2}\left(b' * C_{yy} * b - 1\right) \quad (29-3)$$

حيث أن λ_y , λ_x هما عددان حقيقيان وسيطان يسميان بمضروبي لاغرانج، ثم نشتق التابع L بالنسبة لـa ثم بالنسبة لـ b ونضعهما مساويين للصفر فنحصل على المعادلتين التاليتين[*] :

$$\frac{\partial L}{\partial a} = C_{xy} * b - \lambda_x C_{xx} * a = 0 \qquad (30-3)$$

$$\frac{\partial L}{\partial b} = C_{yx} * a - \lambda_y C_{yy} * b = 0 \qquad (31-3)$$

نضرب الأولى بـ a' والثانية بـ b' فنحصل على العلاقتين:

$$a' * C_{xy} * b - \lambda_x * a' * C_{xx} * a = 0$$
$$b' * C_{yx} * a - \lambda_y * b' * C_{yy} * b = 0$$

ثم نطرح الثانية من الأولى فنحصل على ما يلي:

$$a' * C_{xy} * b - \lambda_x a' * C_{xx} * a - b' * C_{yx} * a + \lambda_y b' * C_{yy} * b = 0$$

ومن العلاقات (3-22) , (3-25) , (3-26) نجد أن:

$a' * C_{xy} * b = b' * C_{yx} * a$ وأن : $a' * C_{xx} * a = b' * C_{yy} * b = 1$ شرطاً

وبالتالي فإننا نحصل على أن: $-\lambda_x + \lambda_y = 0$ أي أن:

$$\lambda_x = \lambda_y = \lambda \quad \text{عدد حقيقي} \qquad (32-3)$$

نعوض قيمة λ الموحدة في المعادلتين (3-30) ,(3-31). ونفترض أن المصفوفة C_{yy} قابلة للعكس (أي C_{yy}^{-1} موجودة) ونضرب المعادلة (3-31) من اليسار بالمقلوب C_{yy}^{-1} فنحصل على أن:

$$C_{yy}^{-1} * C_{yx} * a - \lambda C_{yy}^{-1} * C_{yy} * b = 0$$

ومنها نحصل على أن:

$$C_{yy}^{-1} * C_{yx} * a - \lambda * I * b = \lambda * b$$

وبذلك نجد أن الشعاع b يحسب عن طريق الشعاع a من العلاقة التالية:

$$b = \frac{C_{yy}^{-1} * C_{yx}}{\lambda} * a \qquad (33-3)$$

أي أن الشعاع b يتناسب مع الشعاع a (يرتبط معه تناسبياً)، وبتعويض قيمة الشعاع b من العلاقة (3-33) في معادلة المشتق الأول (3-30) نحصل على أن :

$$C_{xy} \frac{C_{yy}^{-1} * C_{yx}}{\lambda} a - \lambda C_{xx} * a = 0$$

- [*] إن الاشتقاق هنا يتم حسب قواعد اشتقاق المصفوفات المعرفة بالعلاقات التالية:

$$\frac{\partial (a' C_{xy} b)}{\partial a} = \frac{\partial}{\partial a}\left[b' C_{yx} * a\right]' = \left[b' C_{yx}\right]' = C_{xy} * b \qquad 1 - \frac{\partial (c' * x)}{\partial x} = C \Longrightarrow$$

$$\frac{\partial}{\partial b}\left[a' C_{xy} * b\right] = \frac{\partial}{\partial b}\left[\left(C_{yx} a\right)' * b\right] = C_{yx} * a \qquad 2 - \frac{\partial (x' * B * x)}{\partial x} = 2B * x \quad \text{متناظرة } B$$

$$\frac{\lambda_x}{2} \frac{\partial (a' * C_{xx} * a)}{\partial a} = \frac{\lambda_x}{2} * 2C_{xx} * a = \lambda_x C_{xx} * a \qquad 3 - \frac{\partial (X' * X)}{\partial X} = 2X$$

ومنها نجد أن:

$$C_{xy} * C_{yy}^{-1} * C_{yx} * a = \lambda^2 C_{xx} * a \qquad (34-3)$$

وهذا هو الشكل العام لمعادلة القيم الذاتية($Ax = \lambda B * x$)، ولتبسيط الأمور نفترض أن المصفوفة C_{xx} قابلة للعكس (أي أن C_{xx}^{-1} موجودة) ونضرب طرفي العلاقة (3-34) من اليسار بـــ C_{xx}^{-1} فنحصل على المعادلة التالية:

$$C_{xx}^{-1} * C_{xy} * C_{yy}^{-1} * C_{yx} * a = \lambda^2 C_{xx}^{-1} * C_{xx} * a = \lambda^2 I * a$$

أي أن:

$$C_{xx}^{-1} * C_{xy} * C_{yy}^{-1} * C_{yx} * a = \lambda^2 * a \qquad (35-3)$$

وهذا هو الشكل النموذجي لمعادلة القيم الذاتية ($AX = \lambda X$).

وهنا نلاحظ أن نتيجة الجداء المصفوفي الذي في الطرف الأيسر من (3-35) هو عبارة عن مصفوفة مربعة من المرتبة p * p .

وللاختصار نرمز لها بالرمز A فيكون لدينا ما يلي:

$$A_{p*p} = C_{xx}^{-1} * C_{xy} * C_{yy}^{-1} * C_{yx} \qquad (36-3)$$

وعندها تأخذ العلاقة (3-35) شكل معادلة القيم الذاتية بالنسبة لـ a كما يلي:

$$A * a = \lambda^2 * a \qquad (37-3)$$

والتي يمكن كتابتها على الشكل النموذجي كما يلي:

$$[A - \lambda^2 I] * a = 0 \qquad (38-3)$$

وحتى نستطيع الحصول على الشعاع a ، علينا أن نتجنب الحصول على الحل التافه ($a = 0$)، لذلك نفترض أن المصفوفة المربعة $[A - \lambda^2 I]$ هي مصفوفة شاذة. وهذا يعني أن قيمة محددها يجب أن تساوي الصفر، أي يجب أن يكون:

$$|A - \lambda^2 I| = 0 \qquad (39-3)$$

وبما أن مرتبة هذا المحدد تساوي p فإن مفكوكه بالنسبة لـ λ^2 سيشكل معادلة مؤلفة من كثير حدود من المرتبة p للمقادير (λ^2) كما يلي:

$$(\lambda^2)^P + C_1(\lambda^2)^{P-1} + C_2(\lambda^2)^{P-2} + \cdots + C_P = 0 \qquad (40-3)$$

وعند حل هذه المعادلة نحصل على p جذراً موجباً لـ λ^2، وتسمى هذه الجذور بالجذور الكامنة (Roots) للمصفوفة A ويغلب عليها اسم أو مصطلح القيم الذاتية (Eigen values) للمصفوفة A.

وبعد الحصول على هذه القيم الذاتية λ_k^2 نقوم بترتيبها تنازلياً فيكون لدينا:

$$\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \geq \lambda_4^2 \geq \cdots \lambda_p^2 \geq 0 \qquad (41-3)$$

وبعد الحصول على القيم الذاتية λ_k^2 (الجذور السابقة) علينا أن نستخدمها في حساب الشعاعين a , b للتركيبتين (3-6),(3-7)، لذلك نقوم بتعويض كل منها وعلى التوالي في المعادلة الذاتية (3-38) فنحصل في كل مرة على جملة معادلات متجانسة وغير مستقلة للمجهول a، ولتوضيح ذلك نبدأ بتعويض القيمة الذاتية الاولى λ_1^2 في(3-38) فنحصل على المعادلة الذاتية التالية (بفرض أن $p = 3$):

$$[A-\lambda_1^2 I]a = \begin{bmatrix} \alpha_{11}-\lambda_1^2 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22}-\lambda_1^2 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33}-\lambda_1^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = 0 \qquad (42-3)$$

وهي تعطينا جملة معادلات متجانسة بالنسبة للمجهول a كما يلي:

$$\begin{aligned} &(\alpha_{11}-\lambda_1^2)a_1 + \alpha_{12}\ \ a_2 + \alpha_{13} * a_3 = 0 \\ &\alpha_{21}\ \ a_1 + (\alpha_{22}-\lambda_1^2)a_2 + \alpha_{23} * a_2 = 0 \\ &\alpha_{31}\ \ a_1 + \alpha_{32}\ \ a_2 + (\alpha_{33}-\lambda_1^2)a_3 = 0 \end{aligned} \qquad (43-3)$$

وهي جملة معادلات غير مستقلة لأن محدد مصفوفة أمثالها يساوي الصفر، وذلك حسب العلاقة (3-39).

وبالتالي فإن هذه الجملة ستعطينا لانهاية من الحلول المقبولة للشعاع a . وأن كل من هذه الحلول (a^*) له لانهاية من المضاعفات مثل $(C * a^*)$ ويشكل حزمة من الحلول المقبولة . وهذا الحل مع مضاعفاته يحدد لنا فقط الاتجاه العام للشعاع a^*، ولكن طوله أو قياسه أو قيم عناصره تبقى غير محددة . (لأن كل مضاعف يعطينا قياس مختلف عن الآخر مع إن لهم نفس الاتجاه)،(راجع الجزء الأول) .

وحتى نستفيد من المعادلات (3-43) في تحديد قياس الشعاع a، نضيف إليها معادلة الشرط (3-25) ونقوم بحلهما معاً بطريقة التعويض، فنحصل (غالباً) على حل وحيد لها، يعطينا العناصر الأساسية للشعاع $a(a_1, a_2, a_3, \ldots.)$. وبذلك نكون قد توصلنا إلى الهدف المطلوب .

وإذا لم نستطيع الحصول على حل وحيد للشعاع a من المعادلات (3-43) و (3-25) مباشرة (بسبب عدم الاستقلال)، أو إذا كانت دقة الحل غير مضمونة (بسبب كثرة الحسابات)، فإنه يمكننا اتباع طريقة أخرى تتلخص بأن نبحث عن أي حل مقبول للجملة (3-43) مثل (a^*) ونضربه بعدد ثابت k، فنحصل على شعاع جديد مقبول هو $(k * a^*)$، ثم نعوض هذا الحل الأخير في الشرط (3-25) الذي فرضناه على تباين المركب القانوني U فنحصل على معادلة جديدة من الشكل:

$$var(U) = (ka^{*\prime})C_{xx} * (ka^*) = 1 \qquad (44-3)$$

وهي تكافئ المعادلة:

$$k^2 * a^{*1}C_{xx} * a^* = 1 \qquad (44'-3)$$

ومنها نحسب قيمة k^* التي تجعل ذلك الحل a^* محققاً للشرط (3-25)، ونستخدم ذلك العدد المحسوب k^* لتعديل الحل المقبول a^* ، لذلك نقوم بحساب الجداء (k^*a^*) من جديد فنحصل على الشعاع الذاتي الأول والذي سنرمز له أيضاً بـ a^* ونكتبه على الشكل التالي:

$$a^* = \begin{bmatrix} a_1^* \\ a_2^* \\ a_3^* \\ \vdots \end{bmatrix} \qquad (45-3)$$

وللحصول على عناصر الشعاع الذاتي b^* المصاحب لمتحولات المركب القانوني V_1 من الزوج الأول $(U_1\ , V_1)$ نستخدم العلاقة (3-33) فنجد أن :

$$b^* = \begin{bmatrix} b_1^* \\ b_2^* \\ b_3^* \\ \vdots \end{bmatrix} = \frac{C_{yy}^{-1} * C_{yx}}{\lambda_1} * \begin{bmatrix} a_1^* \\ a_2^* \\ a_3^* \\ \vdots \end{bmatrix} \qquad (46-3)$$

وبعد حصولنا على الشعاع b^* نقوم بتعديله ليحقق الشرط (3-26) المفروض على تباين V فنضربه بعدد ثابت k ونعوض الناتج في الشرط (3-26) فنحصل على أن:

$$var\ (V) = k * b^{*\prime} * C_{yy} * kb^* = 1$$

وهي تكافئ المعادلة :

$$k^2 b^{*\prime} C_{yy} * b^* = 1 \qquad (47-3)$$

ومنها نحسب قيمة k^* التي تجعل الشعاع b^* يحقق الشرط (3-26)، لذلك نضرب ذلك الشعاع b^* بـ k^* المحسوبة فنحصل على شعاع جديد($k^*\ b^*$) نرمز له أيضاً بـ b^* ونكتبه على الشكل التالي :

$$b_1^* = \begin{bmatrix} b_1^* \\ b_2^* \\ b_3^* \\ \vdots \end{bmatrix} \qquad (48-3)$$

وبعد الحصول على الشعاعين الذاتيين الأولين b_1^*, a_1^* يمكننا حساب قيمة معامل الارتباط القانوني ρ_1 للزوج الأول $(U_1\ , V_1)$ من العلاقة (3-27) وهي تساوي:

$$\rho_1 = a^{*\prime} * C_{xy} * b^* \qquad (49-3)$$

كما يمكن حساب قيمة ρ_1 من العلاقة التالية (مع وضع الإشارة المناسبة) :

$$\rho_1 = \pm\sqrt{\lambda_1^2} = \pm\lambda_1 \qquad (50-3)$$

وسنحدد إشارته لاحقاً من خلال الشكل البياني للزوج $(U_1\ , V_1)$ أو من خلال حساب معامل الارتباط بينهما، وأخيراً نقوم بتعويض الشعاعين الذاتيين b_1^*, a_1^* في العلاقتين (3-6) و(3-7)، فنحصل على تركيبين خطيين محددين للزوج القانوني الأول $(U_1\ , V_1)$ هما:

$$\tilde{U}_1 = a_1^* X_1 + a_2^* X_2 + a_3^* X_3 + \cdots + a_p^* X_p \qquad (51-3)$$

$$\tilde{V}_1 = b_1^* Y_1 + b_2^* Y_2 + b_3^* Y_3 + \cdots + b_q^* Y_q \qquad (52-3)$$

وبعدها نقوم بحساب القيم النظرية لكل من المركبين $\left(\tilde{U}_1\ , \tilde{V}_1\right)$، وذلك بتعويض قيم المشاهدات الفعلية (في الصفوف) للمتحولات المستقلة X في العلاقة (3-51)، وبتعويض قيم المشاهدات الفعلية (في الصفوف) للمتحولات التابعة Y في العلاقة (3-52) . فنحصل على القيم النظرية المتقابلة للزوج القانوني الأول $\left(\tilde{U}_1\ , \tilde{V}_1\right)$.

ثم نقوم برسم الشكل البياني للقيم النظرية $\left(\tilde{U}_{1i}\ , \tilde{V}_{1i}\right)$ فنحصل على شكل انتشار، ومنه يظهر لنا اتجاه ومتانة الارتباط الخطي بين $(U_1\ , V_1)$.

ومن هنا يمكننا أيضاً حساب القيمة النظرية لمعامل الارتباط القانوني الأول ρ_1 فنحصل على القيمة العظمى للمعامل ρ_1 والتي تعبر عن شدة الارتباط بين مركبي الزوج الأول $\left(\tilde{U}_1\ , \tilde{V}_1\right)$. وإن إشارته

تتحدد من خلال تلك الحسابات أو من خلال اتجاه الشكل البياني(ننصح القارئ بمراجعة المثال العام المحلول في آخر الفصل ثم المتابعة).

ومن جهة أخرى وبشكل عام يمكننا حساب قيمة معامل الارتباط القانوني ρ من العلاقة (3-28)، وذلك بعد تعويض قيمة الشعاع b من العلاقة (3-33) في العلاقة (3-28) فنجد أن:

$$\rho = a' * C_{xy} \frac{C_{yy}^{-1} * C_{yx}}{\lambda} * a \qquad (53-3)$$

$$\rho = \frac{a'}{\lambda} C_{xy} * C_{yy}^{-1} * C_{yx} * a \qquad (54-3)$$

ومن العلاقة (3-34) نجد أن الجداء الأخير من (3-54) يساوي $\lambda^2 * C_{xx} * a$ وبالتعويض نحصل على أن:

$$\rho = \frac{a'}{\lambda} * \lambda^2 C_{xx} * a = \lambda a' * C_{xx} * a$$

وبما أنه لدينا من الشرط (3-25) أن $a' * C_{xx} * a = 1$، فإننا نجد أن:

$$\rho = \lambda = \pm\sqrt{\lambda^2} \qquad (55-3)$$

أي أن قيمة معامل الارتباط ρ_k المقابل لأي زوج قانوني (U_k , V_k) تساوي الجذر التربيعي للقيمة الذاتية λ_k^2 المقابلة لذلك الزوج . وتحدد إشارة ρ_k لكل زوج من خلال الحساب الفعلي لقيمته من القيم النظرية للزوج $\left(\widetilde{U}_k , \widetilde{V}_k\right)$ أو من خلال شكل الانتشار، أو من خلال العلاقة (3−49) .

ومما سبق يمكننا أن نستخلص بعض خواص القيم الذاتية λ_k^2 المحسوبة من معين المصفوفة $[A - \lambda I]$ كمايلي :

بما أن قيمة معامل الارتباط ρ_k كأي معامل ارتباط خطي يحقق المتراجحة المزدوجة التالية:

$$-1 \le \rho_k \le +1 \quad \Longrightarrow \quad 0 \le \rho_k^2 \le 1 \qquad (56-3)$$

ومن العلاقة (3-55) نستنتج $\rho_k^2 = \lambda_k^2$ ومن (3−56) نجد أن:

$$0 \le \lambda_k^2 \le 1 \qquad (57-3)$$

وهذا يعني أن جميع القيم الذاتية λ_k^2 المحسوبة من المصفوفة $[A - \lambda I]$ هي قيم غير سالبة وأصغر أو تساوي الواحد، لذلك يمكننا ترتيبها تنازلياً كما يلي:

$$1 \ge \lambda_1^2 \ge \lambda_2^2 \ge \lambda_3^2 \ge \cdots \lambda_p^2 \ge 0 \qquad (58-3)$$

وتقابلها معاملات التحديد المقابلة للازدواج القانونية (U_k , V_k) :

$$1 \ge \rho_1^2 \ge \rho_2^2 \ge \rho_3^2 \ge \cdots \rho_p^2 \ge 0 \qquad (59-3)$$

* **طريقة أخرى:** يمكن معالجة الموضوع السابق بأسلوب آخر، وذلك بحساب b أولاً ثم حسابa كمايلي:

لنأخذ المشتق الأول المحسوب في العلاقة (3-30) ونضرب طرفي المعادلة (3-30) من اليسار بالمصفوفة المقلوبة C_{xx}^{-1} (بفرض أن C_{xx}^{-1} موجودة) فنحصل على ما يلي:

$$C_{xx}^{-1} * C_{xy} * b = \lambda * C_{xx}^{-1} * C_{xx} * a = \lambda I * a = \lambda a \qquad (60-3)$$

ومنها نستخلص أن الشعاع a يمكن أن يحسب بدلالة الشعاع b من العلاقة :

$$a = \frac{C_{xx}^{-1} * C_{xy}}{\lambda} * b \qquad (61-3)$$

وبتعويض a في معادلة المشتق الثاني (3-31) نجد أن:

$$C_{yx} \frac{C_{xx}^{-1} * C_{xy} * b}{\lambda} - \lambda C_{yy} * b = 0$$

$$C_{yx} * C_{xx}^{-1} * C_{xy} * b = \lambda^2 C_{yy} * b \qquad (62-3)$$

وهو الشكل العام لمعادلة القيم الذاتية $(AX = \lambda BX)$. ولتبسيط الأمور نضرب طرفي (3-62) بالمقلوب C_{yy}^{-1} (بفرض أن C_{yy}^{-1} موجودة) فنحصل على المعادلة الذاتية التالية:

$$C_{yy}^{-1} * C_{yx} * C_{xx}^{-1} * C_{xy} * b = \lambda^2 * b \qquad (63-3)$$

وهي معادلة الشكل النموذجي لمعادلة القيم الذاتية $(AX = \lambda X)$.

وهنا نلاحظ أن نتيجة الجداء المصفوفي في الطرف الأيسر هي عبارة عن مصفوفة مربعة من المرتبة $q * q$، واختصاراً للرموز نرمز لذلك الجداء بالرمز B فيكون لدينا:

$$B = C_{yy}^{-1} * C_{yx} * C_{xx}^{-1} * C_{xy} \qquad (64-3)$$

وبالتعويض في (3-63) نحصل على المعادلة الذاتية التالية:

$$B * b = \lambda^2 * b \qquad (65-3)$$

والتي يمكن كتابتها على الشكل التالي:

$$[B - \lambda^2 I] * b = 0 \qquad (66-3)$$

وهي معادلة الشكل النموذجي لمعادلة القيم الذاتية.

وحتى نتحاشى الحصول على الحل التافه $(b = 0)$ نفترض أن المصفوفة المربعة $[B - \lambda^2 I]$ هي مصفوفة شاذة، وهذا يقتضي أن يكون محددها مساوياً للصفر. فيكون لدينا:

$$|B - \lambda^2 I| = 0 \qquad (67-3)$$

وبما أن مرتبة هذا المحدد تساوي q فإن مفكوكه سيشكل معادلة مؤلفة من كثير حدود من المرتبة q للمقادير λ^2 كما يلي:

$$(\lambda^2)^q + c_1(\lambda^2)^{q-1} + c_2(\lambda^2)^{q-2} + \cdots + c_q = 0 \qquad (68-3)$$

ثم نقوم بإيجاد جذور هذه المعادلة بالنسبة لـ (λ^2) فنحصل على q جذراً. وتسمى هذه الجذور بالجذور الكامنة للمصفوفة B أو بالقيم الذاتية (eigenvalue) للمصفوفة B . ثم نقوم بترتيبها تنازلياً فنحصل على القيم الذاتية التالية:

$$\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \geq \cdots \geq \lambda_q^2 \geq 0 \qquad (69-3)$$

وعلى معادلات التحديد المقابلة لها : $\rho_1^2 \geq \rho_2^2 \geq \rho_3^2 \geq \cdots \geq \rho_q^2 \geq 0$

وللحصول على الشعاع الذاتي الأول b المصاحب لـ V_1 من الزوج (U_1, V_1) نأخذ الجذر الأول λ_1^2 ونقوم بتعويض قيمة λ_1^2 في المعادلة (3-66). ثم نجري الحسابات اللازمة فنحصل على جملة معادلات

خطية غير مستقلة ، تعطينا لا نهاية من الحصول المقبولة للشعاع b، وحتى نستطيع الاستفادة من هذه المعادلات نأخذ المضاعف k لأحد الحلول ونخضعه للشرط الذي وضعناه على تباين V فنحصل على:

$$var(V) = kb' * C_{yy} * kb = K^2 b^1 * C_{yy} * b = 1 \qquad (70-3)$$

ومنها نحصل على قيمة العدد k^* التي تجعل ذلك الحل حلاً يحقق الشرط (3-26)، ومنه نحصل على عناصر الشعاع الأول b^* وهي: $(b_1^*, b_2^*, b_3^*, \dots, b_q^*)$، وللحصول على عناصر الشعاع الأول a نطبق العلاقة (3-61) كما يلي:

$$a^* = \begin{bmatrix} a^* \\ a_1^* \\ \vdots \\ a_p^* \end{bmatrix} = \frac{C_{xx}^{-1} * C_{xy}}{\lambda_1} * \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_q^* \end{bmatrix} \qquad (71-3)$$

ثم نقوم بتعديله حتى يحقق الشرط المفروض عليه وهو الشرط (3-25) كما فعلنا مع b . وبعد ذلك نعوض شعاعي الأمثال b^*, a^* في العلاقتين (3–6) و3–7) فنحصل على نفس التركيبين الخطين المحددين للزوج الأول (U_1 , V_1) وهما:

$$U_1 = a_1^* X_1 + a_2^* X_2 + \cdots + a_p^* X_p \qquad (72-3)$$

$$V_1 = b_1^* Y_1 + b_2^* Y_2 + \cdots + b_q^* Y_q \qquad (73-3)$$

وبعدها نقوم بحساب القيم النظرية لـ V_1 , U_1 وذلك بتعويض القيم الفعلية لــ X ولـ Y في المعادلتين (3-72) و (3-73)، فنحصل على منظومة القيم النظرية المتقابلة $(\widetilde{U}_1 , \widetilde{V}_1)$، ثم نرسم الشكل البياني لهما، الذي يظهر لنا متانة وإشارة معامل الارتباط بين V_1 , U_1، ومنها نحسب قيمة معامل الارتباط القانوني ρ_1 ونحدد إشارته المناسبة. كما يمكن حساب ρ_1 من العلاقة: $\rho_1 = \pm\sqrt{\lambda_1^2}$ ثم تحديد إشارته كما سبق.

*** حساب أمثال الزوج الثاني (U_2 , V_2) وما بعده :**

للحصول على أمثال الزوج الثاني (U_2 , V_2) التي تجعل معامل الارتباط القانوني ρ_2 ثاني أكبر معامل ارتباط، نأخذ القيمة الذاتية الثانية λ_2^2 ونعوضها في المعادلة (3-38)، فنحصل منها على جملة معادلات غير مستقلة، ويكون لها لانهاية من الحلول المقبولة. وحتى نستطيع الحصول على حل وحيد للشعاع a الذي يجعل معامل الارتباط ρ_2 أكبر ما يمكن، نستفيد من الشرطين المشابهين للشرطين السابقين المفروضين على V_1 , U_1 وهما:

$$var(U_2) = 1 \qquad var(V_2) = 1 \qquad (74-3)$$

وهنا يشترط أيضاً أن تكون المركبات القانونية في الزوج (U_2 , V_2) غير مرتبطة خطياً مع المركبات القانونية للزوج الأول V_1 , U_1 . أي أن يكون :

$$\begin{matrix} cov(U_1, U_2) = 0 & cov(V_1, V_2) = 0 \\ cov(U_1, V_2) = 0 & cov(V_1, U_2) = 0 \end{matrix} \qquad (75-3)$$

وبصورة عامة للحصول على أمثال الزوج k وهو $(U_k\ , V_k)$ التي تجعل معامل الارتباط القانوني ρ_k يأخذ قيمة أكبر من المعاملات التالية له، نقوم بتعويض القيمة الذاتية λ_k^2 في المعادلة (3-28) ومنها نحصل على جملة معادلات خطية غير مستقلة. وحتى نستطيع الحصول على حل وحيد للشعاعين b, a نستعين بالشرطين السابقين وهما:

$$var\ (U_k) = 1 \qquad var(V_k) = 1 \qquad (76-3)$$

ونضيف إليهما شروط أن تكون المركبات القانونية $(U_k\ , V_k)$ غير مرتبطة بالمركبات القانونية السابقة لها وهي:

$$\begin{aligned} cov(U_k, V_\ell) &= 0 \quad ,, k \neq \ell \\ cov(U_k, U_\ell) &= 0 \quad ,, k \neq \ell \\ cov(V_k, V_\ell) &= 0 \quad ,, k \neq \ell \\ cov(V_k, U_\ell) &= 0 \quad ,, k \neq \ell \end{aligned} \qquad (77-3)$$

وذلك من أجل قيم $k\ , \ell = 1,2,3 \ldots \ldots p$

وهذا يعني أن معاملات ارتباط المركبين $(U_k\ , V_k)$ تكون معدومة مع جميع المركبات التي تسبقها .

وهكذا نتابع حتى نحصل على أمثال جميع الأزواج الممكنة، ويكون كل زوج مستقل عن الأزواج الأخرى السابقة له، ويمكن توضيح ذلك من خلال عرض التباينات والتباينات المشتركة لجميع الأزواج الممكنة في جدول يتناسب مع الشروط (3–76) و(3-77) كما يلي:

جدول (3-6): التباينات والتباينات المشتركة للمركبات للأزواج الثلاثة الأولى (V_ℓ, U_k) :

المركبات	V_1	V_2	V_3	U_1	U_2	U_3
U_1	ρ_1	0	0	1	0	0
U_2	0	ρ_2	0	0	1	0
U_3	0	0	ρ_3	0	0	1
V_1	1	0	0	ρ_1	0	0
V_2	0	1	0	0	ρ_2	0
V_3	0	0	1	0	0	ρ_3

ملاحظة: إن الشروط (3-77) هي التي تجعل عدد الأزواج القانونية المرتبطة $(U_k\ , V_k)$ يساوي p زوجاً فقط أو يساوي s أصغر العددين (p, q) لأنها تجعل بقية الأزواج معدومة الارتباط .

3–5 اختبار معنوية القيم الذاتية λ_k^2 أو معاملات الارتباط القانونية $\rho_1\ \rho_2 \ldots\ldots \rho_p$

يوجد عدة مؤشرات لاختبار معنوية ρ_k، ونذكر هنا أهم هذه المؤشرات وشروط استخدامها ضمن فرضية العدم التالية: لا توجد علاقة معنوية بين مجموعة المتحولات X ومجموعة المتحولات Y .

أي أن $H_0 = \rho_1 = \rho_2 = \rho_3 = \rho_s = 0$ مقابل أن يكون أحدها غير معدوم $H_1 : \rho_k \neq 0$

1- **اختبار Wilk's – lambda** ويحسب من العلاقة التالية:

$$\bigwedge(\ell) = \prod_{k=\ell}^{s}(1-\rho_k^2) \qquad \mathbf{(78-3)}$$

وهو يخضع للتوزيع F بدرجتي حرية $(p*q\,,n-1-q)$، ويستخدم للدلالة على معنوية جملة معاملات الارتباط المتتالية اعتبار من ρ_ℓ حتى ρ_S . وكلما كانت قيمة هذا المؤشر صغيرة كانت المعنوية كبيرة (وذلك بعكس المؤشرات الأخرى)، وتؤخذ القيمة الحرجة لهλ_S من جدولF المقابلة لدرجتي الحرية $(p*q, n-1-q)$، فإذا كانت قيمة $\wedge(\ell)$ المحسوبة أصغر من القيمة الحرجة λ_S تكون $P \leq \propto$ ، فإننا(وعلى غير العادة) نرفض فرضية العدم H_0 التي تقول أنه لا توجد علاقة معنوية بين مجموعة المتحولات X ومجموعة المتحولات Y .

2- **اختبار $\mathcal{X}^2$ أو Bartlitt**: وهو يحسب اعتماداً على قيمة $\wedge(\ell)$ المعرفة في (3-81) ويعتبر بديلاً عنه ويحسب من العلاقة :

$$\mathcal{X}_{(\ell)}^2 = -\left[n-\frac{1}{2}(p+q+3)\right]ln\bigwedge(\ell) = -\left[n-\frac{1}{2}(p+q+3)\right]\sum_{k=\ell}^{s} ln\left(1-\rho_k^2\right) \qquad \mathbf{(79-3)}$$

وهو يخضع لتوزيع $\mathcal{X}^2$ ذي درجة حرية $df=(p-k+1)(q-k+1)$، فإذا كانت قيمة $\mathcal{X}_{(\ell)}^2$ المحسوبة أكبر من $\mathcal{X}_\alpha^2$ أو $\rho \leq\propto$ نرفض الفرضية H_0، وهو يقيس معنوية جملة المعاملات القانونية من ρ_ℓ حتى ρ_S .

3- **اختبار F**: وهو يعتمد أيضاً على قيمة $\wedge(\ell)$ المعرفة من (78) ويحسب من العلاقة :

$$F(\ell) = \frac{1-\wedge_\ell^{\frac{1}{t}}*df_2}{\wedge_\ell^{\frac{1}{t}}*df_1} \qquad (80-3)$$

وهو يخضع لتوزيع F بدرجتي حرية:

$$df_1 = (p-k+1)(q-k+1)$$

$$df_2 = w*t-\frac{1}{2}\big((p-k+1)\big)\big((q-k+1)\big)+1$$

حيث أن t و w يساويان: $t=\sqrt{\frac{p^2*q^2-4}{p^2+q^2-5}}$ $\qquad w = n-\frac{1}{2}(p+q+3)$

فإذا كانت قيمة $F(\ell)$ المحسوبة أكبر من $F(\alpha)$ أو كانت $\rho \leq\propto$ نرفض فرضية العدم H_0

4- **اختبار Hotelling - lawley trace** : وتحسب قيمته من العلاقة:

$$U(\ell) = \sum_{k=\ell}^{s}\left(\frac{\rho_k^2}{1-\rho_k^2}\right) \qquad \mathbf{(81-3)}$$

ثم نقوم بمقارنة هذه القيمة مع القيمة الحرجة التي تحسب من العلاقة:

$$U_\alpha = U(\ell)\frac{n-q-1}{q}$$

فإذا كانت $U(\ell) \geq U_a$ أو $\rho \leq a$ نرفض فرضية العدم H_0

5-**اختبار Pillia's trace** : وتحسب قيمته من العلاقة التالية:

$$V(\ell) = \sum_{k=\ell}^{s} \rho_k^2 \qquad (82-3)$$

وهو يخضع لتوزيع $\mathcal{X}^2$ بدرجة حرية $(s-\ell)$.

فإذا كانت $V(\ell) \geq V_\alpha$ أو $\rho \leq a$ نرفض فرضية العدم H_0

ملاحظة: إن نتيجة تطبيق هذه الاختبارات على القيم الذاتية λ_k^2 أو على المعاملات ρ_k^2 تفرز لنا تلك المعاملات إلى معاملات معنوية (تقابل $p \leq \propto = 0.05$) ومعاملاته غير معنونه (تقابل $p > 0.05$)، وعندها فإن التحليل الرياضي والإحصائي يقتصر على المعاملات المعنوية وعلى الازدواج القانونية المقابلة لها . وتكون المعاملات المعنوية هي الأولى في جدول الاختبار ويكون المعامل الأول هو أهم المعاملات ثم يليه المعامل الثاني ثم الثالث (إن وجد ..الخ). وإن قيمة المعامل الأول تعبر عن مصداقية ذلك الارتباط (راجع المثال العام المحلول في نهاية هذا الفصل).

6-3 خلاصة التحليل الرياضي:

لحساب الأمثال b_i, a_i في الشعاعين b, a المصاحبين للتركيبين V, U المعرفين بالعلاقتين (3-6) و (3-7) يمكننا أن نتبع إحدى الطرق الحسابية التالية:

- **الطريقة الأولى:** وهي تطبق عندما يكون $p \leq q$ (أي عندما يكون عدد المتحولات في X أقل أو يساوي عدد المتحولات في Y. وذلك لاختصار الحسابات المعقدة اللازمة لذلك)، وتهدف إلى حساب الشعاع a أولا ثم حساب الشعاع b، وتتلخص بما يلي:

1-نقوم بحساب القيم الذاتية λ_k^2 من المعادلة الذاتية:

$$C_{xx}^{-1} * C_{xy} * C_{yy}^{-1} * C_{yx} * a = \lambda^2 * a \qquad (83-3)$$

وذلك باستخدام محدد المعادلة الذاتية:

$$[A - I * \lambda^2] * a = 0 \quad , \quad |A - I * \lambda^2| = 0 \qquad (84-3)$$

ومنها نحصل على القيم الذاتية المرتبة: $\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \dots\dots \lambda_p^2 \geq 0$ ، وعلى المعاملات ρ_k^2 و ρ_k.

2-نقوم بحساب عناصر الشعاع الأول a المقابل للقيمة الأولى λ_1^2 من المعادلة (3-83) وذلك بتعويض قيمة λ_1^2 فيها وحساب عناصر الشعاع a^* ، ثم نقوم بتعديلها لتحقق الشرط المفروض عليها وهو:

$$var\ (U) = a' * C_{xx} * a = 1 \qquad (85-3)$$

3-نقوم بحساب عناصر الشعاع الأول b من عناصر الشعاع a وذلك بتطبيق العلاقة:

$$b^* = \frac{C_{yy} * C_{yx}}{\lambda} * a^* \qquad (86-3)$$

ثم نقوم بتعديلها لتحقق الشرط (3-26) المفروض عليها وهو $var\ (V) = b' * C_{yy} * b = 1$

4-نعوض عناصر الشعاعين b^*, a^* في التركيبين (3-6),(3-7) فنحصل على تركيبين محددين لهما هي:

$$\begin{aligned} \widetilde{U}_1 &= a_1^* x_1 + a_2^* x_2 + \cdots a_p^* x_p \\ \widetilde{V}_1 &= b_1^* y_1 + b_2^* y_2 + \cdots b_q^* y_q \end{aligned} \qquad (87-3)$$

5-نقوم بحساب القيم النظرية للزوج $(\widetilde{U}_1, \widetilde{V}_1)$ ، ثم نرسم شكله البياني ومنهما نحسب معامل الارتباط القانوني الأول ρ_1 لتحديد إشارته .

6-نقوم بإجراء جميع الحسابات اللازمة لتحليل الكفاءة (كما سنرى لاحقاً) .

- **الطريقة الثانية:** وهي تطبق عندما يكون $q \le p$ (أي عندما يكون عدد المتحولات في Y أقل أو يساوي عدد المتحولات في X وذلك لاختصار الحسابات اللازمة لذلك) .

وتهدف إلى حساب الشعاع b أولاً ثم a وتتلخص بما يلي:

1-نقوم بحساب القيم الذاتية λ_i^2 من المعادلة الذاتية :

$$C_{yy}^{-1} * C_{yx} * C_{xx}^{-1} * C_{xy} * b = \lambda^2 * b \qquad (88-3)$$

وذلك باستخدام محدد المعادلة الذاتية:

$$[B - I\lambda^2] * b = 0 \qquad , \qquad |B - I\lambda^2| = 0 \qquad (89-3)$$

فنحصل على الجذور λ_k وعلى المعاملات ρ_k^2 و ρ_k .

2-نقوم بحساب عناصر الشعاع الأول b المقابل للقيمة الأولى λ_1^2 من المعادلة (3-88)، وذلك بتعويض λ_1^2 فيها وحساب عناصر الشعاع الأول b^* ثم تعديله حتى يحقق الشرط المفروض عليه وهو:

$$var(V) = b' * C_{yy} * b = 1 \qquad (90-3)$$

3-نقوم بحساب الشعاع الأول a^* بعد حساب b^* من العلاقة :

$$a^* = \frac{C_{xx}^{-1} * C_{xy}}{\lambda} * b^* \qquad (91-3)$$

ثم نقوم بتعديله ليحقق الشرط (3-25) وهو $var\,(U) = a' * C_{xx} * a = 1$ فنحصل على الشعاع a^*

4-نعوض عناصر الشعاعين a^* وb^* في التركيبين (3-7) و(3-6)، فنحصل على نفس التركيبين المحددين في الطريقة الأولى أو على تركيبين متناسبين معهما وهما:

$$\begin{aligned} \widetilde{U}_1 &= a_1^* X_1 + a_2^* X_2 + \cdots a_p^* X_p \\ \widetilde{V}_1 &= b_1^* Y_1 + b_2^* Y_2 + \cdots b_q^* \mathrm{Y}_q \end{aligned} \qquad (92-3)$$

5-نقوم بحساب القيم النظرية للزوج $(\widetilde{U}_1, \widetilde{V}_1)$ ثم نرسم شكله البياني ونحسب معامل الارتباط القانوني الأول ρ_1 لتحديد إشارته .

6-نقوم بإجراء جميع الحسابات اللاحقة .

- **الطريقة الثالثة:** وهي تطبق عندما يكون p = q كما يلي:

1−نقوم بحساب القيم الذاتية λ_k^2 من مفكوك المحدد :

$$\left|C_{xx}^{-1} * C_{xy} * C_{yy}^{-1} * C_{yx} - \lambda^2 * I\right| = 0 \qquad (93-3)$$

2−ثم نقوم بترتيبها تنازلياً كما يلي:

$$\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \dots \dots \lambda_p^2 \geq 0 \qquad (94-3)$$

وهي عبارة عن معاملات التحديد $\rho_1^2 \geq \rho_2^2 \geq \rho_3^2 \dots \dots \rho_p^2 \geq 0$ المقابلة للأزواج $(U_k\ V_k)$ المتنازلة، ثم نحسب معاملات الارتباط القانونية $\rho_k = \pm\sqrt{\lambda_k^2}$ ويتم تحديد إشارتها حسب طبيعة العلاقة بين U_k و V_k.

3−بعد حساب القيم الذاتية λ_k^2 نقوم بحساب الأشعة الذاتية b, a المقابلة على التوالي لكل قيمة ذاتية λ_k^2 من المعادلتين الذاتيتين على الترتيب :

$$C_{xx} * C_{xy} * C_{yy} * C_{yx} a = \lambda^2 * a \qquad : a \text{ للشعاع } \quad (95-3)$$

$$C_{yy}^{-1} * C_{yx} * C_{xx}^{-1} * C_{xy} * b = \lambda^2 * b \qquad : b \text{ للشعاع } \quad (96-3)$$

ثم نقوم بتعديلهما ليحققان الشرطين (3−25) و(3−26) المفروضين عليهما وبذلك نحصل على الشعاعين b^* , a^* المقابلين لكل قيمة ذاتية λ_k^2، وبالتالي نحصل على التركيبات الخطية :

$$U_k = a_{k1}^* X_1 + a_{k2}^* X_2 + \cdots a_{kp}^* X_p \qquad (97-3)$$

$$V_k = b_{k1}^* Y_1 + b_{k2}^* Y_2 + \cdots b_{kp}^* Y_q \qquad (98-3)$$

4−نقوم بحساب القيم النظرية للأزواج $\left(\widetilde{U}_k\ \widetilde{V}_k\right)$ ونرسم شكله البياني ونحسب معامل الارتباط القانوني له ρ_k لتحديد إشارته .

5−ثم نقوم بإجراء جميع الحسابات اللاحقة .

- **الطريقة الرابعة:**(وتحويل (شوليسكي sholesky):

مما سبق لقد لاحظنا أنه يمكننا حساب الشعاعين b, a من خلال استخدام المعادلتين (3-34),(3-62) أي من خلال المعادلتين التاليتين :

$$\left[C_{xy} * C_{yy}^{-1} * C_{yx} - \lambda^2 C_{xx}\right] * a = 0 \qquad (99-3)$$

$$\left[C_{yx} * C_{xx}^{-1} * C_{xy} - \lambda^2 C_{yy}\right] * b = 0 \qquad (100-3)$$

ولكن تطبيق هاتين المعادلتين يتطلب الكثير من الحسابات المعقدة والمتكررة، وهذا ما يضخم حجم الأخطاء خلال الحسابات ويجعلها تتجاوز مستويات الدقة المطلوبة، ويؤثر على دقة النتائج في تقدير الشعاعين b,a، كما يخل في تناظر المصفوفة الذاتية. لذلك قبل أن نبدأ بحل هاتين المعادلتين بالطرق المعروفة ، سنقوم بتحويل مسألة القيم الذاتية العامة إلى مسألة القيمة الشاذة العامة. وذلك باستخدام تحويل (شوليسكي sholesky) للمصفوفتين المتناظرتين C_{xx} و C_{yy} كما يلي:

إن تحويل (شوليسكي sholesky) يمكن أن يطبق على المصفوفات المتناظرة والمحددة إيجابياً، مثل مصفوفات التباين المشترك C_{xx} و C_{yy} ويكون مستقراً جداً من الناحية العددية حسب مقالة الباحثين

(Golub & Vonloan 1996)، وهو يقتضى أن نفرع المصفوفتين C_{xx} و C_{yy} إلى جداء مصفوفتين مثلثتين كما يلي:

$$C_{xx} = U'_x * U_x \qquad (101-3)$$

$$C_{yy} = V'_y * V_y \qquad (102-3)$$

حيث أن U_x هي المصفوفة المثلثية العليا للمصفوفة C_{xx} فتكون U'_x هي السفلى .

وأن V_y هي المصفوفة المثلثية العليا للمصفوفة C_{yy} فتكون V'_y هي السفلى .

وأن كل منهما ذات عناصر قطرية موجبة

والآن نفترض أن K هي مقلوب U_x أي أن: $K = U_x^{-1}$

كما نفترض أن L هي مقلوب V_y أي أن: $L = V_y^{-1}$

وعندها نجد أن:

$$C_{xx}^{-1} = U_x^{-1}(U_x^{-1})' = K * K' \qquad (103-3)$$

$$C_{yy}^{-1} = V_y^{-1}\left(V_y^{-1}\right)' = L * L' \qquad (104-3)$$

والآن لنأخذ C_{xx}^{-1} و C_{yy} ونعوضهما في المعادلة (3-100) فنجد أن:

$$\left[C_{yx} * K * K' * C_{xy} - \lambda^2 * V'_y * V_y\right] b = 0$$

$$\left[\left(K'C_{xy}\right)' * \left(K' * C_{xy}\right) - \lambda^2 V'_y V_y\right] b = 0 \qquad (105-3)$$

وإذا رمزنا بـ $A = K'C_{xy}$ و $B = V_y$ فإننا نحصل على معادلة من الشكل التالي :

$$(A' * A - \rho^2 B' * B)b = 0 \qquad (106-3)$$

والتي يمكن حلها بدقة عالية بطريقة القيمة الشاذة العامة لفرعي المصفوفتين A , B وبدون أن نحسب الجدائين للمصفوفتين $A' * A$ و $B' * B$ (colub & vanloan 1996+ weenink 1999)

وهكذا سنحصل على المعادلة (3-95) من جداء مصفوفة من فرعي تشوليسكي K' في المصفوفة C_{xy} وكذلك من جداء المصفوفتين المثلثتين V'_y في V_y .

وهذا يسمح لنا في تقدير أفضل للقيم الذاتية مما كان تقديرها من المعادلتين (3-99) , (3-100)، وإن الجذور التربيعية للقيم الذاتية المحسوبة من المعادلة (3-105) تساوي معاملات الارتباط القانوني ρ .

أما الأشعة الذاتية b فتعطينا كيف يتم توليف العمود b لنحصل على الارتباط القانوني الأمثل .

وكذلك الأمر بالنسبة للمعادلة (3-99) يمكن أن نحصل منها على المعادلة التالية:

$$\left[\left(L'C_{yx}\right)' * \left(LC_{yx}\right) - \lambda^2 U'_x U_x\right] a = 0 \qquad (107-3)$$

وهي تتمتع بنفس الخواص للمعادلة (3-96) وتكتبها أيضاً على الشكل :

$$[A' * A - \rho^2 B' * B]a = 0 \qquad (108-3)$$

ومن المعادلتين (3-107) , (3-105) يمكننا حساب الأشعة الذاتية a , b .

- **الطريقة الخامسة:** وهي تعتمد على مصفوفتي البيانات الأولية المركزية(الممعيرة) للمتحولات X وللمتحولات Y والتي نرمز لهما بـ A_x , A_y على الترتيب:

حيث أن: $a_{xij} = \left(X_{ij} - \bar{X}_j\right)$ $\quad a_{yij} = \left(Y_{ij} - \bar{Y}_j\right)$

ونبدأ بتفريع هاتين المصفوفتين بطريقة تفريعات القيمة الشاذة كما يلي :

$$A_x = U_x * D_x * V_x' \qquad (109-3)$$

$$A_y = U_y * D_y * V_y' \qquad (110-3)$$

حيث أن U_x مصفوفة مثلثية عليا و V_y مصفوفة مثلثية سفلى أما D_x فهي مصفوفة قطرية .

ونستخدم هاتين التفريعتين لحساب مصفوفات التباينات المشتركة كما يلي:

$$C_{xx} = A_x' * A_x = (V_x * D_x' * U_x') * U_x * D_x * V_x' = V_x * D_x^2 * V_x' \qquad (111-3)$$

$$C_{yy} = A_y' * A_y = \left(V_y * D_y' * U_y'\right) * U_y * D_y * V_y' = V_y * D_y^2 * V_y' \qquad (112-3)$$

$$C_{xy} = A_x' * A_y = (V_x * D_x' * U_x') * U_y * D_y * V_y' = V_x * D'_x * U_x' * U_y * D_y * V_y' \qquad (\mathbf{113-3})$$

وذلك بالاستفادة من خاصة التعامد حيث أن:

$$\begin{matrix} U_x' * U_x = I & U_y' * U_y = I \\ V_x' * V_x = I & V_y' * V_y = I \end{matrix} \qquad (114-3)$$

وهكذا نجد أن:

$$C_{xx}^{-1} = V_x * D_x^{-2} * V_x' \quad \text{و} \quad C_{yy}^{-1} = V_y * D_x^{-2} * V_y' \qquad (115-3)$$

وبتعويض ذلك في المعادلتين (3-99) , (3-100) نحصل من (3-100) على أن:

$$\left[V_y * D_y * U_y' * U_x * U_x' * U_y * D_y * V_y' - \rho^2 * V_y * D^2 * V_y'\right]b = 0 \qquad (116-3)$$

وبعد ضرب العلاقة (3-116) من اليسار بالمصفوفة $D_y^{-1}V_y'$ نحصل على :

$$\left[U_y' * U_x * U_x' * U_y * D_y * V_y' - \rho^2 * D_y * V_y'\right]b = 0 \qquad (117-3)$$

والتي يمكن إعادة كتابتها كما يلي:

$$\left[\left(U_x' * U_y\right)' * \left(U_x' * U_y\right) - \rho^2 I\right] D_y * V_y' * b = 0 \qquad (118-3)$$

وهي معادلة من الشكل :

$$[A' * A - \rho^2 I]X = 0 \qquad (119-3)$$

والتي يمكن حلها بسهولة بواسطة تعويض فرع القيمة الشاذة للمصفوفة A . حيث نقوم بتعويض فرع القيمة الشاذة $\left(U_x' * U_y = U * D * V'\right)$ في المعادلة (3-118) ونتوصل بعد الإصلاحات إلى أن:

$$(D^2 - \rho^2 I)V' * D_y * V_y' * b = 0 \qquad (120-3)$$

وإن القيم الذاتية للمعادلة (3–120) تساوي D^2 أما الأشعة الذاتية لها فنحصل عليها من أعمدة المصفوفة $V_y' * D_y^{-1} * V$

وبطريقة مشابهة نحصل من المعادلة (3-99) على أن:

$$(D^2 - \rho^2 I)U' * D_x * V_x' * a = 0 \qquad (121-3)$$

وسيكون لها نفس القيم الذاتية D^2 أما الأشعة الذاتية لها فنحصل عليها من أعمدة المصفوفة

$U * D_x^{-1} * V_x'$. وبذلك نكون قد توصلنا إلى عدد الجداءات الضرورية للحصول على القيم الذاتية. ولم نعد بحاجة لحساب الجداءات المذكورة في العلاقة (3–99) بل يكتفي بفرعين للقيمة الشاذة ومصفوفة الجداء $\left(U_x' * U_y\right)$، وإن الجداء الأخير هو مستقر جداً، وذلك لأن الأعمدة في كلتا المصفوفتين متعامدة فيما بينها [David weenink 2003.p.83] .

3-7 الارتباط القانوني المعياري:

وهو حالة خاصة من الارتباط القانوني العام، ويتم استخدامه للتخلص من المشكلات الحسابية التي تنجم عن واحدات القياس المختلفة للمتحولات المستقلة X وللمتحولات التابعة Y، وهو يطبق على المتحولات المعيارية المستخلصة من المتحولات الأصلية (الخام) في كلتا المجموعتين X , Y .

لذلك نقوم بتحويل متحولات المجموعة X إلى متحولات معيارية وفق العلاقة :

$$Z_{xi} = \frac{X_i - \bar{X}_i}{\sigma_{xi}} \qquad (i = 1,2,3, \dots p) \qquad (122-3)$$

فيكون لدينا حكماً إن المتوسط والتباين يساويان:

$$\bar{Z}_{xi} = 0 \qquad var(Z_{xi}) = 1 \qquad (123-3)$$

كما نقوم بتحويل متحولات المجموعة Y إلى متحولات معيارية وفق العلاقة:

$$Z_{yj} = \frac{Y_j - \bar{Y}_j}{\sigma_{yj}} \qquad (j = 1,2,3, \dots q) \qquad (124-3)$$

ويكون لدينا حكماً أن المتوسط والتباين يساويان:

$$\bar{Z}_{yj} = 0 \qquad var\left(Z_{yj}\right) = 1 \qquad (125-3)$$

وبذلك نحصل على مجموعتين من المتحولات المعيارية هما:

$$\begin{aligned} Z_x &: Z_{x1}, Z_{x2}, Z_{x3}, \dots \dots, Z_{xp} \\ Z_y &: Z_{y1}, Z_{y2}, Z_{y3}, \dots \dots, Z_{yp} \end{aligned} \qquad (126-3)$$

إن التعامل مع هذه المتحولات المعيارية يعتبر أكثر سهولة وأفضل دقة من المتحولات الأصلية Y, X، لذلك سنقوم فيما يلي بتطبيق إجراءات الارتباط القانوني على متحولات هاتين المجموعتين ثم نرجع إلى المتحولات الأصلية Y , X ونستخلص منها النتائج الممكنة .

وبطريقة مشابهة لما فعلناه مع المتحولات الأصلية Y , X ، نشكل تركيبين خطيين لهذه المتحولات المعيارية Z_y, Z_x كما يلي:

$$U_{zx} = e_1 Z_{x1} + e_2 Z_{x2} + e_3 Z_{x3} + \dots + e_p Z_{xp} = \grave{e} * Z_x \qquad (127-3)$$

$$V_{zy} = f_1 Z_{y1} + f_2 Z_{y2} + f_3 Z_{y3} + \dots + f_q Z_{yq} = f' * Z_y \qquad (128-3)$$

حيث أن e و f شعاعان عدديان مجهولان، وأن عدد متحولات X هو p وعدد متحولات Yهو q ، ولكن بما أن معامل الارتباط ρ للمركبين (U , V) المعرفين في العلاقتين (3-6),(3-7) لا يتأثر بواحدات القياس وبمركز المحاور، فإنه لا يتأثر بهذه التحويلات المعيارية ويبقى مساوياً لنفسه ومساوياً لمعامل

الارتباط القانوني الجديد للمركبين (U_{zx}, V_{zy})، وبناءً عليه يمكننا حساب قيمته من المركبين (U_{zx}, V_{zy}) وفق العلاقة التالية:

$$\rho(U_{zx}, V_{zy}) = \frac{cov(U_{zx}, V_{zy})}{\sqrt{var(U_{zx})} * \sqrt{var(V_{zy})}} \qquad (129-3)$$

وبنفس الأسلوب يمكننا البرهان على أن:

$$var(U_{zx}) = cov(e' * Z_x) = e' * \ cov(Z_x) * e \qquad (130-3)$$

حيث أن $cov(Z_x)$ هي مصفوفة التباينات والتباينات المشتركة للمجموعة Z_x وإن كل عنصر فيها يساوي التباين المشترك لمتحولين من Z_x ويساوي:

$$\begin{aligned} cov(Z_{xi}, Z_{xj}) &= E\left[(Z_{xi} - 0)(Z_{xj} - 0)\right] \\ &= E\left[\left(\frac{X_i - \bar{X}_i}{\sigma_{xi}}\right)\left(\frac{X_j - \bar{X}_j}{\sigma_{xj}}\right)\right] = r_{x_i x_j} \end{aligned} \qquad (\mathbf{131-3})$$

أي أن كل عنصر في المصفوفة $cov(Z_x)$ هو عبارة عن معامل الارتباط بين المتحولين الأصليين X_i, X_j المقابلين له، أي أن المصفوفة $cov(Z_x)$ ماهي إلا مصفوفة معاملات الارتباط الزوجية لمتحولات المجموعة X الأصلية والتي رمزنا لها بـ R_{xx}، لأنه في الحقيقة إن عملية تحويل المتحولات X والمتحولات Y إلى متحولات معيارية Z_x, Z_y يؤدي إلى تحويل مصفوفات التباينات والتباينات المشتركة $C_{xx}, C_{yy}, C_{xy}, C_{yx}$ إلى المصفوفات الارتباطية $R_{xx}, R_{yy}, R_{xy}, R_{yx}$ على الترتيب:

وبناءً على ذلك نجد أن العلاقة (3-130) تأخذ الشكل التالي:

$$var(U_{zx}) = e' * R_{xx} * e \qquad (132-3)$$

وبطريقة مشابهة نجد أن:

$$var(V_{zy}) = f' * R_{yy} * f \qquad (133-3)$$

كما يمكننا أن نجد أن:

$$cov(U_{zx}, V_{zy}) = e' * R_{xy} * f = f' * R_{yx} * e \qquad (134-3)$$

وبتعويض (3-132) ,(3-133) ,(3-134) في العلاقة (3-129) نجد أن معامل الارتباط القانوني ρ يصبح في هذه الحالة مساوياً لما يلي:

$$\rho(U_{zx}, V_{zy}) = \frac{e' * R_{xy} * f}{\sqrt{e' * R_{xx} * e} * \sqrt{f' * R_{yy} * f}} \qquad (135-3)$$

وحتى نحصل على التركيب الأول أو الزوج الأول (U_{1zx}, V_{1zx}) المقابل لـ ρ_1 علينا الآن أن نقوم بحساب عناصر الشعاعين e , f ، بحيث نجعل قيمة المعامل $\rho(U_{zx}, V_{zy})$ أكبر ما يمكن، وحتى نستطيع حساب عناصر الشعاعين e, f بشكل محدد نضع عليهما الشرطين المشابهين للشرطين (3-25),(3-26) كما يلي:

$$var(U_{1zx}) = e' * R_{xx} * e = 1 \qquad (136-3)$$

$$var(V_{1zy}) = f' * R_{yy} * f = 1 \qquad (137-3)$$

وبذلك نجد أن العلاقة (3-135) تأخذ الشكل التالي:

$$\rho(U_{1zx}, V_{1zy}) = e' * R_{xy} * f = f' * R_{yx} * e \qquad (138-3)$$

وللحصول على أكبر قيمة للمعامل ρ في العلاقة (3-138) نشكل منها مع الشرطين ,(3-137) (3-136) تابع لاغرانج كما يلي:

$$L = e' * R_{xy} * f - \frac{\lambda_x}{2}(e' * R_{xx} * e - 1) - \frac{\lambda_y}{2}(f' * R_{yy} * f - 1) \qquad (\mathbf{139-3})$$

حيث : λ_y , λ_x هما عددان وسيطان يسميان بمضروبي لاغرانج .

نشتق التابع L بالنسبة e ثم بالنسبة لـ f ونضع هذين المشتقين مساويين للصفر (للحصول على القيمة العظمى لـ L) فنجد أن(*):

$$\frac{\partial L}{\partial e} = R_{xy} * f - \lambda_x R_{xx} * e = 0 \qquad (140-3)$$

$$\frac{\partial L}{\partial f} = R_{yx} * e - \lambda_y R_{yy} * f = 0 \qquad (141-3)$$

نضرب الأولى بـ e' والثانية بـ f' فنحصل على المعادلتين:

$$e' * R_{xy} * f - \lambda_x \quad e' * R_{xx} * e = 0 \qquad (142-3)$$

$$f' * R_{yx} * e - \lambda_y \quad f' * R_{yy} * f = 0 \qquad (143-3)$$

ثم نطرح الثانية من الأولى فنحصل على أن:

$$e' * R_{xy} * f - \lambda_x e' * R_{xx} * e - f' * R_{yx} * e + \lambda_y f' * R_{yy} * f = 0 \qquad (\mathbf{144-3})$$

وبما أنه لدينا من (3-136) ,(3-137) , (3-138) أن:

$$e' * R_{xy} * f = f' * R_{yx} * e \quad \text{و} \quad e' * R_{xx} * e = 1\,, \;\; f' * R_{yy} * f = 1$$

فإننا نحصل من (3-144) على أن:

$$-\lambda_x + \lambda_y = 0 \qquad (145-3)$$

أي أن $\lambda_x = \lambda_y = \lambda$، لذلك نعوض قيمة λ الموحدة في العلاقتين (3−140) و(3−141)، ونفترض أن المصفوفة المقلوبة R_{yy}^{-1} موجودة ثم نضرب العلاقة (3-141) من اليسار بـ R_{yy}^{-1} فنحصل على :

$$R_{yy}^{-1} * R_{yx} * e = \lambda R_{yy}^{-1} * R_{yy} * f = \lambda * f \qquad (146-3)$$

ومنها نحصل على الشعاع f بدلالة الشعاع e من العلاقة:

$$f = \frac{R_{yy}^{-1} * R_{yx}}{\lambda} * e \qquad (\mathbf{147-3})$$

وبتعويض قيمة f في العلاقة الأولى (3-140) نحصل على أن:

(*) إن الاشتقاق تم حسب قواعد اشتقاق المصفوفات وهي:

$$\frac{\partial \grave{c}X}{\partial X} = C \qquad , \qquad \frac{\partial(\grave{X}BX)}{\partial X} = 2BX$$

$$R_{xy} * \frac{R_{yy}^{-1} * R_{yx}}{\lambda} * e = \lambda * R_{xx} * e$$

وبإصلاحها نحصل على أن:

$$R_{xy} * R_{yy}^{-1} * R_{yx} * e = \lambda^2 * R_{xx} * e \qquad (148-3)$$

وبفرض أن المصفوفة المقلوبة R_{xx}^{-1} موجودة. نضرب (3-148) من اليسار بالمصفوفة R_{xx}^{-1} فنحصل على أن:

$$R_{xx}^{-1} * R_{xy} * R_{yy}^{-1} * R_{yx} * e = \lambda^2 * e \qquad (149-3)$$

وهي المعادلة النموذجية للقيم الذاتية للمصفوفة R التي ترمز إلى :

$$R_{p*p} = R_{xx}^{-1} * R_{xy} * R_{yy}^{-1} * R_{yx} \qquad (150-3)$$

وأخيراً نكتب (3-149) كما يلي:

$$[R - \lambda^2 I]e = 0 \qquad (151-3)$$

وحتى نتجنب الحصول على الحل التافه وهو $e = 0$ نفترض أن المصفوفة $[R - \lambda^2 I]$ مصفوفة شاذة، وهذا يعني أن قيمة محددها يجب أن يساوي الصفر . أي أن :

$$|R - \lambda^2 I| = 0 \qquad (152-3)$$

وبما أن المحدد (3-152) من المرتبة $p * p$ فإن مفكوكه سيعطينا كثير حدود من المرتبة p بالنسبة للمقادير λ^2 من الشكل :

$$(\lambda^2)^p + c_1(\lambda^2)^{p-1} + c_2(\lambda^2)^{p-2} + \cdots + c_p = 0 \qquad (153-3)$$

نقوم بإيجاد جذور هذه المعادلة بالنسبة لـ (λ^2) فنحصل على p جذراً لها، تسمى بالجذور الكامنة للمصفوفة R أو تسمى بالقيم الذاتية (eigenvalue) للمصفوفة R، ثم نقوم بترتيب هذه الجذور تنازلياً كما يلي:

$$\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \geq \cdots \geq \lambda_{1p}^2 \geq 0 \qquad (154-3)$$

وهي تقابل مربعات معاملات الارتباط القانونية بين الأزواج $\left(U_{zx}, V_{zy}\right)$ المرتبة تنازلياً :

$$\rho_1^2 \geq \rho_2^2 \geq \rho_3^2 \geq \cdots \geq \rho_p^2 \geq 0 \qquad (155-3)$$

وحتى نحصل على الشعاع الذاتي المعياري الأول e المصاحب للتركيب U_{1zx} نأخذ الجذر الأول λ_1^2 ونقوم بتعويضه في المعادلة (3-151) ثم نجري الحسابات اللازمة فنحصل على جملة من المعادلات غير المستقلة تعطينا لا نهاية من الحلول لـ e، وحتى نستطيع الاستفادة منها في حساب عناصر الشعاع الذاتي الأول e المقابل للقيمة الذاتية λ^2 نضيف إليها معادلة الشرط المفروض على e في المعادلة (3-136) ثم نقوم بحل جملة المعادلات الجديدة فنحصل على حل وحيد لـ e وهو يعطينا العناصر الأساسية لذلك الشعاع $e^*(e_1^* e_2^* e_3^* \ldots\ldots)$. ويمكننا اتباع طريقة أخرى وهي تتلخص بأن نأخذ أي حل مقبول e للجملة في (3-151) ونضربه بمقدار ثابت k ثم نجعله يخضع للشرط (3-136) المفروض عليه فنحصل على أن :

$$var(U_{zx}) = ke' * R_{xx} * ke = k^2 e' * R_{xx} * e = 1 \qquad (156-3)$$

ومنها نحصل على قيمة k^* التي تجعل الشعاع e يخضع للشرط (3-136) وبالتالي نحصل على عناصر الشعاع e^* ولتكن : $e_1^*\ , e_2^* , e_3^* , \dots , e_p^*$، وللحصول على عناصر الشعاع الأول f نستخدم العلاقة (3-156) فنجد أن:

$$f^* = \begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ f_q^* \end{bmatrix} = \frac{R_{yy}^{-1} * R_{yx}}{\lambda} * \begin{bmatrix} e_1^* \\ e_2^* \\ \vdots \\ e_q^* \end{bmatrix} \qquad (157-3)$$

ثم نعمل على تعديل عناصر f^* حتى تحقق الشرط المفروض عليها وهو الشرط (3-137) التالي:

$$var(V_{zy}) = f'^* * R_{yy} * \ f^* = 1 \qquad (158-3)$$

وبعد ذلك نقوم بتعويض الشعاعينf^*, e^* في التركيبين الخطيين المعرفين في (3-127),(3-128) فنحصل على أن:

$$\widetilde{U}_{1zx} = \ e_1^* * Z_{x1} + e_2^* * Z_{x2} + e_3^* * Z_{x3} + \cdots + e_p^* * Z_{xp} \qquad (159-3)$$

$$\widetilde{V}_{1zy} = \ f_1^* * Z_{y1} + f_2^* * Z_{y2} + f_3^* * Z_{y3} + \cdots + f_q^* * Z_{yq} \qquad (160-3)$$

ثم نقوم بتحويل هذين التركيبين من المتحولات المعيارية إلى المتحولات الأصلية (الخام Raw) وذلك باستبدال كل Z بما تساويه من العلاقتين (3-122) , (3-124) فنحصل على أن:

$$\widetilde{U}_{1zx} = e_1^*\left(\frac{X_1 - \bar{X}_1}{\sigma_{x1}}\right) + \ e_2^*\left(\frac{X_2 - \bar{X}_2}{\sigma_{x2}}\right) + \cdots + e_p^*\left(\frac{X_p - \bar{X}_p}{\sigma_{xp}}\right) \qquad (161-3)$$

$$\widetilde{V}_{1zx} = f_1^*\left(\frac{Y_1 - \bar{Y}_1}{\sigma_{y1}}\right) + \ f_2^*\left(\frac{Y_2 - \bar{Y}_2}{\sigma_{y2}}\right) + \cdots + f_q^*\left(\frac{Y_q - \bar{Y}_q}{\sigma_{yq}}\right) \qquad (162-3)$$

وبإجراء بعض الإصلاحات نجد أن:

$$\widetilde{U}_{1zx} = \frac{e_1^*}{\sigma_{x1}}X_1 + \frac{e_2^*}{\sigma_{x2}}X_2 + \cdots \frac{e_p^*}{\sigma_{xp}}X_p - \left[\frac{e_1^*}{\sigma_{x1}}\bar{X}_1 + \frac{e_2^*}{\sigma_{x2}}\bar{X}_2 + \cdots \frac{e_p^*}{\sigma_{xp}}\bar{X}_p\right] \quad (163-3)$$

$$\widetilde{V}_{1zx} = \frac{f_1^*}{\sigma_{y1}}Y_1 + \frac{f_2^*}{\sigma_{y2}}Y_2 + \cdots \frac{f_q^*}{\sigma_{yq}}Y_q - \left[\frac{f_1^*}{\sigma_{y1}}\bar{Y}_1 + \frac{f_2^*}{\sigma_{y2}}\bar{Y} + \cdots \frac{f_q^*}{\sigma_{yq}}\bar{Y}_q\right] \quad (164-3)$$

نفرض أن ما ضمن القوسين من (3–163) و(3–164) هو $\bar{V}\ , \bar{U}$ على الترتيب ونضع الناتج فيها. كما يلي:

$$\widetilde{U}_1 = \widetilde{U}_{1zx} + \bar{U} = \ \frac{e_1^*}{\sigma_{x1}}X_1 + \frac{e_2^*}{\sigma_{x2}}X_2 + \cdots \frac{e_p^*}{\sigma_{xp}}\mathrm{X}_p \qquad (165-3)$$

$$\widetilde{V}_1 = \widetilde{V}_{1zy} + \bar{V} = \frac{f_1^*}{\sigma_{y1}}Y_1 + \frac{f_2^*}{\sigma_{y2}}Y_2 + \cdots \frac{f_q^*}{\sigma_{yq}}Y_q \qquad (166-3)$$

وهكذا نكون قد حصلنا على أمثال التركيبين (3-6),(3-7) من خلال أمثال التركيبين المعياريين (3-165),(3-166)، ومن المقارنة بينهما نجد أن:

$$a_1^* = \frac{e_1^*}{\sigma_{x1}} \qquad a_2^* = \frac{e_2^*}{\sigma_{x2}} \ \ \dots\dots\ \ a^* = \frac{e_p^*}{\sigma_{xp}} \qquad (167-3)$$

$$b_1^* = \frac{f_1^*}{\sigma_{y1}} \qquad b_2^* = \frac{f_2^*}{\sigma_{y2}} \quad \dots\dots \; b_q^* = \frac{f_q^*}{\sigma_{yq}} \qquad (168-3)$$

وبعبارة أخرى نقول أنه يمكننا الحصول على عناصر الشعاعين الخام b^*, a^* من عناصر الشعاعين المعيارين e^*, f^* بتقسيمها على الانحرافات المعيارية للمتحولات المصاحبة لها:

كما يمكن كتابة ذلك على شكل مصفوفي كما يلي:

$$a^* = \begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_p^* \end{bmatrix} = (e_1^*, e_2^*, \dots, e_p^*) \begin{bmatrix} \frac{1}{\sigma_{x1}} & 0 & 0 \\ 0 & \frac{1}{\sigma_{x2}} & 0 \\ 0 & 0 & \frac{1}{\sigma_{xp}} \end{bmatrix} = (e^*)' * C_{\sigma_x}^{-\frac{1}{2}} \qquad (169-3)$$

$$b^* = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_q^* \end{bmatrix} = (f_1^*, f_2^*, \dots, f_p^*) \begin{bmatrix} \frac{1}{\sigma_{y1}} & 0 & 0 \\ 0 & \frac{1}{\sigma_{y2}} & 0 \\ 0 & 0 & \frac{1}{\sigma_{yq}} \end{bmatrix} = (f^*)' * C_{\sigma_y}^{-\frac{1}{2}} \qquad (170-3)$$

ملاحظة: كان يمكن معالجة هذا الموضوع بطريقة أخرى وحساب الشعاع f^* أولاً ثم حساب الشعاع e^* وذلك باستخدام المعادلتين:

$$R_{yy}^{-1} * R_{yx} * R_{xx}^{-1} * R_{xy} * f = \lambda^2 * f \qquad (171-3)$$

$$e = \frac{R_{xx}^{-1} * R_{xy}}{\lambda} * f \qquad (172-3)$$

وذلك بالإضافة للشرطين

$$e' * R_{xx} * e = 1 \qquad (173-3)$$

$$f' * R_{yy} * f = 1 \qquad (174-3)$$

ويتم تطبيق هذه الحالة عندما تكون $p > q$.

8-3 حساب التحميلات القانونية Canonical loadings

إن المقصود بالتحميلات القانونية هو جملة معاملات الارتباط الثنائية بين متحولات المجموعة X والمركبات القانونية الخاصة بها U، ومعاملات الارتباط الثنائية بين متحولات المجموعة Y ومركباتها القانونية الخاصة بها V. وتسمي هذه المعاملات بالمعاملات الهيكلية المباشرة . وتشمل هذه التحميلات نوع آخر من المعاملات هو معاملات الارتباط الثنائية بين متحولات المجموعة X والمركبات القانونية المقابلة لها في الطرف الاخر V، ومعاملات الارتباط الثنائية بين متحولات المجموعة Y والمركبات

القانونية المقابلة لها في الطرف الآخر U، وتسمى هذه المعاملات بالمعاملات الهيكلية العابرة (إلى الطرف الآخر) وسنعالج كل منها بفرض أن عدد الأزواج المركبة $(U_k V_k)$ يساوي p، كما يلي:

أ- **التحميلات القانونية المباشرة: وهي مؤلفة من نوعين هما:**

أ-1- **تحميلات كل من متحولات المجموعة X على المركبات القانونية الخاصة بها** $\boldsymbol{U_1, U_2, U_3, \dots, U_p}$، وهي عبارة عن جملة من الأشعة تعبر عن معاملات الارتباط الزوجية بين كل من المتحولات $X_1, X_2, X_3, \dots, X_p$ مع كل من المركبات القانونية $U_1, U_2, U_3, \dots, U_p$ الخاصة بها. وتعطى على شكل جدول منظم كما يلي:

جدول (3-7): **تحميلات** X **على** U :

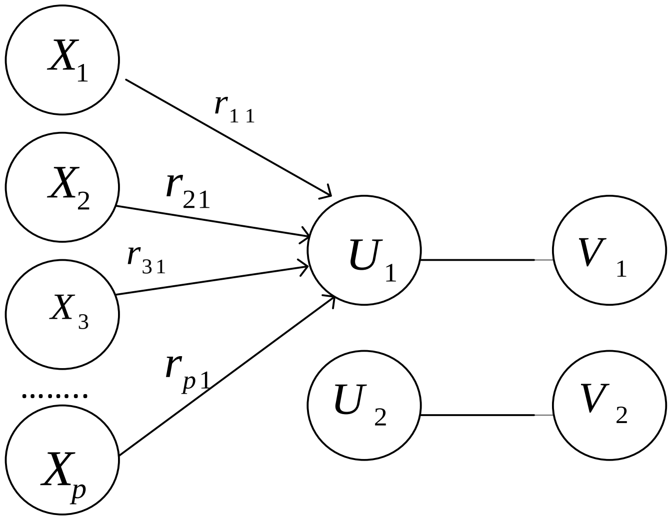
المركبات U / المتحولات X	U_1	U_2	U_3	...	U_p
X_1	r_{11}	r_{12}	r_{13}	...	r_{1p}
X_2	r_{21}	r_{22}	r_{23}	...	r_{2p}
X_3	r_{31}	r_{32}	r_{33}	...	r_{3p}
...	...	...	...	...	...
X_p	r_{p1}	r_{p2}	r_{p3}	...	r_{pp}

ويتم حساب هذه المعاملات حاسوبياً من خلال معادلة (بيرسون) بين كل متحول من X مع كل مركب من U. ولكن عملية الحساب تتم بعد حساب القيم النظرية (التنبؤية) للمركبات $U_1, U_2, U_3 \dots . U_p$ من النماذج الخطية التي تم استخراجها لها، وذلك بالاعتماد على القيم الفعلية للمتحولات X المعطية في جدول شبيه بالجدول (3−1) .

كما يمكن حساب المعاملات r_{ij} في كل شعاع عمود j من جداء المصفوفة الارتباطية R_{xx} في الشعاع المعياري e المتعلق بالمركب الموافق له U_j . ونكتب ذلك كما يلي:

$$\begin{bmatrix} r_{1j} \\ r_{2j} \\ r_{3j} \\ \vdots \\ r_{pj} \end{bmatrix} = R_{xx} \begin{bmatrix} e_{1j} \\ e_{2j} \\ e_{3j} \\ \vdots \\ e_{pj} \end{bmatrix} \qquad (175-3)$$

ويمكن تمثيل تحميلات المتحولات X على المركب القانوني الأول U_1 (مثلاً) كما في الشكل التالي :



الشكل (3-2) تمثيل تحميلات X على U_1 فقط

وكذلك يمكن أن نحصل على أشكال مشابهة لتحميلات X على U_2 ثم على U_3 ...الخ .

وإن هذه التحميلات تعبر عن شدة أو قوة العلاقة بين كل من المتحولات X مع كل من المركبات القانونية الخاصة بها U .

أ-2- **تحميلات كل من متحولات المجموعة Y على مركباتها القانونية الخاصة بها** $\boldsymbol{V_1, V_2, V_3 \ldots . V_p}$ وهي عبارة عن معاملات الارتباط الزوجية بين كل من المتحولات $Y_1, Y_2, Y_3 \ldots Y_q$ مع كل من المركبات القانونية الخاصة بها $V_1, V_2, V_3 \ldots V_p$ وتعطى على شكل جدول منظم كما يلي:

جدول (3-8): تحميلات المتحولات Y على المركبات V

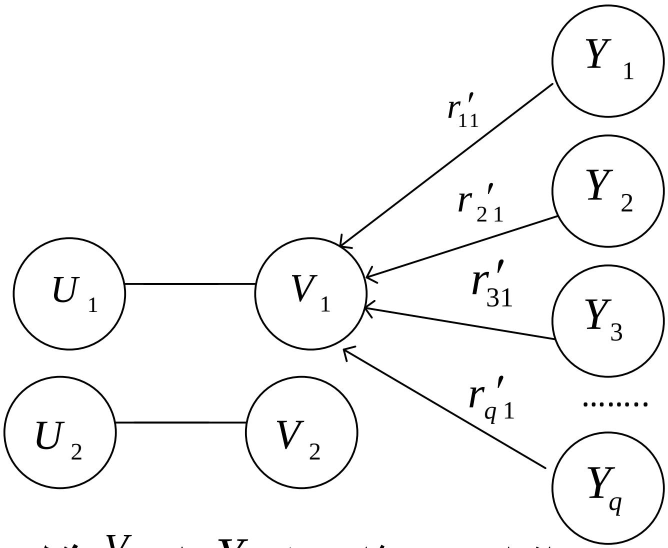
المركبات V / المتحولات Y	V_1	V_2	V_3	...	V_p
Y_1	r'_{11}	r'_{12}	r'_{13}	...	r'_{1p}
Y_2	r'_{21}	r'_{22}	r'_{23}	...	r'_{1p}
Y_3	r'_{31}	r'_{32}	r'_{33}	...	r'_{1p}
...	...	...	...	...	...
Y_q	r'_{q1}	r'_{q2}	r'_{q3}	...	r'_{qp}

ويتم حساب هذه المعاملات حاسوبياً بعد إيجاد القيم النظرية للمركبات القانونية $V_1, V_2, V_3 \ldots V_p$ من النماذج الخطية التي تم استخراجها لها، وذلك بالاعتماد على القيم الفعلية للمتحولات Y المعطية في جدول شبيه بالجدول (3-1) .

كما يمكن حساب المعاملات r'_{ij} في كل شعاع j عمود من جداء المصفوفة الارتباطية R_{yy} في الشعاع المعياري f المتعلق بالمركب الموافق V_j ونكتب ذلك كما يلي:

$$\begin{bmatrix} r'_{1j} \\ r'_{2j} \\ r'_{3j} \\ \vdots \\ r'_{qj} \end{bmatrix} = R_{yy} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_q \end{bmatrix} \qquad (176-3)$$

ويمكن تمثيل تحميلات المتحولات Y على المركب القانوني الأول V_1 (مثلاً) كما في الشكل التالي:



الشكل (3-3) تمثيل تحميلات Y على V_1 فقط

وإن هذه التحميلات تعبر عن شدة أو قوة العلاقة بين كل من المتحولات Y مع كل من المركبات القانونية الخاصة بها V .

ب- **التحميلات القانونية العابرة:** Cross canonical loadings

وهي أيضاً تتألف من نوعين هما:

ب-1- تحميلات متحولات المجموعة X على المركبات القانونية المقابلة لها V في الطرف الآخر .

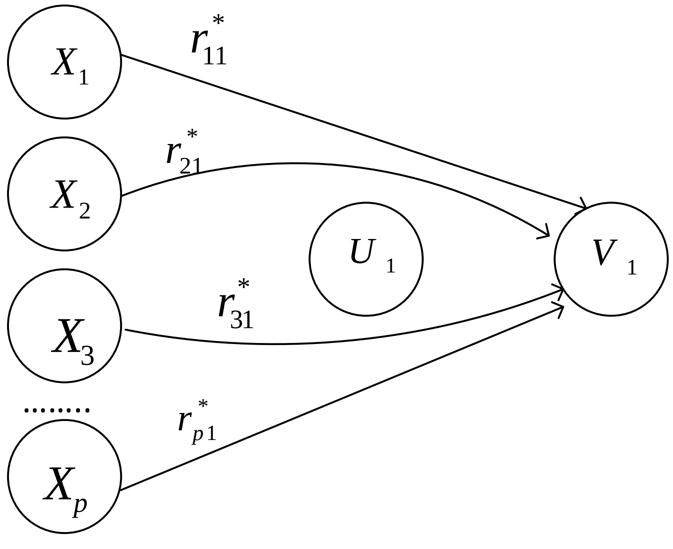
وهي عبارة عن معاملات الارتباط الزوجية بين كل من المتحولات $X_p, \dots, X_3, X_2, X_1$ مع كل من المركبات القانونية المقابلة لها في الطرف الآخر وهي $V_p \dots V_3, V_2, V_1$، وتعطى على شكل جدول منظم كما يلي:

جدول (3-9): تحميلات X العابرة على المركبات V المقابلة لها في الطرف الآخر :

المركبات V / المتحولات X	V_1	V_2	V_3	...	V_p
X_1	r^*_{11}	r^*_{12}	r^*_{13}	...	r^*_{1p}
X_2	r^*_{21}	r^*_{22}	r^*_{23}	...	r^*_{1p}
X_3	r^*_{31}	r^*_{32}	r^*_{33}	...	r^*_{1p}
...	...	...	...	...	...
X_P	r^*_{p1}	r^*_{p2}	r^*_{p3}	...	r^*_{pp}

ويمكن حساب هذه المعاملات من جراء ضرب معاملات التحميلات المباشرة على المركبات U_k في معامل الارتباط القانوني ρ_K المقابل للزوج (U_k, V_k)، وهي تعبر عن حجم تأثير كل من المركبات القانونية V_K على المتحولات المستقلة X، وبالعكس تعبر عن حجم تأثير كل من المتحولات X على المركبات القانونية المقابلة لها V_K، ويمكن أن تحسب من العلاقة: $r_{ik}^* = \rho_k * r_{ik}$.

كما يمكن تمثيل التحميلات العابرة للمجموعة X على أحد المركبات القانونية المقابلة لها (وليكن V_1) كما يلي:



الشكل (3-4): التحميلات العابرة لـــ X على V_1 (مثلا)

ب-2- تحميلات متحولات المجموعة Y على المركبات القانونية المقابلة لها U في الطرف الآخر.

وهي أيضاً عبارة عن معاملات الارتباط الزوجية بين كل من المتحولات $Y_q \dots Y_3, Y_2, Y_1$ مع كل من المركبات القانونية المقابلة لها وهي $U_p, \dots, U_3, U_2, U_1$ وتعطى على شكل جدول منظم كما يلي :

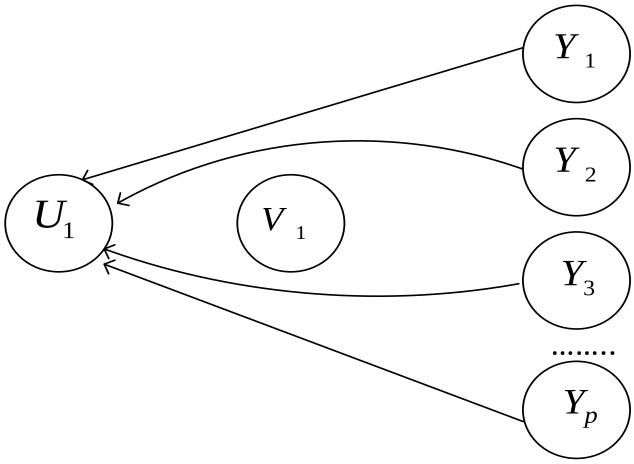
جدول (3-10) **تحميلات** Y **العابرة على المركبات المقابلة لها** U **في الطرف الآخر:**

المركبات U / المتحولات Y	U_1	U_2	U_3	...	U_p
Y_1	r_{11}^{**}	r_{12}^{**}	r_{13}^{**}	...	r_{1p}^{**}
Y_2	r_{21}^{**}	r_{22}^{**}	r_{23}^{**}	...	r_{2p}^{**}
Y_3	r_{31}^{**}	r_{32}^{**}	r_{33}^{**}	...	r_{3p}^{**}
...	...	...	...	...	...
Y_q	r_{q1}^{**}	r_{q2}^{**}	r_{q3}^{**}	...	r_{qP}^{**}

ويمكن حساب هذه المعاملات من جراء ضرب معاملات التحميلات المباشرة على المركبات V_k في معامل الارتباط القانوني ρ_k المقابل للزوج (U_k, V_k)، وهي تعبر عن حجم تأثير كل من المركبات U_k

على كل من المتحولات Y، وبالعكس تعبر عن حجم تأثير كل من المتحولات Y على المركبات القانونية المقابلة لها، ويمكن أن تحسب من العلاقة: $r_{ik}^{**} = \rho_k * r'_{ik}$

ويمكن تمثيل التحميلات العابرة للمجموعة Y على أحد المركبات القانونية المقابلة لها (وليكن U_1) كمايلي:



الشكل (3-5) : التحميلات العابرة لــ Y على U_1 (مثلاً)

3-9 تقييم النماذج القانونية $(\mathbf{U_k, V_k})$ من التحميلات المباشرة :

يتم تقييم النماذج القانونية حسب الأزواج القانونية (U_k, V_k) ولكل مركب U_k أو V_k على حدة . ونستخدم لذلك مربعات التحميلات المباشرة لحساب عدة مؤشرات هي: الكفاءة– التشاركية– الأهمية كمايلي:

أ– **حساب كفاءة النموذج** (U_k, V_k) : Adequacy : ويتم حساب هذه الكفاءة لكل مركب U_k أو V_k على حدة، وتعرف كفاءة المركب U_k بأنها متوسط مربعات التحميلات المباشرة للمتحولات X على U_k ونرمز لها بـ $AD(U_k)$ وتحسب في كل عمود U_k من العلاقة :

$$AD(U_k) = \frac{\sum_{i=1}^{p} r_{ik}^2}{p} \qquad (178-3)$$

كما تعرف كفاءة المركب V_k بأنها متوسط مربعات التحميلات المباشرة للمتحولات Y على V_k ونرمز لها بــ $AD(V_k)$ وتحسب في كل عمود V_k من العلاقة:

$$AD(V_k) = \frac{\sum_{j=1}^{q} \grave{r}_{j\mathrm{k}}^2}{q} \qquad (179-3)$$

ولهذا فإنه لحساب كفاءة المركبات القانونية U_k نربع معاملات الارتباط المباشرة للمتحولات X على U_k. وهي المعاملات الواردة في الجدول (3–6) السابق ونضعها في جدول منظم كالتالي:

جدول (3-11) : مربعات تحميلات X على U

المركبات U / المتحولات X	U_1	U_2	U_3	...	U_p	المجموع = التشاركية
X_1	r_{11}^2	r_{12}^2	r_{13}^2	...	r_{1p}^2	$\sum_{k=1}^{p} r_{1k}^2$
X_2	r_{21}^2	r_{22}^2	r_{23}^2	...	r_{2p}^2	$\sum_{k=1}^{p} r_{2k}^2$
X_3	r_{31}^2	r_{32}^2	r_{33}^2	...	r_{3p}^2	$\sum_{k=1}^{p} r_{3k}^2$
...	...	...	...	...	...	...
X_p	r_{p1}^2	r_{p2}^2	r_{p3}^2	...	r_{pp}^2	$\sum_{k=1}^{p} r_{pk}^2$
المجموع	$\sum_{i=1}^{p} r_{i1}^2$	$\sum_{i=1}^{p} r_{i2}^2$	$\sum_{i=1}^{p} r_{i3}^2$	...	$\sum_{i=1}^{p} r_{ip}^2$	P
المتوسط = كفاءة U_k	$\frac{\sum r_{i1}^2}{p}$	$\frac{\sum r_{i2}^2}{p}$	$\frac{\sum r_{i3}^2}{p}$	...	$\frac{\sum r_{ip}^2}{p}$	1

ومن دراسة هذا الجدول نلاحظ أن كفاءة كل مركب U_k محسوبة وموجودة في السطر الأخير وتساوي متوسط مربعات معاملات ارتباطه مع جميع المتحولات X المعتمدة في الدراسة .

وفي الحالة العامة يكون مجموع كفاءات جميع المركبات U_k مساوياً للواحد (إذا لم يتم الاستغناء عن بعضها بسبب عدم معنوية الجذور الكامنة). ويستفاد منها في تحديد ومعرفة النسبة المئوية التي يفسرها المركب U_k من التباين الكلي للمتحولات الأصلية في المجموعة X. حيث أن كفاءة المركب U_1 تعني النسبة التي يفسرها U_1 من تغيرات مجموعة المتحولات X .

ولحساب كفاءة المركبات القانونية V_k نربع معاملات الارتباط المباشرة للمتحولات Y على V_k وهي المعاملات الواردة في الجدول (3-8) السابق ونضعها في جدول منظم كالتالي:

جدول (3-12) : مربعات تحميلات Y على V

المركبات U / المتحولات Y	V_1	V_2	V_3	...	V_p	المجموع = التشاركية
Y_1	$\grave{r}_{11}^2$	$\grave{r}_{12}^2$	$\grave{r}_{13}^2$	...	$\grave{r}_{1p}^2$	$\sum_{k=1}^{p} \grave{r}_{1k}^2$
Y_2	$\grave{r}_{21}^2$	$\grave{r}_{22}^2$	$\grave{r}_{23}^2$	...	r_{2p}^2	$\sum_{k=1}^{p} \grave{r}_{2k}^2$
...	...	...	...	...	...	...
Y_q	$\grave{r}_{q1}^2$	$\grave{r}_{q2}^2$	$\grave{r}_{q3}^2$	...	r_{qp}^2	$\sum_{k=1}^{p} \grave{r}_{qk}^2$
المجموع	$\sum_{j=1}^{q} \grave{r}_{j1}^2$	$\sum_{j=1}^{q} \grave{r}_{j2}^2$	$\sum_{j=1}^{q} \grave{r}_{j3}^2$	...	$\sum_{j=1}^{q} \grave{r}_{jq}^2$	Q
المتوسط = كفاءة V_k	$\frac{\sum_{j=1}^{q} \grave{r}_{j1}^2}{q}$	$\frac{\sum^{q} \grave{r}_{j2}^2}{q}$	$\frac{\sum^{q} \grave{r}_{j3}^2}{q}$	...	$\frac{\sum^{q} \grave{r}_{jq}^2}{q}$	1

وعند دراسة هذا الجدول نلاحظ أن كفاءة كل مركب V_k محسوبة وموجودة في السطر الأخير وتساوي متوسط مربعات معاملات ارتباطه مع جميع المتحولات Y المعتمدة في الدراسة .

وفي الحالة العامة يكون مجموع كفاءات جميع المركبات V_k مساوياً للواحد (إذا لم يكن قد تم الاستغناء عن بعضها)، ويستفاد من هذه الكفاءة في تحديد النسبة المئوية التي يفسرها المركب V_k من تباين المتحولات الأصلية للمجموعة Y . حيث أن كفاءة المركب V_1 تعني النسبة المئوية التي يفسرها V_1 من تغيرات مجموعة المتحولات Y .

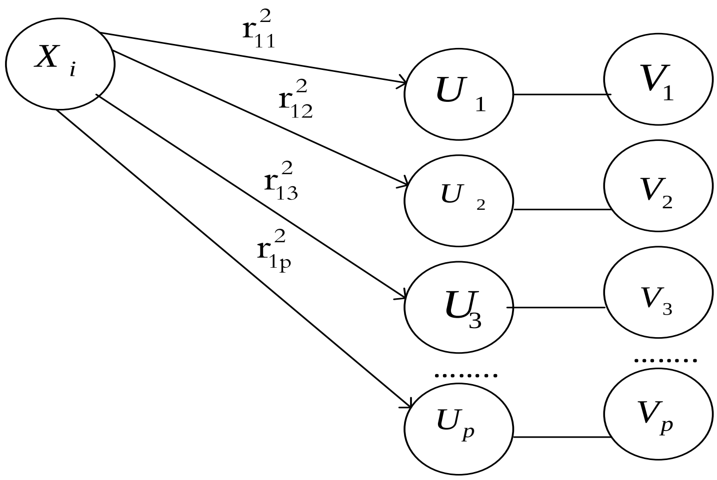
ب- **حساب تشاركيه المتحولات X والمتحولات** Y communilty:

ويتم حساب تشاركيه كل متحول من X على حدة وكل متحول من Y على حدة .

وتعرف تشاركية المتحول X_i بأنها تساوي مجموع مربعات معاملات ارتباط المتحول المحدد X_i مع جميع المركبات U_k. ويرمز لها بالرمز $com\ (X_i)$ وتحسب من العلاقة:

$$com(X_i) = \sum_{k=1}^{p} r_{ik}^2 \qquad (180-3)$$

وهي المقادير المحسوبة على يمين الجدول (3-11) السابق، ويبرهن على أنه إذا كان التحليل القانوني يشمل جميع المركبات U_k (ولم يتم الاستغناء عن بعضها بسبب عدم معنوية الجذور) فإن تشاركية كل متحول (X_i) تساوي الواحد، وأن مجموع تشاركيات المجموعة X يساوي p (مجموع العمود الأخير)، ويمكن تمثيل تشاركية المتحول X_i (مثلاً) بيانياً كمايلي:



الشكل (3–6): تشاركية X_i (مثلاً) على المركبات U_k

وهي تعبر عن نسبة مساهمة جميع المركبات القانونية $U_1, U_2, \dots, U_p$ وفي تغيرات المتحول الأصلي المحدد X_i وكم هو $\mathrm{X_i}$ مفيداً لنا في التحليل القانوني .

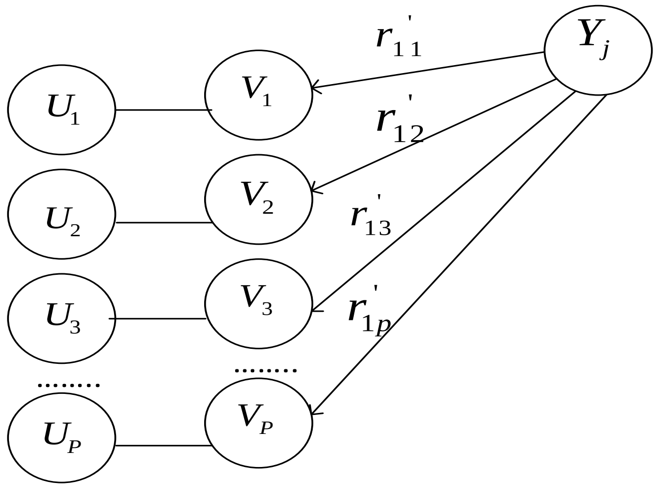
كما تعرف تشاركية المتحول Y_j بأنها تساوي مجموع مربعات معاملات ارتباط المتحول Y_j مع جميع المركبات V_k ويرمز لها بالرمز $com\left(Y_j\right)$ وتحسب من العلاقة :

$$com\left(Y_j\right) = \sum_{k=1}^{q} \grave{r}_{jk}^2 \qquad (181-3)$$

وهي المقادير المحسوبة على يمين الجدول (3-12) السابق .

ويبرهن على أنه إذا كان التحليل القانوني يشمل جميع المركبات V_k (ولم يتم الاستغناء عن بعضها بسبب عدم معنوية الجذور الكامنة) فإن تشاركية كل متحول Y_j تساوي الواحد .

ويمكن تمثيل تشاركية المتحول Y_j (مثلاً) بيانياً كما يلي :



الشكل (3-7): تشاركية Y_j على المركبات V_k

وهي تعبر عن نسبة مساهمة جميع المركبات القانونية $V_1, V_2, \dots, V_q$ في تغيرات المتحول الأصلي المحدد Y_j . وكم هو Y_j مفيداً لنا في التحليل القانوني .

ج- حساب الأهمية النسبية للمتحولات في النموذج (U_k, V_k) :

تعرف هذه الأهمية لكل متحول X على حدة ولكل متحول Y على حدة، بأنها نسبة مربع معامل الارتباط الخاص بالمتحول X_i أو Y_j على إجمالي التشاركية لذلك المتحول. وتحسب من العلاقات التالية:

$$R(X_{ik}) = \frac{r_{ik}^2}{\sum_{k=1}^{p} r_{ik}^2} \qquad (i: 1, 2, 3, \dots . p) \qquad (182-3)$$

$$R(Y_{jk}) = \frac{\grave{r}_{jk}^2}{\sum_{k=1}^{q} r_{jk}'^2} \qquad (i: 1, 2, 3, \dots q) \qquad (183-3)$$

ويستفاد منها في تحديد نسبة مساهمة أو أهمية كل من X_i أو Y_j في تباين كل من المركبات V_k أو U_k.

3-10 تقيم النماذج القانونية $(\mathbf{V_k}, \mathbf{U_k})$ من التحميلات العابرة :

يتم التقيم هنا لكل مركب V_k أو U_k من التحميلات العابرة إليه من الطرف الآخر، ومنها نحسب نفس المؤشرات السابقة: الكفاءة العابرة- التشاركية العابرة- الأهمية العابرة كما يلي:

أ- حساب الكفاءة العابرة للنموذج (V_k, U_k) Redundancy :

يتم حساب هذه الكفاءة لكل مركب V_k أو U_k على حدة، وتعرف الكفاءة العابرة للمركب U_k بأنها متوسط مربعات التحميلات العابرة للمتحولات Y على المركب U_k، ونرمز لها بالرمز $Red(U_k)$ وتحسب من العلاقة :

$$Red(U_k) = \frac{\sum_{j=1}^{q} r_{jk}^{*2}}{q} \qquad (184-3)$$

كما تعرف الكفاءة العابرة للمركب V_k بأنها متوسط مربعات التحميلات العابرة للمتحولات X على V_k، ونرمز لها بالرمز $Red(V_k)$ وتحسب من العلاقة :

$$Red(V_k) = \frac{\sum_{j=1}^{p} r_{ik}^{**2}}{p} \qquad (185-3)$$

ويمكن حسابها من جداول مشابهة للجدولين (10),(11)، ومن جهة أخرى يمكن حساب الكفاءة العابرة للمركبين V_k أو U_k من خلال الجداء المتتالي للكفاءة المباشرة لـ V_k أو U_k في مربع معامل الارتباط القانوني ρ_k^2 ، أي يمكن حسابها من العلاقتين :

$$Red(U_k) = AD(U_k) * \rho_k^2 \qquad (186-3)$$

$$Red(V_k) = AD(V_k) * \rho_k^2 \qquad (187-3)$$

ب-حساب التشاركية العابرة للمتحولات X و Y: وتعرف بأنها مجموع مربعات التحميلات العابرة لكل متحول على المركبات المقابلة له، ويتم حسابها لكل متحول X_i على المركبات المقابلة له من الطرف الآخر $V_1, V_2, \dots, V_p$ من العلاقة:

$$com^*(X_i) = \sum_{k=1}^{p} r_{ik}^{*2} \qquad i: 1,2,3, \dots p \qquad (188-3)$$

ويتم حسابها لكل متحول Y_j على المركبات المقابلة له من الطرف الاخر $U_1, U_2, \dots, U_p$ من العلاقة :

$$com^*(Y_j) = \sum_{k=1}^{p} r_{jk}^{**2} \qquad j: 1,2,3, \dots q \qquad (189-3)$$

ج- حساب الأهمية النسبية العابرة للمتحولات X و Y

وهي تساوي نسبة مربع معامل الارتباط الخاص بالمتحول X_i أو Y_j على التشاركية العابرة الكلية له وتحسب من العلاقتين:

$$R^*(X_i) = \frac{r_{ik}^{*2}}{\sum_{k=1}^{*} r_{ik}^{*2}} \qquad (190-3)$$

i: 1 , 2, 3, ... p

$$R^*(Y_j) = \frac{r_{jk}^{**2}}{\sum_{k=1}^{p} r_{jk}^{**2}} \qquad (191-3)$$

j: 1 , 2, 3, ... q

وهي تتمتع بنفس المعاني ولها نفس التفسيرات المذكورة في مؤشرات التحميلات المباشرة، ولكنها أقل أهمية منها .

3-1 مثال عام:

ادرس الارتباط القانوني لمجموعة من المتحولات المستقلة X_1, X_2, X_3 مع مجموعة من المتحولات التابعة Y_1, Y_2, Y_3 ، علماً بأن مصفوفات الارتباط بين متحولات كل مجموعة R_{yy}, R_{xx} وكذلك R_{xy} و R_{yx} بين المتحولات X مع المتحولات Y هي كما في الجدول المنظم التالي:

جدول (1): مصفوفة معاملات الارتباط للمجموعتين X , Y .

المتحولات	X_1	X_2	X_3	Y_1	Y_2	Y_3
X_1	1	0.4036	−0.2438	0.2249	−0.3387	−0.0221
X_2	0.4036	1	−0.0819	0.0961	−0.1676	0.0754
X_3	−0.2438	−0.0819	1	−0.1312	0.0755	0.0872
Y_1	0.2249	0.0961	−0.1312	1	−0.1133	0.0570
Y_2	−0.3387	−0.1676	0.0755	−0.1133	1	0.0222
Y_3	−0.0221	0.0754	0.0872	−0.0570	0.0222	1

الحل: نقوم اولاً بإيجاد القيم المميزة λ_k^2 من معادلة المحدد $|R - \lambda^2 I| = 0$ حيث أن R هي المصفوفة $R = R_{xx}^{-1} * R_{xy} * R_{yy}^{-1} * R_{yx}$، ثم نقوم بحساب الأشعة المعيارية المصاحبة لمركبين الخطيين المعياريين للمجموعة Y, X المعرفين بالعلاقتين:

$$U_{zx} = e_1 Z_{x1} + e_2 Z_{xe2} + e_3 Z_{x3} \qquad : \quad Z_{xi} = \frac{X_i - \bar{x}_i}{\sigma_{xi}} \qquad (1)$$

$$V_{zy} = f_1 Z_{y1} + f_2 Z_{y2} + f_3 Z_{y3} \qquad : \quad Z_{yj} = \frac{Y_j - \bar{Y}_j}{\sigma_{yj}} \qquad (2)$$

لذلك نقوم بحساب الجداء المصفوفي التالي: $R = R_{xx}^{-1} * R_{xy} * R_{yy}^{-1} * R_{yx}$ كما يلي:

$$R_{xx} = \begin{bmatrix} 1 & 0.4036 & -0.2438 \\ 0.4036 & 1 & -0.0819 \\ -0.2438 & -0.0819 & 1 \end{bmatrix} \Rightarrow R_{xx}^{-1} = \begin{bmatrix} 1.262 & -0.4874 & 0.26776 \\ -0.4874 & 1.195 & 0.02096 \\ 0.26776 & -0.02096 & 1.06356 \end{bmatrix}$$

$$R_{yy} = \begin{bmatrix} 1 & -0.1133 & -0.0570 \\ -0.1133 & 1 & 0.0222 \\ -0.0570 & -0.0222 & 1 \end{bmatrix} \Rightarrow R_{yy}^{-1} = \begin{bmatrix} 1.01606 & 0.11389 & 0.0553871 \\ 0.11389 & 1.01326 & -0.0160026 \\ 0.0553871 & -0.0160026 & 1.00351 \end{bmatrix}$$

ثم نقوم بحساب الجداء الأول فنجد أن:

$$R_{xx}^{-1} * R_{xy} = \begin{bmatrix} 1.262 & -0.4874 & 0.26776 \\ -0.4874 & 1.195 & 0.02096 \\ 0.26776 & -0.02096 & 1.06356 \end{bmatrix} * \begin{bmatrix} 0.2249 & -0.3387 & -0.0221 \\ 0.0961 & -0.1676 & 0.0754 \\ -0.1312 & 0.0755 & 0.0872 \end{bmatrix}$$

وبعد الحساب نحصل على أن الناتج A يساوي:

$$R_{xx}^{-1} * R_{xy} = \begin{bmatrix} 0.2018535 & -0.3255330 & -0.0412928 \\ 0.00797055 & -0.0367781 & 0.0990473 \\ -0.0813353 & -0.0068771 & 0.0852448 \end{bmatrix} = \text{A} \qquad (3)$$

ثم نقوم بحساب الجداء الثاني فنجد أن:

$$R_{yy}^{-1} * R_{yx} = \begin{bmatrix} 1.01606 & 0.11389 & 0.0553871 \\ 0.11389 & 1.01326 & -0.0160026 \\ 0.0553871 & -0.0160026 & 1.00351 \end{bmatrix} * \begin{bmatrix} 0.2249 & 0.0961 & -0.1312 \\ -0.3387 & -0.1676 & 0.0755 \\ 0.0221 & 0.0754 & 0.0872 \end{bmatrix}$$

وبعد الحساب نحصل على أن الناتج B يساوي:

$$R_{yy}^{-1} * R_{yx} = \begin{bmatrix} 0.188713 & 0.0827317 & -0.1198787 \\ -0.3172233 & -0.1600840 & 0.0601632 \\ -0.004301 & 0.0836696 & 0.0790313 \end{bmatrix} = \text{B} \qquad (4)$$

وبعد ذلك نقوم بحساب الجداء الرباعي للمصفوفات الارتباطية فنجد أن الناتج هو المصفوفة R التالية:

$$\text{R} = R_{xx}^{-1} * R_{xy} * R_{yy}^{-1} * R_{yx} = A * B = \begin{bmatrix} 0.141537 & 0.0653574 & -0.0470465 \\ 0.012745 & 0.0148343 & 0.00465965 \\ -0.0135341 & 0.0015043 & 0.0160736 \end{bmatrix} \quad (\mathbf{5})$$

ثم نقوم بحساب القيم الذاتية للمصفوفة النهائية R ، ولذلك نشكل المعادلة المميزة لهذه المسألة كما يلي:

$$[R - \lambda^2 I]e = 0 \qquad (6)$$

ولإيجاد حلول مقبولة (غير الصفرية) لهذه المعادلة بالنسبة لـ e نفترض أن المصفوفة $[R - \lambda^2 I]$ شاذة فيكون محددها معدوماً . لذلك نضع المحدد التالي المشكل منها مساوياً للصفر :

$$|R - \lambda^2 I| = \begin{vmatrix} (0.141537 - \lambda^2) & 0.0653574 & 0.0470465 \\ 0.012745 & (0.0148343 - \lambda^2) & 0.00465965 \\ -0.0135341 & 0.0015043 & (0.0160736 - \lambda^2) \end{vmatrix} = 0 \quad (\mathbf{7})$$

ثم نقوم بحساب مفكوك هذا المحدد حسب طريقة (سايروس) فنحصل على كثير حدود من الدرجة الثالثة بالنسبة لـ λ^2 كما يلي :

$$|R - \lambda^2 I| = (0.141537 - \lambda^2)(0.0148343 - \lambda^2)(0.0160736 - \lambda^2) \\ -(0.0653574)(0.00465965)(0.0135341)$$

$$
\begin{aligned}
&-(0.0470465)(0.012745)(0.0015043)\\
&-(0.0470465)(0.0148343-\lambda^2)(0.0135341)\\
&-(0.00465965)(0.0015043)(0.141537-\lambda^2)\\
&-(0.0160736-\lambda^2)(0.0653574)(0.012745)=0
\end{aligned}
$$

$$
\begin{aligned}
|R-\lambda^2 I| = {}& 0.000033748-0.00461305\lambda^2+0.1724449\lambda^4-\lambda^6\\
&-0.00000489757+0.000636732\lambda^2\\
&-0.00000944547+0.0000070095\lambda^2\\
&-0.0000099211+0.00083298006\lambda^2\\
&-0.00001338899=0
\end{aligned}
$$

$$|R-\lambda^2 I| = 0.00000502386-0.003136328(\lambda^2)+0.1724449(\lambda^2)^2-(\lambda^2)^3=0 \quad (\mathbf{8})$$

وهي معادلة من الدرجة الثالثة بالنسبة للوسيط (λ^2) ، ولحلها وإيجاد جذورها نستخدم خوارزمية معادلات الدرجة الثالثة المبرمجة في الحواسيب والآلات الحاسبة فنحصل على الجذور الثلاثة (المرتبة تنازلياً) التالية:

$$\lambda_1^2 = 0.152033, \quad \lambda_2^2 = 0.018639, \quad \lambda_3^2 = 0.00177287$$

وهذه الجذور تسمى الجذور الكامنة أو القيم الذاتية eigenvalues للمصفوفة R وهي التي تقابل ثلاثة معاملات للارتباط القانوني للأزواج الثلاثة الممكنة وهي:

- الزوج الأول: $(U_1,\ V_1)$ ومعامل ارتباطه القانوني: $\rho_1 = \sqrt{\lambda_1^2} = 0.3899$
- الزوج الأول: $(U_2,\ V_2)$ ومعامل ارتباطه القانوني: $\rho_2 = \sqrt{\lambda_2^2} = 0.1365$
- الزوج الأول: $(U_3,\ V_3)$ ومعامل ارتباطه القانوني: $\rho_3 = \sqrt{\lambda_3^2} = 0.0421$

وهنا نلاحظ أن الزوج الأول هو أهم من الزوجين الآخرين لأنه يقابل معامل الارتباط القانوني الأكبر .

ولاختبار معنوية هذه المعاملات (أو القيم) نستخدم اختباري $wilk's$ وX^2 للتحقق من فرضية العدم التالية: إن معاملات الارتباط القانونية ρ_1, ρ_2, ρ_3 تساوي الصفر: $H_0: \rho_1 = \rho_2 = \rho_3 = 0$ مقابل الطريقة البديلة : أحدهما غير معدوم : H_1 ونحصل على الجدول التالي:

جدول (2): نتيجة اختبار معنوية معاملات الارتباط القانونية :

الاختبار / الجذور	Wilk's	Chi-SQ	DF د . ح	Sig = p	القرار
$\rho_1+\rho_2+\rho_3$	0.831	206.694	9	0.000	**معنوي**
$\rho_2+\rho_3$	0.980	22.956	4	0.000	**معنوي**
ρ_3	0.998	1.926	1	0.165	**غير معنوي**

من خلال قراءة السطر الأول في هذا الجدول نلاحظ أن قيمة p أقل من مستوى الدلالة (0.05) . وبما أن هذا السطر يدرس معنوية المعاملات الثلاثة من ρ_1 وحتى ρ_3 $(\rho_1$ وما يليه): لذلك نرفض H_0 ونستنتج أن بعض هذه المعاملات معنوي. أي أن بعضها يختلف عن الصفر .

ومن خلال قراءة السطر الثاني نلاحظ أن قيمة p أيضاً أقل من مستوى الدلالة (0.05) . وبما أن هذا السطر يدرس معنوية المعاملين ρ_2 و ρ_3 (ρ_2 وما يليه). لذلك فإننا نرفض H_0 ونستنتج أن أحد هذين المعاملين يختلف عن الصفر (معنوي) .

ومن خلال دراسة السطر الثالث نلاحظ أن قيمة p أكبر من مستوى الدلالة (0.05). ولكن بما أن هذا السطر يدرس معنوية ρ_3 فقط، نستنتج أن قيمة ρ_3 لا تختلف جوهرياً عن الصفر . ونستنتج أن الارتباط بين مركبي الزوج الثالث غير معنوي . ويمكن الاستغناء عنه. ولكننا سنحتفظ به حتى النهاية لأغراض التحليل .

ومما سبق نستنتج أن المعاملين ρ_1 و ρ_2 هما المعاملان المعنويان وأن الزوجين $(U_1,\ V_1)$ و $(U_2,\ V_2)$ هما الزوجان اللذان يعتمد عليهما في هذه الدراسة لتقدير العلاقة بين متحولات المجموعتين Y,X أو بين مركبيها القانونيين V, U .

وبعد ذلك ننتقل إلى حساب الأشعة f, e للتركيبين المعياريين (1),(2) المقابلين لكل قيمة من القيم الذاتية λ_1^2 ثم λ_2^2 ثم λ_3^2 وذلك من خلال المعادلة الذاتية (6) على التوالي:

فعندما نأخذ القيمة الذاتية الأولى $\lambda_1^2 = 0.152033$ ونعوضها في المعادلة (6) وبعد أن نطرح قيمة λ_1^2 من عناصر القطر الرئيسي لـ R نحصل على أن:

$$[R - I\lambda_1^2]e = \begin{bmatrix} -0.010496 & 0.0653574 & -0.0470465 \\ 0.012745 & -0.1371987 & 0.00465965 \\ -0.0135341 & 0.0015043 & -0.1359594 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (9)$$

ومنها نحصل على ثلاث معادلات خطية متجانسة وغير مستقلة (لأن المصفوفة شاذة ومحددها يساوي الصفر) وهي:

$$\begin{aligned} -0.010496e_1 + 0.0653574e_2 - 0.0470465e_3 &= 0 \\ 0.012745e_1 - 0.1371987e_2 + 0.00465965e_3 &= 0 \\ -0.0135341e_1 + 0.0015043e_2 - 0.1359594e_3 &= 0 \end{aligned} \quad (10)$$

وهي تعطينا لا نهاية من الحلول المقبولة للشعاع (e)، وحتى نحصل على شعاع محدد منها نستخدم معادلة الشرط المفروض على ذلك الشعاع (e) وهو أن يكون تباين U_{1zx} مساوياً للواحد أي أن:

$$var(U_{1zx}) = \grave{e} * R_{xx} * e = 1 \quad (11)$$

نعالج الشرط (11) فنجد أنه يساوي:

$$(e_1 e_2 e_3) * \begin{bmatrix} 1 & 0.4036 & -0.2438 \\ 0.4036 & 1 & -0.0819 \\ -0.2438 & -0.0819 & 1 \end{bmatrix} * \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = 1$$

$$(e_1 e_2 e_3) * \begin{bmatrix} e_1 & + & 0.4036e_2 & + & (-0.2438e_3) \\ 0.4036e_1 & + & e_2 & + & (-0.0819e_3) \\ -0.2438e_1 & + & (-0.0819e_2) & + & e_3 \end{bmatrix} = 1$$

$$e_1^2 + 0.4036e_1e_2 - 0.2438e_1e_3 + 0.4036e_2e_1 + e^2 - 0.0819e_2e_3$$
$$-0.2438e_1e_3 - 0.0819e_2e_3 + e_3^2 = 1$$

وبعد الاصلاح نجد أن الشرط (11) يأخذ الشكل التالي :

$$e_1^2 + e_2^2 + e_3^2 + 2(0.4036)e_1e_2 - 2(0.2438)e_1e_3 - 2(0.0819)e_2e_3 = 1 \quad (12)$$

نعود الآن إلى المعادلات (10) ونحسب من الأولى العنصر e_1 فنحصل على أن:

$$e_1 = 6.2268864e_2 - 4.4823266e_3 \qquad (13)$$

ثم نعوض ذلك في المعادلة الثانية (أو في الثالثة) فنحصل على أن:

$$-0.057837033e_2 - 0.0524676e_3 = 0 \qquad (14)$$

ومنها نحصل على أن:

$$e_2 = -0.907162717e_3 \qquad (15)$$

وبالتعويض في (13) نحصل على أن:

$$e_1 = -10.13112579e_3 \qquad (16)$$

ثم نقوم بتعويض e_1 و e_2 في معادلة الشرط (12) فنحصل على أن:

$$102.6397098e_3^2 + 0.82294419e_3^2 + e_3^2 + 7.41863585e_3^2$$
$$+0.4939936935e_3^2 + 0.148593253e_3^2 = 1$$

ومنها نحصل على أن:

$$116.96982 * e_3^2 = 1 \qquad (17)$$

$$e_3^2 = \frac{1}{116.96982} = 0.008592$$

وأخيراً نجد أن العنصر e_3 يساوي (نأخذ الاشارة الموجبة)

$$e_3 = +\sqrt{0.008592} = 0.09246 \qquad (18)$$

وبعد التعويض في (15) و(16) نحصل على أن:

$$\begin{matrix} e_2 = -0.083878 \\ e_1 = -0.936743 \end{matrix} \qquad (19)$$

وهكذا نكون قد حصلنا على عناصر الشعاع (e) المصاحب للمركب القانوني المعياري U_{1zx} ونكتبه بعد التقريب على الشكل التالي:

$$e^* = \begin{bmatrix} -0.937 \\ -0.084 \\ 0.092 \end{bmatrix} \qquad (20)$$

ولحساب الشعاع (f) المصاحب للمركب القانوني المعياري V_{1zy} نطبق العلاقة:

$$f^* = \frac{R_{yy}^{-1} * R_{yx}}{\lambda_1} * e^*$$

ومن الحسابات السابقة نجد أن:

$$f^* = \frac{1}{\sqrt{0.152033}} * \begin{bmatrix} 0.188713 & 0.0827317 & -0.1198787 \\ -0.3172233 & -0.1600840 & 0.0601632 \\ -0.004301 & 0.0836696 & 0.0790313 \end{bmatrix} * \begin{bmatrix} -0.937 \\ -0.084 \\ 0.092 \end{bmatrix}$$

$$f^* = \frac{1}{0.389914}\begin{bmatrix}-0.19480\\0.31622\\0.0042727\end{bmatrix} = \begin{bmatrix}-0.500\\-0.811\\0.011\end{bmatrix} \qquad (\mathbf{21})$$

وهكذا نكون قد حصلنا على الشعاعين e^* و f^* المصاحبين للزوج الأول المعياري (U_{1zx}, V_{1zy}) وبذلك يمكننا كتابة هذين المركبين القانونيين المعياريين كما يلي:

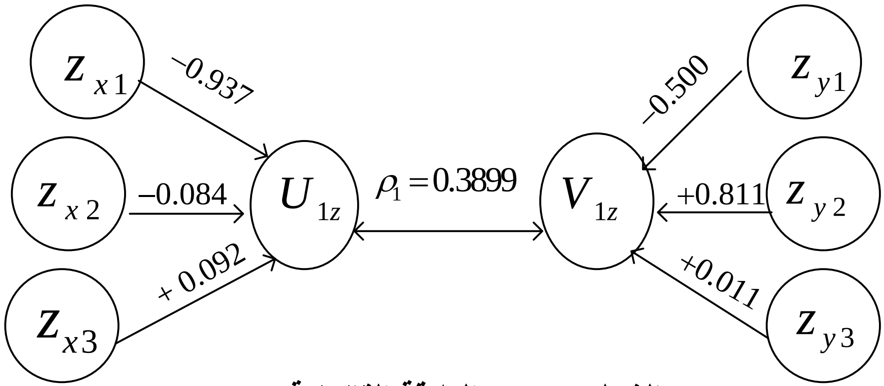
$$U_{1zx} = -0.937z_{x_1} - 0.084z_{x_2} + 0.092z_{x_3} \qquad (22)$$

$$V_{1zy} = -0.500z_{y_1} - 0.811z_{y_2} + 0.011z_{y_3} \qquad (23)$$

علماً بأن معامل الارتباط القانوني له يساوي:

$$\rho_1 = \sqrt{\lambda_1^2} = 0.3899$$

وبذلك يمكننا تمثيل العلاقة بين هذه المتحولات والمركبات للزوج الأول كما يلي:



الشكل (3-8): العلاقة القانونية

وهكذا نجد أنه أصبح بإمكاننا حساب القيم النظرية لمركبي الزوج الأول $(\tilde{U}_{1z}, \tilde{V}_{1z})$ ثم رسم الشكل البياني لهما وتحديد إشارة معامل الارتباط بينهما .

ولمتابعة عمليات التحليل علينا حساب الشعاعين المعياريين (f , e) المصاحبين للزوج الثاني (U_{2zx}, V_{2zy}) وذلك بتعويض القيمة الذاتية الثانية $\lambda_2^2 = 0.018639$ في المعادلة الذاتية (6) وإجراء الحسابات اللازمة لذلك، كما فعلنا في حالة الزوج الأول، ثم علينا حساب الشعاعين المعيارين (f,e) المصاحبين للزوج الثالث (U_{3zx}, V_{3zy}) وذلك بتعويض القيمة الذاتية الثالثة $\lambda_3^2 = 0.00177287$ في المعادلة الذاتية (6) وإجراء الحسابات اللازمة لذلك .

وفي النهاية نحصل على الأشعة المصاحبة للأزواج المعيارية الثلاثة والتي يطلق عليها اسم المعاملات القانونية المعيارية standard canonical coefficients .

ويتم تصنيفها حسب كل مجموعة في جدولين منفصلين كما يلي:

جدول (3): المعاملات القانونية المعيارية للمجموعة X (الأشعة e^*)

المتحولات \ المركبات	U_{1z}	U_{2z}	U_{3z}
Z_{x1}	−0.937	0.073	−0.161
Z_{x2}	−0.084	−0.699	0.836
Z_{x3}	0.092	−0.780	−0.668

جدول (4): المعاملات القانونية المعيارية للمجموعة Y (الأشعة f^*)

المتحولات \ المركبات	V_{1z}	V_{2z}	V_{3z}
Z_{y1}	−0.500	0.362	0.797
Z_{y2}	0.811	0.306	0.512
Z_{y3}	0.011	−0.881	0.477

ولحساب المعاملات القانونية a^* و b^* المصاحبة للمتحولات الأصلية X و Y نقوم بتقسيم كل عنصر من عناصر الأشعة e^* و f^* المصاحب للمتحول Z_{xi} على الانحراف المعياري للمتحول X_i وللمتحول Y_j المقابل له على التوالي، أي إننا نطبق العلاقتين التاليتين:

لكل سطر من الجدول (3)

$$a_{ik}^* = \frac{e_{ik}^*}{\sigma_{xi}^*} \qquad i = 1,2,3 \qquad (24)$$

لكل سطر من الجدول (4)

$$b_{jk}^* = \frac{f_{jk}^*}{\sigma_{yi}^*} \qquad j = 1,2,3 \qquad (25)$$

كما يمكننا حساب الأشعة a^* و b^* لجميع الأزواج القانونية الممكنة باتباع نفس الإجراءات المحددة في القسم النظري العام . وإجراء الحسابات المعقدة اللازمة لذلك .

وفي النهاية نحصل على المعاملات القانونية الخام للأزواج الثلاثة المصاحبة للمتحولات الأصلية في المجموعتين Y,X كما يلي :

جدول (5): المعاملات القانونية للمتحولات الأصلية X (Raw) للأشعة a^*

المتحولات \ المركبات	U_1	U_2	U_3
X_1	−0.310	0.024	−0.204
X_2	−0.016	−0.131	0.156
X_3	0.082	0.690	−0.591

جدول (6): المعاملات القانونية للمتحولات الأصلية Y (Raw) للأشعة b^*

المركبات / المتحولات	V_1	V_z	V_3
Y_1	−0.463	0.336	0.739
Y_2	0.662	0.250	0.418
Y_3	0.010	−0.803	0.435

بناءً على هذه النتائج يمكننا كتابة الزوج الأول (U_1 , V_1) بدلالة المتحولات الأصلية Y, X كما يلي:

$$U_1 = -0.310X_1 - 0.016X_2 + 0.082X_3 \qquad (26)$$

$$V_1 = -0.463Y_1 + 0.662Y_2 + 0.010Y_3 \qquad (27)$$

وإن قيمة معامل الإرتباط القانوني بينهما تساوي نفس القيمة السابقة :

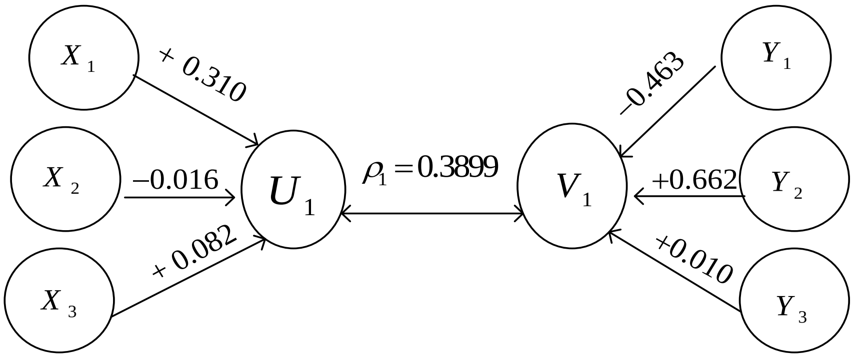
$$\rho_1^* = \sqrt{\lambda_1^2} = \sqrt{0.152033} = 0.3866$$

ومن المعادلتين (26) و(27) يمكننا حساب القيم النظرية (التنبؤية) للمركبين القانونين $(\widetilde{U}_1 , \widetilde{V}_1)$ وذلك بتعويض قيم المتحولات Y, X المعلومة من البيانات المتوفرة فنحصل على عمودين نظريين $(\widetilde{U}_1 , \widetilde{V}_1)$ مقابلين لبيانات Y, X .

ثم نقوم برسم شكل الانتشار للزوج $(\widetilde{U}_1 , \widetilde{V}_1)$ ونحسب قيمة معامل الارتباط القانوني النظري بينهما ونحدد إشارة ρ_1^* من تلك القيمة أو من شكل الانتشار .

وأخيراً يمكننا التعليق على المعادلتين (26) و(27) كما يلي:

- إنه إذا ازداد X_1 بمقدار الواحد فإن ذلك سيؤدي لانخفاض U_1 بمقدار (0.310)، وإذا ازداد X_2 بمقدار الواحد فإن U_1 ينخفض بمقدار (0.016)، أما إذا ازداد X_3 بمقدار الواحد فإن U_1 سيزداد بمقدار (0.082) .
- إنه إذا ازداد Y_1 بمقدار الواحد فإن V_1 سينخفض بمقدار (0.463)، ولكن إذا ازداد Y_2 بمقدار الواحد فإن V_1 سيزداد بمقدار (0.662)، وإذا ازداد Y_3 بمقدار الواحد فإن V_1 سيزداد بمقدار (0.010) فقط .
- كما يمكننا تمثيل العلاقة بين U_1 , V_1 على الشكل التالي :



الشكل (3-9): العلاقة القانونية بدلالة المتحولات الأصلية

وبطريقة مشابهة يمكننا معالجة الزوجين الآخرين (U_2, V_2) و (U_3, V_3) فنجد أنهما يكتبان كما يلي:

$$U_2 = 0.024X_1 - 0.131X_2 + 0.690X_3 \qquad (28)$$

$$V_2 = 0.336Y_1 + 0.250Y_2 - 0.803Y_3 \qquad (29)$$

وإن قيمة معامل الارتباط القانوني بينهما تساوي:

$$\rho_2^* = \sqrt{\lambda_2^2} = \sqrt{0.018639} = 0.1365$$ **(وهو معنوي)**

ولتحديد إشارة ρ_2^* علينا حساب القيم النظرية لـ $(\tilde{U}_2, \tilde{V}_2)$، نقوم برسم شكل الانتشار لهما وحساب معامل الارتباط بينهما .

أما الزوج الثالث فيكتب كمايلي:

$$U_3 = -0.204X_1 + 0.156X_2 - 0.591X_3 \qquad (30)$$

$$V_3 = 0.739Y_1 + 0.418Y_2 + 0.435Y_3 \qquad (31)$$

وإن قيمة معامل الارتباط القانوني لهما تساوي:

$$\rho_3^* = \sqrt{\lambda_3^2} = \sqrt{0.00177287} = 0.0421$$ **(وهو غير معنوي)**

ولتحديد إشارة ρ_3^* علينا حساب القيم النظرية لـ $(\tilde{U}_3, \tilde{V}_3)$ ثم رسم شكل الانتشار لهما وحساب معامل الارتباط بينهما ويمكننا متابعة تحليل كل زوج على حدة كما فعلنا مع الزوج الأول (U_1, V_1) ونترك ذلك للقارئ .

وحتى نتأكد من صحة الحسابات ومن دقة النماذج القانونية السابقة، علينا أن نتأكد من تحقيق الشروط (3−76) و(3−77) المفروضة على المركبات (U_k, V_ℓ) بأشكالها المختلفة، لذلك نقوم بحساب التباينات المشتركة لجميع الحالات الممكنة فنحصل على جدول كالجدول (6) السابق، الذي يتضمن 4 مصفوفات قطرية. ويتم حساب هذه التباينات المشتركة المختلفة اعتماداً على القيم النظرية للمركبات U_k و V_ℓ المحسوبة من النماذج القانونية المستخلصة .

ويمكن حسابها من علاقات مشابهة للعلاقة (22) والتي نصيغ أحدها كما يلي :

$$cov(U_k, V_\ell) = e_k' * R_{xy} * f_\ell \qquad k: 1,2,3 \ldots\ldots p$$

$$\ell: 1,2,3, \ldots\ldots p$$

وبناء على ذلك تم حساب عناصر المصفوفة القطرية الأولى فكانت كما يلي:

جدول ($6'$): التباينات المشتركة لـ U و V

المركبات	V_1	V_2	V_3
U_1	$\rho_1 = 0.3899$	0.0002	0.0001
U_2	0.0016	$\rho_2 = 0.1365$	0.0001
U_3	0.0045	0.0019	$\rho_3 = 0.0042$

وهي مصفوفة شبه قطرية تدل على تحقق الشروط (3-77) في الحالات المدروسة وإن عناصر قطرها الرئيسي تساوي معاملات الارتباط القانوني ρ_1, ρ_2, ρ_3 وذلك بسبب الشرط (3-76) .

حساب التحميلات المباشرة لمتحولات كل مجموعة:

أ- **التحميلات المباشرة لمتحولات المجموعة X** : هي عبارة عن معاملات الارتباط الزوجية بين تلك المتحولات X والمركبات القانونية الخاصة بها U_1, U_2, U_3 ولحساب هذه المعاملات لابد من حساب القيم النظرية للمركبات القانونية من العلاقات (26),(28),(30) وتنظيم جدول مناسب لهذه القيم النظرية مقابل القيم الفعلية لقيم المتحولات X_1, X_2, X_3 ثم القيام بحساب مصفوفة معاملات الارتباط الزوجي بين كل متحول من X مع المركبات القانونية U_1, U_2, U_3 فنحصل في النهاية على جدول يسمى جدول تحميلات X على U .

ولكن من جهة أخرى يمكن حساب هذه المعاملات الارتباطية لكل من X_1, X_2, X_3 مع المركب U_1 على انفراد من خلال جداء المصفوفة R_{xx} في الشعاع e^* المقابل له. أي أنه يمكننا أن نحصل على شعاع المعاملات مع U_1 من تطبيق العلاقة :

$$U_1 \Rightarrow \begin{bmatrix} r_{x11} \\ r_{x21} \\ r_{x31} \end{bmatrix} = R_{xx} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

وكذلك الأمر بالنسبة للمركبين U_2 و U_2، وبتطبيق ذلك على بيانات هذا المثال نجد أن شعاع معاملات ارتباط U_1 مع المتحولات X_1, X_2, X_3 يساوي :

$$U_1 \Rightarrow \begin{bmatrix} r_{x11} \\ r_{x21} \\ r_{x31} \end{bmatrix} = \begin{bmatrix} 1 & 0.4036 & -0.2438 \\ 0.4036 & 1 & -0.0819 \\ -0.2438 & -0.0819 & 1 \end{bmatrix} \begin{bmatrix} -0.937 \\ -0.084 \\ 0.092 \end{bmatrix} = \begin{bmatrix} -0.9933 \\ -0.4696 \\ 0.3273 \end{bmatrix}$$

أما شعاع معاملات ارتباط U_2 مع X_1, X_2, X_3 فتحسب من العلاقة :

$$U_2 \Rightarrow \begin{bmatrix} r_{x11} \\ r_{x21} \\ r_{x31} \end{bmatrix} = \begin{bmatrix} 1 & 0.4036 & -0.2438 \\ 0.4036 & 1 & -0.0819 \\ -0.2438 & -0.0819 & 1 \end{bmatrix} \begin{bmatrix} -0.073 \\ -0.699 \\ 0.780 \end{bmatrix} = \begin{bmatrix} -0.01895 \\ -0.60566 \\ -0.74055 \end{bmatrix}$$

وكذلك نحسب عناصر شعاع معاملات ارتباط U_3 مع X_1, X_2, X_3 ، من العلاقة :

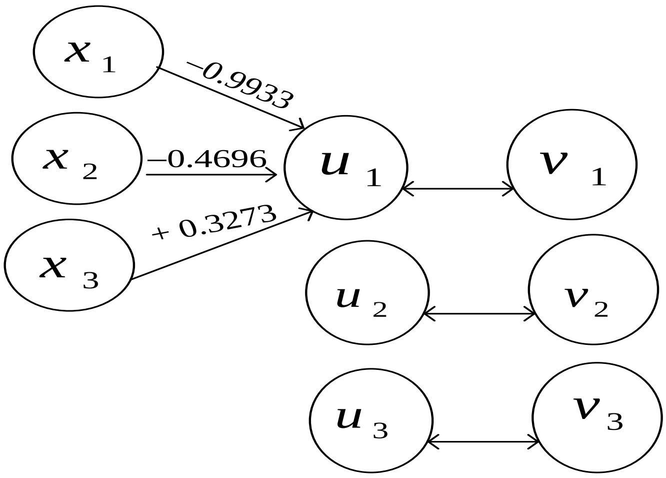
$$U_3 \Rightarrow \begin{bmatrix} r_{x11} \\ r_{x21} \\ r_{x31} \end{bmatrix} = \begin{bmatrix} 1 & 0.4036 & -0.2438 \\ 0.4036 & 1 & -0.0819 \\ -0.2438 & -0.0819 & 1 \end{bmatrix} \begin{bmatrix} -0.616 \\ 0.836 \\ -0.668 \end{bmatrix} = \begin{bmatrix} -0.11573 \\ 0.64209 \\ -0.58641 \end{bmatrix}$$

وبعد تقريب هذه المعاملات نحصل على تحميلات X على U كالتالي:

جدول (7) **تحميلات** X **على** U :

المركبات U / المتحولات X	U_1	U_2	U_3
X_1	−0.9933	−0.0190	−0.1157
X_2	−0.4696	−0.3057	0.6421
X_3	0.3273	−0.7406	−0.5864

وهكذا نجد أن المتحول الأكثر ارتباطاً بالمركب U_1 هو X_1 وبمعامل ارتباط سالب (-0.9933) وأكثرها ارتباطاً بـ U_2 هو X_3 وأكثرها ارتباطاً بـ U_3 هو X_3 ويمكن تمثيل تحميلات X على U_1 (فقط) كما يلي:



الشكل (3-10): تحميلات X على U_1 فقط

ب-**التحميلات المباشرة لمتحولات المجموعة Y:** وهي عبارة عن معاملات الارتباط الزوجية بين تلك المتحولات Y والمركبات القانونية الخاصة بها وهي V_1, V_2, V_3 ولحساب هذه المعاملات يجب حساب القيم النظرية للمركبات V_1, V_2, V_3 من العلاقات (27) , (29) , (31) وتنظيم جدول مناسب لها مقابل لقيم المتحولات Y ثم القيام بحساب مصفوفة معاملات الارتباط بينها فنحصل على جدول يسمى جدول تحميلات Y على V .

ولكن من جهة أخرى يمكن حساب هذه المعاملات الزوجية لكل من Y_1, Y_2, Y_3 مع المركب V_1 على حدة من خلال جداء المصفوفة R_{yy} في الشعاع f^* المقابل له . أي من خلال العلاقة :

$$V_1 \Rightarrow \begin{bmatrix} \grave{r}_{y11} \\ \grave{r}_{y21} \\ \grave{r}_{y31} \end{bmatrix} = R_{yy} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$$

وكذلك الأمر بالنسبة للمعاملات مع المركبين V_2 , V_3 وبتطبيق ذلك على بيانات هذا المثال نجد أن شعاع معاملات ارتباط V_1 مع المتحولات Y_1 , Y_2 , Y_3 يساوي :

$$V_1 \Longrightarrow \begin{bmatrix} \grave{r}_{y11} \\ \grave{r}_{y21} \\ \grave{r}_{y31} \end{bmatrix} = \begin{bmatrix} 1 & 0.4036 & -0.0570 \\ -0.1133 & 1 & 0.0222 \\ -0.0570 & -0.0819 & 1 \end{bmatrix} \begin{bmatrix} -0.500 \\ 0.811 \\ 0.011 \end{bmatrix} = \begin{bmatrix} -0.5925 \\ 0.8679 \\ 0.0575 \end{bmatrix}$$

أما شعاع معاملات ارتباط V_2 مع Y_1 , Y_2 , Y_3 فتحسب من العلاقة :

$$V_2 \Longrightarrow \begin{bmatrix} \grave{r}_{y11} \\ \grave{r}_{y21} \\ \grave{r}_{y31} \end{bmatrix} = \begin{bmatrix} 1 & 0.4036 & -0.0570 \\ -0.1133 & 1 & 0.0222 \\ -0.0570 & -0.0819 & 1 \end{bmatrix} \begin{bmatrix} 0.362 \\ 0.306 \\ -0.881 \end{bmatrix} = \begin{bmatrix} -0.3775 \\ 0.2454 \\ 0.8948 \end{bmatrix}$$

وإن شعاع معاملات ارتباط V_3 مع Y_1 , Y_2 , Y_3 فتحسب من العلاقة :

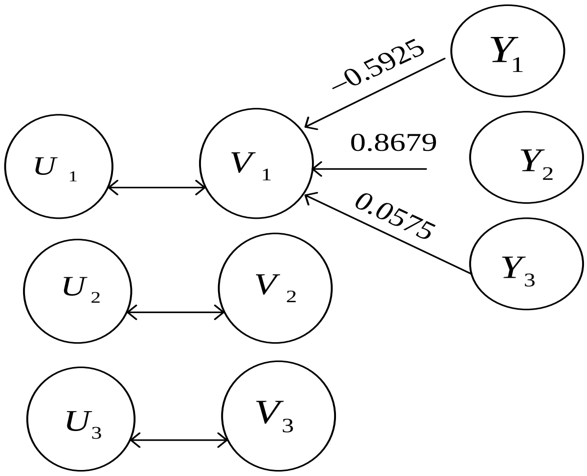
$$V_3 \Longrightarrow \begin{bmatrix} \grave{r}_{y11} \\ \grave{r}_{y21} \\ \grave{r}_{y31} \end{bmatrix} = \begin{bmatrix} 1 & 0.4036 & -0.0570 \\ -0.1133 & 1 & 0.0222 \\ -0.0570 & -0.0819 & 1 \end{bmatrix} \begin{bmatrix} 0.797 \\ 0.512 \\ 0.477 \end{bmatrix} = \begin{bmatrix} 0.7118 \\ 0.4323 \\ 0.4429 \end{bmatrix}$$

وبذلك نحصل على تحميلات Y على V ونضعها في جدول كالتالي:

جدول (8): تحميلات Y على V :

المركبات V / المتحولات Y	V_1	V_2	V_3
$\boldsymbol{Y_1}$	−0.5925	0.3775	0.7118
$\boldsymbol{Y_2}$	0.8679	0.2454	0.4323
$\boldsymbol{Y_3}$	0.0575	−0.8948	0.4429

ويمكن تمثيل تحميلات Y على V_1 (فقط) كما يلي :



الشكل (3-11): تحميلات Y على V_1 فقط

ج- **حساب التحميلات العابرة لـ X على V** : وهي عبارة عن معاملات الارتباط الزوجية بين المتحولات X مع كل من المركبات المقابلة لها من الطرف الآخر V_1 , V_2 , V_3 وهي تحسب بعدة طرق هي:

- بعد حساب القيم النظرية للمركبات لـ V, V_2, V_3 يتم حساب معاملات ارتباطها مع Y_1, Y_2, Y_3 وتوضع في جدول منظم .
- يمكن حسابها من خلال الجداء $[r_{xv}] = R_{xy} * f_v$.
- ولكن الطريقة الأسهل لحسابها هو ضرب كل شعاع $[r_{xv}]$ من التحميلات المباشرة لـ U في معامل الارتباط القانوني ρ_k المقابل للزوج U_k, V_k وهكذا نجد أن تحميلات X على المركب V_1 نحسب من العلاقة:

$$\begin{bmatrix} r *_{x1} V_1 \\ r *_{x2} V_1 \\ r *_{x3} V_1 \end{bmatrix} = \rho_1 * \begin{bmatrix} r_{x11} \\ r_{x21} \\ r_{x31} \end{bmatrix} = 0.3899 * \begin{bmatrix} -0.993 \\ -0.4696 \\ 0.3273 \end{bmatrix} = \begin{bmatrix} -0.3873 \\ -0.1831 \\ 0.1276 \end{bmatrix}$$

وكذلك نحصل على التحميلات العابرة لـ X على V_2 حيث نجد أن:

$$\begin{bmatrix} r *_{x1} V_2 \\ r *_{x2} V_2 \\ r *_{x3} V_2 \end{bmatrix} = \rho_2 * \begin{bmatrix} r_{x12} \\ r_{x22} \\ r_{x32} \end{bmatrix} = 0.1365 * \begin{bmatrix} -0.0190 \\ -0.6057 \\ -0.7406 \end{bmatrix} = \begin{bmatrix} -0.0026 \\ -0.0827 \\ 0.1011 \end{bmatrix}$$

وكذلك نحصل على التحميلات العابرة لـ X على v_3 من العلاقة :

$$\begin{bmatrix} r *_{x1} V_3 \\ r *_{x2} V_3 \\ r *_{x3} V_3 \end{bmatrix} = \rho_3 * \begin{bmatrix} r_{x13} \\ r_{x23} \\ r_{x33} \end{bmatrix} = 0.0421 * \begin{bmatrix} -0.1157 \\ 0.6421 \\ -0.5864 \end{bmatrix} = \begin{bmatrix} -0.0049 \\ 0.0270 \\ -0.0247 \end{bmatrix}$$

وبذلك نكون قد حصلنا على التحميلات العابرة لـ X على V ونضعها في جدول مناسب كما يلي:

جدول (9): التحميلات العابرة لـ X على V

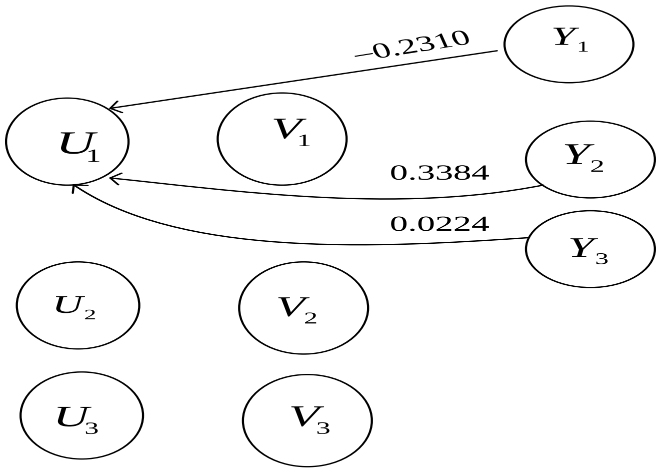
المركبات V / المتحولات Y	V_1	V_2	V_3
Y_1	−0.4873	−0.0026	−0.0049
Y_2	−0.1831	−0.0827	0.0270
Y_3	0.1276	0.1011	−0.0247

د- **التحميلات العابرة لـ Y على U ونحصل عليها بنفس الطريق السابقة** وإذا ضربنا أشعة التحميلات المباشرة لـV بمعاملات الارتباط القانوني الموافقة لها ρ_k نحصل على تحميلات Y على U وهي كما في الجدول التالي:

جدول (10): التحميلات العابرة لـ Y على U

المركبات V / المتحولات Y	U_1	U_2	U_3
Y_1	−0.2310	0.0515	0.0300
Y_2	0.3384	0.0335	0.0182
Y_3	0.0224	−0.1221	0.0186

ويمكن تمثيل هذه التحميلات العابرة لـ Y على U_1 (فقط) كمايلي:



الشكل (3-12) تحميلات Y على U_1 فقط

تقييم نماذج الأزواج U_k, V_k من التحميلات المباشرة :

نستخدم التحميلات المباشرة لتقييم نماذج الأزواج القانونية U_k, V_k، ويتم تقييم كل مركب U_k على حدة من خلال متوسط مربعات تحميلات هذه المركبات على المتحولات X .

كما يتم تقييم كل مركب V_k على حدة من خلال متوسط مربعات تحميلات هذه المركبات على المتحولات Y . وهناك عدة مؤشرات لتقييم النموذج المركب وهي الكفاءة والتشاركية والأهمية .

أ- **حساب كفاءة وأهمية المركبات U_k في النموذج U_k, V_k (Adequacy)**

لحساب هذه الكفاءة نحسب مربعات تحميلات X على U من الجدول (7) ونضعها في جدول كالتالي :

جدول (11): مربعات التحميلات X على U

المركبات V / المتحولات Y	U_1	U_2	U_3	المجموع= التشاركية
X_1	0.9866	0.00036	0.1339	1
X_2	0.2205	0.36687	0.41229	1
X_3	0.1071	0.54849	0.34386	1
المجموع= $\sum_{i=1}^{2} r_{ik}^2$	1.3142	0.91572	0.76954	3
كفاءة $U_k =$ $= \frac{\sum r_{ik}^2}{3}$	0.4381	0.30524	0.25651	1

لقد حسبنا على هامش هذا الجدول كفاءات المركبات U_k ووضعناها في السطر الأخير حيث قمنا بحساب متوسط مربعات تحميلات كل U_k على حدة من العلاقة :

$$Ad(U_k) = \frac{\sum_{i=1}^{3} r_{ik}^2}{3} \qquad k: 1,2,3$$

وأخيراً نلاحظ أن مجموع كفاءات المركبات U_k يساوي الواحد (تقريباً) وإن مجموع مربعات التحميلات في كل سطر X_i يساوي الواحد (تقريباً) وهذا يعني أن تشاركية كل متحول X_i في المركبات U_1 , U_2 , U_3 هي تشاركية كاملة. وإن المركبات U_1 , U_2 , U_3 تفسر كامل التغيرات في تغيرات كل X_i، ومن نفس الجدول يمكننا استخلاص الأهمية النسبية لكل متحول X_i بالنسبة لكل مركب U_k من العلاقة:

$$R(X_i, U_k) = \frac{r^2_{x_i u_k}}{\sum_{k=1} r^2_{ik}} = r^2_{x_i u_k} \quad : (\sum r^2_{ik} = 1 \text{ لأن})$$

أي أن الأهمية النسبية للمتحول X_1 تساوي (0.9866) على U_1 وتساوي (0.00036) على U_2 وتساوي (0.01339) على u_3 .

أما الأهمية النسبية لـ X_2 فهي تساوي (0.2205) على U_1 وتساوي (0.36688) على U_2 وتساوي (0.41229) على U_3 .

ولكن الأهمية النسبية لـ X_3 فهي تساوي (0.1071) على U_1 وتساوي (0.54849) على U_2 وتساوي (0.34386) على U_3 .

نلاحظ: إذا كان مجموع مربعات المعاملات في الأسطر X_i لا يساوي الواحد. فيجب عند حساب الأهمية النسبية لـ X_i تطبيق العلاقة السابقة وتقسيم مربع كل معامل $r^2_{x_i u_2}$ على مجموع مربعات المعاملات $\left(\sum_{k=1}^{p} r^2_{ik}\right)$ في سطر X_i .

ولكننا في هذا المثال لم نحتاج إلى ذلك لأن مجموع مربعات معاملات كل سطر X_i كان يساوي الواحد، وذلك لأننا لم نستغنى عن الزوج الثالث بسبب عدم معنويته وتركناه حتى الأخير لادخاله في عمليات التحليل .

ب- حساب كفاءة وأهمية المركبات V في النموذج (U_k , V_k) (Adequancy)

لحساب هذه الكفاءة نحسب مربعات تحميلات Y على V من الجدول (8) ونضعها في جدول كالتالي:

جدول (12): مربعات التحميلات Y على V

المركبات V / المتحولات Y	U_1	U_2	U_3	المجموع= التشاركية
Y_1	0.3511	0.1425	0.5067	1
Y_2	0.7533	0.0602	0.1869	1
Y_3	0.0033	0.8007	0.1962	1
المجموع $\sum r^2_{jk}$	1.1077	1.0034	0.8898	3
كفاءة $V_k = \frac{\sum r^2_{jk}}{3}$	0.3692	0.3345	0.2966	1

ولقد حسبنا على هامش هذا الجدول كفاءات المركبات V_k ووضعناها في السطر الأخير. حيث قمنا بحساب متوسط مربعات تحميلات كل V_k على حدة من العلاقة :

$$Ad(V_k) = \frac{\sum_{j=1}^{3} r_{jk}^2}{3} \qquad k: 1,2,3$$

وأخيراً نلاحظ أن مجموع كفاءات المركبات V_k يساوي الواحد وأن مجموع مربعات التحميلات في كل سطر Y_j يساوي الواحد . وهذا يعني أن تشاركية كل متحول Y_j في المركبات V_1, V_2, V_3 هي تشاركية كاملة . وأن المركبات V_1, V_2, V_3 تفسر كامل التغيرات في كل متحول Y_j .

ومن نفس الجدول يمكننا استخلاص الأهمية النسبية لكل متحول Y_j بالنسبة V_k من العلاقة :

$$R(Y_i, V_k) = \frac{r_{y_j v_k}^2}{\sum r_{jk}^2} = r_{y_j v_k}^2 \quad : (\sum_{k=1}^{2} r_{jk}^2 = 1 \text{ لأن})$$

أي أن الأهمية النسبية للمتحول Y_1 تساوي (0.3511) على V_1 وتساوي (0.1425) على V_2 وتساوي (0.5067) على V_3 . وهكذا بالنسبة للمتحولين Y_2 و Y_3، كما في السطرين الثاني والثالث .

وختاماً يمكننا استخلاص كفاءات الأزواج المركبة (U_k, V_k) ووضعها في جدول خاص كالتالي:

جدول (13): كفاءات المركبات في الأزواج (U_k, V_k)

رقم الزوج k	كفاءة U_k	كفاءة V_k
1	0.4381	0.3692
2	0.30524	0.3345
3	0.25651	0.2966
المجموع	1	1

إن العمود الأول يعبر عن نسبة التباين المفسر للمجموعة الأولى حسب المركبات U الخاصة بها، أما العمود الثاني فيعبر عن نسبة التباين المفسر للمجموعة الثانية حسب المركبات V الخاصة بها .

وهكذا نجد أن كفاءة الزوج الأول (U_1, V_1) هو الأفضل وأن كفاءة U_1 أفضل من كفاءة V_1، ويليه الزوج الثاني (U_2, V_2) الذي فيه كفاءة U_2 أقل كفاءة من V_2 ، ثم الزوج الثالث الذي فيه كفاءة U_3 أقل كفاءة من V_3. علما أن الزوج الثالث كان يجب الاستغناء عنه من البداية، لأن عملية اختبار الجذور الكامنة أو القيم الذاتية λ_k^2 اعطتنا أن القيمة الثالثة المقابلة له غير معنوية، ولكننا احتفظنا به حتى النهاية لتوضيح بعض جوانب التحليل القانوني .

تقييم نماذج الأزواج (U_k, V_k) من التحميلات العابرة cross loadings

لتقييم هذه النماذج من التحميلات العابرة يمكننا اتباع نفس الأسلوب الذي اتبعناه في التحميلات المباشرة والحصول على جدول يتضمن مربعات تلك التحميلات لكل مجموعة على مركباتها .

ولكننا هنا سنقوم بحساب الكفاءات العابرة (Redundancy) للمركبات في الأزواج القانونية (U_k, V_k) مباشرة من الكفاءات المباشرة السابقة باستخدام العلاقتين التاليتين :

الكفاءة العابرة لـ U_k $\qquad Red(U_k) = Ad(U_k) * \rho_k^2$

الكفاءة العابرة لـ V_k $\qquad Red(V_k) = Ad(V_k) * \rho_k^2$

وبما أنه لدينا مما سبق أن قيم ρ_k^2 تساوي:

$$\rho_3^2 = \lambda_3^2 = 0.00177287, \quad \rho_2^2 = \lambda_2^2 = 0.018639, \quad \rho_1^2 = \lambda_1^2 = 0.152033$$

واعتماداً على الجدول (13) السابق نحصل على أن الكفاءات العابرة لـ (U_k, V_k)

جدول (14) الكفاءات العابرة للمركبات V_k, U_k في الأزواج (U_k, V_k)

رقم الزوج k	U_k	V_k
1	0.0666	0.05613
2	0.0057	0.00623
3	0.00045	0.00053
المجموع	$\neq 1$	$\neq 1$

وهنا نلاحظ أن مجموع كل منها لا يساوي الواحد. وهي تعبر عن كفاءات عابرة ضعيفة للمتحولات X أو Y على المركبات المقابلة لها في الطرف الآخر، أي تعبر عن نسبة التباين المفسر في كل مجموعة X أو Y حسب المركبات القانونية المقابلة لها .

المــراجـــع

1- ANDERSON T.W. An Introduction to multivariate statistical Analysis. Sec. Ed. John Wiley and sous- New York 2002.
2- Johnson, R. Wichern, D. Applied multivariate stattcal Analysis 1988 P. Hall. IE.
3- Richard J. Harris. A primer of multivariate statictics edu.3-2001 L.E.A London.
4- David R. Hardoon, Sandor Szedmak, John shawetoylar : canonical correlation; An overview with application to learning mothods department of computer Science- University of London Technical Report. CSD- TR- 03- 08/2003.
5- David weenink: Canonical correlation Analysis, Institute of phontic science, university of Amsterdam, IFA Proceeding 25-2003.
6- Magnus Borga: Canonical correlation a Tutorial January 12, 2001- LU- Sweden.
7- Francis R. Bach, Michael I. Jordan: A probabilistic Interpretation of canonical correlation. Analysis, computer science division, university California, Berkeley CA 9444 USA. Lancaster, K: Mathematical Economics, London Mossow New york cof, pal 1972.
8- Mons Thalin; Analysis of Factors and canonical correlation; Department of mathematics, uppsal university thalin @. Math. Uu.Se.-2011.

9-الطويل، مجدي: المصفوفات (النظرية والتطبيق)- جامعة القاهرة- كلية الهندسة عام .

10- العشعوش، أيمن+ العربيد، عدنان: الاقتصاد القياسي- كلية الاقتصاد- جامعة تشرين- سورية- 2015 .

11- العلي، إبراهيم+ عكروش، محمد: الاحصاء التطبيقي- كلية الاقتصاد- جامعة تشرين- سورية 2005 .

12- العلي، إبراهيم+ كابوس، أمل: الاحصاء الرياضي- كلية الاقتصاد- جامعة حلب- سورية- عام 1986 .

13- عناني، محمد عبد السميع: التحليل القياسي والاحصائي للعلاقات الدولية- الدار الجامعية- الاسكندرية- مصر 2009 .

أسس
التحليل الإحصائي متعدد المتغيرات

الجزء الرابع
التحليل العاملي Factor Analysis

الفصل الرابع : طرائق حساب العوامل العامة F_P بدلالة المتحولات Z_P:

المراجع المعتمدة:

الفصل الأول

مفاهيم التحليل العاملي Factor Analysis

1-1: تمهيد:

يُعتبر التحليل العاملي (العواملي) فرعاً من فروع الإحصاء الرياضي، وأسلوباً مهماً من الأساليب المستخدمة في التحليل متعدد المتغيرات. وهو يتناول بشكل خاص دراسة العلاقات الارتباطية بين عدة متحولات، منتمية إلى بيئة واحدة ومرتبطة مع بعضها البعض بصورة مباشرة أو غير مباشرة، مثل $X_1, X_2, X_3, \dots \dots, X_n$ (حيث n عدد المتحولات)، ويعمل على التعبير عن تلك المتحولات بدلالة عدد قليل من المتحولات الافتراضية (السرية وغير المعروفة) تسمى بالعوامل العامة (Common Factors)، ويستخدم لذلك نماذج أو معادلات خطية محددة، دون أن يخسر كثيراً من المعلومات المتوفرة عن تلك المتحولات . ويقسم التحليل العاملي إلى قسمين [Okon, Jan p.23] هما:

- التحليل العاملي الاستكشافي Exploratory Factor Analysis :

ويستخدم كثيراً في التطبيقات العملية وهو يقوم بدراسة العلاقات بين المتحولات عبر تلك النماذج دون استخدام فرضيات مسبقة عنها. ثم يستخدم النتائج التي يتم التوصل إليها لاستكشاف العلاقات الموجودة بين تلك المتحولات واستخلاص العوامل المؤثرة عليها .

- التحليل العاملي التوكيدي Confirmatory Factor Analysis :

وهو يعتمد على فرضيات مسبقة عن تلك المتحولات والعلاقات بينها والعوامل المعبرة عنها والعمل على اختبار تلك الفرضيات واستخلاص النتائج. ولن نتعرض له في هذا الكتاب.

ويهدف التحليل العاملي إلى استخلاص تلك المتحولات الافتراضية السرية (العوامل)، وتحديد النماذج الرياضية الخطية التي تعبر عن المتحولات الأساسية $X_1, X_2, X_3, \dots \dots, X_n$ بدلالة تلك العوامل أفضل تعبير، ثم حساب قيم العوامل بدلالة المتحولات المدروسة .

وهنا تبرز أمامنا الأسئلة التالية:

- ماهي هذه العوامل الافتراضية (Factors) وكم يجب أن يكون عددها حتى تعبر بأعلى دقة ممكنة عن العلاقات المشاهدة بين المتحولات المدروسة؟
- كيف يمكن أن نتعامل أو نعالج كميات كبيرة من البيانات العددية المأخوذة من المتحولات المدروسة لتحويلها ما أمكن إلى تكوين بسيط وبأقل خسارة ممكنة من المعلومات؟
- ماهي أبسط الصيغ الرياضية لتمثيل علاقة المتحولات المدروسة $X_1, X_2, X_3, \dots \dots, X_n$ بتلك العوامل الافتراضية المجهولة والتي سنرمز لها بـ $F_1, F_2, F_3, \dots \dots, F_p$ ؟

إن الاجابة على هذه الأسئلة وغيرها سترد في الفقرات اللاحقة.

ويعتمد التحليل العاملي على البيانات المتوفرة عن المتحولات المدروسة، وإن أهم تلك البيانات هي عناصر المصفوفة الارتباطية R بين المتحولات الكمية المدروسة X. ويمكن استبدالها بعناصر مصفوفة التباين والتباين المشترك Covariance، ويعمل على الاستفادة من عناصر المصفوفة R في الاجابة على الأسئلة السابقة وإيجاد العلاقات الرياضية، التي تربط تلك المتحولات X بالعوامل الافتراضية F .

مثال (1-1):

لتوضيح ذلك نفترض أنه لدينا أربعة متحولات كمية عن بعض صفات الإنسان البالغ [Uberla,p.55] ولتكن (بتصرف)الصفات التالية:

X_1 : طول الشخص البالغ (سم)

X_2 : وزن الشخص البالغ (كغ)

X_3 : مقدار ضغط الدم الانقباضي (ملم)

X_4 : مقدار ضغط الدم الانبساطي (ملم)

ولنفرض إننا سحبنا في زمان ومكان معين عينة عشوائية كبيرة من الأشخاص (بحجم N)، وقمنا بقياس هذه المتحولات عندهم ، فحصلنا على جدول من البيانات العددية يضم القيم المتقابلة للمتحولات الأربعة المدروسة ولجميع هؤلاء الأشخاص، ثم قمنا بحساب القيم الوصفية لهذه المتحولات وحسبنا المصفوفة الارتباطية لها، واختصاراً لنفترض أنه عند حساب المصفوفة الارتباطية لهذه البيانات الميدانية حصلنا على المصفوفة الارتباطية التالية :

$$R = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{bmatrix} 1 & 0.7225 & 0.0850 & 0.0800 \\ 0.7225 & 1 & 0.0800 & 0.0750 \\ 0.0850 & 0.0800 & 1 & 0.5625 \\ 0.0800 & 0.0750 & 0.5625 & 1 \end{bmatrix} \end{array}$$

وعند دراسة عناصر هذه المصفوفة نلاحظ أن المتحول X_1 يرتبط مع X_2 بعلاقة قوية (0.7225)، ولا يرتبط مع المتحولين X_4, X_3 . كما إن المتحول X_3 يرتبط مع المتحول X_4 بعلاقة ضعيفة (0.5625) ولا يرتبط مع المتحولين X_2, X_1، وبما أن المصفوفة الارتباطية هي مصفوفة متناظرة فإنه يمكننا أن نعيد مضمون الملاحظة السابقة على نفس المتحولات بطريقة معاكسة (أو متناظرة) ونقول أن X_2 مرتبط مع X_1، وأن X_4 مرتبط بعلاقة ضعيفة مع X_3 .

والسؤال الآن: هل يمكننا التعبير عن عناصر هذه المصفوفة الارتباطية R بطريقة أبسط من ذلك، بحيث نحصل على تقدير لـ R بأقل خسارة ممكنة؟

إن الإجابة على هذا السؤال هي موضوع التحليل العاملي. لذلك سنحاول أن نعبر عن المصفوفة السابقة R من خلال جداء شعاع عمود a بمنقوله a' المعرف كما يلي:

$$a.a' = \begin{bmatrix} 0.90 \\ 0.80 \\ 0.50 \\ 0.05 \end{bmatrix} [0.90 \quad 0.80 \quad 0.50 \quad 0.05]$$

وبإجراء عمليات ضرب عناصر الأسطر بعناصر الأعمدة نحصل على المصفوفة $\breve{R}_1$ التالية:

$$a.a' = \begin{bmatrix} (0.810) & 0.720 & 0.450 & 0.045 \\ 0.720 & (0.640) & 0.400 & 0.040 \\ 0.450 & 0.400 & (0.250) & 0.025 \\ 0.045 & 0.040 & 0.025 & (0.003) \end{bmatrix} = \breve{R}_1$$

وهي مصفوفة لا تعبر بشكل جيد عن المصفوفة الأصلية R، بسبب اختلاف العناصر القطرية فيها عن الواحد واختلاف بعض العناصر الأخرى . لذلك سنكرر المحاولة ونأخذ شعاعين عمودين أخريين a_2, a_1 ونضعهما ضمن مصفوفة مستطيلة نرمز لها بـ A ثم نحسب جداء A بمنقولها A' فنحصل على مصفوفة جديدة نرمز لها بـ $\breve{R}_2$ كما يلي:

$$A.A' = \begin{array}{c} \begin{matrix} a_1 & a_2 \end{matrix} \\ \begin{bmatrix} 0.90 & 0.05 \\ 0.80 & 0.05 \\ 0.05 & 0.80 \\ 0.05 & 0.70 \end{bmatrix} \end{array} \begin{bmatrix} 0.90 & 0.80 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.80 & 0.70 \end{bmatrix}$$

وبإجراء عمليات ضرب الأسطر بالأعمدة نحصل على المصفوفة المقدرة $\breve{R}_2$ التالية :

$$A.A' = \begin{bmatrix} (0.8125) & 0.7225 & 0.0850 & 0.0800 \\ 0.7225 & (0.6425) & 0.0800 & 0.0750 \\ 0.0850 & 0.0800 & (0.6425) & 0.5625 \\ 0.0800 & 0.0750 & 0.5625 & (0.4925) \end{bmatrix} = \breve{R}_2 = R_h$$

علماً بأنه تم حساب عناصر المصفوفة $\breve{R}_2$ من جراء حساب الجداء المصفوفي العادي (مجموع جداء عناصر السطر بعناصر العمود). فمثلاً العنصر الأول من السطر الأول في $\breve{R}_2$ تم حسابه من جداء عناصر السطر الأول من A بعناصر العمود الأول من A' كما يلي:

$$(0.90*0.90+0.05*0.05=0.8125)$$

وحساب العنصر الثاني من السطر الأول من $\breve{R}_2$ كما يلي:

$$(0.09 * 0.80 + 0.05 * 0.05 = 0.7225)$$

وهكذا دواليك .

وعند مقارنة المصفوفة $\breve{R}_2$ مع المصفوفة الأصلية R نجد أن معظم عناصرها تتطابق تقريباً مع عناصر المصفوفة R (وذلك باستثناء عناصر القطر الرئيسي التي وضعناها ضمن أقواس لأنها مازالت تختلف عن الواحد الذي في R الأصلية)، وإن عناصر هذا القطر الرئيسي الموضوعة ضمن قوسين تسمى بالتشاركيات (Commonalty) المقابلة لكل متحول من المتحولات المدروسة X_1, X_2, X_3, X_4 ، وإن

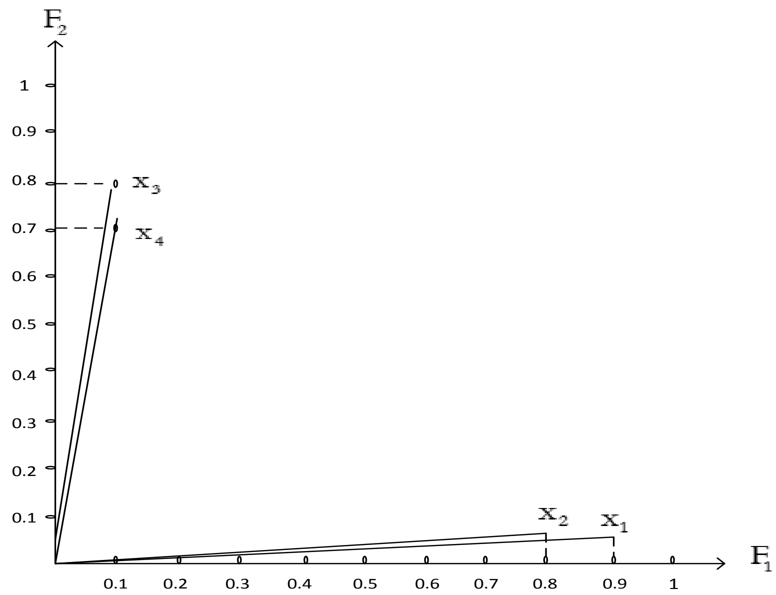
فروقات هذه التشاركيات عن الواحد يعبر عن مقدار الخسارة أو الخطأ المرتكب في تقدير R من خلال $\breve{R}_2$ وما شابهها.

وهكذا نجد أنه يمكن استخدام مصفوفة صغيرة ذات شعاعين فقط مثل A في التعبير عن المصفوفة الارتباطية R (باستثناء العناصر القطرية) ونكتبها كما يلي:

$$A = \begin{array}{c c} & \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} & \begin{bmatrix} 0.90 & 0.05 \\ 0.80 & 0.05 \\ 0.05 & 0.80 \\ 0.05 & 0.70 \end{bmatrix} \end{array}$$

كما يمكن رسم أسطر هذه المصفوفة على شكل بياني، وذلك باعتبار المتحولات المدروسة كأشعة سطرية في A وإن عناصر كل سطر في A يعبر عن مسقطي المتحول المقابل له X_i على محورين احداثيين افتراضيين نرمز لهما بـ F_1 و F_2 بدلاً من الشعاعين a_1 و a_2 .

وبفرض أن هذين المحورين الافتراضين F_1 و F_2 متعامدان نقوم برسم هذه المتحولات عليهما فنحصل على الشكل التالي:



الشكل (1-1): التمثيل البياني للمتحولات على العوامل

وبذلك نحصل على أربعة أشعة هي:

$X_1(0.90\,,0.05)$, $X_2(0.80\,,0.05)$

$X_3(0.05\,,0.80)$, $X_4(0.05\,,0.70)$

وهنا نلاحظ أن المتحولين X_1 و X_2 محمولان تقريباً على محور العامل F_1، أي أنهما يشكلان عاملاً خاصاً يمكن أن نسميه عامل الخواص الجسمية، وأن المتحولين X_3 و X_4 محمولان تقريباً على محور العامل F_2، أي يشكلان عاملاً خاصاً يمكن أن نسميه عامل خواص ضغط الدم، وبصورة عامة قد نحصل على أشعة متعددة وقد تكون موزعة على المستوى أو الفضاء المستخدم بشكل آخر .

وهكذا نجد أنه يمكننا إيجاد مصفوفة مثل A لتقدير معظم عناصر المصفوفة الارتباطية R (باستثناء عناصر قطرها الرئيسي). ولإيجاد حل لمشكلة هذه العناصر الرئيسية كان لابد من تعديل الأسلوب المتبع في عملية التقدير ليشمل جميع عناصر المصفوفة.

ولكن بما أن الجداء $A.A'$ يعطينا مصفوفة أخرى $\breve{R}_2$ تختلف عن المصفوفة الأصلية R، بأن عناصر قطرها الرئيسي تساوي قيم مختلفة عن الواحد وتسمى بالتشاركيات الخاصة بالمتحولات المدروسة $\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \mathrm{X}_4$ ، فإنه يمكننا استخدام الجداء $A.A'$ لتقدير مصفوفة أخرى، معدلة عن المصفوفة الارتباطية R، وذلك باستبدال عناصر قطرها الرئيسي المساوية للواحد بقيم التشاركيات لتلك المتحولات ، و تسمي المصفوفة المؤلفة بهذه الطريقة بالمصفوفة المخففة Reduction matrix ونرمز لها بـ R_h .

ولذلك فإن التحليل العاملي ينطلق من المصفوفة المخففة R_h المحسوبة من المصفوفة الارتباطية R باستبدال عناصر قطرها الرئيسي (المساوية للواحد) بقيم التشاركيات المحسوبة أو المقدرة لكل متحول مدروس. ويعمل على تقدير عناصر هذه المصفوفة المخففة بواسطة مصفوفة أخرى A أقل رتبة. وسنتعرض لاحقاً لأساليب حساب أو تقدير هذه التشاركيات لتعويضها في المصفوفة R ولاستخلاص العوامل منها.

ولكن العوامل المستخلصة من المصفوفة A للتعبير عن R_h ليست وحيدة، لابل هناك لانهاية من الحلول الممكنة لذلك . وحتى نستخلص حلاً وحيداً لهذه المسألة، نقوم بمعالجة الأمر من خلال تدوير المحاور الاحداثية عدداً كبيراً من المرات حتى نحصل على أبسط وأفضل حل لها ، ونعتبره الحل النهائي لهذه المسألة .

وأخيراً ينتقل التحليل العاملي إلى دراسة علاقات العوامل الافتراضية F مع المتحولات المدروسة X والحصول على المصفوفة الهيكلية، ثم يعمل على تقدير قيم العوامل نفسها المقابلة لقيم المتحولات المدروسة .

وهكذا نجد أنه عند معالجة أية مسألة عاملية تواجهنا عدة مشكلات هي:

1- مشكلة جمع البيانات عن المتحولات المدروسة ومعالجتها حسب الحاجة .
2- مشكلة حساب المصفوفة الارتباطية أو مصفوفة التباينات المشتركة .
3- مشكلة حساب أو تقدير قيم التشاركيات القطرية في المصفوفة المخففة R_h .
4- مشكلة تحديد عدد العوامل ثم تحديد نوعية كل منها .
5- مشكلة حساب المصفوفة A لتقدير المصفوفة R_h .
6- مشكلة تدوير المحاور المتعامدة والمائلة للحصول على الحل النهائي .
7- مشكلة حساب قيم العوامل بدلالة قيم المتحولات المدروسة .
8- مشكلات أخرى خاصة .

وهذا ما سنعالجه في الفقرات اللاحقة بعد استعراض بعض المفاهيم الرياضية للتحليل العاملي .

1−2: المفاهيم الرياضية لتحليل العوامل (العواملي) ([Uberla,p.61]، بتصرف):

لنفترض أنه لدينا n متحولاً كمياً، عن ظاهرة ما في مجتمع معين، ونرمز لها بـ:Xn, ..., X_1, X_2, X_3،
وإنه جمعنا عن كل منها N مشاهدة متقابلة وكانت قيمها ومتوسطاتها وتبايناتها وانحرافاتها المعيارية، هي المعروضة في الجدول التالي:

جدول (1−1): البيانات الأساسية للمتحولات الأصلية للمتحولات: $\mathbf{X_1, X_2, \dots, X_n}$.

المشاهدات / المتحولات	1	2	3	....	N	المتوسطات $\bar{x}_j$	التباينات $S^2_{(j)}$
X_1	x_{11}	x_{12}	x_{13}	....	x_{1N}	$\bar{x}_1$	$S^2_{(1)}$
X_2	x_{21}	x_{22}	x_{23}	....	x_{2N}	$\bar{x}_2$	$S^2_{(2)}$
X_3	x_{31}	x_{23}	x_{33}	....	x_{3N}	$\bar{x}_3$	$S^2_{(3)}$
⋮	⋮	⋮	⋮	....	⋮	⋮	⋮
X_n	x_{n1}	x_{n2}	x_{n3}	....	x_{nN}	$\bar{x}_n$	$S^2_{(n)}$

ويفترض أن يكون كل من هذه المتحولات خاضعاً لتوزيع طبيعي محدد $N(\mu_j\ \sigma_j^2)$ توقعه μ_j وتباينه σ_j^2 في المجتمع المدروس،(نلفت الانتباه إلى أن الرمز n هو عدد المتحولات، بينما N هو حجم العينة).

علماً بأن متوسطات العينة لهذه المتحولات تحسب من العلاقة :

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji} \qquad (j{:}\ 1\ \ 2\ \ 3 \ldots\ldots n) \qquad (1-1)$$

وأن التباينات والانحرافات المعيارية لعينة المشاهدة تحسب من العلاقة :

$$S_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_{ji} - \bar{x}_j\right)^2 \qquad (j{:}\ 1\ \ 2\ \ 3 \ldots\ldots n) \qquad (2-1)$$

$$S_j = +\sqrt{S_j^2} \qquad (j{:}\ 1\ \ 2\ \ 3 \ldots\ldots n) \qquad (3-1)$$

وأن التباين الإجمالي S^2 يساوي مجموع تباينات هذه المتحولات :

$$S^2 = \sum_{j=1}^{n} S_j^2 \qquad (4-1)$$

وبناءً على هذه البيانات المأخوذة من عينة المشاهدات يتم حساب جميع معاملات الارتباط الزوجية بين أي متحولين X_j و X_k من العلاقة :

$$r_{jk} = \frac{\sum^{N}\left(x_{ji} - \bar{x}_j\right)\left(x_{ki} - \bar{x}_k\right)}{(N-1) * S_j * S_k} \qquad (5-1)$$

وبالتالي نحصل على المصفوفة الارتباطية R التي تأخذ الشكل التالي:

$$R = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & & X_n \end{array} \\ \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & 1 & \dots & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{bmatrix} \end{array} \qquad (6-1)$$

وهي مصفوفة من المرتبة $n*n$، وتتميز بأنها متناظرة وإن جميع عناصر قطرها الرئيسي تساوي الواحد.
وحتى نتخلص من تأثير واحدات القياس المختلفة لهذه المتحولات نقوم بحساب القيم المعيارية $\mathfrak{z}_{ji}$ للمشاهدات x_{ij} الواردة في الجدول السابق وذلك باستخدام علاقة المعيرة التالية :

$$\mathfrak{z}_{ji} = \frac{x_{ji} - \bar{x}_j}{S_j} \qquad \begin{array}{l} i = 1\ 2\ 3 \ldots..N \\ j = 1\ 2\ 3 \ldots..n \end{array} \qquad (7-1)$$

وبذلك نحصل على جدول مشابه للجدول (1−1) ويتضمن القيم المعيارية $\mathfrak{z}_{ij}$ لجميع المتحولات المعيارية المقابلة للمتحولات الأصلية ونرمز لها بـ $Z_1, Z_2, , Z_3, \dots, Z_n$.

جدول (1−2): قيم المتحولات المعيارية Z_j المقابلة للمتحولات الأصلية x_j :

المشاهدات / المتحولات المعيارية	1	2	3	....	N	$\bar{Z}_j$	التباينات $S^2_{(Zj)}$
Z_1	$\mathfrak{z}_{11}$	$\mathfrak{z}_{12}$	$\mathfrak{z}_{13}$	....	$\mathfrak{z}_{1N}$	0	1
Z_2	$\mathfrak{z}_{21}$	$\mathfrak{z}_{22}$	$\mathfrak{z}_{23}$	....	$\mathfrak{z}_{2N}$	0	1
Z_3	$\mathfrak{z}_{31}$	$\mathfrak{z}_{23}$	$\mathfrak{z}_{33}$	....	$\mathfrak{z}_{3N}$	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$	$\vdots$
Z_n	$\mathfrak{z}_{n1}$	$\mathfrak{z}_{n2}$	$\mathfrak{z}_{n3}$	....	$\mathfrak{z}_{nN}$	0	1

وبما أننا افترضنا أن المتحولات الأصلية خاضعة لتوزيعات طبيعية، فإن كل من المتحولات المعيارية $Z_1, Z_2, Z_3 \ldots\ldots..Z_n$ تكون خاضعة للتوزيع الطبيعي المعياري المعروف بالرمز $N(0,1)$، أي أن متوسطات هذه المتحولات المعيارية Z_j تساوي الصفر وإن تبايناتها وانحرافاتها المعيارية تساوي الواحد .

أي أنه سيكون لدينا نتيجة التحويل الشروط التالية:

$$\bar{Z}_j = 0 \quad , \qquad \boldsymbol{S^2_{(Zj)}} = 1 \qquad : (j{:}\,1\,2\,3..\ n) \qquad (8-1)$$

وبذلك نجد أن التباين الإجمالي S^2_Z للمتحولات المعيارية يساوي :

$$S^2_Z = \sum_{j=1}^{n} \boldsymbol{S^2_{(Zj)}} = n \qquad (9-1)$$

أي أن التباين الإجمالي للمتحولات المعيارية Zj يساوي n (حيث أن n عدد المتحولات)

أما معامل الارتباط لأي زوج من المتحولات المعيارية Z_j و Z_k فيساوي نفس معامل الارتباط المحسوب من الزوج الأصلي المقابل X_j و X_k لأن :

$$r(Z_j\,,Z_k)=\frac{\sum^N(\mathfrak{z}_{ji}-\bar{\mathfrak{z}}_j)(\mathfrak{z}_{ki}-\bar{\mathfrak{z}}_k)}{(N-1)*S(Z_j)*S(Z_k)}=\frac{\sum^N(\mathfrak{z}_{ji}-0)(\mathfrak{z}_{ki}-0)}{(N-1)*1*1} \qquad (10-1)$$

$$=\frac{\sum \mathfrak{z}_{ji}\mathfrak{z}_{ki}}{N-1}=\frac{\sum_{i=1}^{N}\left(\frac{x_{ji}-\bar{x}_j}{S_i}\right)\left(\frac{x_{ki}-\bar{x}_k}{S_k}\right)}{(N-1)}=r(x_j,x_k)=r_{jk}$$

وعند حساب هذه المعاملات لجميع المتحولات المعيارية Z_k و Z_j (أو الأصلية X_k و X_j) نحصل على المصفوفة الارتباطية نفسها والتي رمزنا لها في (6–1) بالرمز R .

ومن (10–1) يمكننا أن نستنتج أن كل عنصر r_{jk} في المصفوفة الارتباطية R يحسب من العلاقة التالية (بدلالة قيم Z_j و Z_k) :

$$r_{jk}=\frac{1}{N-1}\sum_{i=1}^{N}\mathfrak{z}_{ji}*\mathfrak{z}_{ki} \qquad (11-1)$$

وإذا وضعنا قيم المتحولات المعيارية Z_j الواردة في الجدول *(2–1)* على شكل مصفوفة $\mathfrak{z}$ ذات n سطراً وN عموداً، ويتألف كل سطر فيها من N قيمة لمتحول المعياري المقابل Z_j يكون لدينا :

$$\mathfrak{z}_{(N*n)}=\begin{array}{c}Z_1\\Z_2\\\vdots\\Z_n\end{array}\begin{bmatrix}\mathfrak{z}_{11}&\mathfrak{z}_{12}&\cdots&\mathfrak{z}_{1N}\\\mathfrak{z}_{21}&\mathfrak{z}_{22}&\cdots&\mathfrak{z}_{2N}\\\vdots&\vdots&\vdots&\vdots\\\mathfrak{z}_{n1}&\mathfrak{z}_{n2}&\cdots&\mathfrak{z}_{nN}\end{bmatrix} \qquad (12-1)$$

وعندها نجد أن كل عنصر r_{jk} من المصفوفة R يساوي جداء السطر $\mathfrak{z}_j$ في منقول السطر $\mathfrak{z}_k$ مقسوماً على *(N–1)*، أي أن:

$$r_{jk}=\frac{1}{N-1}[\mathfrak{z}_{j1},\mathfrak{z}_{j2},\mathfrak{z}_{j3},\dots,\mathfrak{z}_{jN}]*\begin{bmatrix}\mathfrak{z}_{k1}\\\mathfrak{z}_{k2}\\\vdots\\\mathfrak{z}_{kN}\end{bmatrix}=\frac{1}{N-1}\mathfrak{z}_j*\mathfrak{z}_k' \qquad (13-1)$$

وعندها نجد من العلاقة *(13–1)* أنه يمكننا أن نعبر عن المصفوفة R كلها بدلالة المصفوفة $\mathfrak{z}$ المعرفة في *(12–1)* كما يلي:

$$R=\frac{1}{N-1}\mathfrak{z}*\mathfrak{z}'=C \qquad (14-1)$$

ولقد رمزنا لهذه المصفوفة أيضاً بـ C، لأنها تساوي مصفوفة التباينات المشتركة للمتحولات المعيارية Z_j، ويمكن كتابة *(14–1)* بشكل مصفوفي كما يلي :

$$R=\frac{1}{N-1}\begin{bmatrix}\mathfrak{z}_{11}&\mathfrak{z}_{12}&\mathfrak{z}_{13}&\cdots&\mathfrak{z}_{1N}\\\mathfrak{z}_{21}&\mathfrak{z}_{22}&\mathfrak{z}_{23}&\cdots&\mathfrak{z}_{2N}\\\vdots&\vdots&\vdots&\vdots&\vdots\\\mathfrak{z}_{n1}&\mathfrak{z}_{n2}&\mathfrak{z}_{n3}&\cdots&\mathfrak{z}_{nN}\end{bmatrix}*\begin{bmatrix}\mathfrak{z}_{11}&\mathfrak{z}_{21}&\cdots&\mathfrak{z}_{n1}\\\mathfrak{z}_{12}&\mathfrak{z}_{22}&\cdots&\mathfrak{z}_{n2}\\\mathfrak{z}_{13}&\mathfrak{z}_{23}&\cdots&\mathfrak{z}_{n3}\\\vdots&\vdots&\vdots&\vdots\\\mathfrak{z}_{1N}&\mathfrak{z}_{2N}&\cdots&\mathfrak{z}_{nN}\end{bmatrix}=C \qquad (15-1)$$

ومما سبق نستخلص ما يلي :

إذا قمنا بتحويل قيم المتحولات الأصلية $X_1 , X_2 , \dots , X_n$ إلى متحولات معيارية وفق العلاقة :

$$\mathfrak{z}_{ji} = \frac{x_{ji} - \bar{x}_j}{S_j} \qquad (16-1)$$

فإننا سنحصل على المصفوفة $\mathfrak{z} = [\mathfrak{z}_{ji}]$ المعرفة بالعلاقة *(12-1)*، ولكن علينا أن نتذكر دائماً أن عناصر هذه المصفوفة تخضع لـ *(2n)* شرطاً هي:

- إن متوسط قيم كل متحول فيها (كل سطر) يساوي الصفر وتشكل n شرطاً هي:

$$\bar{\mathfrak{z}}_j = \frac{1}{N}\sum_{i=1}^{N} \mathfrak{z}_{ij} = 0 \quad : \ (j: 1 2 3 \dots n) \qquad (17-1)$$

- إن تباين قيم كل متحول فيها (كل سطر) يساوي الواحد وتشكل n شرطاً أخرى هي:

$$S^2_{(Zj)} = \frac{1}{N-1}\sum \mathfrak{z}^2_{ij} = 1 \quad : \quad (j: 1 2 3 \dots n) \qquad (18-1)$$

وهذه الشروط هي القيم المذكورة على يمين الجدول *(1-2)* السابق . علماً بأن مجموع هذه التباينات يساوي التباين الإجمالي للمتحولات *Zj* ويساوي *n* .

والسؤال الآن هو: كيف يمكننا الاستفادة من المصفوفة الارتباطية *R* لدراسة خواص المتحولات $X_1 , X_2 , \dots , X_n$، وتجميعها ضمن مجموعات متجانسة، واستخلاص العوامل التي تمثل هذه المجموعات؟

إن الإجابة على ذلك ستتم من خلال استعراض نماذج التحليل العاملي واستخلاص العلاقات التي تساعدنا على ذلك . وذلك باستخدام الطرق الرياضية المناسبة للتحليل العاملي .

1-3: النماذج الخطية للتحليل العاملي: [Harman p. 26]

إن الهدف الأساسي للتحليل العاملي هو التعبير عن المتحولات المعيارية Zj بدلالة متحولات أخرى افتراضية سرية غير معروفة يطلق عليها اسم العوامل (Factors). وحتى نحصل على تلك العوامل مرتبة حسب الأهمية علينا العمل على تحقيق الهدفين التاليين:

- جعل تباينات تلك العوامل تشكل تنازلياً وعلى التوالي أكبر حصة من التباين الإجمالي S^2 والذي يساوي n (بعدد المتحولات المعيارية) .
- تحقيق أفضل تقريب أو تقدير للمصفوفة الارتباطية R بأبسط الأشكال الممكنة .

وإن أبسط نموذج للتعبير عن كل متحول Zj بدلالة عدة عوامل افتراضية هو النموذج الخطي، الذي أصبح معتمداً في التحليل العاملي. ولكن هذا لا يستبعد اعتماد نماذج أخرى غير خطية لتحقيق أهداف التحليل. وسنستعرض فيما يلي أهم النماذج الخطية المعتمدة في التحليل العاملي كما يلي:

1-3-1 : نموذج المكونات الأساسية Principal components :

إن أول من تناول هذا الموضوع كان (Person. K. 1901) حيث اقترح طريقة حدسية لاختصار وضغط الجداول الكبيرة والمعلومات الكثيرة مع إظهار أكبر نسبة من التباين الإجمالي . ولكن (Hoteling 1944) قام بتطوير هذه الطريقة وأطلق عليها طريقة المكونات الأساسية (Principal components) أو تحليل المكونات (Components analysis) وقد وضع نموذجه الخطي الأول لكل متحول Z_j على الشكل التالي:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \cdots . a_{jn}F_n \qquad (j: 1\ 2\ 3 \ldots n) \qquad (19-1)$$

وهو يتألف من n معادلة خطية تتضمن n مجهولاً، أطلق عليها اسم المكونات الأساسية (وليس العوامل)، أي أن كل متحول معياري Zj يرتبط خطياً بـ n من المكونات المجهولة والمستقلة عن بعضها البعض (تسمى المكونات الأساسية) وعددها بعدد المتحولات Zj ويرمز لها بـ $F_1, F_2, F_3, \ldots, F_n$. وتعتمد هذه الطريقة على تقدير المصفوفة R بحيث يكون نصيب التباين المحسوب للمكون الأول F_1 من التباين الاجمالي S^2 لتلك المتحولات أكبر من نصيب F_2 ، ونصيب المكون F_2 أكبر من المكون F_3 ... الخ، وفي الدراسات الميدانية يستغنون عن المكونات الأخيرة، لأن نصيب تباينها من التباين الاجمالي يكون صغيراً ومهملاً، ويحتفظون فقط بعدد قليل من تلك المكونات، التي يشكل نصيبها نسبة كبيرة من التباين الاجمالي S^2 للمتحولات Zj، ولكن لأمور تتعلق بدقة تقريب المصفوفة الارتباطية R بين المتحولاتZj يمكن الاحتفاظ بجميع المكونات الضرورية .

2-3-1 : نموذج العوامل (المحاور) الرئيسية Princepal (Axis) Factors:

يعتمد هذا النموذج على مدخل آخر لمعالجة هذه المسألة، وهو يتطلب أن تحقق النماذج الخطية أفضل تقريب أو تقدير للمصفوفة الارتباطية R، وبحيث تجعل حصص تباينات العوامل تأخذ على التوالي أكبر حصة ممكنة من التشاركية الإجمالية للمتحولات المدروسة Zj (والتي سنعرفها لاحقاً). ويتضمن هذا النموذج متحولات خاصة تسمى عوامل التميز U_i .

لذلك يتم تحديد حد أعلى للدقة (أو للخطأ المرتكب) في عملية التعبير أو التقدير، وذلك لأجل الاقتصار على عدد قليل من العوامل التي تحقق تلك الدقة . وإن النموذج الأساسي للتحليل العاملي في هذه الحالة يأخذ الشكل التالي :

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots a_{jm}F_m + d_jU_j \qquad (20-1)$$

حيث أن: $(j: 1\ 2\ 3 \ldots n)$ وأن: U_j هو عامل التميز للمتحول Z_j .

وهذا يعني أن كل متحول Z_j يرتبط خطياً بـ m عاملاً فقط $(m < n)$ هي $F_1, F_2, \ , F_m$، وتسمى هذه العوامل بالعوامل العامة (Common factors)، مضافاً إليها عامل إضافي U_j يسمى بعامل التميز للمتحول Z_j (Character factor). أما الأعداد المجهولة $a_{j1}, a_{j2} \ldots a_{jm}$ وكذلك d_j، فما هي إلا أمثال هذه العوامل التي تعبر -كما سنرى- عن الجذر التربيعي لنصيب كل منها في التشاركية الإجمالية

للمتحول Zj . والآن علينا حساب الأمثال a_{jp} و d_j بطرائق تختلف عما نعرفه في الانحدار المتعدد ، وذلك لأن قيم العوامل العامة $F_1, F_2, ,, \ , F_m$ مجهولة وليست معلومة كما في حالة الانحدار المتعدد .

وبما أن المتحولات Z_j هي متحولات معيارية . فإنه وبدون أن نخل بالنموذج الأساسي $(1-20)$ نفترض أيضاً أن العوامل $F_1, F_2,, \ , F_m$ و U_j هي أيضاً متغيرات معيارية ومستقلة (رغم إنها مازالت مجهولة) وإن متوسطاتها معدومة $\left(\bar{F}_p = 0\right)$ وإن تباين كل منها يساوي الواحد $\left(\sigma^2_{Fp} = 1\right)$.

كما نفترض أيضاً أن عوامل التميز U_j مستقلة عن بعضها البعض وأنها غير مرتبطة بالعوامل العامة F_p، وأن متوسطاتها معدومة $\left(\bar{u}_j = 0\right)$ وتبايناتها تساوي الواحد .

وبناء على العلاقة $(1-20)$ يمكننا كتابة المتحولات المعيارية Z_j بدلالة العوامل الافتراضية F_p (العوامل العامة) وبدلالة عوامل التميز U_j كما يلي:

$$\begin{aligned}
Z_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + d_1U_1 \\
Z_2 &= a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + d_2U_2 \\
Z_3 &= a_{31}F_1 + a_{32}F_2 + \cdots + a_{3m}F_m + d_3U_3 \qquad (21-1) \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
Z_n &= a_{n1}F_1 + a_{n2}F_2 + \cdots + a_{nm}F_m + d_nU_n
\end{aligned}$$

ويسمى هذا الشكل بالصورة النموذجية لتحليل العوامل (Factor pattern) ، ويفترض فيها أن تكون عوامل التميز U_j مستقلة ومعيارية، أي أن $r\left(u_j, u_k\right) = 0$ (حيث $j \neq k$)، وإن متوسطاتها تساوي الصفر $\left(\bar{u}_j = 0\right)$ وتبايناتها تساوي الواحد مهما يكن j ، كما يفترض أن تكون المتحولات U_j غير مرتبطة مع العوامل العامة F_p، أي أن $r\left(u_j, F_p\right) = 0$ مهما كانت (j:1 2 ..n أو p:1 2 ...m)،

ويمكن وضع المعادلات $(1-21)$ في جدول احصائي كما يلي :

جدول $(1-3)$ الصورة النموذجية للنموذج الرئيسي $(1-21)$

العوامل المتحولات Z_j		F_1	F_2	F_3	....	F_m		U_1	U_2	U_3	....	U_n
Z_1		a_{11}	a_{12}	a_{13}	....	a_{1m}		d_1	....	....	....	....
Z_2	=	a_{21}	a_{22}	a_{23}	....	a_{2m}	+	....	d_2	....	....	....
Z_3		a_{31}	a_{32}	a_{33}	....	a_{3m}		....	....	d_3	....	....
....		....	....	....	....	....		....	....	....	....	....
Z_n		a_{n1}	a_{n2}	a_{n3}	....	a_{nm}		....	....	....	....	d_n

وهنا نشير أيضاً إلى أن العوامل العامة $F_1, F_2, \dots , F_m$ يمكن أن تكون مرتبطة أو غير مرتبطة مع بعضها البعض، كما يمكن أن يكون بعض أمثالها a_{jp} معدوماً . وهذا يعني أن بعض المتحولات Z_j يمكن أن يعبر عنها بعدد من العوامل أقل من العدد m المذكور. وإن عدد العوامل العامة اللازمة للتعبير عن المتحول Z_j تسمى درجة التعقيد Complexity .

1-4: الشكل المصفوفي للنماذج الخطية في تحليل العوامل:

إذا رمزنا لعمود المتحولات Z_j بالرمز Z ، ولعمود العوامل العامة F_p بالرمز F ، ولعمود عوامل التميز بالرمز U ، ولمصفوفة الأمثال $[a_{jp}]$ بالرمز A ، ولمصفوفة الأمثال $[d_j]$ بالرمز D ، فيكون لدينا :

$$Z=\begin{bmatrix}Z_1\\Z_2\\Z_3\\\vdots\\Z_n\end{bmatrix},\quad F=\begin{bmatrix}F_1\\F_2\\F_3\\\vdots\\F_m\end{bmatrix},\quad U=\begin{bmatrix}U_1\\U_2\\U_3\\\vdots\\U_n\end{bmatrix}\qquad\mathbf{(22-1)}$$

$$A_{n*m}=\begin{bmatrix}a_{11}&a_{12}&a_{13}&\dots&a_{1m}\\a_{21}&a_{22}&a_{23}&\dots&a_{2m}\\a_{31}&a_{32}&a_{33}&\dots&a_{3m}\\\dots&\dots&\dots&\dots&\dots\\a_{n1}&a_{n2}&a_{n3}&\dots&a_{nm}\end{bmatrix},\quad D_{n*n}=\begin{bmatrix}d_1&0&0&0&0\\0&d_2&0&0&0\\0&0&d_3&0&0\\\dots&\dots&\dots&\dots&\dots\\0&0&0&0&d_n\end{bmatrix}$$

وبناء على ذلك يمكننا كتابة المعادلات (1−21) على شكل مصفوفي كما يلي :

$$\begin{bmatrix}Z_1\\Z_2\\Z_3\\\vdots\\Z_n\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}&\dots&a_{1m}\\a_{21}&a_{22}&a_{23}&\dots&a_{2m}\\a_{31}&a_{32}&a_{33}&\dots&a_{3m}\\\dots&\dots&\dots&\dots&\dots\\a_{n1}&a_{n2}&a_{n3}&\dots&a_{nm}\end{bmatrix},\begin{bmatrix}F_1\\F_2\\F_3\\\vdots\\F_m\end{bmatrix}+\begin{bmatrix}d_1&0&0&0&0\\0&d_2&0&0&0\\0&0&d_3&0&0\\\dots&\dots&\dots&\dots&\dots\\0&0&0&0&d_n\end{bmatrix}\begin{bmatrix}U_1\\U_2\\U_3\\\vdots\\U_n\end{bmatrix}$$

ويمكن كتابة هذا الشكل المصفوفي بشكل مختصر كما يلي:

$$Z=A*F+D*U\qquad(23-1)$$

ومن الشكل المصفوفي (1−23) يمكننا أن نحصل على القيم العددية $\mathfrak{z}_{ij}$ لكل متحول Z_j بدلالة الأمثال a_{jp} و d_j والقيم العددية للعوامل العامة F_{ip} وعوامل التميز U_j، وذلك من خلال جداء المصفوفة A في مصفوفة القيم العددية للعوامل العامة F_p، مضافاً إليها جداء المصفوفة D في مصفوفة القيم العددية لعوامل التميز U_j كمايلي :

$$\mathfrak{z}=\begin{bmatrix}\mathfrak{z}_{11}&\mathfrak{z}_{12}&\dots&\mathfrak{z}_{1N}\\\mathfrak{z}_{21}&\mathfrak{z}_{22}&\dots&\mathfrak{z}_{2N}\\\dots&\dots&\dots&\dots\\\mathfrak{z}_{i1}&\mathfrak{z}_{i2}&\dots&\mathfrak{z}_{iN}\\\dots&\dots&\dots&\dots\\\mathfrak{z}_{n1}&\mathfrak{z}_{n2}&\dots&\mathfrak{z}_{nN}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&\dots&a_{1m}\\a_{21}&a_{22}&\dots&a_{2m}\\\dots&\dots&\dots&\dots\\a_{i1}&a_{i2}&\dots&a_{im}\\\dots&\dots&\dots&\dots\\a_{n1}&a_{n2}&\dots&a_{nm}\end{bmatrix}*\begin{bmatrix}\mathfrak{f}_{11}&\mathfrak{f}_{12}&\dots&\mathfrak{f}_{1j}&\dots&\mathfrak{f}_{1N}\\\mathfrak{f}_{21}&\mathfrak{f}_{22}&\dots&\mathfrak{f}_{2j}&\dots&\mathfrak{f}_{2N}\\\dots&\dots&\dots&\dots&\dots&\dots\\\mathfrak{f}_{m1}&\mathfrak{f}_{m2}&\dots&\mathfrak{f}_{mj}&\dots&\mathfrak{f}_{nm}\end{bmatrix}$$

$$+\begin{bmatrix}d_1&0&0&0&0\\0&d_2&0&0&0\\\dots&\dots&\dots&\dots&\dots\\0&0&d_j&0&0\\\dots&\dots&\dots&\dots&\dots\\0&0&0&0&d_n\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_n\end{bmatrix}\qquad\mathbf{(24-1)}$$

كما يمكن التعبير عن هذه العلاقة وبدلالة مصفوفات القيم العددية للمتحولات والعوامل كما يلي :

$$\mathfrak{z}=A*f+D*u\qquad(25-1)$$

وعندها نجد أن كل قيمة $\mathfrak{z}_{ij}$ للمتحول Z_j تحسب من خلال جداء السطر i للمصفوفة A في العمود j للمصفوفة $[\mathfrak{f}_{ij}]$ ثم إضافة حد الخطأ $d_j u_{ij}$ كما يلي :

$$\mathfrak{z}_{ij} = a_{i1}\mathfrak{f}_{1j} + a_{i2}\mathfrak{f}_{2j} + a_{ip}\mathfrak{f}_{pj} + \cdots + a_{in}\mathfrak{f}_{mj} + d_j u_{ij} \qquad (26-1)$$

حيث أن: $i = 1\ 2\ 3\ \ldots.N$ وأن $j = 1\ 2\ 3\ \ldots.n$

وحيث أن: $\mathfrak{f}_{ip}$ هي القيمة العددية الافتراضية للعامل F_p

وهنا نلاحظ أن كل حد $\left(a_{ip}\mathfrak{f}_{pj}\right)$ في المجموع $(26-1)$ يعكس درجة تأثير العامل F_p على قيمة المتحول $\mathfrak{z}_{ij}$. أما الحد $d_j u_{ij}$ فيعكس مقدار الخطأ المرتكب في تقدير القيمة $\mathfrak{z}_{ij}$ للمتحول Z_j .

5-1: تحليل مكونات التباين S_j^2 للمتحول المعياري Z_j [Harman p.28]

لقد أشرنا سابقاً إلى أن Z_j هو متحول معياري وإن متوسطه $\bar{Z}_j = 0$ وتباينه $S_j^2 = 1$، وانطلاقاً من تعريف التباين S_j^2 المعرف بالعلاقة $(18-1)$ للمتحول Z_j ، فإنه يمكننا حساب هذا التباين بدلالة العوامل $F_1, F_2, \ldots, F_m$، وذلك اعتماداً على النموذج الخطي $(20-1)$، الذي يربط Z_j بتلك العوامل، حيث نجد أن:

$$1 = S_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathfrak{z}_{ij} - 0\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\mathfrak{z}_{ij}^2 =$$

$$1 = S_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[a_{j1}\mathfrak{f}_{1j} + a_{j2}\mathfrak{f}_{2j} + \cdots + a_{jm}\mathfrak{f}_{mj} + d_{ij}u_{ij}\right]^2$$

$$1 = S_j^2 = \sum_{p=1}^{m} a_{jp}^2\left(\frac{\sum_{i=1}^{N}\mathfrak{f}_{ip}^2}{N-1}\right) + d_j^2 * \sum_{i=1}^{N}\left(\frac{u_{ij}^2}{N-1}\right) \qquad (27-1)$$

$$+2\sum_{p<q=1}^{m} a_{jp} * a_{jq}\sum_{i=1}^{N}\left(\frac{\mathfrak{f}_{ip} * \mathfrak{f}_{iq}}{N-1}\right) + 2d_j\sum_{j=1}^{m} a_{jp}\left(\frac{\sum^{N}\mathfrak{f}_{ip} * u_{ij}}{N-1}\right)$$

وبما أن تباين المتحول المعياري S_j^2 يساوي الواحد، وإن جميع المتحولات Z_j والعوامل F_p و u_j هي متحولات معيارية، فإنه يمكننا تحويل العلاقة السابقة إلى الشكل التالي:

$$1 = S_j^2 = \sum_{j=1}^{m} a_{jp}^2 * 1 + d_j^2 * 1 + 2\sum_{p<q=1}^{m} a_{jp} * a_{jq} * r(F_p * F_q) \qquad (28-1)$$

$$+2d_j\sum a_{jp} * r(F_p * u_j)$$

ولكن بما أن عوامل التميز U_j غير مرتبطة مع العوامل العامة F_p فإن $r\left(F_p * u_j\right) = 0$. وإذا افترضنا (هذا ما سنعرضه لاحقاً) أن العوامل العامة F_p غير مرتبطة مع بعضها البعض، فإنه يكون لدينا $r\left(F_p * F_q\right) = 0$ ، وبذلك تأخذ العلاقة $(28-1)$ الشكل المختصر التالي:

$$S_j^2 = a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots + a_{jm}^2 + d_j^2 = 1 \qquad (29-1)$$

وهنا نلاحظ أن التباين الواحدي S_j^2 يتألف من قسمين رئيسيين هما:

- قسم يسمى التشاركية (Communality) الخاصة بالمتحول Z_j ، وهي تساوي مجموع مربعات أمثال العوامل F_p، ويتم حسابها من سطر Z_j (أفقياً) على جميع العوامل العامة . ويرمز لها بـ h_j^2 وتساوي :

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots + a_{jm}^2 \qquad (30-1)$$

- قسم آخر يسمى تميز المتحول Z_j (Uniqueness) وهو الحد d_j^2 المتمم لـ h_j^2 ، وبذلك يمكننا أن نكتب S_j^2 كما يلي :

$$S_j^2 = h_j^2 + d_j^2 = 1 \qquad (31-1)$$

وهنا نشير إلى أن مربع كل من الأمثال a_{jp}^2 في (1−29) يعبر عن حصة العامل F_p من التباين الواحدي S_j^2 للمتحول Z_j ، أي أن :

a_{j1}^2 هي حصة تباين العامل الأول F_1 من التباين S_j^2 للمتحول Z_j فقط.

و a_{j2}^2 هي حصة تباين العامل الثاني F_2 من التباين S_j^2 للمتحول Z_j فقط. وهكذا....

وإن d_j^2 هي حصة تباين عامل التميز U_j من التباين S_j^2 للمتحول Z_j فقط.

وإذا أخذنا مجموع حصص العامل F_p من جميع المتحولات Z_j (نجمعها عمودياً)، فإننا سنحصل على الحصة الإجمالية للعامل F_p من قيمة التباين الإجمالي S^2 لجميع المتحولات Z، ونرمز لها بـ V_p وهي تساوي :

$$V_p = \sum_{j=1}^{n} a_{jp}^2 \qquad (p: 1\ 2 \ldots . m) \qquad (32-1)$$

كما يمكننا حساب الحصة الكلية لجميع العوامل $F_1, F_2, \ldots, F_m$ من التباين الإجمالي، ونرمز لها بـ V ، وإننا نحصل عليها من العلاقة التالية :

$$V = \sum_{p=1}^{m} V_p = \sum_{p=1}^{m}\sum_{j=1}^{n} a_{jp}^2 = \sum_{j=1}^{n}\sum_{p=1}^{m} a_{jp}^2 = \sum_{j=1}^{n} h_j^2 \qquad (33-1)$$

ويسمى المقدار $V = \sum h_j^2$ بالتشاركية الإجمالية للعوامل العامة المعيارية الداخلة في النموذج .

وأخيراً يمكننا أن نحسب نسبة حصة تباين كل عامل F_p من التباين الإجمالي من العلاقة التالية:

$$P_p = \frac{V_p}{n} \qquad (33-1)'$$

وأن نحسب النسبة الكلية للتشاركية الاجمالية من التباين الاجمالي لتلك العوامل من العلاقة :

$$P = \frac{V}{n} = \frac{\sum_1^m V_p}{n} \qquad (34-1)$$

ويمكن أن نوضح ذلك جدولياً على الشكل التالي:

جدول (1–4) التشاركية والتميز والخطأ والموثوقية .

$S_j^2 = 1 = 100\% = Z_j$ تباين									
a_{j1}^2	a_{j2}^2	a_{j3}^2	.....	.....	a_{jp}^2	.....	a_{jm}^2	d_j^2	
التشاركية h_j^2								d_j^2	
a_{j1}^2	a_{j2}^2	a_{j3}^2	.....	a_{jp}^2	......	.....	a_{jm}^2	b_j^2	e_j^2
الموثوقية P_j^2									e_j^2

وهناك من يقوم بتقسيم عامل التميز U_j إلى قسمين مستقلين كما يلي:

- قسم يعكس خصوصية العامل (Specialty) Z_j ونرمز له بـ SP_j ولأمثاله بـ b_j .
- قسم يعكس الخطأ العشوائي المرتكب في تقدير تباين Z_j (error variance) ونرمز له بـ E_j، ولأمثاله بـe_j ، ويسمى تباين الخطأ العشوائي لـ Z_j أو معامل الشك (unreliability) في تقدير تباين Z_j .

وبناء على ذلك يمكننا كتابة النموذج (1–20) السابق كما يلي:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \cdots + a_{jm}F_m + b_jSP_j + e_jE_j \qquad (35-1)$$

حيث أن : (j: 1 2 3 ..... n)

وبما أن عاملي الخصوصية SP_j والخطأ العشوائي E_j غير مرتبطين فإنه يكون لدينا:

$$d_j^2 = b_j^2 + e_j^2 \qquad (36-1)$$

وبذلك تأخذ العلاقة (1–26) السابقة :

$$S_j^2 = h_j^2 + b_j^2 + e_j^2 = 1 \qquad (37-1)$$

وهكذا نجد أن التباين الخاص S_j^2 أصبح يتألف من ثلاثة أجزاء هي :

- جزء يعبر عن تشاركية العوامل في Z_j ورمزنا لها بـ h_j^2 .
- جزء يعبر عن حصة خصوصية المتحول Z_j المصاحب لـ SP_j يرمز له بـ b_j^2 .
- جزء يعبر عن حصة عامل الخطأ العشوائي في تقدير تباين Z_j ويرمز له بـ e_j^2 .

كما ويسمى الجزء الأخير e_j^2 بدرجة الشك (unreliability)، ويسمى متممه :($P_j^2 = 1 - e_j^2$) بدرجة الموثوقية (reliability) . والجدول (1–4) يوضح ذلك .

علماً بأن الطرائق المستخدمة في تحليل العوامل تعطينا قيمة كل من التشاركية h_j^2 والتميز d_j^2 لكل متحول معياري Z_j . ثم تقوم بتقسيم d_j^2 إلى جزأيها b_j^2 و e_j^2 ، وإن عملية التقسيم ليست لها علاقة بنوع الحل العاملي الذي نحصل عليه .

فمثلاً إذا كانت درجة الموثوقية المطلوبة لتجربة ما معلومة وتساوي P_j^2 ، فإن درجة الشك المساوية لـ e_j^2 تحسب من العلاقة :

$$e_j^2 = 1 - P_j^2 \qquad (38-1)$$

وإذا علمنا مقدار الخطأ المرتكب d_j^2 فإنه يمكننا حساب مقدار حصة الخصوصية b_j^2 من العلاقة (1−36) كما يلي :

$$b_j^2 = d_j^2 - e_j^2 \qquad (39-1)$$

وهكذا نجد من(1−38) إن درجة الموثوقية P_j^2 هي متمم الخطأ العشوائي e_j^2 وتحسب من العلاقة (−37 1) فنجد أن:

$$P_j^2 = 1 - e_j^2 = h_j^2 + b_j^2 \qquad (40-1)$$

ومنها نستنتج أن التشاركية h_j^2 تساوي :

$$h_j^2 = P_j^2 - b_j^2 \leq P_j^2 \qquad (41-1)$$

أي أن التشاركية h_j^2 لأي متحول Z_j لا يمكن أن تتجاوز درجة الموثوقية P_j^2 وهي تساويها فقط عندما تكون حصة الخصوصية معدومة $(b_j^2 = 0)$.

واعتماداً على نموذج العوامل الرئيسي (1−20) يمكننا كتابة العلاقات بين مكونات التباين الواحدي S_j^2 للمتحول Z_j حسب أهميتها كما يلي :

$$
\left.
\begin{array}{ll}
\text{تباين } Z_j: & S_j^2 = h_j^2 + d_j^2 = h_j^2 + b_j^2 + e_j^2 = 1 \\
\text{الموثوقية}: & P_j^2 = h_j^2 + b_{j2}^2 = 1 - e_j^2 \\
\text{التشاركية}: & h_j^2 = \sum_{p=1}^{m} a_{jp}^2 = 1 - d_j^2 \\
\text{التميز}: & d_j^2 = b_j^2 + e_j^2 = 1 - h_j^2 \\
\text{الخصوصية}: & b_j^2 = d_j^2 - e_j^2 \\
\text{درجة الخطأ (الشك)}: & e_j^2 = 1 - P_j^2
\end{array}
\right\} \qquad (42-1)
$$

وأخيراً: يمكننا تعريف وحساب كثافة (العوملة) C_j للمتحول Z_j كما يلي :

$$C_j = \frac{h_j^2}{h_j^2 + b_j^2} 100 = \frac{\text{التشاركية}}{\text{الموثوقية}} 100 \qquad (43-1)$$

كما يمكن تعريف وحساب نسبة الكثافة الكلية (للعوملة) من العلاقة :

$$C = \frac{\sum_1^n C_j}{n} \qquad (44-1)$$

وهي تأخذ قيمة أقل من 100% ، ولا تساوي 100% إلا إذا كان عامل الخصوصية b_j^2 معدوماً (وهذا قد لا يحدث) .

ملاحظة: إن بعض الباحثين المختصين لا يأخذون بفرضية تقسيم الحد العشوائي الأخير U_j إلى حدين مستقلين (الخصوصية SP_j والخطأ E_j)، لأنهم يعتبرون أن تقسيم العامل U_j إلى عاملين SP_j و E_j يكون مفيداً فقط عندما تكون المتحولات معبرة عن عناصر المجتمع تماماً أو معبرة عنها بدقة عالية ، وعندها تكون الأمثال $e_j^2 = 0$ ، وهذا قد لا يكون متوفراً .

مثال (1−2):

وكتطبيق على حساب هذه المؤشرات ، لنفترض أننا استطعنا أن نمثل (5) متحولات معيارية Z_1, Z_2, Z_3, Z_4, Z_5 بدلالة عاملين F_1 و F_2 حسب النموذج (1−21) وحصلنا على الأمثال a_{jp} المبينة في الجدول التالي:

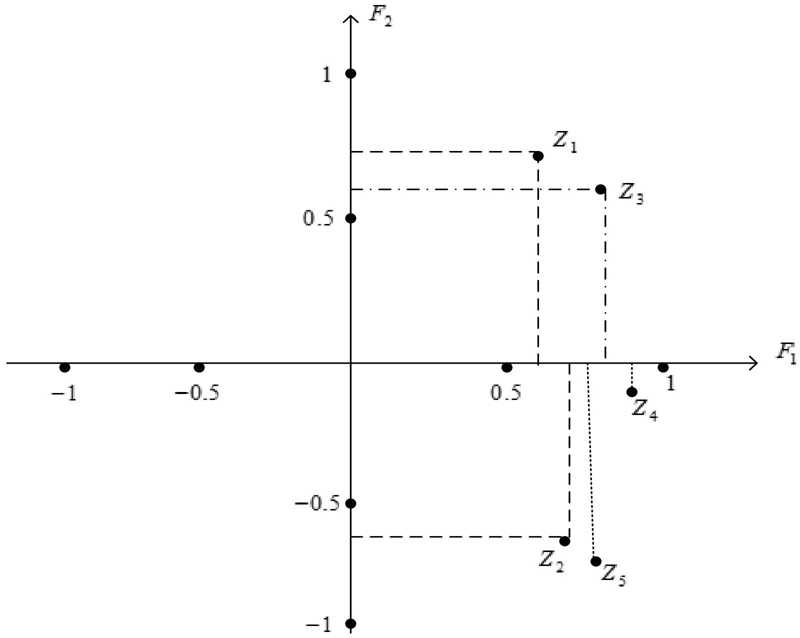
جدول (1−5) الأمثال العددية a_{j1} و a_{j2} للعاملين F_1 و F_2 :

العوامل / المتحولات	F_1	F_2	التشاركيات $h_j^2 = a_{j1}^2 + a_{j2}^2$	$d_j^2 = 1 - h_j^2$	d_j
Z_1	0.5810	0.8064	0.98784	0.01216	0.1103
Z_2	0.7671	−0.5448	0.88525	0.11475	0.3387
Z_3	0.6724	0.7260	0.97920	0.02080	0.1442
Z_4	0.9324	−0.1043	0.88025	0.11975	0.3460
Z_5	0.7911	−0.5582	0.93743	0.06257	0.2501
حصة العامل F_p $V_P = \sum_{j=1}^{5} a_{jp}^2$	2.8733	1.7966	4.66997	0.33003	———
النسبة المئوية% $P = \frac{V_{p*100}}{n}$	57.46	35.93	93.39	6.61	———

وهكذا يمكننا كتابة النموذج العاملي لهذه المتحولات الخمسة (بعد حساب الجذور d_j) كما يلي:

$$Z_1 = 0.5810\, F_1 + 0.8064\, F_2 + 0.1103\, U_1$$
$$Z_2 = 0.7671\, F_1 - 0.5448\, F_2 + 0.3387\, U_2$$
$$Z_3 = 0.6724\, F_1 + 0.7260\, F_2 + 0.1442\, U_3$$
$$Z_4 = 0.9324\, F_1 - 0.1043\, F_2 + 0.3460\, U_4$$
$$Z_5 = 0.7911\, F_1 - 0.5582\, F_2 + 0.2501\, U_5$$

وهو يرسم الشكل البياني التالي:



الشكل (1−2): توزيع المتحولات على المستوى $F_1 0\, F_2$

ومن الجدول السابق تم حساب التشاركيات h_j^2 لكل متحولات Z_j حسب العلاقة (1−30) كما يلي:

$$h_1^2 = (0.5810)^2 + (0.8064)^2 = 0.98784$$
$$h_2^2 = (0.7671)^2 + (-0.5448)^2 = 0.88525$$
$$h_3^2 = (0.6724)^2 + (0.7260)^2 = 0.97920$$
$$h_4^2 = (0.9324)^2 + (-0.1043)^2 = 0.86948$$
$$h_5^2 = (0.7911)^2 + (-0.5582)^2 = 0.93743$$

وبذلك نجد أن التشاركية الإجمالية لهذه المتحولات الخمسة تساوي:

$$\sum_{j=1}^{5} h_j^2 = 4.66997$$

وهي أقل من التباين الإجمالي لها والذي يساوي: $S^2 = 5$

كما تم حساب مربعات أمثال متحولات التميز U_j من العلاقة (1−31) وتم وضعها في العمود الأخير من الجدول (1−5) السابق، حيث نجد منها أن:

$$d_j^2 = 1 - h_j^2$$

ولحساب الحصة الإجمالية لكل عامل F_p في التأثير على هذه المتحولات طبقنا العلاقة (1−32) ، وأخذنا مجموع مربعات الأمثال في كل عمود ووضعناه في السطر الأخير من الجدول السابق فوجدنا أن:

$$V_1 = 2.873 \qquad V_2 = 1.7966$$

وأن النسبة المئوية لهاتين الحصتين من التباين الإجمالي ($S^2 = 5$) تساويان:

$$P_1 = 57.46\ \% \qquad P_2 = 35.93\ \%$$

وأن مجموع حصتي هذين العاملين V_1 و V_2 يساوي 4.6699 ويشكل حوالي 93.39 % من التباين الإجمالي S^2 .

أما حصة عوامل التميز فتساوي 0.33003 وتشكل ما نسبته 6.61 % من التباين الإجمالي S^2 . وهي تمثل حصة عوامل الخصوصية والعوامل العشوائية معاً.

ومما سبق نستنتج أن النموذج العاملي المحسوب بدلالة العاملين F_1 و F_2 يعبر بشكل جيد عن المتحولات الخمسة المدروسة Z_j، وإن هذه المتحولات تتوزع على المستوي المشكل من هذين العاملين كما في الشكل (1−2)، وسنتابع عمليات تحليل هذه المسألة في الفقرات التالية .

1−6: الصورة العاملية والهيكل العاملي [Harman P.31] :

بعد أن نقوم بالتعبير عن جملة المتحولات Z_j بدلالة العوامل F_p ، نقوم بتعريف بعض القضايا المتعلقة (بعوملة) البيانات التجريبية، وذلك انطلاقاً من النموذج (1−21) الذي يربط كل متحول Z_j بعدد محدد m من العوامل العامة F_p مضافاً إليها عوامل التميز U_j والذي يأخذ الشكل التالي:

$$\begin{aligned} Z_1 &= a_{11}F_1 + a_{12}F_2 + \cdots a_{1m}F_m + d_1U_1 \\ Z_2 &= a_{21}F_1 + a_{22}F_2 + \cdots a_{2m}F_m + d_2U_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Z_n &= a_{n1}F_1 + a_{n2}F_2 + \cdots a_{nm}F_m + d_nU_n \end{aligned} \qquad (45-1)$$

إن جملة المعادلات (1−45) تسمى الصورة العاملية (Factor pattern)، أو باختصار نسميها الصورة، وإن مصفوفة الأمثال $[a_{jp}]$ تسمى مصفوفة الصورة ونرمز لها بـ A ، ونرمز لمصفوفة الأمثال $[d_j]$ بالرمز D، ويمكن عرض هذه الصورة من خلال جدول منظم تضم خلاياه فقط الأمثال a_{jp} و d_j المصاحبة للعوامل العامة وعوامل التميز، وبحيث يتم تخصيص كل عمود فيه لعامل واحد F_p أو U_j (كما في الجدول (1−3))، وبصورة عامة فإن العوامل العامة $F_1, F_2, \ldots, F_m$ يمكن أن تكون مستقلة عن بعضها البعض (غير مرتبطة) أو تكون مرتبطة مع بعضها البعض. أما عوامل التميز U_j ، فيفترض أنها غير مرتبطة بين بعضها البعض، وغير مرتبطة أيضاً مع العوامل العامة F_p . وإنه يمكن لبعض الأمثال a_{jp} للمتحولات Z_j أن تكون معدومة، وهذا يعني أنه يمكن التعبير عن أي متحول Z_j بعدد من العوامل أقل من m ، لذلك يسمى عدد العوامل العامة اللازمة للتعبير عن Z_j بدرجة التعقيد complexly .

ولكن تحليل العوامل لا يكتفي منا بالحصول على الصورة العاملية A ، بل يتطلب منا الحصول على ما يسمى بمصفوفة الهيكل العاملي (Factor structure) ، وهي عبارة عن مصفوفة معاملات الارتباط بين المتحولات Z_j والعوامل F_p ، وسنرمز لها بالرمز S . ولإيجاد هذه المصفوفة الهيكلية واستخراج العلاقة التابعية بين مصفوفة الصورة ومصفوفة الهيكل ، نقوم بضرب كل معادلة من (1−45) بالعوامل العامة $F_1, F_2, \ldots, F_m$ على التوالي، ثم نأخذ مجاميع هذه الجداءات على جميع المشاهدات التي عددها N ، ثم نقسمها على (N−1) ، فنحصل على معاملات الارتباط المختلفة بين المتحولات Z_j والعوامل F_p ، وبين العوامل العامة F_p نفسها (بفرض أنها مرتبطة مع بعضها) وفق العلاقات التالية:

$$\begin{aligned}
r_{(Z_jF_1)} &= a_{j1} + a_{j2}r_{(F1F_2)} + a_{j3}r_{(F_1F_3)} + \cdots + a_{jm}r_{(F_1F_m)} + 0 \\
r_{(Z_jF_2)} &= a_{j1}r_{(F2F_1)} + a_{j2} + a_{j3}r_{(F_2F_3)} + \cdots + a_{jm}r_{(F_2F_m)} + 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
r_{(Z_jF_p)} &= a_{j1}r_{(FpF_1)} + a_{j2}r_{(FpF_2)} + \cdots + a_{jp} \ldots + a_{jm}r_{(F_pF_m)} + 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
r_{(Z_jF_m)} &= a_{j1}r_{(FmF_1)} + a_{j2}r_{(FmF_2)} + \cdots + a_{jm}r_{(F_mF_{m-1})} + a_{jm} + 0
\end{aligned} \qquad (46-1)$$

حيث أن (j: 1 2 3 ….n)

وإذا ضربنا كل Z_j بـ U_j وأخذنا المجاميع على N وقسمنا الناتج على (N−1) لحصلنا على أن معامل الارتباط بينهما يساوي: $r_{(Z_j,u_j)} = d_j * r_{(u_j,u_j)} = d_j * 1$ ومنها نجد أن:

$$r_{(Z_j,u_j)} = d_j \qquad (j: 1\ 2\ 3 \ldots\ldots n) \qquad (47-1)$$

ومن العلاقة الأخيرة (1−47) نستنتج مباشرة أن معامل ارتباط أي متحول Z_j مع عامل التميز U_j يساوي دائماً قيمة الأمثال d_j المصاحبة للعامل U_j في النموذج الرياضي (1−21) .

ولكننا عند الحديث عن المصفوفة الهيكلية العاملية S ، نقصد بها معاملات الارتباط الزوجية بين المتحولات Z_j والعوامل العامة $F_1, F_2, \ldots, F_m$ فقط ، أي نقصد بها جدول المعاملات $S_{jp} = r_{(Z_j,F_p)}$ المعرفة في الطرف الأيسر من (1−46) .

وهنا نلاحظ أن المعادلات (1−46) تشكل جملة من المعادلات الخطية مؤلفة من n.m معادلة، وتتضمن m.n متحولاً مجهولاً هي الأمثال a_{jp} (حيث j:1 2 3 … n و p:1 2 3 … m) وذلك مع افتراض أن المعاملات في الطرف الأيسر $r_{(Z_j,F_p)}$ معلومة .

كما نلاحظ من المعادلات (1−46) أن معاملات الارتباط الهيكلية $r_{(Z_j,F_p)}$ لا تساوي بصورة عامة الأمثال المقابلة لها a_{jp} .أي أن المصفوفة الهيكلية تختلف بشكل عام عن مصفوفة الصورة .

كما إن حساب عناصر المصفوفة الهيكلية [Sjp] المؤلفة من المعاملات r (Zj, Fp) ، يقتضي أن تكون قيم Fp المقابلة معلومة، وهذا غير متوفر من البيانات المتاحة .

إلا أنه إذا كانت العوامل العامة F_p غير مرتبطة بين بعضها البعض، فعندها تكون المعاملات الارتباطية بينها معدومة ، أي يكون لدينا كل: $r_{(Fp,F_q)} = 0$ (حيث $p \neq q$)، وبالتالي فإن المعادلات (1−46) تتحول إلى الشكل البسيط التالي :

$$r_{(Z_j,F_p)} = a_{jp} \qquad \begin{matrix} (j : 1\ 2\ 3 \ldots. n) \\ (p: 1\ 2\ 3 \ldots. m) \end{matrix} \qquad (48-1)$$

أي أن معاملات الارتباط الهيكلية تتطابق مع الأمثال a_{jp} الموافقة لها في صورة النموذج فقط وفقط عندما تكون العوامل العامة F_p غير مرتبطة . وهذا يعني أنه إذا كان التحليل العاملي يجري مع افتراض عدم ارتباط العوامل العامة ، فإن المصفوفة الهيكلية تصبح مساوية لمصفوفة الصورة $A = [a_{jp}]$.

أي أنه لتطبيق التحليل العاملي بشكل عام يلزمنا الحصول على مصفوفتي الهيكلية والصورة، لأن مصفوفة الهيكلية تمكننا من إظهار معاملات الارتباط بين المتحولات Z_j والعوامل F_p، التي تعتبر ضرورية لتحديد أي العوامل أكثر تأثيراً على المتحولات المدروسة وعلى تقديراتها اللاحقة.

أما الصورة فهي تكون على شكل معادلات انحدار تعبر عن العلاقات الخطية بين تلك المتحولات Z_j والعوامل F_p . وهي تستخدم لدراسة الارتباط بين المتحولات عند اختبار معنوية الحل الأخير. وهي ستكون مفيدة عند مقارنة نظم مختلفة من العوامل لتحليل عدد معين من المتحولات .

وبصورة عامة فإننا نرمز للمصفوفة الهيكلية المؤلفة من معاملات ارتباط المتحولات Z_j والعوامل F_p، بالرمز S ونكتبها كما يلي :

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \dots & S_{1m} \\ S_{21} & S_{22} & S_{23} & \dots & S_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ S_{n1} & S_{n2} & S_{n3} & \dots & S_{nm} \end{bmatrix} = \qquad (49-1)$$

حيث رمزنا بـ S_{jp} لمعامل الارتباط بين المتحول Z_j والعامل العام F_p أي أن :

$$S_{jp} = r_{(Z_j,F_p)} \qquad \begin{matrix} (j : 1\ 2\ 3 \dots .n) \\ (p: 1\ 2\ 3 \dots .m) \end{matrix} \quad \text{حيث أن} \qquad (50-1)$$

وإذا اقتصرنا في المعالجة على العوامل العامة F_p فإن صورة النموذج (1−25) تأخذ الشكل التالي :

$$\mathfrak{z} = A * \mathfrak{f} \qquad (51-1)$$

وإذا ضربنا من اليمين طرفي هذه العلاقة بمنقول مصفوفة القيم العددية $\mathfrak{f}'$ وقسمنا الناتج على (N−1) نحصل على أن:

$$\mathfrak{z} * [\mathfrak{f}'/(N-1)] = A * \left(\mathfrak{f} * \mathfrak{f}'/(N-1)\right) \qquad (52-1)$$

وهنا نلاحظ أن الطرف الأيسر $\left(\mathfrak{z} * \frac{\mathfrak{f}'}{N-1}\right)$ هو نفس المصفوفة الهيكلية S المعرفة في (1−49) أي أن:

$$S = \mathfrak{z} * \frac{\mathfrak{f}'}{N-1} = A * \left(\frac{\mathfrak{f} * \mathfrak{f}'}{N-1}\right) \qquad (53-1)$$

كما نلاحظ أن الجداء $\left(\frac{\mathfrak{f}*\mathfrak{f}'}{N-1}\right)$ ماهو إلا مصفوفة الارتباطات الزوجية بين العوامل العامة F_p نفسها، ونرمز لها بـ ϕ وهي تأخذ الشكل التالي:

$$\phi = \left(\frac{\mathfrak{f} * \mathfrak{f}'}{N-1}\right) = \begin{bmatrix} 1 & r_{(F1F_2)} & r_{(F1F_3)} & \dots & r_{(F1F_m)} \\ r_{(F2F_1)} & 1 & r_{(F2F_3)} & \dots & r_{(F2F_m)} \\ \dots & \dots & \dots & \dots & \dots \\ r_{(FmF_1)} & r_{(FmF_2)} & \dots & \dots & 1 \end{bmatrix} \qquad (54-1)$$

وبتعويض (1−54) في العلاقة (1−53) نحصل على أن:

$$S = A * \phi \qquad (55-1)$$

وهكذا نكون قد حصلنا على علاقة تابعية تربط بين مصفوفة الصورة A ومصفوفة الهيكل S , وهي تساوي ناتج جداء مصفوفة الصورة A في مصفوفة معاملات الارتباط بين العوامل العامة ϕ .

والآن لنضرب من اليمين طرفي العلاقة (1−55) بالمقلوب ϕ^{-1} ، فنحصل على علاقة لحساب مصفوفة الصورة A وهي :

$$A = S * \phi^{-1} \qquad (56-1)$$

وباستخدام هذه العلاقة التي تربط مصفوفة الصورة مع المصفوفة الهيكلية، يمكننا استخراج عدة علاقات لحساب عناصر المصفوفة الارتباطية R ، حيث نجد أنه حسبما ذكرنا سابقاً في (1−14) أن المصفوفة الارتباطية للمتحولات Z_j تعرف بالعلاقة التالية :

$$R = \mathfrak{z} * \mathfrak{z}'/(N-1) \qquad (57-1)$$

وبما أن لدينا من (1−51) أن: $\mathfrak{z} = A * \mathfrak{f}$ ، وبتعويض ذلك في (1−57) نحصل على أن تقدير المصفوفة R يساوي:

$$\tilde{R} = \frac{1}{N-1} A * \mathfrak{f} * \mathfrak{f}' * A' = A\left[\frac{1}{N-1} * \mathfrak{f} * \mathfrak{f}'\right] * A' \qquad (58-1)$$

ومن العلاقة (1−54) نحصل على أن:

$$\tilde{R} = A * \phi * A' \qquad (59-1)$$

وإذا أخذنا بعين الاعتبار العلاقة (1−55) نحصل على أن:

$$\tilde{R} = S * A' \qquad (60-1)$$

ولكن بما أن المصفوفة ϕ متناظرة فإن: $\phi' = \phi$ ، وبما أن: $S = A * \phi$، فإن:

$S' = \phi' A' = \phi * A'$ ، وبالتعويض في *(1−59)* نحصل على التقدير من العلاقة الهامة التالية :

$$\tilde{R} = A * S' \qquad (61-1)$$

وإذا كانت جميع العوامل العامة F_p غير مرتبطة، فإن : ($r(F_p,F_q)=0 \quad :p\neq q$) ، وإن المصفوفة ϕ تصبح مصفوفة واحدية $(\phi = I)$ وعندها فإن العلاقة (1−59) تأخذ الشكل المخفف التالي :

$$\tilde{R}_h = A * A' \qquad (62-1)$$

وهذه هي العلاقة الأساسية التي تستخدم في حساب أو تقدير الأمثال في مصفوفة الصورة A من خلال المصفوفة المخففة R_h (كما رأينا سابقاً)، وهذا يعني أنه يمكننا تقدير معاملات المصفوفة R_h من خلال جداء مصفوفة الصورة A في منقولها A' فنحصل على المصفوفة المخففة R_h .

ومن العلاقة (1−54) نلاحظ أنه إذا كانت العوامل F_p غير مرتبطة (فإن $r(F_p,F_q)=0 \quad :p\neq q$)، وإن المصفوفة ϕ تصبح واحدية (ϕ=I)، وإن عناصر المصفوفة الهيكلية S تصبح مساوية لعناصر المصفوفة الصورية ،أي يكون لدينا :

$$S=A \qquad (a62-1)$$

وبالمقابل نجد أن العلاقة (1−62) تعني أنه يمكننا أن نحسب عناصر المصفوفة A من خلال المصفوفة الارتباطية المخففة $\tilde{R}_h$. فإذا كتبنا تلك العلاقة على الشكل التالي: $\tilde{R}_h = A * A'$ ، أو على الشكل المفصل كما يلي :

$$\tilde{R}_h = \begin{bmatrix} h_1^2 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & h_2^2 & r_{23} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & h_n^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix} * \begin{array}{c} \begin{matrix} a'_1 & a'_2 & a'_3 & \dots & a'_n \end{matrix} \\ \begin{bmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & a_{n3} & \dots & a_{nm} \end{bmatrix} \end{array}$$

فإنه وبإجراء الجداء الذي في الطرف الأيمن ومطابقته مع عناصر المصفوفة المخففة $\tilde{R}_h$ ، نحصل بالنسبة لعناصر السطر الأول على n معادلة من الشكل التالي (حيث a_i هو السطر i في المصفوفة A و a'_i هو منقوله العمود في A') :

$$\begin{aligned} h_1^2 &= a_1 * a'_1 = a_{11}^2 + a_{12}^2 + a_{13}^2 + \dots + a_{1m}^2 \\ r_{12} &= a_1 * a'_2 = a_{11}a_{21} + a_{12}a_{22} + a_{13}a_{23} + \dots + a_{1m} * a_{2m} \\ r_{13} &= a_1 * a'_3 = a_{11}a_{31} + a_{12}a_{32} + a_{13}a_{33} + \dots + a_{1m} * a_{3m} \qquad (63-1) \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \\ r_{1n} &= a_i * a'_n = a_{11}a_{n1} + a_{12}a_{n2} + a_{13}a_{n3} + \dots + a_{1m} * a_{nm} \end{aligned}$$

وبسبب التناظر يكون لدينا $\frac{n(n+1)}{2}$ معادلة مستقلة تقابل العناصر القطرية وغير القطرية للمصفوفة المخففة R_h ، وتكتب على الشكل التالي:

$$r_{jk} = a_j * a'_k = \sum_{p=1}^{m} a_{jp} * a_{kp} \qquad \begin{matrix} (j \neq k) \\ (j.k = 1\ 2\ 3 \dots .n) \end{matrix} \qquad (64-1)$$

$$r_{jj} = h_j^2 = \sum_{p=1}^{m} a_{j.p}^2 \qquad (j = 1\ 2\ 3 \dots n) \quad (\text{للعناصر القطرية}) \qquad (65-1)$$

ولحل هذه المعادلات نحتاج إلى معالجات خاصة وفي مقدمتها إيجاد تقديرات خاصة للتشاركيات h_j^2 كما سنرى في الفقرة التالية .

وهنا نلاحظ أن الجداء $A * A'$ يعطينا أن العناصر القطرية في المصفوفة $\tilde{R}_h$ تكون مساوية للتشاركيات h_j^2 وغير مساوية للواحد، وإن هذا يعني أن النموذج يكون مقتصراً على العوامل العامة F_p فقط، ولا يتضمن عوامل التميز U_j ، وفي هذه الحالة لا نضع فرضية حول وجود عوامل التميز U_j .

ولكن في بعض الحالات قد يكون ضرورياً أن نجعل العناصر القطرية في $\tilde{R}$ مساوية للواحد ، لذلك نضيف لكل تشاركية h_j^2 مقدار التميز d_j^2 ، فنحصل على مصفوفة جديدة نرمز لها بـ $[\tilde{R}_h + D^2]$ وهي تساوي المصفوفة الارتباطية الأصلية R ، أي أن:

$$R = R_h + D^2 \qquad (66-1)$$

وعندها فإن العلاقة (1-59) تأخذ شكلاً موسعاً بدلالة المصفوفة الموسعة M كما يلي :

$$\tilde{R} = [\tilde{R}_h + D^2] = M * \begin{bmatrix} \phi & 0 \\ 0 & I \end{bmatrix} * M' \qquad (67-1)$$

حيث أن M هي المصفوفة الموسعة المشكلة من المصفوفتين A و D كما يلي:

$$M = [A \backslash D] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} & d_1 & 0 & 0 & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & d_2 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & 0 & d_n \end{bmatrix}$$

و يمكن كتابة العلاقة(1−67) على الشكل التالي :

$$\tilde{R}=\begin{bmatrix}(h_1^2+d_1^2) & r_{12} & \dots & r_{1n}\\ r_{21} & (h_2^2+d_2^2) & \dots & r_{2n}\\ \dots & \dots & \dots & \dots\\ r_{n1} & r_{n2} & \dots & (h_n^2+d_n^2)\end{bmatrix}=\begin{bmatrix}a_{11}a_{12} & \dots & a_{1m}d_1000\\ a_{21}a_{22} & \dots & a_{2m}0d_200\\ \dots & \dots & \dots\\ a_{n1}a_{n2} & \dots & a_{nm}000\ d_n\end{bmatrix}*$$

$$\begin{bmatrix}1 & r_{F_1F_2} & \dots & r_{F_1F_m} & 00000\\ r_{F_2F_1} & 1 & \dots & r_{F_2F_m} & 00000\\ \dots & \dots & \dots & \dots & \dots\\ r_{F_mF_1} & r_{F_mF_2} & \dots & 1 & 10000\\ 0 & 0 & \dots & \dots & 01000\\ \vdots & \vdots & \dots & \dots & 00100\\ 0 & 0 & \dots & \dots & 00001\end{bmatrix}*\begin{bmatrix}a_{11} & a_{21} & \dots & a_{n1}\\ a_{12} & a_{22} & \dots & a_{n2}\\ \vdots & \vdots & \dots & \vdots\\ a_{1m} & a_{2m} & \dots & a_{nm}\\ d_1 & 0 & 0 & 0\\ 0 & d_2 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & d_3\end{bmatrix}\quad(68-1)$$

7−1: أساليب تقدير قيم التشاركيات $\mathbf{h_j^2}$:

لقد تعرفنا من (1−30) على أن تشاركية المتحول Z_j والتي رمزنا لها بـ h_j^2 تساوي مجموع مربعات أمثال المتحول Z_j على جميع العوامل العامة F_p، أي أنها تساوي :

$$h_j^2=a_{j1}^2+a_{j2}^2+a_{j3}^2+\dots+a_{jm}^2 \qquad (69-1)$$

وإن قيمتها تشكل جزءاً من تباين Z_j المساوي للواحد ($S_j^2=1$) ، لذلك فإنها تحقق العلاقة التالية :

$$0\le h_j^2\le 1 \qquad (70-1)$$

كما وجدنا أن قيم هذه التشاركيات تحتل القطر الرئيسي في المصفوفة الارتباطية المخففة الناجمة عن الجداء $R_h=A*A'$، وإن هذه القيم مازالت مجهولة لأن المصفوفة A مازالت مجهولة . ومن هنا ظهرت مشكلة التشاركية التي تقتضي إيجاد تقدير لهذه التشاركيات قبل المباشرة في حل المسألة العاملية واستخلاص العوامل الرئيسية. وإن إيجاد تقديرات للتشاركيات h_j^2 ووضعها في المصفوفة المخففة R_h يجعل جميع عناصر تلك المصفوفة معلومة .

ولهذا كان لابد من إيجاد تقديرات أولية لهذه التشاركيات ووضعها على قطر المصفوفة R_h ثم متابعة الحسابات للحصول على أمثال العوامل الرئيسية a_{jp} في المصفوفة الصورة A .

وأخيراً نشير إلى أن معظم الطرائق المستخدمة في التحليل العاملي تعتمد على المصفوفة المخففة R_h، التي تتضمن قيماً جاهزة للتشاركيات h_j^2 موضوعة على القطر الرئيسي للمصفوفة R_h .

وهناك عدة أساليب لتقدير هذه التشاركيات [Harman p. 97] هي باختصار كما يلي:

1− اختيار أكبر عنصر في كل عمود(أوسطر) من المصفوفة الأصلية R (بدون القطر)، لذلك ندرس قيم المعاملات الارتباطية في المصفوفة الارتباطية الأصلية، ونحذف كل عنصر قطري فيها (المساوي للواحد) ونستبدله بأكبر عنصر في عموده (أو سطره) فنحصل على مصفوفة جديدة مخففة R_h، نستخدمها لمتابعة التحليل العاملي. ولكن هذا الأسلوب لا يصلح إذا كان عدد المتحولات n صغيراً.

2- تقدير التشاركيات h_j^2 عن طريق حساب متوسط معاملات ارتباط كل متحول Z_j مع المتحولات الأخرى (غير القطرية) ووضعها على القطر. باستخدام العلاقة التالية:

$$h_j^2 = \frac{1}{n-1}\sum_{k\neq j=1}^{n} r_{jk} \qquad (k \neq j) \qquad (71-1)$$

3- حساب تقديرات لـ h_j^2 من الثلاثية التالية :

$$h_{jj}^2 = \frac{r_{jk} * r_{j\ell}}{r_{k\ell}} \qquad (72-1)$$

حيث أن k و ℓ هما دليلا المتحولين Z_k و Z_ℓ اللذين يكون معامل ارتباطهما مع المتحول Z_j أكبر من ارتباطهما من المتحولات الأخرى .

4- لقد تم تطوير الطريقة السابقة وحساب تقديرات h_j^2 من العلاقة التالية التي تدعى بمتوسط الثلاثيات وهي :

$$h_j^2 = \frac{1}{v}\sum_{k<l=1}^{n}\left[\frac{r_{jk} * r_{j\ell}}{r_{k\ell}}\right] \qquad (k, \ell \neq j) \qquad (73-1)$$

حيث v هو عدد الثلاثيات المختلفة التي يمكن تشكيلها لمعاملات المتحول Z_j مع المتحولات الأخرى k و ℓ أي أن $(k, \ell \neq j)$

5- حساب التقدير التام للتشاركيات h_j^2 : يتم حساب هذا التقدير بعد استبدال العناصر القطرية الواحدية في المصفوفة الارتباطية R بأكبر عنصر من كل عمود (أو سطر)، وذلك حسب الأسلوب (1) ، فنحصل على مصفوفة مخففة أولى R_{h1} ، ونستخدمها لإيجاد تقديرات جديدة للتشاركيات h_j^2 وحسابها من العلاقة التالية :

$$h_j^2 = \frac{\left(\sum_{k=1}^{n} r_{jk}\right)^2}{\sum_{\ell=1}^{n}\sum_{k=1}^{n} r_{\ell k}} = \frac{Z_j \text{ مربع مجموع عناصر السطر}}{R_h \text{ المجموع الكلي لجميع عناصر المصفوفة}} \qquad (74-1)$$

6- حساب تقدير h_j^2 بطريقة المعدل العام، بعد حذف العناصر القطرية من المصفوفة الأصلية، واستبدالها بالأعداد h_j^2 المحسوبة من العلاقة التالية :

$$h_j^2 = \frac{n}{n-1} * \frac{\left(\sum_{k=1}^{n} r_{jk}\right)^2}{\sum_{\ell=1}^{n}\sum_{k=1}^{n} r_{\ell k}} \qquad (\mathrm{k} \neq \mathrm{j}) \quad (\ell \neq j) \qquad (75-1)$$

7- حساب مربع معامل الارتباط المتعدد (معامل التحديد) R_j^2 لعلاقة الانحدار المتعدد، بين المتحول Z_j مع المتحولات الأخرى. لذلك نحسب مقلوب المصفوفة الارتباطية الأصلية R فنحصل على المصفوفة R^{-1} والتي سنرمز لعناصرها بالرموز r^{jk}، وعندها يمكن تقدير التشاركية h_j^2 للمتحول Z_j من العلاقة :

$$\tilde{h}_j^2 = R^2_{j123(j-1),(j+1)\ldots n} = 1 - \frac{1}{r^{jj}} \qquad (76-1)$$

وتعد المعاملات R_j^2 أفضل تقديرات للتشاركيات h_j^2، ولكنها لا تساويها تماماً. بل إن قيمها تبقى أقل من قيمة h_j^2، بمقدار ما، يعبر عن أمثال الخطأ d_j، أي أن :

$$d_j^2 = h_j^2 - R_j^2 \quad \text{وأن} \quad R_j^2 \le h_j^2 \qquad (77-1)$$

8- يمكن تقدير قيم التشاركيات h_j^2 من نموذج المكونات الأساسية المعرف بالنموذج العام (1–19)، ولذلك نأخذ المصفوفة الارتباطية الأصلية R كما هي (نتركها تتضمن واحدات على قطرها الرئيسي). ونحسب منها أمثال المكونات الأساسية a_{kp}. فنحصل على n مكوناً أساسياً (حسب الفقرة 2–1)، ثم نقوم بحساب القيم الذاتية أو الجذور الكامنة λ_p المقابلة للمكونات الأساسية من العلاقة :

$$\lambda_p = \sum_{k=1}^{n} a_{kp}^2 \quad \text{(نربعها ونجمعها عمودياً)} \quad (p: 1\ 2\ 3 \ldots n) \qquad (78-1)$$

وأخيراً نقوم بحذف كل المكونات التي تقابلها قيماً ذاتية أصغر من الواحد: أي نحذف المكون الأساسي F_p إذا كان: $\lambda_p < 1$ ، وبذلك يبقى لدينا m عاملاً أساسياً ويتحول نموذج المكونات الأساسية إلى نموذج العوامل الرئيسية (1–20)، ومن النموذج الأخير نحسب مجموع مربعات أمثال كل متحول Z_j على العوامل الرئيسية المتبقية فنحصل على تقديرات للتشاركيات h_j^2 ونحسبها من العلاقة :

$$h_j^2 = \sum_{p=1}^{m} a_{jp}^2 \quad \text{(نربعها ونجمعها أفقياً)} \quad (j: 1\ 2\ 3 \ldots n) \qquad (79-1)$$

ثم نضع هذه التشاركيات على القطر الرئيسي للمصفوفة R ونتابع عمليات التحليل حسب نموذج العوامل الرئيسية (1–20) .

9- أسلوب المعاودة: وهو ينطلق من وضع أي أرقام تقديرية لـ h_j^2 على القطر الرئيسي للمصفوفة R (بشرط أن تكون $0 \le h_j^2 \le 1$ فنحصل على مصفوفة مخففة R_{h1} . ثم نقوم بحل النموذج الناتج، فنحصل على أمثال العوامل الرئيسية الداخلة فيه ، ومنها نحسب قيم تشاركيات كل المتحولات Z_j فنحصل على قيم محدودة لها مثل h_j^2 .

نعوض قيم هذه التشاركيات من جديد في المصفوفة R فنحصل على مصفوفة مخففة أخرى R_{h2} . ثم نقوم بحل النموذج من جديد فنحصل على أمثال جديدة للعوامل الرئيسية الداخلة فيه، ومنها نحسب قيم التشاركيات الجديدة المقابلة للمتحولات Z_j فنحصل على قيم ثانية للتشاركيات h_j^2 . وتكون قريبة جداً من القيم الأولى .

نعوض قيم التشاركيات الثانية من جديد في المصفوفة R فنحصل على مصفوفة مخففة R_{h3} . ثم نقوم بحل النموذج من جديد فنحصل على أمثال جديدة للعوامل الرئيسية الداخلة فيه، ومنها نحسب القيم الثالثة للتشاركيات h_j^2 . وهكذا نكرر هذه العمليات حتى تستقر قيم التشاركيات عند قيم لا يتجاوز الفرق بين القيم في التجربتين الاخيرتين حد الدقة (الخطأ) المحدد بـ(0.001 أو 0.0001) . وأخيراً نحصل على القيم الفعلية أو الحقيقية للتشاركيات h_j^2، فنعتمدها ونعتمد الحل الناتج عنها في المصفوفة الأخيرة .

ملاحظة: إن أسلوب المعاودة هو أفضل الأساليب لتقدير قيم التشاركيات h_j^2 ، لأنه يمكن أن يعطينا قيماً تقديرية أقرب ما تكون للقيم الحقيقية. وهي تتقارب كلما زدنا من عدد مرات المعاودة .ولكن هذه الطريقة تحتاج إلى حسابات كثيرة وتنفذ بواسطة الحواسيب.

أما الأساليب الأخرى فتعطينا تقديرات محددة ولا يمكننا معرفة مدى دقتها .

مثال (1−3): لنفترض أنه لدينا المصفوفة الارتباطية لـ 6 المتحولات المعروضة في الجدول التالي:

جدول (1−5) معاملات الارتباط لستة متحولات :[Harman p.103]

المتحولات	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	$\sum r_{jk}$
Z_1	1	0.72	0.75	0.49	0.42	0.28	2.66
Z_2	0.72	1	0.78	0.42	0.36	0.24	2.52
Z_3	0.75	0.78	1	0.35	0.30	0.20	2.38
Z_4	0.49	0.42	0.35	1	0.42	0.28	1.96
Z_5	0.42	0.36	0.30	0.42	1	0.24	1.74
Z_6	0.28	0.24	0.20	0.28	0.24	1	1.24

عند دراسة معاملات الارتباط في هذا المثال نجد أن المتحولات الثلاثة الأولى $Z_1Z_2Z_3$ مرتبطة مع بعضها بعلاقات قوية، أما المتحولات الثلاثة الأخرى $Z_4Z_5Z_6$ ضعيفة الارتباط فيما بينها، وإذا حسبنا مرتبة هذه المصفوفة لوجدناها range=2، لذلك نفترض أنه يمكننا التعبير عن هذه المتحولات الستة بدلالة عاملين أساسيين فقط هما: F_1 و F_2 .

وسنقوم بتقدير التشاركيات h_j^2 للمتحولات Z_j حسب الأساليب المختلفة المذكورة أعلاه . وقبل كل شيء نقوم بحذف العناصر القطرية (الواحدية) من المصفوفة الارتباطية الأصلية وذلك لتسهيل عمليات حساب التشاركيات h_j^2 ولوضعها مكانها على القطر الرئيسي فنجد أن :

1− إن تقديرات h_j^2 حسب أكبر عنصر في كل سطر (أو عمود) فيها تساوي كما يلي:

$$h_j^2: 0.75 , 0.78 , 0.78 , 0.49 , 0.42 , 0.28$$

ولقد تم وضعها في العمود (1) من الجدول (1−7) .

2- يمكننا تقدير التشاركيات h_j^2 بواسطة متوسطات المعاملات غير القطرية في كل سطر (أو عمود) وحساب h_j^2 من العلاقة (1-71) كما يلي:

$$h_j^2 = \frac{1}{n-1}\sum_{k=1}^{n} r_{jk} \qquad (k \neq j)$$

وكمثال على ذلك نجد أن: $h_1^2 = \frac{1}{5}(0.72 + 0.75 + 0.49 + 0.42 + 0.28) = 0.53$

ولقد وضعنا هذه القيم في العمود (2) من الجدول (1-7)

جدول (1-7) قيم التشاركيات h_j^2 المحسوبة بمختلف الأساليب

الاسلوب / المتحولات	1	2	3	4	5	6	7
	اسلوب أكبر عنصر في القطر	اسلوب متوسط الاسطر	اسلوب الثلاثية	متوسط الثلاثيات	التقدير التام	المعدل العام	معامل التحديد R_j^2
Z_1	0.75	0.53	0.69	0.69	0.73	0.68	0.66
Z_2	0.78	0.50	0.75	0.75	0.68	0.61	0.66
Z_3	0.78	0.48	0.81	0.81	0.62	0.54	0.69
Z_4	0.49	0.39	0.39	0.49	0.38	0.37	0.32
Z_5	0.42	0.35	0.36	0.36	0.29	0.29	0.25
Z_6	0.28	0.25	0.16	0.16	0.14	0.14	0.12

3- أما تقديرات h_j^2 حسب الثلاثيات فحسبناها من العلاقة (1-72) ووضعناها في العمود (3) من الجدول (1-7) وكمثال على ذلك نذكر إننا حسبنا h_1^2 كما يلي :

$$h_1^2 = \frac{r_{13}r_{12}}{r_{23}} = \frac{0.75 * 0.72}{0.78} = 0.69$$

ولكن واجهتنا مشكلة في حساب h_4^2، لأن أكبر معامل ارتباط من سطر Z_4 يساوي 0.49 ولأن قيمتا المعاملين التاليين له متساويتان $r_{42} = r_{45} = 0.42$ ، لذلك قمنا بحساب قيمتين لـ h_4^2 ثم أخذنا متوسطيهما كما يلي :

$$h_4^2 = \frac{r_{41}r_{42}}{r_{12}} = \frac{0.49 * 0.42}{0.72} = 0.29$$

$$h_4^2 = \frac{r_{41}r_{45}}{r_{15}} = \frac{0.49 * 0.42}{0.42} = 0.49$$

وإن متوسطيهما يساوي : $\bar{h}_4 = \frac{0.29+0.49}{2} = 0.39$

كما قمنا بحساب تقديرات h_j^2 بواسطة العلاقة (1-73) التي تعبر عن متوسطات الثلاثيات ووضعناها في العمود (4) ، كما تم حساب تقديرات h_j^2 بطريقة التقدير التام (1-74) من المصفوفة المعدلة ووضعناها في العمود (5) فمثلاً كانت (بعد وضع أكبر عنصر على القطر) .

$$h_1^2 = \frac{\sum\left(r_{jk}\right)^2}{\sum\sum r_{k\ell}} = \frac{(2.66+0.75)^2}{16.00} = \frac{(3.41)^2}{16.00} = 0.73$$

كما تم حساب تقديرات لـ h_j^2 حسب المعدل العام من(1−75) ووضعناها في العمود (6) ، حيث حسبنا h_1^2 من العلاقة:

$$h_1^2 = \frac{n}{n-1} * \frac{\sum\left(r_{jk}\right)^2}{\sum\sum r_{jk}} = \frac{6}{5}\frac{(2.66)^2}{12.50} = 0.68$$

كما حسبنا تقديرات لـ h_j^2 من معاملات التحديد R_j^2 للانحدار المتعدد ووضعناها في العمود (7) .

وهنا نلاحظ أن تقديرات h_j^2 حسب هذه الأساليب متقاربة، ولكنها غير متطابقة ، وإن أسوأها هو أسلوب متوسط الأسطر ، لذلك يتم اللجوء إلى طريقة المعاودة للوصول إلى القيم الحقيقية لهذه التشاركيات .

الفصل الثاني

طرائق استخلاص العوامل وحساب الأمثال عليها

Methods of Extraction Factors

2-1 : تمهيد:

لقد وجدنا في الفصل السابق أن العلاقة (1-62) هي العلاقة الأساسية لحساب الأمثال a_{jp} في مصفوفة الصورة العاملية A، وهي تشترط أن تكون العوامل F_p معيارية وغير مرتبطة مع بعضها البعض، وتكون مستقلة عن عوامل التميز المعيارية U_j ، وهناك عدة طرائق لحساب هذه الأمثال، إلا أن كل منها يعطينا لا نهاية من الحلول المقبولة، وعلينا البحث من بينها عن الحل المناسب .

ولقد أشرنا سابقاً إلى أن كل متحول Z_j يمكن تمثيله في فضاء العوامل F_p على شكل شعاع محدد طوله h_j، وعندها تكون مساقطه على المحاور الاحداثية المتعامدة في فضاء العوامل F_p هي الأمثال a_{jp} .

ولتبسيط التوصيف الرياضي وتسهيل تفسير مضمونة يفضل أن تكون تلك المساقط موزعة على المحاور الاحداثية بشكل بسيط ومناسب. لذلك يتم تدوير المحاور الاحداثية عدة مرات حتى نحصل على الحل البسيط المناسب الذي يعطينا أبسط هيكل لذلك التوصيف، ولذلك قام (Thurstone 1947) بوضع عدة مواصفات للهيكل البسيط S نلخصها بما يلي :

1- كل سطر من مصفوفة الهيكل العاملي S يجب أن يتضمن صفراً واحداً على الأقل .
2- كل عمود من مصفوفة الهيكل العاملي S يجب أن يتضمن m صفراً على الأقل .
3- لكل زوج من الأعمدة يمكن أن نجد عدة متحولات، بحيث تكون عناصر الهيكل العاملي المقابلة لها في أحد العمودين مساوية للصفر بينما تكون غير معدومة في العمود الآخر .
4- إذا كان عدد العوامل m أكبر أو يساوي (4)، ففي أي زوج من الأعمدة يجب أن تكون نسبة كبيرة من المتحولات Z_j لها معاً أمثال صفرية .
5- في كل زوج من الأعمدة يوجد عدد قليل من المتحولات Z_j بحيث تكون عناصرها في العمودين معاً غير معدومة .

وأخيراً نشير إلى أن مبدأ الهيكل البسيط يطبق على مختلف الحلول العاملية المتعامدة أو المائلة. ولذلك سنقوم بعدة عمليات لتدوير المحاور الإحداثية بحثاً عن ذلك الهيكل البسيط . وهكذا يمكن اعتبار عملية التدوير هي وسيلة لتخفيض درجة تعقيد تمثيل المتحولات، وتهدف إلى الحصول على الحل الذي يتم فيه تمثيل كل متحول بعامل واحد، وبحيث تكون درجة تعقيد أي متحول مساوية للواحد – له مسقط واحد – وهو الحل الذي يحقق شروط الهيكل البسيط والمبين في الشكل(2-1) التالي:

العوامل/المتحولات	F_1	F_2	F_3	U_1	U_2	U_3	U_4	U_5	U_6
Z_1	X			x					
Z_2	X				X				
Z_3	X					x			
Z_4		X					x		
Z_5		X						X	
Z_6			x						X

الشكل (2−1) أحد أشكال الهيكل البسيط

وهناك عدة طرائق لحساب أمثال النماذج العاملية a_{jp} والحصول على الحلول الأولية المتعامدة أو المائلة وسنعرضها كما يلي:

أ− طريقة المكونات الأساسية: وتعتمد على النموذج العام (1−19) المؤلف من n مكوناً.

ب− الطرائق التي تتطلب تقديرات للتشاركيات h_j^2 وتعتمد على النموذج المختصر (1−20) .

1− طريقة العوامل الرئيسية: وتكون حلولها مؤلفة من m عاملاً فقط ، حيث $(m < n)$.

2− طريقة المجموعات: وتكون حلولها مارة من مراكز المجموعات وعدد عواملها بعدد المجموعات.

ج− الطرائق التي تتطلب تقدير عدد العوامل العامة m وأهمها:

1− طريقة الإمكانية العظمى .

2− طريقة البواقي الصغرى .

3− طرائق النماذج البسيطة (الأحادي+ الثنائي+ المزدوج+ المثلثي +المتعدد) .

د− طرائق العوامل المتعددة: وهي تعتمد على وجود حلول أولية وتستخدم في الإحداثيات المتعامدة أو المائلة.

2−2 طريقة المكونات الأساسية (Principal components) :

تعتمد طريقة المكونات الأساسية (PC) على النموذج العاملي المعرف في العلاقة (1−19) ، الذي يعطينا n متحولاً معيارياً Z_j بدلالة n عاملاً معيارياً F_p كما يلي:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \cdots + a_{jn}F_n \quad : (j: 123 \dots n) \qquad (1-2)$$

ويشترط فيه أن تكون العوامل $F_1, F_2, \dots, F_n$ غير مرتبطة مع بعضها البعض، أي أنها متعامدة وتشكل فضاءً خاصاً في R^n تنسب إليه جميع أشعة المتحولات Z_j .ويمكن كتابة هذه المعادلات بشكل مصفوفي كما يلي:

$$Z = A * F \qquad 1(@-2)$$

كما يمكن عرض الصورة العاملية لهذا النموذج وتباينات المتحولات والعوامل في جدول كما يلي :

جدول (2−1) الصورة العاملية للنموذج (2−1):

المتحولات \ العوامل	F_1	F_2	F_3	……	F_n	تباينات Z_j	التشاركيات h_j^2
Z_1	a_{11}	a_{12}	a_{13}	……	a_{1n}	$S_1^2=1$	1
Z_2	a_{21}	a_{22}	a_{23}	……	a_{2n}	$S_2^2=1$	1
Z_3	a_{31}	a_{32}	a_{33}	……	a_{3n}	$S_3^2=1$	1
⋮	⋮	⋮	⋮	……	⋮	⋮	⋮
Z_n	a_{n1}	a_{11}	a_{11}	……	a_{nn}	$S_n^2=1$	1
تباينات العوامل V_p	V_1	V_2	V_3	……	V_n	$S^2=n$	n
النسبة المئوية من التباين الكلي	P_1	P_2	P_3	……	P_n	100	100

وتهدف هذه الطريقة إلى جعل تباينات العوامل V_p تأخذ على التوالي أكبر حصة ممكنة من التباين الاجمالي S^2 للمتحولات Z_j .

أي أن هذه الطريقة تهدف إلى جعل تباين العامل الأول V_1 يشكل أكبر حصة ممكنة من التباين الاجمالي S^2 (المساوي لـ n) ، ثم جعل تباين العامل الثاني V_2 يشكل أكبر حصة ممكنة من باقي التباين الاجمالي (S^2-V_1)، ثم جعل تباين العامل الثالث V_3 يشكل أكبر حصة ممكنة من باقي التباين الاجمالي $(S^2-V_1-V_2)$… وهكذا نتابع هذا الأمر لجميع العوامل الأخرى، حتى يستهلك كامل التباين الاجمالي S^2.

وعندها نحصل على التباينات المرتبة: $V_1>V_2>V_3\ldots>V_n$ والتي تحقق العلاقة:

$$\sum_{p=1}^{n} V_p = S^2 = n \qquad (2-2)$$

ومن جهة أخرى نجد من الجدول (2−1) أن التباين لكل متحول معياري Z_j معرف في النموذج (2−1) يساوي التشاركية h_j^2 ويساوي الواحد (في كل سطر). أي أن:

$$S_j^2 = \sum_{p=1}^{n} a_{jp}^2 = h_j^2 = 1 \qquad (3-2)$$

وإن التباين الاجمالي S^2 يساوي مجموع تباينات المتحولات Z_j، أي يساوي $S^2=\sum S_j^2 = n$.

ومن جهة أخرى نجد أن:

$$S^2 = \sum_{j=1}^{n} S_j^2 = \sum_{j=1}^{n}\sum_{p=1}^{n} a_{jp}^2 = \sum_{p=1}^{n}\left(\sum_{j=1}^{n} a_{jp}^2\right) = n \qquad (4-2)$$

وبالمقارنة مع العلاقة (2−2) نجد أن :

$$S^2 = \sum_{p=1}^{n}\left(\sum_{j=1}^{n} a_{jp}^2\right) = \sum_{p=1}^{n} V_p \qquad (5-2)$$

ومنها نستنتج مرة ثانية أن التباين V_p للعامل F_p كما في (1−32)يساوي :

$$V_p = \sum_{j=1}^{n} a_{jp}^2 \qquad : \quad (P{:}\,1\ \ 2\ \ 3\ldots n) \qquad (6-2)$$

أي أن التباين V_p للعامل F_p يساوي مجموع مربعات أمثال الصورة A في عمود العامل F_p مأخوذاً عمودياً على جميع المتحولات Z_j ، وهو يختلف عن التشاركيات h_j^2 التي تساوي مجموع مربعات أمثال سطر المتحول Z_j مأخوذاً أفقياً على جميع العوامل ،حيث أن:

$$h_j^2 = \sum_{p=1}^{n} a_{jp}^2 \qquad (6\,a-2)$$

والسؤال الآن كيف سنحسب أمثال العمود F_1 في الجدول (2−1) لنجعل التباين V_1 يأخذ أكبر حصة ممكنة من S^2 ؟

للجواب على هذا السؤال؟ نفترض أن العامل F_1 يرتبط مع المتحولات Z_j بعلاقة خطية (عمودية) من الشكل [Frank wood 2009 DCA]

$$F_1 = b_{11}Z_1 + b_{21}Z_2 + b_{31}Z_3 + \cdots + b_{n1}Z_n \qquad (7-2)$$

والتي يمكن كتابتها مصفوفياً كما يلي:

$$F_1 = b_1' * Z = (b_{11}b_{21}\ b_{31}\ldots b_{n1}\ ) * \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ \vdots \\ Z_n \end{bmatrix} \qquad (8-2)$$

ومنها يمكن حساب تباين F_1 كما يلي:

$$\begin{aligned} var\,(F_1) &= var(b_1' * Z) = E[(b_1' * Z)(b_1' * Z)'] = E[b_1' * Z * Z' * b_1] \\ &= b_1'E(Z * Z') * b_1 \qquad (9-2) \end{aligned}$$

وبما أن المتحولات Z_j هي متحولات معيارية فإن توقع الجداء $(Z * Z')$ حسب (1−14) يساوي :

$$E(Z * Z') = \frac{1}{N-1}ZZ' = R_{nn} = C_{nn} \qquad (10-2)$$

حيث أن: R_{nn} هي المصفوفة الارتباطية للمتحولات Z_j ، وإن: C_{nn} هي مصفوفة التباينات المشتركة للمتحولات Z_j وهما من المرتبة $n * n$ ، وعندها نجد أن التباين $var\,(F_1)$ يساوي:

$$var\,(F_1) = b_1' * R * b_1 = b_1' * C * b_1 \qquad (11-2)$$

ويصبح هدفنا الآن جعل المقدار الأخير $b_1' * R * b_1$ يأخذ أكبر حصة ممكنة من التباين الاجمالي S^2 .

أي أن هدفنا أصبح حساب عناصر الشعاع b_1 التي تعظم المقدار :

$$b_1' * R * b_1 \Rightarrow \max \qquad (12-2)$$

ولكن هذا الهدف غير محدد ويعطينا لا نهاية من الحلول المقبولة، لذلك لابد من وضع شرط عليه حتى تتمكن من الحصول على حل محدد. وإن أفضل وأسهل شرط هو أن نشترط أن يكون مربع طول الشعاع b_1 مساوياً للواحد . وهذا يعني أن يكون :

$$\|b_1\|^2 = \sum_{j=1}^{n} b_{j1}^2 = b_1' * b_1 = 1 \qquad (13-2)$$

وبذلك يتحول هدفنا إلى تعظيم المقدار $(b_1' * R * b_1)$ مع وجود الشرط $(b_1' * b_1 = 1)$، ولتحقيق ذلك الهدف نستخدم طريقة مضروب (لاغرانج) ونشكل من (2–12) والشرط (2–13) تابع (لاغرانج) كما يلي:

$$L_1 = b_1' * R * b_1 - \lambda(b_1' b_1 - 1) \qquad (14-2)$$

ثم نشتق التابع L_1 بالنسبة لـ b_1 (حسب قواعد اشتقاق المصفوفات*) ونضعه مساوياً للصفر فنحصل على أن:

$$\frac{\partial L_1}{\partial b_1} = 2R * b_1 - 2\lambda b_1 = 0 \qquad (15-2)$$

ومنها نحصل على المعادلة التالية :

$$R * b_1 = \lambda b_1 \qquad (16-2)$$

والتي يمكن كتابتها على الشكل التالي:

$$(R - \lambda I) * b_1 = 0 \qquad (17-2)$$

وهي الشكل النموذجي للمعادلة المميزة للمصفوفة R، وهذا يعني أن λ هي القيمة الذاتية للمصفوفة الارتباطية R، وأن الشعاع b_1 هو الشعاع الذاتي المقابل لها، والذي يحقق العلاقة (2–17) .[انظر الجزء الأول الفصل الثاني: القيم الذاتية للمصفوفات الشاذة]

وحتى نستبعد الحل التافه $(b_1 = 0)$ للمعادلة (2–17) نعتبر أن المصفوفة $(R - \lambda I)$ هي مصفوفة شاذة، وبذلك تكون قيمة محددها D معدومة، أي يكون لدينا:

$$D = |R - \lambda I| = 0 \qquad (18-2)$$

وبعدها نقوم بفك أو نشر هذا المحدد حسب الأسطر أو الأعمدة فنحصل على كثير حدود من الدرجة n بالنسبة لـ λ ، ويسمى المعادلة المميزة لـ λ ، وهو يكون من الشكل التالي :

$$\lambda^n + C_1\lambda^{n-1} + C_2\lambda^{n-2} + \cdots + C_{n-1}\lambda + C_n = 0 \qquad (19-2)$$

* ⁻ قواعد الاشتقاق في المصفوفات هي كما يلي:

$$\frac{\partial C'x}{\partial x} = C \quad , \quad \frac{\partial (x' * x)}{\partial x} = 2x \quad , \quad \frac{\partial (x' * B * x)}{\partial x} = 2Bx$$

وبعدها نقوم بحساب جذور هذه المعادلة فنحصل على n جذراً حقيقياً موجباً، ثم نرتبها تنازلياً ونرمز لها بالرموز $\lambda_1 \ \lambda_2 \ \lambda_3 \dots \lambda_n$ ، ونكتبها مرتبة كما يلي:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_n \geq 0 \qquad (20-2)$$

ثم نقوم بتعويض القيمة الكبرى λ_1 في المعادلة المميزة (2−17) فنحصل على جملة معادلات خطية متجانسة وغير مستقلة عن بعضها البعض (لأن محددها D معدوم)، ويكون لهذه المعادلات لا نهاية من الحلول المقبولة. وللحصول على أحدها نعطي أحد عناصر الشعاع b_1 قيمة محددة (مثل $b_{11} = 1$) ثم نعود ونحل المعادلات فنحصل على حل خاص لها نرمز له بـ $\propto_1$ حيث:

$$\propto_1 (\propto_{11}= 1, \propto_{21}, \propto_{31}, \dots, \propto_{n1}) \qquad (21-2)$$

وحتى نجعل هذا الحل الخاص يحقق الشرط (2−13) نحسب طول الشعاع $\propto_1$ من العلاقة :

$$\|\propto_1\| = \sqrt{\propto_1^2+\propto_2^2+\cdots+\propto_n^2} \qquad (22-2)$$

ثم نقسم عناصر الشعاع الخاص $\propto_1$ على طوله $\|\propto_1\|$ ، فنحصل على الشعاع الخاص الواحدي b_1 التالي:

$$b_1 = \left[\frac{\propto_{11}}{\|\propto_1\|}, \frac{\propto_{21}}{\|\propto_1\|}, \frac{\propto_{31}}{\|\propto_1\|}, \dots\dots, \frac{\propto_{n1}}{\|\propto_1\|}\right] \qquad (23-2)$$

ونرمز له بالرمز : $b_1 = [\ b_{11} \quad b_{21} \quad b_{31} \ \dots\dots \ b_{n1}\ ]$

وهو يحدد لنا الاتجاهات العامة لذلك الشعاع b_1 ولا يحدد لنا قيم عناصر الحل المطلوب b_1 من بين مضاعفاته المختلفة .

والآن لنبحث عن العلاقة التي تربط تباين العامل الأول $var(F_1)$ مع القيمة الذاتية λ_1، فنلاحظ أن قيمة λ_1 وشعاع الحل الخاص الواحدي b_1 المقابل لها يحققان المعادلة (2−16)، ومنها نجد أن:

$$R\, b_1 = \lambda_1\, b_1 \qquad (24-2)$$

ثم نضرب طرفي العلاقة (2−24) من اليسار بـ b_1' فنجد أن :

$$b_1' * R * b_1 = b_1' * \lambda_1 * b_1 = \lambda_1 b_1' * b_1$$

ومن العلاقتين (2−13) و (2−11) نحصل على أن :

$$var(F_1) = \lambda_1 \qquad (25-2)$$

أي أن تباين العامل الأول F_1 يساوي القيمة الذاتية المقابلة له λ_1، أو يساوي القيمة الذاتية الكبرى λ_1 للمصفوفة R . ولكننا في الجدول (2−1) رمزنا لتباين F_1 بالرمز V_1 وهو يساوي حسب (2−6) ما يلي:

$$V_1 = \sum_{j=1}^{n} a_{j1}^2 \qquad \text{(عمودياً)} \qquad (26-2)$$

وبالمقارنة مع (2−25) نجد أن:

$$var(F_1) = \lambda_1 = V_1 = \sum_{j=1}^{n} a_{j1}^2 = \|a_1\|^2 \qquad (27-2)$$

أي أن λ_1 تساوي مربع طول الشعاع العمودي a_1 في الجدول (2−1) وإن طول ذلك الشعاع يساوي:

$$\|a_1\| = \sqrt{\sum a_{j1}^2} = \sqrt{\lambda_1} \qquad (28-2)$$

وبما أن الاتجاهات العامة للحل معطية من خلال اتجاهات الشعاع الواحدي b_1 في (2−23)، وحتى نستطيع الحصول على قيم عناصر الشعاع a_1 والذي طوله يساوي $\sqrt{\lambda_1}$. نضرب عناصر الشعاع الواحدي b_1 بجذر القيمة الذاتية $\sqrt{\lambda_1}$ فنحصل على الشعاع a_1 كما يلي :

$$a_1 = \left[\frac{\propto_{11}\sqrt{\lambda_1}}{\|\propto_1\|}, \frac{\propto_{21}\sqrt{\lambda_1}}{\|\propto_1\|}, \frac{\propto_{31}\sqrt{\lambda_1}}{\|\propto_1\|}, \dots, \frac{\propto_{n1}\sqrt{\lambda_1}}{\|\propto_1\|}\right] \qquad (29-2)$$

وهو شعاع أمثال المتحولات Z_j على العامل F_1 وطوله يساوي $\sqrt{\lambda_1}$.

وبصورة عامة نحصل مقابل كل قيمة ذاتية λ_p للمصفوفة R على شعاع ذاتي خاص $\propto_p$ ومنه نحصل على الشعاع الواحدي b_p ، ومنه نحصل على شعاع الأمثال a_p الذي يقابل العامل F_p، ويكون تباين العامل F_p مساوياً .

$$V_p = var(F_p) = \lambda_p = \sum_{j=1}^{n} a_{jp}^2 \qquad (30-2)$$

ولتوضيح كيفية الحصول على عناصر العامل الثاني، (نضع p = 2) ونفترض أن العامل F_2 يرتبط مع المتحولات Z بالعلاقة :

$$F_2 = b_2' * Z \qquad (31-2)$$

وبنفس الطريقة السابقة نجد أن تباينه يساوي :

$$var(F_2) = b_2' * R * b_2 \qquad (32-2)$$

ولكن عناصر b_2 يجب أن تحقق شرط الطول الواحدي المشابه لـ (2−13) التالي:

$$b_2' * b_2 = 1 \qquad (33-2)$$

وأن تحقق شرطا إضافياً آخر موضوع على العامل الثاني F_2، وهو أن يكون F_2 غير مرتبط مع العامل الأول F_1 ، هذا يعني أن التباين المشترك لـ F_1 و F_2 يجب أن يكون معدوماً، أي أن:

$$cov(F_1, F_2) = 0 \qquad (34-2)$$

وبناء على (2−34) يمكن البرهان على أن هذا الشرط سيؤدي إلى العلاقات التالية:

$$cov(b_1'Z, b_2'Z) = E[b_1'Z * Z' * b_2] = b_1'E(Z * Z') * b_2 = 0$$

ومنها ومن (2−10) و(2−24) نستخلص العلاقات التالية:

$$\begin{aligned} cov(b_1'Z, b_2'Z) &= b_1' * R * b_2 = b_2' * R * b_1 = 0 \qquad (35-2) \\ &= b_1' * \lambda_2 * b_2 = b_2' * \lambda_1 * b_1 = 0 \\ &= \lambda_2 * b_1' * b_2 = \lambda_1 * b_2' * b_1 = 0 \end{aligned}$$

ومن هذه العلاقات نستنتج أن الشعاعين b_1 و b_2 يجب أن يحققا الشرطين المتكافئين :

$$b_2' * b_1 = 0 \qquad (36-2)$$
$$b_1' * b_2 = 0 \qquad (37-2)$$

وهما يعبران عن شرط واحد هو عدم ارتباط F_2 مع F_1 (أو هو شرط تعامد F_1 مع F_2) .

ولحساب عناصر الشعاع b_2 نقوم بتشكيل تابع (لاغرانج) الثاني من العلاقة (2−32) ومن الشرطين (2−33) و (2−36) كما يلي:

$$L_2 = b_2' * R * b_2 - \lambda(b_2' * b_2 - 1) - \mu(b_2' * b_1) \qquad (38 - 2)$$

ثم نشتق L_2 بالنسبة لـ b_2 ونضعه مساوياً للصفر فنجد أن:

$$\frac{\partial L_2}{\partial b_2} = 2R * b_2 - 2\lambda * b_2 - 2\mu b_1 = 0$$

ومنها نجد أن :

$$Rb_2 - \lambda b_2 - \ \mu b_1 = 0 \qquad (39 - 2)$$

ثم نقوم بضرب طرفي هذه العلاقة من اليسار بـ b_1' فنحصل على :

$$b_1' * R * b_2 - \lambda b_1'^{*} b_2 - \ \mu b_1' * b_1 = 0 \qquad (40 - 2)$$

ومن العلاقات المتعددة في (2−35) نجد أن:

$$0 - \ 0 - \ \mu * 1 = 0 \qquad (41 - 2)$$

ومنها نستنتج أن قيمة المضروب $\mu = 0$، وعندها فإن العلاقة (2−39) تأخذ الشكل التالي:

$$Rb_2 - \ \lambda b_2 = 0 \qquad (42 - 2)$$

والتي يمكن كتابتها على الشكل التالي :

$$(R - \lambda I)b_2 = 0 \qquad (43 - 2)$$

وهي الصيغة النموذجية للمعادلة المميزة للمصفوفة R، وهذا يعني أن قيمة λ فيها هي القيمة الذاتية الثانية λ_2 للمصفوفة الارتباطية R، وإن الشعاع b_2 هو الشعاع الذاتي الواحدي للمصفوفة R والمقابل لـ λ_2 والذي يحقق المعادلة (2−43) بعد تعويض λ_2 فيها .

لذلك نقوم بحساب قيمة λ_2 (أو نأخذها من الحسابات السابقة عند حساب λ_1 من (2−19) والحصول على جميع القيم المرتبة λ_p) وتعويضها في المعادلة (2−43)، وبعدها نقوم بإيجاد حل خاص للمعادلات (2−43) فنحصل على الحل الخاص $\propto_2$ ثم نقوم بتحويله إلى شعاع واحدي b_2 (كما فعلنا مع $\propto_1$) ثم نضرب عناصر b_2 بـ $\sqrt{\lambda_2}$ فنحصل على عمود الأمثال a_2 المقابل للعامل F_2 وللمتحولات Z_j في الجدول (2−1) . وهكذا نتابع العمل ونكرر نفس الخطوات والاجراءات والحسابات من F_1 حتى F_n .

وأخيراً نحصل على n شعاعاً عمودياً متعامداً للأمثال a_p هي :

$$a_1 \ \ a_2 \ \ a_3 \ \ldots\ldots\ldots \ a_n$$

وهي تقابل القيم الذاتية :

$$\lambda_1 \ \ \lambda_2 \ \ \lambda_3 \ \ldots\ldots\ldots \ \lambda_n$$

وهي تشكل الصورة العاملية المذكورة في الجدول (2−1) ، علماً بأن تباين كل عامل من هذه العوامل يساوي (عمودياً) ما يلي:

$$V_p = \lambda_p = \sum_{j=1}^{n} a_{jp}^2 \qquad (44 - 2)$$

وأخيراً نجد أن مجموع تباينات جميع هذه العوامل يساوي :

$$\sum_{p=1}^{n} V_p = \sum_{p=1}^{n} \lambda_p = S^2 = n \qquad (45-2)$$

وإن الحصة النسبية لتباين كل عامل F_p من التباين الاجمالي تساوي :

$$P = \frac{\lambda_p}{S^2} = \frac{\lambda_p}{n} \qquad (46-2)$$

ملاحظة: إذا أخذنا الأشعة المتعامدة $a_1\ a_2\ a_3\ \dots\ a_n$ التي حصلنا عليها مقابل الجذور الكامنة $\lambda_1\ \lambda_2\ \lambda_3\ \dots\ \lambda_n$ على الترتيب نجد أن كل منها يحقق المعادلة $R * a_p = \lambda_p * a_p$ وإذا شكلنا من هذه الأشعة المتعامدة مصفوفة A ، فيمكننا كتابتها على الشكل التالي :

$$R * (a_1 , a_2\ , a_3\ \dots\ a_n) = (\lambda_1 a_1 , \lambda_2 a_2 , \lambda_3 a_3\ \dots\ \lambda_n a_n)$$

$$R * A = A * \Lambda \qquad (47-2)$$

حيث أن:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad , \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

وبما أن الأشعة العمودية $a_1 , a_2 , \dots a_n$ مستقلة خطياً فإن المحدد $|A|$ يكون غير معدوم، وإنه يوجد مقلوب للمصفوفة A نرمز له بـ A^{-1} . لذلك نضرب (2–47) من اليسار بـ A^{-1} فنجد أن :

$$A^{-1} * R * A = \Lambda \qquad (48-2)$$

وبذلك تتحول المصفوفة الارتباطية R إلى مصفوفة قطرية Λ ، وتكون عناصرها القطرية مساوية للقيم الذاتية λ_p ، وتكون الأشعة العمودية في المصفوفة A هي الأشعة الذاتية لـ R .

ولكن بسبب تناظر المصفوفة الارتباطية R ، فإنه عندما نأخذ منقول الطرفين في (2–48) نجد أن:

$$A' * R * (A^{-1})' = \Lambda \qquad (49-2)$$

وهذا يعني أن أسطر المصفوفة $(A^{-1})'$ هي نفسها أصبحت الأشعة الذاتية $a_1 , a_2 , \dots a_n$ ، وذلك لأن المصفوفة A مصفوفة متعامدة فإنها تحقق العلاقة :

$$A * A' = \mathrm{I} \qquad (50-2)$$

وهذا يعني أن:

$$A^{-1} = A' \Longrightarrow (A^{-1})' = A \qquad (51-2)$$

ومنهما بسبب تناظر R تتحول العلاقة (2–49) إلى الشكل التالي:

$$A' * R * A = \Lambda \qquad (52-2)$$

وهذه المعادلة تسمى بالمعادلة الطيفية للتحليل العاملي.

وهي تعني أنه يمكننا بمساعدة أي مصفوفة متعامدة A ، أن نحول المصفوفة المربعة المتناظرة R إلى مصفوفة قطرية Λ بحيث تكون عناصرها القطرية λ_p حقيقية وتساوي القيم الذاتية لـ R ، وذلك من خلال الجداء $A' * R * A$، المعرف في (2−52) .

وتكون الأشعة العمودية a_p هي الأشعة الذاتية للمصفوفة R وهي أشعة مستقلة خطياً لأنها متعامدة .

مثال (2−1) :

لنفترض إننا نريد دراسة التمثيل العاملي لمتحولين X و Y (X علامة الرياضيات، Y علامة الفيزياء)، ولنفترض أن البيانات التي جمعناها من عينة من الطلاب أعطتنا أن المصفوفة الارتباطية بينهما كانت تساوي:

$$R = \begin{array}{c c} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \end{array}$$

لنقم الآن بتحويل X و Y إلى متحولين معياريين Z_1 و Z_2 حسب العلاقة (1−7)، ثم نفترض أن هذين المتحولين يمكن تمثيلها، بواسطة عاملين افتراضيين F_1 و F_2 حسب النموذج (1−19) كما يلي :

$$Z_1 = a_{11}F_1 + a_{12}F_2$$
$$Z_2 = a_{21}F_1 + a_{22}F_2$$

ولحساب الشعاع العمودي الأول لأمثال F_1 نحسب فوراً القيم الذاتية لـ R من المعادلة المميزة التالية :

$$(R - \lambda I)b_1 = 0$$

ولحساب القيم الذاتية λ نعتبر المصفوفة $(R - \lambda I)$ شاذة ونضع محددها مساوياً للصفر فنجد أن :

$$|R - \lambda I| \quad = 0$$
$$\begin{vmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{vmatrix} = 0$$
$$(1-\lambda)^2 - 0.25 = 0$$
$$\lambda^2 - 2\lambda + 0.75 = 0$$

نحل المعادلة الأخيرة فنجد أن يوجد لها جذران حقيقيان وموجبان هما:

$$\lambda_1 = 1.5 \qquad\qquad \lambda_2 = 0.5$$

ولإيجاد الشعاع الأول b_1 المقابل لـ λ_1 نعوض $\lambda_1 = 1.5$ في المعادلة المميزة فنجد أن:

$$\begin{bmatrix} 1-1.5 & 0.5 \\ 0.5 & 1-1.5 \end{bmatrix}\begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = 0$$

ومنها نحصل على المعادلتين المتجانستين :

$$-0.5\, b_{11} + 0.5\, b_{21} = 0$$
$$0.5\, b_{11} + 0.5\, b_{21} = 0$$

وهما معادلتان غير مستقلتين (لأن أحدهما تنتج عن الأخرى بضربها بـ -1)، لذلك نأخذ أحدهما، فنجد منها أن :

$$b_{11} = b_{21} \qquad\qquad \text{(متساويان ولهما نفس الإشارة)}$$

وهذا يجعلنا أمام لا نهاية من الحلول، لذلك نأخذ حلاً خاصاً لها ونضع $b_{11} = 1$ فنجد أن $b_{21} = 1$

وبذلك نحصل على حل خاص نرمز له بـ $\propto_1$ ويساوي :

$$\propto_1 = \begin{bmatrix} \propto_{11} \\ \propto_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

ولتحويل هذا الشعاع إلى شعاع واحدي نحسب طوله فنجد أن :

$$\|\propto_1\| = \sqrt{1^2 + 1^2} = \sqrt{2}$$

ثم نقسم عناصر الشعاع $\propto_1$ على $\sqrt{2}$ فنحصل على الشعاع الواحدي b_1 التالي:

$$b_1 = \begin{bmatrix} \frac{\propto_{11}}{\|\propto_1\|} \\ \frac{\propto_{21}}{\|\propto_1\|} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

وللحصول على الشعاع العمودي a_1 المقابل للعامل F_1 نضرب عناصر الشعاع الواحدي b_1 بجذر λ_1 أي بجذر $\sqrt{1.5}$ فنحصل على أن:

$$a_1 = \begin{bmatrix} \frac{\propto_{11}\sqrt{\lambda_1}}{\|\propto_1\|} \\ \frac{\propto_{21}\sqrt{\lambda_1}}{\|\propto_1\|} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{1.5}}{\sqrt{2}} \\ \frac{\sqrt{1.5}}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0.866 \\ 0.866 \end{bmatrix}$$

وللتأكد من الحسابات نحسب مجموع مربعات عناصر a_1 فنجد أن:

$$(0.866)^2 + (0.866)^2 = 1.5 = \lambda_1$$

ننتقل الآن لحساب عناصر الشعاع a_2 ونقوم بتعويض $\lambda_2 = 0.5$ في المعادلة المميزة فنجد أن:

$$\begin{bmatrix} 1-0.5 & 0.5 \\ 0.5 & 1-0.5 \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix} = 0$$

ومنها نحصل على المعادلتين المتجانستين غير المستقلين :

$$0.5\, b_{12} + 0.5\, b_{22} = 0$$
$$0.5\, b_{12} + 0.5\, b_{22} = 0$$

وهما عبارة عن معادلة واحدة. نأخذ أحدهما فنحصل منها على أن :

$$b_{12} = -b_{22} \qquad \text{(مختلفان بالإشارة)}$$

وهذا يجعلنا أيضاً أمام لا نهاية من الحلول المقبولة، لذلك نأخذ حلاً خاصاً لها بوضع $b_{22} = 1$ فنجد أن $b_{12} = -1$ وبالتالي نحصل على الحل الخاص $\propto_2$ التالي : $\propto_2 = \begin{bmatrix} -1 \\ +1 \end{bmatrix}$

ولتحويل هذا الشعاع إلى شعاع واحدي نحسب طوله فنجد أن:

$$\|\propto_2\| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}$$

ثم نقسم عناصر الشعاع $\propto_2$ على طوله $\sqrt{2}$ فنحصل على الشعاع الواحدي b_2 كما يلي :

$$b_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

وللحصول على شعاع الأمثال a_2 المقابل لـ F_2 نضرب عناصر الشعاع b_2 بجذر القيمة الذاتية λ_2 ، أي بجذر $\sqrt{0.5}$ ، فنجد أن :

$$a_2 = \begin{bmatrix} -\frac{\sqrt{0.5}}{\sqrt{2}} \\ +\frac{\sqrt{0.5}}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -0.5 \\ +0.5 \end{bmatrix}$$

وبذلك نحصل على الصورة العاملية التالية: $Z_1 = 0.866\, F_1 - 0.5\, F_2$

$$Z_2 = 0.866\, F_1 + 0.5\, F_2$$

ونلخص الحل النهائي في الجدول التالي:

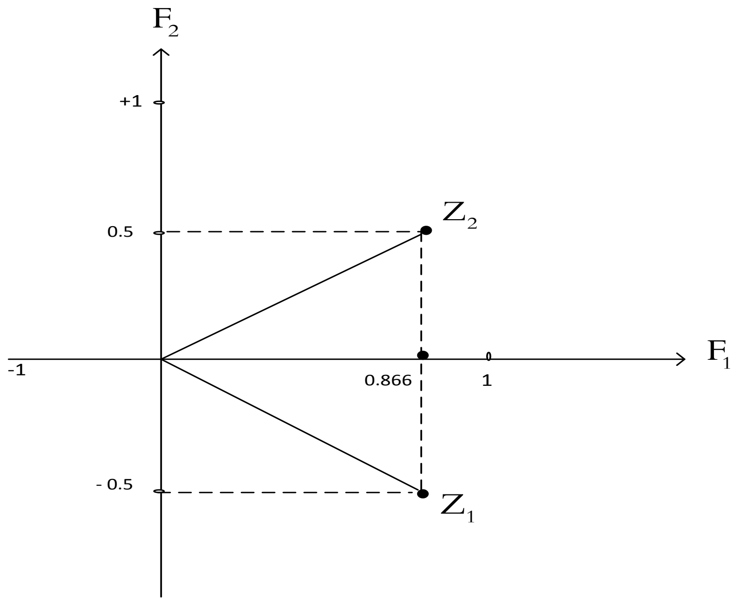
جدول (2−2) الحل النهائي للمثال (2−1)

العوامل / المتحولات	F_1	F_2	التشاركيات h_j^2
Z_1	0.866	−0.5	1
Z_2	0.866	+0.5	1
V_p التباين $=\lambda_p$	1.5	0.5	2
النسبة المئوية %	%75	%25	100

وهذا يعني أن العامل الأول يربط بين المتحولين Z_1 و Z_2 وإنه يفسر 75% من التباين الاجمالي لهما ، ويمكن تسمية هذا العامل المشترك بين الرياضيات والفيزياء باسم افتراضي مثل (قوة المنطق) .

أما العامل الثاني فيفسر 25% من التباين الإجمالي وهو يعود إلى التأثيرات الأخرى في تحصيل الرياضيات والفيزياء .

وأخيراً يمكن تمثيل هذه الصورة بيانياً كما يلي :



الشكل (2−2) الصورة العاملين للمثال

مثال (2−2) :

لنفترض أنه لدينا البيانات العددية لخمسة متحولات عن أحوال السكان في 12 محافظة كما في الجدول التالي [البيانات مقتبسة بتصرف من Harmon ص25] .

جدول (2−3) :البيانات الأصلية (الخام) للمتحولات الخمسة:

المحافظة / المتحولات	1	2	3	4	5	6	7	8	9	10	11	12	المتوسط $\bar{x}_i$	التباين S_j^2
عدد السكان X_1	5700	1000	3400	3800	4000	8200	1200	9100	9900	9600	9600	9400	6242	3440
متوسط سنوات التعليم X_2	12.8	10.9	8.8	13.6	12.8	8.3	11.4	11.5	12.5	13.7	9.6	11.4	11.4	1.8
عدد العاملين المهنيين X_3	2500	600	1000	1700	1600	2600	400	3300	3400	3600	3300	4000	2333	1241
عدد المختصين X_4	270	10	10	140	140	60	10	60	180	390	80	100	121	115
متوسط ثمن الشقة في مكان السكن X_5	25000	10000	9000	25000	25000	12000	16000	14000	18000	25000	12000	13000	17000	6368

وعندما قمنا بحساب المصفوفة الارتباطية لهذه المتحولات حصلنا على الجدول التالي:

جدول (2−4): **المصفوفة الارتباطية R للمثال (2−2):**

المتحولات	X_1	X_2	X_3	X_4	X_5
X_1	1.00000	0.00975	0.97245	0.43887	0.02241
X_2	0.00975	1.00000	0.15428	0.69141	0.86307
X_3	0.97243	0.15428	1.00000	0.51472	0.12193
X_4	0.43887	0.69141	0.51472	1.00000	0.77765
X_5	0.02241	0.86307	0.12193	0.77765	1.00000

وللتعبير عن هذه المصفوفة الارتباطية R بواسطةالمكونات الأساسية(principal components) نحول المتحولات X إلى متحولات معيارية Z ونضعها في النموذج *(1−19)*، الذي يضم *5* مكونات أساسية (بقدر عدد المتحولات) وهي F_1, F_2, F_3, F_4, F_5 وبعد إجراء حسابات معقدة (شرحناها سابقاً) نحصل على عناصر المصفوفة A المعروضة في الجدول التالي (للمتحولات المعيارية Z_j).

جدول (2−5): عناصر الحل النهائي حسب المكونات الاساسية :

المكونات الأساسية / المتحول Z_j	F_1	F_2	F_3	F_4	F_5	التباين S_j^2	التشاركية h_j^2 للعاملين F_1 و F_2 فقط
Z_1	0.5810	0.8064	0.0276	−0.0645	−0.0852	1.0000	0.98783
Z_2	0.7671	−0.5448	0.3193	0.1118	−0.0216	1.0002	0.88515
Z_3	0.6724	0.7260	0.1149	−0.0072	0.0862	0.9999	0.97930
Z_4	0.9324	−0.1043	−0.3078	0.1582	0.0000	1.0000	0.88023
Z_5	0.7911	−0.5582	−0.0647	−0.2413	0.0102	0.9990	0.93746
V_p حصة العامل F_p من التباين الإجمالي S^2	2.8733	1.7966	0.2148	0.1000	0.0153	$S_1^2 = 5$	*4.66997*
النسبة المئوية % للحصة V_p من التباين الإجمالي	57.5	35.9	4.3	2.0	0.3	100.0	

وهذا يعطينا خمس معادلات أفقية (معادلة لكل متحول Z_j) هي:

$$Z_1 = 0.5810 * F_1 + 0.8064 * F_2 + 0.0276 * F_3 - 0.0645 * F_4 - 0.0852 * F_5$$
$$Z_2 = 0.7671 * F_1 - 0.5448 * F_2 + 0.3193 * F_3 + 0.1118 * F_4 - 0.0216 * F_5$$
$$Z_3 = 0.6724 * F_1 + 0.7260 * F_2 + 0.1149 * F_3 - 0.0072 * F_4 + 0.0862 * F_5$$
$$Z_4 = 0.9324 * F_1 - 0.1043 * F_2 - 0.3078 * F_3 + 0.1582 * F_4 + 0.000 * F_5$$
$$Z_5 = 0.7911 * F_1 - 0.5592 * F_2 - 0.0647 * F_3 - 0.2413 * F_4 + 0.0102 * F_5$$

ومن السطرين الأخيرين في الجدول السابق نلاحظ أن نسبة حصة العامل الأول V_1 تشكل 57.5% من التباين الاجمالي (المساوي لـ 5) . وحصة العامل الثاني V_2 تشكل 35.9% من التباين الاجمالي . وإنهما معاً يشكلان حوالي 93.4% من التباين الاجمالي .

وهذا يعني أن العاملين F_1 و F_2 يمكن أن يعبران عن العلاقات بين المتحولات الخمسة المدروسة بدقة كافية، ويفسران (93.4%) من التباين الاجمالي بخطأ قدره 6.6% منه، ولقد حسبنا تشاركيات المتحولات Z_j المقابلة للعاملين F_1 و F_2 فقط ووضعنا في العمود الأخير، فكانت التشاركية الإجمالية (4.66997).

ملاحظة: نلاحظ من أعمدة المصفوفة في الجدول السابق أن أمثال العوامل الأخيرة F_3 و F_4 ثم F_5 تصبح أصغر فأصغر لدرجة أنه يمكن اهمال هذه العوامل في عمليات التحليل . علماً بأن هذه الأمثال تعبر أيضاً عن معاملات الارتباط الزوجية بين المتحولات Z_j والعوامل F_p ، وذلك حسب العلاقة (1−48) وبفرض أن العوامل العامة F_p غير مرتبطة مع لعضها البعض. وهذا يعني أنه يمكننا دون أن نتحمل خسارة كبيرة اختصار عدد العوامل في النموذج العاملي إلى m=2 عاملاً فقط (m<n) ، فنحصل على النموذج المختصر المعرف بالعلاقة (1−20) والمتضمن عوامل التميز U_j، والذي يسمى بنموذج العوامل الرئيسية، وهناك عدة أساليب لتحديد عدد العوامل العامة m، التي يمكن أن تدخل في النموذج العاملي وتحافظ على حد الدقة المطلوب، كما سنرى في الفقرة التالية:

2-2-1 كيفية تحديد عدد العوامل m الداخلة في نموذج المكونات الأساسية :

يمكن اختصار وتحديد عدد العوامل الداخلة في نموذج المكونات الأساسية (1-19) بعدة أساليب هي:

1- بعد الحصول على أمثال المكونات الأساسية أو قبل الحصول عليها، يمكننا أن نشترط عند المعالجة على الحاسوب أن يتم تجاهل أو حذف الأمثال التي تقل قيمتها عن قيمة محددة (مثلاً عن 0.20 أو 0.30 أو 0.40 أو غير ذلك)، وهذا قد يؤدي آلياً إلى حذف بعض المكونات الأساسية كلياً (المكونات المتأخرة) والابقاء على بعض المكونات التي سنسميها بالعوامل الرئيسية ونرمز لعددها بـ m عاملاً .

2- كما يمكن أن نقوم بحساب القيم الذاتية (المميزة) للمصفوفة R من المعادلة $(R-\lambda I)a=0$ ونرمز لها بـ λ_k ثم نرتبها تنازلياً كما يلي :

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \geq 0 \qquad (53-2)$$

وهي تقابل العوامل $F_1\,,\,F_2\,,\,F_3\,,\ldots\,,F_n$

ثم نقوم بحذف العوامل التي تقابلها قيمة ذاتية $\lambda_k < 1$ (لأنها تشكل نسبة صغيرة أو مهملة من التباين الاجمالي) ، ونحتفظ بالعوامل العامة التي تقابل قيماً ذاتية $\lambda_k \geq 1$ ونرمز لعددها بـ m ونسميها بالعوامل الرئيسية .

3- كما يمكننا أن نقوم بحساب رتبة range المصفوفة R ولتكن مساوية لـ m وعندها يكون عدد العوامل الرئيسية m مساوياً لرتبة المصفوفة R .

4- يمكننا الاعتماد على متراجحة أو جدول (Thurstone 1963) الذي برهن أن عدد العوامل الرئيسية m يحقق المتراجحة المزدوجة التالية:

$$\left[\frac{(2n+1)-\sqrt{8n+1}}{2}\right] \leq m \leq \left[\frac{(2n+1)+\sqrt{8n+1}}{2}\right] \qquad (54-2)$$

واعتماداً على هذه المتراجحة تم استخلاص الجدول التالي :

جدول (2-6): قيم m المقابلة لقيم n :

عدد المتحولات n	2	3	4	5	6	7	8	9	10	11	12	13	14	15
عدد العوامل m	1	1	2	3	3	4	5	6	6	7	8	9	10	10

5- مع أن الجدول (2-6) يشير إلى أن عدد العوامل العامة m لا يتجاوز $\frac{2}{3}n$ ، ولكن التجارب العملية المختلفة أظهرت أن عدد العوامل العامة F_p الرئيسية m لا يتجاوز نصف عدد المتحولات Z_j المدروسة . أي أظهرت أن:

$$m \leq \frac{n}{2} \qquad (55-2)$$

2–3 : طريقة العوامل (المحاور) الرئيسية ((principal factors(Axis)) :

تعتمد هذه الطريقة على النموذج العاملي المختصر المعرف بالعلاقة (1–20) والذي نكتبه كما يلي:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + d_jU_j \qquad (56-2)$$

حيث أن: U_j هو حد الخطأ العشوائي وأن (j: 1 2 3 … n) وأن: $(m < n)$.

وهو يتألف من n معادلة خطية تتضمن m * n عدداً مجهولاً هي الأمثال a_{jp}، أما d_j فتحسب من a_{jp} بعد حسابها، ويشترط في هذا النموذج أن تكون العوامل $F_1 , F_2 , \ldots , F_m$ معيارية (بمتوسط معدوم وتباين يساوي الواحد) وغير مرتبطة مع بعضها البعض، أي أنها متعامدة مع بعضها ، وتشكل فضاءً خاصاً في R^m تنسب إليه أشعة المتحولات Z_j . وتكون الصورة العاملية لهذا النموذج كما في الجدول التالي :

جدول (2–7): الصورة العاملية للنموذج (2–56)

العوامل / المتحولات	F_1	F_2	F_3	....	F_m	التشاركيات h_j^2	مربع الأمثال d_j^2
Z_1	a_{11}	a_{12}	a_{13}	....	a_{1m}	h_1^2	$d_1^2 = 1 - h_1^2$
Z_2	a_{21}	a_{22}	a_{23}	....	a_{2m}	h_2^2	$d_2^2 = 1 - h_2^2$
Z_3	a_{31}	a_{32}	a_{33}	....	a_{3m}	h_3^2	$d_3^2 = 1 - h_3^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$	$\vdots$
Z_n	a_{n1}	a_{n2}	a_{n3}	....	a_{nm}	h_n^2	$d_1^2 = 1 - h_n^2$
حصة تباين العامل F_p	V_1	V_2	V_3	....	V_m	$\frac{\sum^n h_j^2}{\sum^m V_p}$	$S^2 = h^2 + d^2$
النسبة المئوية	P_1	P_2	P_3	....	P_m		

كما يشترط أن تكون المتحولات U_j معيارية (بمتوسط معدوم وتباين يساوي الواحد) وغير مرتبطة مع العوامل $F_1 , F_2 , \ldots , F_m$، وأن تكون U_j مستقلة عن بعضها البعض، أما الأمثال dj فتحسب بعد حساب h_j^2 كما في العمود الأخير من الجدول السابق(2–7). علماً بأن تباين أي متحول Z_j يساوي:

$$S_j^2 = h_j^2 + d_j^2 = 1 \qquad \text{ومنها نجد أن:} \qquad d_j^2 = 1 - h_j^2$$

وأن التباين الاجمالي يساوي :
$$S^2 = \sum_1^n h_j^2 + \sum_1^n d_j^2 = h^2 + d^2$$

حيث رمزنا بـ h^2 لـ $h^2 = \sum_{j=1}^n h_j^2$ ، ورمزنا بـ d^2 لـ $d^2 = \sum_{j=1}^n d_j^2$

وتختلف هذه الطريقة عن طريقة المكونات الأساسية بأنها تهدف إلى جعل تباينات العوامل العامة تأخذ على التوالي أكبر حصة ممكنة من التشاركية الاجمالية $\left(h^2 = \sum_1^n h_j^2\right)$ وليس من التباين الاجمالي S^2 ، لأن التباين الاجمالي S^2 يتضمن حداً مجهولاً هو d^2 . كما هو مبين في الجدول (2–7) وملحقاته .

أي أنها تهدف إلى حساب الأمثال a_{jp} في كل عمود بحيث يشكل تباين العامل الأول F_1 أكبر حصة ممكنة V_1 من التشاركية الاجمالية h^2، ثم بحيث يشكل تباين العامل الثاني F_2 أكبر حصة ممكنة V_2 من باقي h^2، ثم بحيث يشكل تباين العامل الثالث F_3 أكبر حصة ممكنة V_3 من باقي h^2 ...الخ

وهكذا نتابع حتى نحسب جميع الأمثال a_{jp} في جميع الأعمدة ونستهلك الجزء الأكبر من التشاركية الاجمالية (حوالي 80% أو 90% منه)، وعندها نحصل على الصورة العاملية المختصرة الواردة في الجدول (2–7)، وهي تختلف كثيراً عن الصورة العاملية لطريقة المكونات الأساسية، التي تشمل n عاملاً F_p ولا تأخذ بعين الاعتبار المتحولات العشوائية U_j .

وإذا أخذنا مجموع مربعات الأمثال a_{jp} للمتحول Z_j على العوامل F_p (أفقياً)، فإننا سنحصل على تشاركية المتحول Z_j والتي رمزنا لها بـ h_j^2 وتساوي :

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots + a_{jm}^2 = \sum_{p=1}^{m} a_{jp}^2 \qquad \text{(أفقياً)} \qquad (57-2)$$

أما إذا أخذنا مجموع مربعات الأمثال a_{jp} المقابلة للعامل F_p على المتحولات Z_j (عمودياً)، فإننا سنحصل على حصة تباين العامل F_p من التشاركية الإجمالية h^2 ، والذي رمزنا له بـ V_p ويساوي :

$$V_p = a_{1p}^2 + a_{2p}^2 + a_{3p}^2 + \cdots + a_{np}^2 = \sum_{j=1}^{n} a_{jp}^2 \qquad \text{(عمودياً)} \qquad (58-2)$$

ويمكننا أن نجد من الجدول (2–7) أن مجموع حصص العوامل يساوي :

$$V = \sum_{p=1}^{m} V_p = \sum_{p=1}^{m}\sum_{j=1}^{n} a_{jp}^2 = \sum_{j=1}^{n}\left(\sum_{p=1}^{m} a_{jp}^2\right) \qquad (59-2)$$

ومن (2–57) نجد أن التشاركية الاجمالية تساوي:

$$V = \sum_{j=1}^{n} h_j^2 = h^2 \qquad (59\text{–}2)$$

وبما أنه لدينا من (1–31) أن: $S^2 = h^2 + d^2$ (حيث أن: $d^2 = \sum_{j=1}^{n} d_j^2$) ، فإننا نجد أيضاً أن التباين الاجمالي يساوي:

$$S^2 = V + d^2 \qquad (60-2)$$

ومنها نستنتج أن مجموع حصص العوامل V سيكون أصغر من S^2، وإن الفرق بين S^2 و V يعود إلى العوامل الأخرى (بما فيها العوامل U_j)، أي أن : $V = S^2 - d^2 < S^2$

وبما أن مجموع حصص العوامل V سيكون أقل من التباين الاجمالي S^2، وهذا يعني أن هذه الطريقة تتخلى عن جزء من الدقة مقابل تسهيل الحسابات .

والآن نعود إلى قضية حساب الأمثال a_{jp} لكل عامل على حدة في النموذج (2–56). وسنعمل على شرح ما قدمه [*Harman P. 154*] كما يلي :

لنأخذ أولاً أمثال العامل الأول F_1 وهي a_{j1}، فنجد أن هدفنا سيكون حساب الأمثال a_{j1}، التي تجعل الحصة V_1 للعامل F_1 تأخذ أكبر قيمة ممكنة من التشاركية الاجمالية h^2، أي ان هدفنا سيكون تعظيم قيمة الحصة V_1 المعرفة في (2−58) كما يلي :

$$V_1 = a_{11}^2 + a_{21}^2 + a_{31}^2 + \cdots + a_{n1}^2 \Rightarrow Max \qquad (61-2)$$

وذلك ضمن الشروط المفروضة على جميع الأمثال a_{jp} في الصورة العاملية الواردة في الجدول (2−7) .

ولكننا وجدنا من الفقرات السابقة في (1−6) أن أمثال هذه الصورة هي عناصر المصفوفة A، وإن هذه العناصر مرتبطة مع عناصر المصفوفة الارتباطية المخففة (إذا كانت العوامل F_p غير مرتبطة) المعرفة حسب (1−62) بالعلاقة التالية :

$$R_h = A_{n*m} * A'_{m*n} \qquad (62-2)$$

وهي تفرض على الأمثال a_{jp} جملة من الشروط الخطية يمكن أن نستخلصها من المطابقة بين طرفي الجداء التالي: [Harman P. 154]

$$\begin{bmatrix} h_1^2 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & h_2^2 & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & h_3^2 & \dots & r_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & h_{nn}^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ a_{31} & a_{32} & \dots & a_{3m} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} * \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ a_{13} & a_{23} & \dots & a_{n3} \\ \vdots & \vdots & \dots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix} \qquad (63-2)$$

وعند إجراء جداء عناصر $A * A'$ ومطابقتها مع عناصر R_h ، نحصل على عدد من المعادلات الخطية (معادله مقابل كل معامل ارتباط r_{jk})، فإذا أخذنا عناصر السطر الأول من المصفوفة R_h نجد أنها تعطينا الشروط التالية :

$$h_1^2 = a_{11}^2 + a_{12}^2 + a_{13}^2 + \cdots a_{1m}^2 = \sum_{p=1}^{m} a_{jp} * a_{jp} \qquad : (j = k = 1)$$

$$r_{12} = a_{11}a_{21} + a_{12}a_{22} + \cdots + a_{1m}a_{2m} = \sum_{p=1}^{m} a_{jp}a_{kp} \qquad : (j = 1, k = 2)$$

$$(64-2)$$

$$r_{13} = a_{11}a_{31} + a_{12}a_{32} + \cdots + a_{1m}a_{3m} = \sum_{p=1}^{m} a_{jp}a_{kp} \qquad : (j = 1, k = 3)$$

$$\dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots$$

$$r_{1n} = a_{11}a_{n1} + a_{12}a_{n2} + \cdots + a_{1m}a_{nm} = \sum a_{jp}a_{kp} \qquad : (j = 1, k = n)$$

وهي تشكل n معادلة وتتضمن m*n مجهولاً هي a_{jp} . وإذا أخذنا عناصر السطر الثاني لـ R_h فإننا سنحصل على n معادلة أخرى، وإذا أخذنا عناصر جميع الأسطر فإننا سنحصل على n^2 معادلة خطية. ويمكننا صياغة جميع معادلات هذه المعاملات r_{jk} في معادلة واحدة هي :

$$r_{jk} = \sum_{p=1}^{m} a_{jp}a_{kp} \qquad : (j, k{:}\, 1\ \ 2\ \ 3 \dots n) \qquad (65-2)$$

ولكن بما أن المصفوفة R_h متناظرة فإن المعاملات $\left(r_{jk}=r_{kj}\right)$ وأن $\left(r_{jj}=\mathrm{h}_j^2\right)$ فإن عدد المعادلات المستقلة في (2−64) يساوي عدد المعاملات r_{jk} المستقلة ويساوي $\frac{n(\mathrm{n}+1)}{2}$ معادلة .

ولحساب الأعداد a_{jp} نشكل تابع (لاغرانج) من التابع V_1 المعرف في (2−61) والشروط المعرفة في (2−64)، فنجد بالنسبة للعامل الأول F_1 أنه يساوي ما يلي :

$$L_1=V_1-\sum_{j=1}^{n}\sum_{k=1}^{n}\mu_{jk}\left(\sum_{p=1}^{m}a_{jp}a_{kp}-r_{jk}\right) \qquad (66-2)$$

حيث أن μ_{jk} هي مضاريب لاغرانج، وهي تحقق خاصة التناظر $\mu_{jk}=\mu_{kj}$، وإن r_{jk} هي مقادير معلومة من المصفوفة R_h، وإن $h_j^2=r_{jj}$ مقدرة فيها، وإن المجاميع على j و k متكافئة . والآن نقوم بكتابة ذلك التابع كما يلي :

$$L_1=\sum_{j=1}^{n}a_{j1}^2-\sum_{j=1}^{n}\sum_{k=1}^{n}\mu_{jk}\sum_{p=1}^{m}a_{jp}*a_{kp}+\sum_{j=1}^{n}\sum_{k}^{n}\mu_{jk}r_{jk} \qquad (67-2)$$

ثم نأخذ المشتق الجزئي لـ L_1 بالنسبة لـ a_{j1} فنجد أنه (عندما نضع(p = 1)) أن:

$$\frac{\partial L_1}{\partial a_{j1}}=2a_{j1}-\sum_{j=1}^{n}\sum_{k=1}^{n}\mu_{jk}a_{k1}+0=0 \qquad (68-2)$$

وبما أن المجموع نفسه مكرر مرتين على k ثم على j ، فإننا نحصل من (2−68) على أن:

$$2a_{j1}-2\sum_{k=1}^{n}\mu_{jk}a_{k1}=0 \qquad (69-2)$$

أو على المعادلات التالية:

$$a_{j1}-\sum_{k=1}^{n}\mu_{jk}a_{k1}=0 \qquad : \quad (j\colon 1\ 2\ 3\ \dots n) \qquad (70-2)$$

حيث أن الأعداد a_{j1} هي أمثال عمود العامل F_1 ، وهي تشكل n معادلة خطية تتضمن الأمثال a_{j1} والأمثال و a_{k1} مثقلة بالمضاريب المجهولة μ_{jk} .

ولإدخال الأمثال الأخرى من a_{jp} في المعالجة نقوم باشتقاق L_1 المعرف في (2−67) بالنسبة a_{jp} (حيث $P\neq 1$) ونضعه مساوياً للصفر فنحصل على أن (علماً بأن $\mu_{jk}=\mu_{kj}$) :

$$\frac{\partial L_1}{\partial a_{jp}}=0-\sum_{j=1}^{n}\sum_{k=1}^{n}\mu_{jk}a_{kp}=-2\sum_{k=1}^{n}\mu_{jk}a_{kp}=0$$

ومنها نحصل على المعادلات التالية:

$$-\sum_{k=1}^{n}\mu_{jk}*a_{kp}=0 \qquad (71-2)$$

حيث أن $(P\neq 1)$ وأن (j: 1 2 3…. n)

وهي تشكل n معادلة، كل منها تتضمن n مجهولاً مأخوذاً على k . وهنا نلاحظ أنه يمكننا دمج العلاقات في (2−70) والعلاقات (2−71) في علاقة واحدة ، وذلك باستخدام رمز (كرونكر δ_{1p})، الذي يأخذ إحدى قيمتين هما : $\delta_{1p} = 1$ عندما $(p = 1)$، و $\delta_{1p} = 0$ عندما $(p \neq 1)$ ، فنحصل على المعادلة الموحدة التالية :

$$\delta_{1p} * a_{j1} - \sum_{k=1}^{n} \mu_{jk} a_{kp} = 0 \qquad (72-2)$$

حيث أن: (p: 1 2 3… n)، فإذا كانت $(p = 1)$ نحصل على $(2-70)$ ، وإذا كانت $(p \neq 1)$ نحصل على $(2-71)$، والآن علينا التخلص من المضاريب μ_{jk} ، لذلك نقوم بضرب طرفي العلاقة $(2-72)$ بـ a_{j1} ثم نأخذ المجموع بالنسبة لـ j فنحصل على أن:

$$\delta_{1p} \sum_{j=1}^{n} a_{j1}^2 - \sum_{k=1}^{n} \left(\sum_{j=1}^{n} \mu_{jk} a_{j1} \right) a_{kp} = 0 \qquad (73-2)$$

وقياساً على العلاقة $(2-70)$ نلاحظ أن المجموع ضمن القوسين يساوي a_{k1} لأن $\mu_{jk} = \mu_{kj}$ ولأن:

$$\sum_{j=1}^{n} \mu_{jk} a_{j1} = \sum_{j=1}^{n} \mu_{kj} a_{j1} = a_{k1} \qquad (74-2)$$

ولنرمز الآن للمجموع $\sum_{j=1}^{n} a_{j1}^2$ الذي في العلاقة $(2-73)$ بالرمز λ_1 فيكون لدينا :

$$\lambda_1 = \sum_{j=1}^{n} a_{j1}^2 = V_1 = \qquad F_1 \text{ حصة تباين العامل} \qquad (76-2)$$

ثم نعوض ذلك في العلاقة $(2-73)$ فنحصل على أن:

$$\delta_{1p} * \lambda_1 - \sum_{k=1}^{n} a_{k1} a_{kp} = 0 \qquad (77-2)$$

ثم نضرب طرفي العلاقة $(2-77)$ بالمقدار a_{jp} ونأخذ مجموع النتائج على جميع العوامل (أي نجمع علىp) فنحصل على أن:

$$\lambda_1 \sum_{p=1}^{m} \delta_{1p} a_{jp} - \sum_{p=1}^{m} \sum_{k=1}^{n} a_{jp} a_{k1} a_{kp} = 0 \qquad (78-2)$$

وبما أن $\delta_{1p} = 0$ عندما $(p \neq 1)$ وأن $\delta_{11} = 1$ عندما $(p = 1)$، فإن هذا يعني أن معظم حدود المجموع $\sum_{p=1}^{n} \delta_{1p} a_{jp}$ ستنعدم ويبقى منها الحد الأول فقط . وعندها فإن العلاقة $(2-78)$ تأخذ الشكل التالي:

$$\lambda_1 a_{j1} - \sum_{k=1}^{n} a_{k1} \left(\sum_{p=1}^{m} a_{jp} * a_{kp} \right) = 0 \qquad (79-2)$$

وبملاحظة أن المجموع ضمن القوسين يساوي حسب العلاقة $(65-2)$ معامل الارتباط r_{jk} وبالتالي يمكننا كتابة العلاقة $(79-2)$ على الشكل التالي :

$$\lambda_1 a_{j1} - \sum_{k=1}^{n} a_{k1} * r_{jk} = 0 \qquad (80-2)$$

ثم نكتبها على الشكل المناسب التالي :

$$\sum_{k=1}^{n} r_{jk} * a_{k1} - \lambda_1 a_{j1} = 0 \qquad (81-2)$$

حيث أن (j: 1 2 3... n)

وهي تشكل n معادلة خطية، مقابل قيم j أو مقابل المتحولات Z_j ، وتضم قيم الأمثال العمودية a_{j1} المقابلة للعامل الأول F_1 فقط .

فإذا وضعنا $(j=1)$ (للمتحول الأول Z_1) نحصل على أن:

$$\sum_{k=1}^{n} r_{1k} * a_{k1} - \lambda_1 a_{11} = 0 \qquad (82-2)$$

والتي يمكن نشرها وكتابتها بشكل مفصل كما يلي:

$$(r_{11}a_{11} - \lambda_1 a_{11}) + r_{12}a_{21} + r_{13}a_{31} + \cdots + r_{1n}a_{n1} = 0 \qquad (83-2)$$

وباستبدال الرمز λ_1 بـ λ وكل عنصر قطري r_{jj} بـ h_j^2 كما في العلاقة $(83-2)$ ، وبإعطاء j جميع القيم الممكنة في العلاقة $(81-2)$ ، نحصل منها على n معادلة خطية متجانسة تتضمن n مجهولاً هي الأمثال a_{j1}، وهي امثال العامل الأول F_1 في معادلات المتحولات المعيارية $Z_1, Z_2, \dots, Z_n$، وتأخذ معادلات أمثال العامل الأول F_1 الشكل التالي :

$$\begin{aligned}
&(h_1^2 - \lambda)a_{11} + r_{12}a_{21} + r_{13}a_{31} + \cdots + r_{1n}a_{n1} = 0 \\
&r_{21}a_{11} + (h_2^2 - \lambda)a_{21} + r_{23}a_{31} + \cdots + r_{2n}a_{n1} = 0 \qquad (84-2) \\
&r_{31}a_{11} + r_{32}a_{21} + (h_3^2 - \lambda)a_{31} + \cdots + r_{3n}a_{n1} = 0 \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&r_{n1}a_{11} + r_{n2}a_{21} + r_{n3}a_{31} + \cdots + (h_n^2 - \lambda)a_{n1} = 0
\end{aligned}$$

والتي يمكن كتابتها مصفوفياً على الشكل التالي:

$$(R_h - \lambda I)a_1 = 0 \qquad (85-2)$$

حيث a_1 هو الشعاع العمودي لأمثال F_1 .

وهكذا تحولت مسألة تعظيم التابع $(66-2)$، ضمن الشروط $(81-2)$ ، إلى جملة معادلات مؤلفة من n معادلة خطية متجانسة وتتضمن n مجهولاً هي الأمثال a_{j1} الخاصة بالعامل الأول F_1 فقط، وهي الأمثال المقابلة للمتحولات Z_j . ومعلوم من نظرية المصفوفات أن الشرط اللازم والكافي للحصول على حل عملي (غير الحل التافه المعدوم) لهذه المعادلات الخطية المتجانسة هو أن يكون محدد المصفوفة $(R_h - \lambda I)$ مساوياً للصفر، أي أن يكون المحدد التالي معدوماً: $|R_h - \lambda I| = 0$ أي أن يكون:

$$\Delta = \begin{vmatrix} (h_1^2-\lambda) & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & (h_2^2-\lambda) & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & (h_3^2-\lambda) & \dots & r_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & (h_n^2-\lambda) \end{vmatrix} = 0 \qquad (86-2)$$

حيث أن: r_{jk} معلومة و h_j^2 مقدرة ، أما المجهول الجديد فهو الوسيط λ . وحتى تتحقق العلاقة $(2-86)$ علينا أن نحسب قيمة أو قيم λ التي تجعل ذلك المحدد معدوماً . لذلك نقوم بنشر أو فك هذا المحدد حسب الأسطر أو الأعمدة فنحصل على معادلة كثير حدود تتضمن الوسيط λ من الدرجة n ، وتسمى بالمعادلة المميزة للجملة $(2-84)$ ، ويكون لها الشكل التالي:

$$\lambda^n + b_1 * \lambda^{n-1} + b_2 * \lambda^{n-2} + \cdots b_n = 0 \qquad (87-2)$$

وعندما نقوم بحل هذه المعادلة فإننا سنحصل على n جذراً حقيقياً لـ λ نرمزها ونرتبها تنازلياً كما يلي:

$$\lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \lambda_n \qquad (88-2)$$

وتسمى هذه الجذور بالجذور الكامنة أو بالقيم المميزة أو الذاتية للمصفوفة المخففة R_h .

ملاحظة: يمكن أن تكون بعض هذه الجذور مكررة وخاصة عندما تكون مرتبة المصفوفة للمحدد Δ أقل من n .

وإذا بدلنا λ التي في المعادلات $(2-83)$ بأكبر الجذور الكامنة (أو القيم المميزة) التي حصلنا عليها وهو λ_1 ، فإننا سنحصل على جملة محددة من المعادلات الخطية المتجانسة، ومنها نحصل على الأشعة الذاتية a_{j1} المقابلة للقيمة λ_1(كما فعلنا في طريقة المكونات الرئيسية) . وسيكون لدينا لانهاية من الحلول المقبولة المتناسبة ومنها نختار أحد الحلول الخاصة المقبولة، ولكن أي حل منها مقابل لـ $\lambda 1$ يجب أن يحقق الهدف $(2-76)$ المطلوب منه وهو :

$$\lambda_1 = \sum_{j=1}^{n} a_{j1}^2 = V_1 \Rightarrow Max \quad \Rightarrow \quad \left\|a_{j1}\right\|^2 = \lambda_1 \qquad (89-2)$$

وهذا يعني أن قيمة حصة التباين V_1 تساوي أكبر جذور المعادلة المميزة للمحدد $|R_h - \lambda I|$ ، وتساوي تحديداً الجذر الأكبر λ_1 للمصفوفة المخففة R_h .

ومن جهة أخرى نلاحظ أن قيمة λ_1 تساوي مربع طول الشعاع a_{j1} المطلوب حسابه، لأنها حسب العلاقة $(2-89)$ تساوي مجموع مربعات عناصره .

وهكذا يمكننا إيجاد حل للمسألة وحساب قيم الأمثال a_{j1} الخاصة بالعامل الأول F_1 ، آخذين بعين الاعتبار أن يكون له النصيب الأكبر من التشاركية الاجمالية $\left(\sum \mathrm{h}_j^2\right)$.

لذلك نعوض أكبر قيمة لـ λ وهي λ_1 في المعادلات $(2-84)$ ونحلها فنحصل على أحد الحلول الخاصة الممكنة بالنسبة لـ a_{j1} وليكن الشعاع التالي:

$$\alpha_1\ (\alpha_{11}, \alpha_{21}, \alpha_{31}, \dots \dots\ \alpha_{n1}) \qquad (90-2)$$

وحتى نحصل على شعاع يضمن لنا تعظيم V_1 في $(2-89)$، نقوم بتحويل شعاع الحل الخاص إلى شعاع واحدي ، لذلك نقوم بحساب طوله من العلاقة: $\|\propto_1\| = \sqrt{\propto_{11}^2+\propto_{12}^2+\cdots+\propto_{n1}^2}$ ثم نقسم كل من عناصره $\propto_{j1}$ على طوله $\|\propto_1\|$ فنحصل الشعاع الواحدي التالي :

$$e_1[e_{11}..e_{n1}] = e_1\left[\frac{\propto_{11}}{\|\propto_1\|},\frac{\propto_{21}}{\|\propto_1\|},\frac{\propto_{31}}{\|\propto_1\|},\dots,\frac{\propto_{n1}}{\|\propto_1\|}\right] \qquad (91-2)$$

ولإيجاد شعاع الأمثال a_{j1} نضرب عناصر الشعاع الواحدي e_1 بطول الشعاع ، والذي يساوي حسب $(2-89)$ المقدار $\sqrt{\lambda_1}$، وبذلك نجد أن العناصر a_{j1} تحسب من العلاقة:

$$a_{j1} = e_{j1}\sqrt{\lambda_1} = \frac{\propto_{j1}*\sqrt{\lambda_1}}{\|\propto_1\|} = \frac{\propto_{j1}*\sqrt{\lambda_1}}{\sqrt{\propto_{11}^2+\propto_{12}^2+\cdots+\propto_{n1}^2}} \qquad (92-2)$$

وهي تعطينا قيم الأمثال العمودية للعامل الأول F_1، والتي تعبر عن معاملات ارتباط العامل F_1 مع المتحولات Z_j، (حسب العلاقة $(2-76)$)، وهي التي تجعل حصة تباين F_1 من التشاركية الاجمالية أكبر ما يمكن .

ثم ننتقل إلى حساب قيم أمثال العوامل الأخرى بحيث نجعل حصة كل منها بالتدريج مما تبقى من التشاركية الاجمالية أكبر ما يمكن .

وهكذا نجد أن المسألة التالية تكمن في البحث عن الشعاع a_{jp}، الذي يشكل أكبر حصة من التشاركية المتبقية. وحتى نستطيع معالجة هذه المسألة يجب علينا حساب المصفوفة المتبقية من مصفوفة معاملات الارتباط R (بعد استبعاد أثر العامل الأول فقط) ...الخ .

وهكذا نقوم بتكرار عملية حساب المصفوفة الارتباطية المتبقية لاستبعاد أثر عاملين أو ثلاثة عوامل أو (m−1) عاملاً . وهذا يجعلنا بحاجة إلى استخدام رموز خاصة بذلك .

لذلك فإننا سنرمز لمعامل الارتباط المتبقي من المعامل الأصلي r_{jk} بعد استبعاد S عاملاً بالرمز sr_{jk} وعندها فإننا سنحصل بعد استبعاد العامل الأول F_1 على المعامل المتبقي كما يلي:

$$1r_{jk} = r_{jk} - a_{j1}*a_{k1} = a_{j2}*a_{k2} + a_{j3}*a_{k3} + \cdots + a_{jm}*a_{km} \qquad (93-2)$$

وذلك لأنه لدينا من العلاقة $(2-65)$ أن:

$$r_{jk} = \sum_{p=1}^{m} a_{jp}*a_{kp} = a_{j1}*a_{k1} + a_{j2}*a_{k2} + \cdots + a_{jm}*a_{km}$$

ولحساب كامل عناصر المصفوفة الارتباطية المتبقية، نحسب المصفوفة R_1^* الناتجة عن جداء الشعاع الذاتي a_1 المقابل لـ F_1 ولـ λ_1 في منقوله a_1 ، فنحصل على مصفوفة مربعة ومتناظرة من المرتبة $(n*n)$ ونرمز لها بـ R_1^* وتساوي :

$$R_1^* = a_1 * a_1' = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} * [a_{11}, a_{21}, \dots, a_{n1}] \qquad (94-2)$$

ثم نقوم بحساب المصفوفة الارتباطية المتبقية R1 (بعد استبعاد F_1 فقط) من العلاقة:

$$R_1 = R_h - R_1^* = R_h - a_1 * a_1' \qquad (95-2)$$

ثم ننتقل إلى حساب الأمثال a_{j2} المقابلة للعامل الثاني F_2 بحيث نجعل حصة العامل F_2 من التشاركية المتبقية أكبر ما يمكن، أي نقوم بحساب a_{j2} التي تعظم قيمة التابع :

$$V_2 = a_{12}^2 + a_{22}^2 + a_{32}^2 + \cdots a_{n2}^2 \qquad (96-2)$$

وذلك ضمن الشروط المفروضة على a_{j2} والمشابهة للشروط $(2-84)$.

وبالاعتماد على المعادلات المميزة $(2-85)$ يمكننا حساب القيمة الذاتية λ_2 لهذه المعادلات وحساب الشعاع الذاتي المقابل لها a_{j2}، ثم متابعة المعالجة حتى نشمل جميع العوامل الأخرى .

ملاحظة: نجد من جهة أخرى أنه ليس ضرورياً أن نقوم بحساب λ_2 التي تعظم حصة F_2 من التشاركية المتبقية باستخدام المصفوفة في (2–95)، وذلك لأن قيمة λ_2 نكون قد حسبناها سابقاً من (2–87)، وهذا يعني أن λ_2 هي القيمة الثانية التي حصلنا عليها من جذور المعادلة المميزة $(2-87)$، لأن أكبر القيم المميزة للمصفوفة المتبقية R_1 هي نفسها القيمة المميزة الثانية λ_2 التي حصلنا عليها للمصفوفة المخففة $\tilde{R}_h$ ، وهذا يعني أنه للحصول على أمثال العامل F_2 في جميع المعادلات التي تعظم V_2، يكفي أن نأخذ القيمة المميزة الثانية λ_2 للمصفوفة الأصلية $\tilde{R}_h$ ونعوضها في المعادلات المميزة $(2-84)$ ثم نحسب الشعاع a_2 من الأشعة المميزة المقابلة لها فنحصل على أمثال F_2 ، وهكذا نتابع العمل للحصول على أمثال العوامل الأخرى من F_3 حتى F_m .

مثال (2–3):

لنأخذ المثال السابق $(2-2)$ الوارد في طريقة المكونات الأساسية، والمتضمن خمسة متحولات، ولنقم بمعالجته بطريقة العوامل الرئيسية .

لذلك نقوم باستبدال العناصر القطرية المساوية $\left(r_{jj} = 1\right)$ في المصفوفة الارتباطية R بقيم التشاركيات h_j^2 المحسوبة للعاملين F_1 و F_2 فقط والمأخوذة من الجدول (2–5) والتي كانت كما يلي :

$$h_j^2\text{: } 0.98783 \ , 0.88515 \ , 0.97930 \ , 0.88023 \ , 0.93746$$

فنحصل على المصفوفة المخففة R_h التالية :

$$R_h = \begin{bmatrix} (0.98783) & 0.00975 & 0.97245 & 0.43887 & 0.02241 \\ 0.00975 & (0.88515) & 0.15428 & 0.69141 & 0.86307 \\ 0.97245 & 0.15428 & (0.97930) & 0.51472 & 0.12193 \\ 0.43887 & 0.69141 & 0.51472 & (0.88023) & 0.77765 \\ 0.02241 & 0.86307 & 0.12193 & 0.77765 & (0.93746) \end{bmatrix}$$

ثم نقوم بتشكيل المصفوفة $(R - \lambda I)$ ونجعل محددها يساوي الصفر، ثم نفكه فنحصل على المعادلة $(2-86)$ وهي معادلة كثير حدود من الدرجة الخامسة بالنسبة لـ λ .

وعند القيام بإيجاد جذور هذه المعادلة، نحصل على الجذور الكامنة (القيم الذاتية لـ λ) والمرتبة تنازلياً كما يلي:

$$\lambda_1 = 2.79652 \, , \lambda_2 = 1.75496 \, , \lambda_3 = 0.10642 \, , \lambda_4 = 0.02094 \, , \lambda_5 = 0.00888$$

وبذلك نجد أن قيمة التشاركية الاجمالية $\sum \mathrm{h}_j^2$ لهذه العوامل حسب الجذور السابقة تساوي ما يلي:

$$\sum \mathrm{h}_j^2 = \sum \lambda_j = 4.68775$$

وإذا أخذنا العاملين F_1 و F_2 نجد أن مجموع حصتيهما يساوي 4.55148، وهذا يعني أنه يمكننا التعبير عن هذه المتحولات الخمسة بدلالة العاملين الأول والثاني فقط (لأنها يشكلان 97.09% من التشاركية الاجمالية)، ، أي يمكننا الاقتصار على العاملين F_1 و F_2 ، ولهذا نقوم بحساب القيمتين الذاتيتين λ_1 و λ_2 من العلاقة (2−86) ، ثم نحدد الشعاعين الذاتيين المقابلين لهما a_1 و a_2 من العلاقة (2 − 84)، واللذين يعبران عن أمثال العاملين F_1 و F_2 ، ويعطيانا أكبر حصة لـ F_1 ومن التشاركية الاجمالية $\left(100 * \frac{2.79652}{4.68775} = \%59.66\right)$ وأكبر حصة لـ F_2 من التشاركية المتبقية أو ثاني أكبر حصة من التشاركية الاجمالية$\left(100 * \frac{1.75496}{4.68775} = \%37.44\right)$.

كما يمكننا الاستفادة من الحل السابق وإعادة تكرار الحسابات من جديد وعندها نقوم بحساب القيم التقديرية لتشاركيات هذا الحل $\left(\mathrm{h}_j^2\right)$، ونعتبرها قيماً أولية لحل جديد ونستبدلها في المصفوفة R، فنحصل على مصفوفة جديدة مخففة R_h^*، ثم نقوم بإعادة الحسابات من جديد فنحصل على جذور كامنة جديدة λ_j' (قريبة من الجذور السابقة)، ويمكن أن نكرر ذلك حتى تستقر النتائج المحسوبة لـ h_j^2 في التجربتين الأخيرتين ويكون الفرق بينهما أقل من الدقة المحددة (بـ 0.0001.أو بـ 0.00001) .

وفي هذا المثال تم تكرار هذه العمليات 14 مرة على الحاسوب، وفي النهاية حصلنا على تقديرات مستقرة للتشاركيات الأخيرة كما يلي : $\mathrm{h}_j^2 = 1.00000\,, 0.76687\,, 0.95562\,, 0.79837\,, 0.97142$

وحصلنا على قيم الأمثال a_{jp} للعاملين F_1 و F_2 في التجربة الأخيرة كما هي مبينة في الجدول التالي :

جدول (2−8): الحل النهائي بطريقة العوامل الرئيسية:

المتحولات	طريقة العوامل الرئيسية		التشاركية h_j^2 $h_j^2 = \sum_{k=1}^{2} a_{jk}^2$	الأمثال $d_j^2 = 1 - h_j^2$
	العامل F_1	العامل F_2		
عدد السكان Z_1	0.622	0.785	1.000	0.000
التعليم Z_2	0.702	−0.524	0.767	0.233
العمالة Z_3	0.701	0.681	0.956	0.044
العمالة المهنية Z_4	0.882	−0.145	0.798	0.202
قيمة المسكن Z_5	0.779	−0.604	0.971	0.029
V_p حصة العامل F_p من التشاركية الكلية $V_p = \sum_{j=1}^{5} a_{jp}^2$	2.756	1.740	$4.492 = \sum h_j^2$	$0.508 = \sum d_j^2$
النسبة المئوية لحصة تباين F_p من التشاركية الإجمالية %	61.3	38.7	100	
النسبة من التباين الإجمالي= (5)	0.5512	0.348	0.8992	0.1016

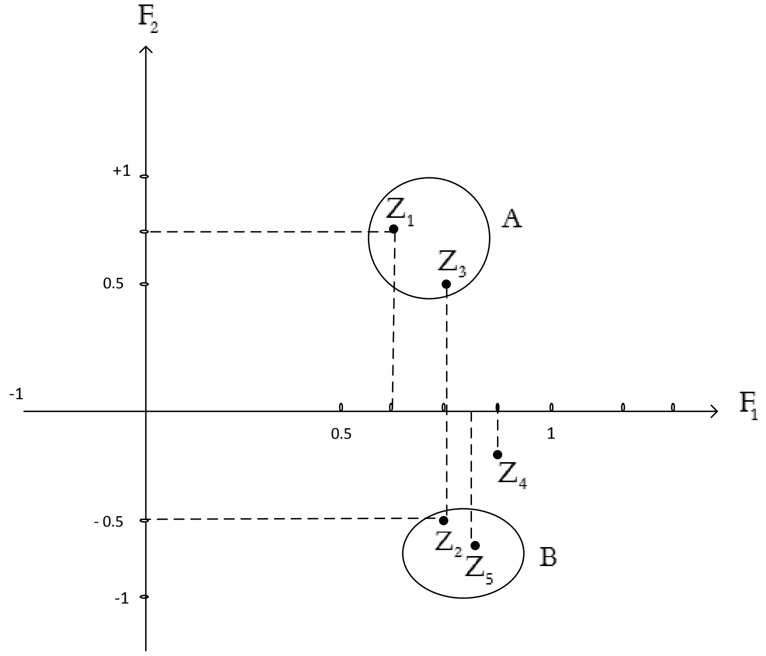
هذا يعني أنه يمكن التعبير عن هذه المتحولات بدلالة عاملين فقط، وإن التشاركية الاجمالية أصبحت تساوي 4.492 بدلاً من 4.6877، أي بخسارة صغيرة جداً، وإن تباين F_1 يشكل 61.3% من التشاركية الاجمالية لهما (4.492)، وتباين F_2 يشكل 38.7% من التشاركية الاجمالية. وإنهما معاً يساويان (4.492) ويشكلان من التباين الاجمالي (المساوي لـ 5) نسبة 89.92%.

وإن قيمة الخطأ المتبقي = $S^2 - h^2 = 5 - 4.492 = 0.508$

وهو يشكل ما نسبته 10.16% من التباين الإجمالي S^2 .

وهنا نلاحظ إن توزع الأمثال a_{j1} و a_{j2} في الجدول على العاملين *F1* و *F2* لا تساعدنا كثيراً على تفسير العلاقة بين المتحولات Z_j .

لذلك نقوم برسم المتحولات Z_j بدلالة العاملين F_1 و F_2 على شكل بياني بحيث تمثل كل متحول Z_j بشعاع مركباته $Z_j(a_{j1}\ a_{j2})$ على المستوى المحدد بالعاملين المتعامدين F_1 و F_2 كما يلي :



الشكل (2−3) التمثيل البياني للمثال (2−3)

ومن الشكل (2−3) نلاحظ أن المتحولين Z_1 و Z_3 متقاربان ويقعان في الربع الأول . وهما يشكلان عاملاً خاصاً مؤلفا من عدد السكان ونسبة العمالة فقط ويشكلان المجموعة *A* ، وكذلك نجد أن المتحولين Z_2 و Z_5 متقاربان ويقعان في الربع الرابع ويشكلان عاملاً آخر مؤلف من نسبة التعليم وقيمة المسكن فقط ويشكلان المجموعة *B* ، أما المتحول Z_4 يبدو أنه متحول منعزل ويشكل لوحده عاملاً مؤلفاً من العمالة المهنية فقط ، ويمكن دمجه مع المتحولين الأخيرين ووضعه في المجموعة *B*.

وحتى نستطيع إظهار هذه العوامل الجديدة علينا أن نقوم بتدوير المحاور الإحداثية العاملية وتحويلها من إحداثيات متعامدة إلى إحداثيات مائلة لتمر من مركزي المجموعتين *A* و *B* كما سنرى لاحقاً .

4-2 : طريقة البواقي الصغرى (المربعات الصغرى):

تعتمد هذه الطريقة على نموذج العوامل الرئيسية المعرف بالعلاقة (2 – 20) ، الذي يتطلب تحديد عدد العوامل m أولاً، ثم القيام بمقارنة المصفوفة الارتباطية الأصلية R (التي عناصر قطرها الرئيسي تساوي الواحد)، مع المصفوفة المخففة R_h (التي تكون عناصر قطرها الرئيسي مؤلفة من التشاركيات h_j^2)، ودراسة الفروقات بينهما وسنعرضها ونشرحها حسب [Harman P. 204] كما يلي :

لقد وجدنا أن عناصر المصفوفة المخففة مرتبطة بمصفوفة الأمثال A المصاحبة للعوامل الرئيسية بالعلاقة:

$$R_h = A_{n*m} * A'_{m*n} \qquad (97-2)$$

وتسعى هذه الطريقة إلى جعل الفرق بين المصفوفتين R و R_h أصغر ما يمكن، لذلك علينا أن نقوم بحساب مجموع مربعات الفروقات بين عناصرهما المتقابلة وأن نحاول معالجته وجعله أصغر ما يمكن .

وهنا نلاحظ أن عناصر المصفوفة الارتباطية الأصلية R معلومة وتساوي r_{jk}، ولكن عناصر المصفوفة المخففة R_h غير معروفة ويجب حسابها من الجداء $A * A'$، بحيث تكون قيمها أقرب ما يمكن من قيم عناصر المصفوفة R .

وإذا رمزنا لعناصر R_h بالرموز $\tilde{r}_{jk}$ فإنها تحسب من العلاقة (2 – 97) كما يلي:

$$\tilde{r}_{jk} = \sum_{p=1}^{m} a_{jp} a_{kp} \qquad (k, j = 1\ 2\ 3 \dots n) \qquad (98-2)$$

ثم نقوم بصياغة تابع المربعات الصغرى للفروقات $\left(r_{jk} - \tilde{r}_{jk}\right)$ كما يلي:

$$\mathrm{F} = \sum_{k=1}^{n} \sum_{j=1}^{n} \left(r_{jk} - \tilde{r}_{jk}\right)^2 \Rightarrow \min \qquad (99-2)$$

وبتعويض (2 – 98) في (2 – 99)، مع أخذ التناظر في R و R_h بعين الاعتبار وبعد الاستغناء عن العناصر القطرية فيها يمكننا كتابة العلاقة (2 – 99) كما يلي:

$$\mathrm{F} = \sum_{k<j}^{n} \sum_{j=1}^{n} \left[r_{jk} - \sum_{p=1}^{m} a_{jp} a_{kp} \right]^2 \Rightarrow \min \qquad (100-2)$$

ولقد تم حذف العناصر القطرية في $(A * A')$ لأنها عبارة عن التشاركيات h_j^2 المرتبطة بالأمثال a_{jp} ، وتحسب منها لاحقاً باستخدام العلاقة : $h_j^2 = \sum_{p=1}^{m} a_{jp}^2$.

وتهدف طريقة البواقي الصغرى إلى تصغير قيمة التابع F والذي يتضمن $\frac{n(n-1)}{2}$ معاملاً واقعاً تحت (أو فوق) القطر الرئيسي للمصفوفة A ، من خلال الحصول على قيم مناسبة لأمثال العوامل الرئيسية المحددة مسبقاً بـ m عاملاً، ومنها نحسب قيم التشاركيات h_j^2 المقابلة للمتحولات Z_j من: $h_j^2 = \sum_{p=1}^{m} a_{jp}^2$ ، فنحصل على جميع عناصر المصفوفة المخففة R_h بنتيجة تلك الحسابات .

ولكن هذه الطريقة تصطدم– في بعض الحالات– بحصولنا على قيم لبعض التشاركيات h_j^2 أكبر من الواحد. وهذا أمر يمكن معالجته بسهولة . وذلك بإضافة n شرطاً عليها، وهو أن يكون :

$$h_j^2 = \sum_{p=1}^{m} a_{jp}^2 \leq 1 \qquad (j: 1\ 2\ 3 \ldots n) \qquad (101-2)$$

وبذلك لا تخرج قيم h_j^2 عن مجال قيمها هو $[0\ ,1]$

وهناك معالجات رياضية معقدة للتابع $(2-100)$ مع الشروط $(2-101)$ للحصول على الأمثال a_{jp} بواسطة طريقة المعاودة. ويتم تنفيذها على الحواسيب حصراً .

مثال (2–4) :

لنأخذ المثال (2–2) المعروض في الفقرة (2–2) ولنطبق عليه طريقة البواقي الصغرى وفق الخطوات التالية:

1– ندخل البيانات العددية للمتحولات الخمسة X_1, X_2, X_3, X_4, X_5 .

2– نحسب المصفوفة الارتباطية الأصلية R (ذات الواحدات على القطر الرئيسي) .

3– نحدد عدد العوامل العامة الممكنة m وليكن $\text{m} = 2$ (أو نحدد مجال تغيره من m_1 إلى m_2) .

4– نحدد العدد الأقصى لمرات المعاودة خلال الحسابات وليكن $t = 25$ مرة .

5– نحدد دقة التقريب ولتكن $d = 0.005$.

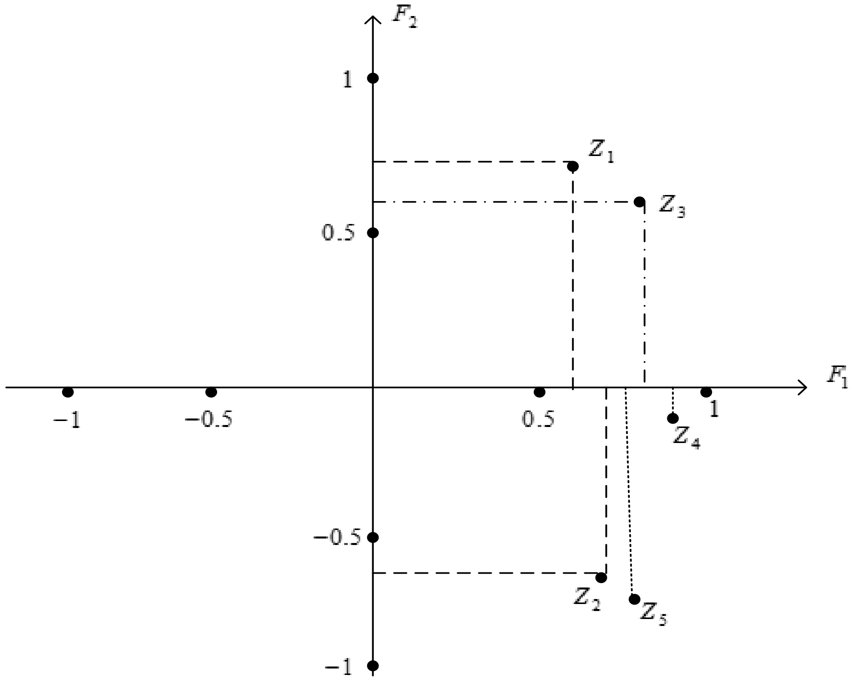
6– نعطي أمر البداية في الحسابات فنحصل على الحل التالي:

جدول (2–9): الحل النهائي بطريقة البواقي الصغرى:

المتحولات	F_1	F_2	التشاركية h_j^2
السكان Z_1	0.621	+0.782	1.000
التعليم Z_2	0.701	−0.522	0.764
العمالة Z_3	0.702	+0.683	0.958
المهنية Z_4	0.881	−0.144	0.797
قيمة المنزل Z_5	0.781	−0.605	0.976
حصة العامل F_5 $V_P = \sum_{J=1}^{5} a_{jp}^2$	2.756	1.739	4.495

وإذا قمنا بمقارنة هذا الحل مع الحل الذي حصلنا عليه بطريقة المكونات الأساسية لنفس المثال، نلاحظ من الجدول (2–5) أن العاملين الرئيسيين F_1 و F_2 فيها، كانا يجسدان حصة كبيرة من التباين الاجمالي (4.670/5=0.934)، ولكن إذا حسبنا مجموع مربعات بواقي العناصر غير القطرية لذلك الحل لوجدنا أنه يساوي($\text{F} = 0.01217$) وهذه القيمة لا تمثل أصغر قيمة للتابع F .

أما الحل الذي حصلنا عليه بواسطة طريق البواقي الصغرى والمعروض في الجدول السابق $(2-9)$ فإنه يعطينا قيمةً للتابع الهدف F أصغر بكثير من ذلك(F=0.00098)، ومنه نستنتج رغم أن هذين العاملين F_1 و F_2 اللذين يجسدان حصة من التباين الاجمالي $(0.90 \approx 4.495/5)$ أقل من حصتهما في الحل السابق ، فإنهما يعكسان صورة أوضح للمعاملات غير القطرية في المصفوفة الارتباطية، ويسمحان بتجسيد حوالي 90% من التباين الاجمالي للمتحولات الخمسة . وأخيراً يمكننا رسم المتحولات بدلالة الحل النهائي على المحورين المتعامدين F_1 و F_2 كما يلي:



الشكل (2−4)التمثيل البياني للحل النهائي بطريقة البواقي الصغرى

2−5 : طريقة الامكانية العظمى (Maximum Likelihood method)

إن مبدأ طريقة الإمكانية العظمى بسيط جداً، ولكن الحسابات والتحويلات الرياضية المتعلقة بها معقدة جداً، ولهذا فإننا سنقتصر على بعض المفاصل الهامة فيها [Harman. P.229] .

إن هذه الطريقة تعتمد على نموذج العوامل الرئيسية المعرف بالعلاقة $(2-20)$ والذي يتطلب تحديد عدد العوامل m أولاً، ويأخذ الشكل التالي:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots a_{jm}F_m + d_jU_j \qquad (102-2)$$

وتشترط أن تكون المتحولات Z_j متحولات معيارية. أي أن متوسطاتها تساوي الصفر وتبايناتها تساوي الواحد. كما تفترض أن تكون جميع العوامل الرئيسية (المجهولة) $F_1, F_2, \dots, F_m$ وكذلك المتحولات الخاصة $U_1, U_2, \dots, U_n$ متحولات مستقلة عن بعضها البعض. وإن كل منها يخضع للتوزيع الطبيعي المعياري $N(0,1)$ الذي متوسطه يساوي الصفر وتباينه يساوي الواحد .

وهذا يجعلنا أن نعتبر أن عدد العوامل m هو عبارة عن عينة مسحوبة من أصل n عاملاً عاماً ولكل منها N قيمة عددية، وهي ستتيح لنا تقدير أمثال تلك العوامل في المجتمع الكلي من بيانات العينة.

وإذا افترضنا أن هذه العينة مسحوبة عشوائياً من مجتمع خاضع لتوزيع طبيعي متعدد المتحولات في المجتمع والعينة، فإنه يمكننا أن نجد تابع التوزيع المشترك بدلالة عناصر مصفوفة التباين المشترك للمتحولات Z_j وبالتالي يمكننا حساب تقديرات الأمثال a_{jp} التي فيه .

ولقد حصل [ويشارت 1928] على التوزيع المشترك لتلك العناصر فكانت صيغته حسب ما ورد في [Harman 1972 p.334] كما يلي :

$$L = K|G|^{-\frac{N-1}{2}} * |S|^{-\frac{N-n-2}{2}} * e^{-\frac{N-1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\sigma^{jk} * S_{jk}} \qquad (103-2)$$

حيث أن: K هو عدد ثابت له علاقة بـ N و بـ n فقط ، ويستخدم لجعل قيمة تكامل L على D تساوي الواحد .

وأن: $|G|$ هو محدد مصفوفة التباينات المشتركة في المجتمع للمتحولات Z_j .

وإن: $|S|$ هو محدد مصفوفة التباينات المشتركة لـ Z_j في العينة المسحوبة .

وإن σ^{jk} هي عناصر مقلوب مصفوفة التباينات G^{-1} للمتحولات Z_j في المجتمع .

وإن S_{jk} هي عناصر مصفوفة التباينات المشتركة للمتحولات Z_j في العينة .

وإذا نظرنا إلى العلاقة (2–103) لوجدنا أن L هو تابع معقد للعناصر σ^{jk} المجهولة . لذلك يجب تحديد أو تقدير العناصر σ^{jk} التي تجعل هذا التابع يأخذ أكبر قيمة ممكنة له (المقابلة للمنوال) . ولتبسيط هذه العلاقة نأخذ لوغاريتم الطرفين فنجد أن:

$$\ln L = -\frac{N-1}{2}\left(ln|G| + \sum_{j=1}^{n}\sum_{k=1}^{n}\sigma^{jk} * S_{jk}\right) + c\left[\text{مقدار ليس له علاقة بـ } G\right] \quad (104-2)$$

ولتحسين شكل هذه العلاقة نقسم طرفيها على $-\frac{N-1}{2}$ ونشكل منها تابعاً جديداً كما يلي:

$$F = -\frac{2}{N-1}\ln L = ln|\mathrm{G}| + \sum\sum\sigma^{jk} * S_{jk} + \mathrm{C} \qquad (\text{ثابت}) \qquad (105-2)$$

وبذلك يصبح هدفنا تحديد العناصر σ^{jk} التي تجعل التابع $\ln L$ أكبر ما يمكن، أو التي تجعل التابع F في (2–105) أصغر ما يمكن . ولإيجاد العناصر σ^{jk} التي تجعل F أصغر ما يمكن، نقوم بحساب المشتقات الجزئية لـ F بالنسبة للعناصر σ^{jk} فنحصل على المعادلات التالية :

$$\frac{\partial F}{\partial\sigma^{jk}} = 0 \qquad (j\,, k = 1\ 2\ 3 \ldots n) \qquad (106-2)$$

وبسبب التناظر فهي تشكل $\frac{n(n+1)}{2}$ معادلة وتتضمن $\frac{n(n+1)}{2}$ مجهولاً. نقوم بحل هذه المعادلات (بواسطة الحاسوب) فنحصل على تقديرات لكل من σ^{jk}، ثم نستخدمها في حساب الأمثال a_{jp} وفق معالجات رياضية معقدة .

ولكننا من جهة أخرى نجد أن المصفوفة الارتباطية الأصلية R تساوي حسب العلاقة (1−65) ما يلي:

$$R = \tilde{R}_h + D^2 \qquad (107-2)$$

وبما أن المصفوفة المخففة $\tilde{R}_h$ تساوي $\tilde{R} = A * A'$ فإن العلاقة السابقة تصبح كما يلي:

$$R = A * A' + D^2 \qquad (108-2)$$

والآن تصبح مهمتنا إيجاد تقدير لعناصر المصفوفتين A و D^2 اللتين تحققان العلاقة (2−108) وتجعلان قيمة التابع F في (2−105) أصغر ما يمكن :

لمعالجة هذه المسألة نضرب العلاقة (2−108) من اليسار بـ $(A' * D^{-2})$ فنجد أن:

$$\begin{aligned} A' * D^{-2} * R &= A' * D^{-2} * A * A' + A' * D^{-2} * D^2 \\ A' * D^{-2} * R &= (A' * D^{-2} * A + I) * A' \end{aligned} \qquad (109-2)$$

ولنرمز بـ J للجداء الداخلي ونكتبه كما يلي:

$$J = A' * D^{-2} * A \qquad (109\,a-2)$$

ونعوضه في العلاقة السابقة فنجد أن:

$$A' * D^{-2} * R = (J + I) * A' = J * A' + A' \qquad (110-2)$$

ومنها نجد أن:

$$J * A' = A' * D^{-2} * R - A'$$

وأن:

$$A' = J^{-1} * (A' * D^{-2} * R - A')$$

ومنها نجد بعد أخذ منقول الطرفين أن :

$$A = \left(R' * D^{-2'} * A - A\right) * \ (J^{-1})' \qquad (111-2)$$

وبما أن R و J و J^{-1} و D^{-2} هي مصفوفات متناظرة، فإن $R' = R$ وإن $J^{-1} = (J^{-1})'$، $D^{-2'} = D^{-2}$ وبذلك نجد أن المصفوفة A تساوي :

$$A = (R * D^{-2} - I) * A * J^{-1} \qquad (111-2)$$

ومنها نحسب عناصر مصفوفة الأمثال A المصاحبة للعوامل الرئيسية والمقابلة لكل المتحولات Z_j بطريقة المعاودة ، ويفضل اتباع الخطوات التالية:

1− نأخذ أي مصفوفة أولية للأمثال ونرمز لها بـ $A_{\frac{1}{2}}$ (عادة يأخذونها من أعمدة العوامل m الأولى المتعامدة)، (إن معنى الدليل $\frac{1}{2}$ سيتضح لاحقاً، وهو يدل على المصفوفة الوسيطة السابقة للمعاودة رقم i) .

2− نحسب المصفوفة القطرية D^2 من العلاقة :

$$D^2 = d_{iag}\left[I - A_{i-\frac{1}{2}} * A'_{i-\frac{1}{2}}\right] \quad : \ \left(i: 1\ 2\ 3 \ldots \text{رقم المعاودة}\right) \qquad (113-2)$$

3− نحسب المصفوفة $J_{i-\frac{1}{2}}$ من العلاقة :

$$J_{i-\frac{1}{2}} = A'_{i-\frac{1}{2}} * D^{-2} * A_{i-\frac{1}{2}} \qquad (114-2)$$

4- نحسب المصفوفة القطرية J_i من العلاقة (2-52) التي تعطينا أن :

$$J_i = Q_i' * J_{i-\frac{1}{2}} * Q_i \qquad (115-2)$$

حيث Q_i هي المصفوفة المتعامدة التي تعطينا مصفوفة قطرية من المصفوفة $J_{i-\frac{1}{2}}$.

5- نحسب مصفوفة الأمثال A_i من العلاقة:

$$A_i = A_{i-\frac{1}{2}} * Q_i \left(\text{تدوير } A_{i-\frac{1}{2}} \text{ لتصبح } J_i \text{ قطرية} \right) \qquad (116-2)$$

6- نحسب المصفوفة $A_{i+\frac{1}{2}}$ من العلاقة :

$$A_{i+\frac{1}{2}} = [R * D_i^{-2} - I] * A_i * J_i^{-1} \qquad (117-2)$$

7- نتأكد من تحقيق التقارب بحساب الفروقات :

$$\left| A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}} \right| < \varepsilon \qquad (118-2)$$

وإذا تحقق مستوى الدقة نتوقف عن الحساب، وإذا لم يتحقق نعيد الحسابات من جديد ... وهكذا .

8- نكرر المعاودة حتى يتحقق مستوى الدقة المطلوب، إن الرمز $\frac{1}{2}$ يعبر عن رقم المصفوفة الوسيطة، التي تظهر خلال عملية المعاودة رقم i، وهو لا يظهر في دليل المصفوفة نتيجة المعاودة .

وعملياً فإن العمليات تبدأ بحساب المصفوفة J_i من الخطوة 4، وهذا يتم بواسطة النظرية الأساسية (التي تنص على أنه يمكن تحويل المصفوفة المتناظرة $J_{i-\frac{1}{2}}$ إلى مصفوفة قطرية باستخدام مصفوفة تحويل متعامدة مثل Q_i (علماً بأن $Q_i * Q_i' = I$) ، وعندها تكون المصفوفة القطرية الجديدة J_i هي مصفوفة القيم الذاتية (المميزة) للمصفوفة $J_{i-\frac{1}{2}}$ ، وإن أعمدة المصفوفة Q_i هي الأشعة الذاتية الواحدية للمصفوفة $J_{i-\frac{1}{2}}$.

9- بعد حساب المصفوفة الجديدة لأمثال العوامل a_{jk} نقوم بالتحقيق من اختبار التقارب (2-118) . وإذا لم يتحقق ذلك الشرط نكرر عمليات المعاودة من 1-7 حتى يصبح أكبر الفروقات بين الأمثال a_{jk} أقل من القيمة المفروضة للتقارب (ولتكن 0.001) .

مثال (2-5):

سنتابع معالجه المثال (2-2) لخمس المتحولات السابق باستخدام طريقة الإمكانية العظمى .

وهنا علينا أن نحدد مسبقاً عدد العوامل m المطلوبة في الحل، وفي هذه الحالة يمكننا أن نختار m=1 أو m=2 أو m=3، ولكننا سنختار العدد m=2 حتى نستطيع مقارنة الحل الناتج عن هذه الطريقة مع الحلول السابقة الناتجة عن طرائق أخرى .

وبعد معالجة بيانات هذا المثال وفق الخوارزمية المذكورة على الحاسوب وتحويل الحل الناتج إلى الشكل القانوني (سنعرفه لاحقاً) حصلنا على الحل المبين في الجدول التالي:

جدول (2−10) : الحل النهائي للمثال المدروس:

العوامل / المتحولات	الحل الأولى لعاملين		الشكل القانوني للحل لعاملين		التشاركية
	F_1	F_2	F_1	F_2	h_j^2
Z_1	0.999	−0.008	0.621	0.783	0.998
Z_2	0.019	0.899	0.711	−0.550	0.809
Z_3	0.974	0.109	0.697	0.689	0.961
Z_4	0.446	0.785	0.891	−0.147	0.815
Z_5	0.030	0.960	0.766	−0.580	0.923
F_p حصة V_p	2.147	2.360	2.759	1.748	4.507

وبمقارنة الشكل القانوني لهذا الحل مع الحل الناتج عن طريق البواقي الصغرى، نلاحظ أنهما متقاربان جداً. وإن التشاركية الاجمالية تشكل حوالي 91% من التباين الإجمالي لهذه المتحولات الخمسة (لأن: $\frac{4.507}{5} = 0.91$). وهذا أمر جيد ونادر ما يحدث .

وعند مقارنة هذا الحل مع الحل بطريقة العوامل الرئيسية نلاحظ بعض الفروقات . وخاصة في حصص العوامل V_p من التشاركية الاجمالية، وذلك لأن طريقتا البواقي الصغرى والإمكانية العظمى تهدفان إلى تعظيم حصص العوامل العامة، وليس إلى تعظيم تشاركيات المتحولات Z_j كما في طريقة العوامل الرئيسية .

2−5−1: اختبار معنوية تقدير عدد العوامل m:

لقد أشرنا إلى أن بعض طرائق التحليل العاملي (البواقي الصغرى والإمكانية الصغرى) تشترط أن يتم مسبقاً تحديد أو تقدير عدد العوامل m الداخلة في النموذج العاملي. وتستخدم لذلك العلاقة التالية [Harman p.239]

$$\left[\frac{(2n+1)-\sqrt{8n+1}}{2}\right] \le m \le \left[\frac{(2n+1)+\sqrt{8n+1}}{2}\right] \qquad (119-2)$$

لذلك كان من الضروري العمل على اختبار صحة هذا التقدير من خلال اختبار فرضية العدم، التي تقول أن عدد العوامل يساوي m_0 أي أن: $H_0: m = m_0$، ويستخدم لذلك مؤشر اختبار خاص يرمز له بـ U_m ويحسب من العلاقة :

$$U_m = -2\ln\lambda = N\ln\frac{|\tilde{R}|}{|R|} \qquad (120-2)$$

حيث أن λ هي نسبة الإمكانية العظمى (Wilks) وأن $0 < \lambda < 1$

وأن: $|\tilde{R}|$ هي قيمة محدد المصفوفة الارتباطية المحسوبة أو المقدرة .

وأن: $|R|$ هي قيمة محدد المصفوفة الارتباطية الأصلية .

وأن N هو حجم عينة المشاهدات ويفترض أن يكون كبيراً ، وكلما تناقصت قيمة λ تزايدت قيمة U_m .
علماً بأن U_m يخضع تقاربياً للتوزيع X^2 وبدرجة حرية تساوي :

$$v = \frac{1}{2}[(n-m)^2 + n - m] \qquad (121-2)$$

ولقد قام (Barlitt) بتعديل مؤشر هذا الاختبار للعينات متوسطة الحجم فتوصل إلى المؤشر التالي:

$$U_m = \left[N - \frac{n}{2} - \frac{3m}{2} - \frac{11}{6}\right] \ln \frac{|\tilde{R}|}{|R|} \qquad (122-2)$$

ولتبسيط العمليات الحسابية تم إيجاد تقدير تقريبي لنسبة المحددين $|\tilde{R}|$ و $|R|$ ، و تمت صياغة U_m على الشكل التالي:

$$U_m = N \sum_{j<k} \sum_{k=1}^{n} \frac{\bar{r}_{jk}^2}{d_j^2 * d_k^2} \qquad (123-2)$$

حيث أن: $\bar{r}_{jk}$ تساوي:

$$\bar{r}_{jk} = r_{jk} - r'_{jk} \qquad (124-2)$$

وإن: r'_{jk} هي قيم المعاملات الارتباطية المحسوبة وهي عناصر المصفوفة $\tilde{R}$.
أما d_j و d_k فهما أمثال عوامل التميز U_j و U_k في النموذج.
وبعد حساب قيمة U_m نقارنها مع القيمة الجدولية لـ $X_\propto^2(v)$ عند مستوى الدلالة $\propto$ ودرجة الحرية v، فإذا كانت $U_m \leq X_\propto^2$ (أو كان $sig > \propto$)، فإننا نقبل فرضية العدم التي تقول أن عدد العوامل العامة يساوي m_0 . أي أن m_0 عاملاً عاماً كافياً لتوصيف البيانات المدروسة، أما إذا كانت $U_m > X_\propto^2$ (أو كانت $sig < \propto$)، فإننا نرفض فرضية العدم القائلة بأن عدد العوامل العامة يساوي m_0 عاملاً، ونقول إنه لتوصيف البيانات المدروسة نحتاج إلى عدد من العوامل m أكبر من m_0 المفترضة. ولذلك يمكن الاستغناء عن العلاقة (2−120)، ونعيد الاختبار من جديد ونفترض أن عدد العوامل العامة وهو $m' = (m_0 + 1)$ ونقوم بإجراء الحسابات اللازمة لحساب قيمة U_m ثم نقارنها بـ $X_\propto^2$ من جديد، ونتخذ قراراً حول $(m' = m + 1)$ ... الخ، وهكذا نكرر هذه الخطوات حتى نتوصل إلى أول قيمة لـ U_m تحقق المتراجحة $U_m < X_\propto^2$ وعندها نضع عدد العوامل يساوي m' الأخيرة .

مثال (2−6):

لنأخذ بتصرف الدراسة التي قام بها (فرنسيس مولين 1939) حول العوامل المؤثرة في نمو الفتيات من عمر 7 إلى 17 عاماً، والتي تضمنت ثمانية متحولات مورفولوجية، وتشكل مجموعتين مختلفتين من المتحولات (متحولات عن القامة ومتحولات عن البدانة) وطبقها على عينة مؤلفة من 305 فتاة في المدارس الابتدائية والاعدادية. وبعد جمع البيانات وحساب المصفوفة الارتباطية لهذه المتحولات كانت كما يلي:

جدول (2−11) المصفوفة الارتباطية للمثال المدروس (مقتبسة من *Harman P. 241*)

المتحولات	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
الطول Z_1	1	0.846	0.805	0.859	0.473	0.398	0.391	0.382
تباعد الزراعين Z_2 (المباع)	0.846	1	0.881	0.826	0.376	0.326	0.277	0.415
أطوال الساعدين Z_3 (الزراع)	0.805	0.881	1	0.801	0.380	0.319	0.237	0.347
أطوال الأرجل Z_4 (الساق)	0.859	0.826	0.801	1	0.436	0.329	0.327	0.365
الوزن Z_5	0.473	0.376	0.380	0.436	1	0.762	0.730	0.629
محيط الحوض Z_6	0.398	0.326	0.319	0.329	0.762	1	0.583	0.577
محيط الصدر Z_7	0.391	0.277	0.237	0.327	0.730	0.583	1	0.539
عرض الصدر Z_8	0.382	0.415	0.347	0.365	0.629	0.577	0.539	1

إن دراسة أولية لمعاملات الارتباط المعروضة في الجدول السابق تبين لنا أن المتحولات الأربعة الأولى ترتبط بقوة مع بعضها البعض، وهي تشكل المتحولات التي تقيس حالة القامة (العامل الأول)، وإن معاملات ارتباطها مع المتحولات الأربعة الثانية ضعيفة. ومن جهة أخرى نجد أن المتحولات الأربعة الثانية ترتبط بقوة مع بعضها البعض، وهي تشكل المتحولات التي تقيس حالة البدانة (العامل الثاني)، وإن معاملات ارتباطها مع المتحولات الأربعة الأولى ضعيفة .

ولهذا يمكننا أن نفترض أن عدد العوامل العامة ($m = 2$)، ونحل هذه المسألة بواسطة طريقة الامكانية العظمى أو بواسطة طريقة البواقي الصغرى. ولكننا سنفترض أن عدد العوامل العامة يمكن أن يكون ($m = 3$). وسنحل هذه المسألة مرتين: مرة للحالة $m = 2$ ومرة للحالة $m = 3$. وبعد إجراء الحسابات المعقدة على الحاسوب بطريقة البواقي الصغرى حصلنا على الجدول التالي:

جدول (2−12) الحل النهائي بطريقة البواقي للحالتين $m = 2$ و $m = 3$:

المتحولات	عدد العوامل $m = 2$			عدد العوامل $m = 3$			
	F_1	F_2	h_j^2	F_1	F_2	F_3	h_j^2
Z_1	0.856	−0.324	0.838	0.860	−0.322	−0.160	0.868
Z_2	0.848	−0.412	0.889	0.867	−0.432	0.242	0.998
Z_3	0.808	−0.409	0.821	0.803	−0.396	0.031	0.803
Z_4	0.831	−0.342	0.808	0.835	−0.340	−0.163	0.839
Z_5	0.750	0.571	0.889	0.751	0.583	−0.113	0.915
Z_6	0.631	0.492	0.640	0.626	0.492	0.019	0.635
Z_7	0.569	0.510	0.583	0.565	0.508	0.001	0.577
Z_8	0.607	0.351	0.493	0.611	0.362	0.182	0.537
حصة العامل F_p من التباين V_p	4.449	1.510	5.959	4.480	1.533	0.158	6.171
النسبة المئوية من التباين الكلي =8	55.61	18.88	74.49	58.00	19.16	1.98	77.14

ومن هذا الجدول يتضح لنا أن الحل الأول المؤلف من عاملين F_1 و F_2 يفسر حوالي (74.45%) من التباين الإجمالي (=8) لهذه المتحولات.

بينما نلاحظ أن الحل الثاني المؤلف من ثلاثة عوامل F_1 و F_2 و F_3 يفسر حوالي (77.14%) من التباين الإجمالي لهذه المتحولات، ومنه نستنتج بدون برهان أن العامل الثالث F_3 لم يكن مؤثراً في التحليل وأن الحل الأول المؤلف من العاملين F_1 و F_2 يعتبر حلاً مقبولاً .

وأخيراً نقوم باختبار فرضية العدم: عدد العوامل $m = 2$، ونقوم بحساب قيمة U_m من العلاقة(2−120) فنجد أن:

$$U_m = 305 \ln \left[\frac{0.00105459}{0.00096740}\right] = 26.32$$

ثم نبحث عن القيمة الجدولية لـ X^2 المقابلة لمستوى دلالة $\propto = 0.05$ ولدرجة حرية v تساوي :

$$v = \frac{1}{2}[(8-2)^2 + 8 - 2] = 21$$

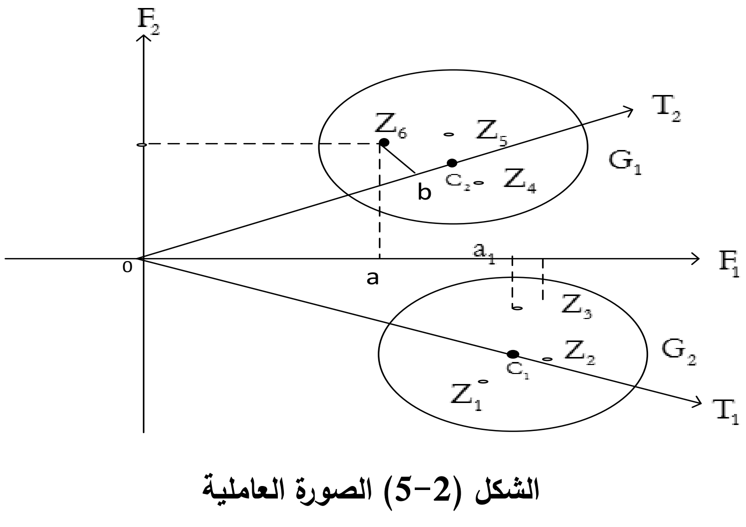
فنجد أن: $X_\propto^2 = 32.671$، وبالمقارنة نجد أن: $U_m < X_\propto^2$ ، لذلك نقبل فرضية العدم $(m = 2)$ التي تقول أن عدد العوامل يساوي 2 ....اي أن توصيف البيانات المدروسة لا يحتاج إلى أكثر من عاملين.

ويمكن التاكد من هذه النتيجة بدراسة قيم مصفوفة البواقي الثالثة، التي تدل على عدم أهمية تلك البواقي .

2−6 : طريقة المجموعات العاملية:

لقد لاحظنا من الطرائق السابقة أن الحل النهائي يكون على شكل مصفوفة A مؤلفة من أعمدة متعامدة في الفضاء الجزئي R^m، وتسمى بمصفوفة الصورة العاملية . ولكن تفسير معاني تلك الصورة قد يكون صعباً أو مستحيلاً بسبب تجمع أو تناثر المتحولات مع بعضها البعض .

فمثلاً عندما تكون الصورة العاملية لستة متحولات Z_j ومؤلفة من عاملين غير مرتبطين (متعامدين) F_1 و F_2 ، يمكننا رسم صورة الحل النهائي لها باستخدام المحاور الإحداثية المتعامدة F_1 0 F_2 كما يلي:



الشكل (2−5) الصورة العاملية

وعندها تظهر لنا مواقع المتحولات Z_j وتبرز أمامنا تجمعاتها المختلفة على المستوى $F_1\ 0\ F_2$. وغالباً ما نحصل على عدة مجموعات متقاربة مع وجود بعض النقاط المنعزلة، وللاستفادة من هذه الوضعية في تحليل وتفسير معاني تلك التجمعات سنربطها بعوامل أخرى مائلة وتمر من مراكز تلك التجمعات. ولكننا قبل التعرض لها نستعرض قضية تجميع المتحولات Z_j في مجموعات متجانسة ، ثم نشرح كيفية تطبيق هذه الطريقة وتحويل الحل المائل إلى حل متعامد حسب ما قدمه [Haman p. 254]. .

2-6-1: تجميع المتحولات Z_j في مجموعات متجانسة :

إن طريقة المجموعات ترتكز على تجميع المتحولات Z_j ضمن مجموعات جزئية متجانسة نرمز لها بـ G_p، كل منها يشمل عدة متحولات تعبر عن عامل معين . وإن نجاح العملية البحثية يعتمد على جودة عملية تجميع تلك المتحولات ضمن المجموعات، وتكون عملية التجميع سهلة إذا كانت تشترط على المتحولات في كل مجموعة G_p أن تكون ارتباطاتها مع بعضها البعض أقوى من ارتباطاتها مع المتحولات الأخرى، ولقياس هذا الشرط تم استنباط مؤشر خاص لذلك يسمى (معامل الانتماء) ويدعى أحياناً (المعامل B) وهو يعتمد على عناصر المصفوفة الارتباطية المخففة R_h . ويعرف بالعلاقة التالية :

$$\text{معامل الانتماء للمتحول } Z_j = \frac{\text{متوسط معاملات الارتباط للمتحول } Z_j \text{ مع متحولات المجموعة المحددة } G_p}{\text{متوسط معاملات ارتباط المتحول } Z_j \text{ مع المتحولات الأخرى من غيرالمجموعة } G_p} \times 100$$

$$B_{(j)} = \frac{\bar{r}_S}{\bar{r}_T}100 \qquad (125-2)$$

حيث أن المتوسطان $\bar{r}_S$ و $\bar{r}_T$ يشملان قيم معاملات المتحولات مع نفسها rjj المستبدلة بقيم التشاركيات h_i^2 من العلاقة: $r_{jj} = h_i^2$

والآن لنعرف المجاميع التالية:

+ مجموع معاملات ارتباط المتحول Z_j مع بقية متحولات المجموعة G_p. ونرمز له بـ S ونحسبه من العلاقة:

$$S = \sum r_{jk} \qquad (j, K \in G_p, \quad j < k) \qquad (126-2)$$

+ مجموع معاملات ارتباط المتحول Z_j مع المتحولات الأخرى خارج المجموعة G_p ونرمز له بـ T ونحسبه من العلاقة:

$$T = \sum r_{jk} \qquad (j \in G_p, \quad k \notin G_p) \qquad (127-2)$$

ولنفترض أن عدد المعاملات في المجموع S هو n_s وإن عدد المعاملات في المجموع T هو n_T ، فعندها يأخذ المعامل $B_{(j)}$ الشكل التالي :

$$B_{(j)} = \frac{\frac{S}{n_s}}{\frac{T}{n_T}}100 = \frac{n_T * S}{n_s * T}100 \quad : (j \in G_p\,, P = 1\ 2\ 3 \dots m) \quad (128-2)$$

وإذا كان عدد المتحولات في المجموعة المحددة Gp يساوي v ، فإن عدد المعاملات الزوجية فيها سيكون مساوياً لعدد الأزواج الممكن تشكيلها من v ،أي يساوي ما يلي:

$$n_s = C_v^2 = \frac{v(v-1)}{2} \qquad (129-2)$$

وعندها يكون عدد المعاملات الأخرى (خارج تلك المجموعة) ومساوياً لجداء العدد v بعدد المتحولات الأخرى ($n-v$) أي يساوي :

$$n_T = v(n-v) \qquad (130-2)$$

وبعد التعويض نجد أن العلاقة $(2-128)$ تأخذ الشكل التالي:

$$B(j) = \frac{\dfrac{S}{\dfrac{v(v-1)}{2}}}{\dfrac{T}{v(n-v)}} 100 \qquad (131-2)$$

وبعد الاختصار على v نحصل على أن معامل إنتماء المتحول Z_j إلى المجموعة يعطى بالعلاقة:

$$B(j) = 2 * \frac{(n-v) * S}{(v-1) * T} * 100 \qquad (132-2)$$

وهناك عدة أساليب أخرى لحساب هذا المعامل لكل متحول Z_j، وأهمها أن نبدأ بأخذ المتحولين الأكثر ارتباطاً بين المتحولات Z_j ونشكل منهما المجموعة الأولى ونحسب لهما قيمة المعامل B . ثم نبحث عن أكبر معامل ارتباط لهما مع المتحولات الأخرى، ونحدد المتحول الثالث المرتبط بشدة معهما ونضيفه إلى تلك المجموعة، ثم نبحث عن أكبر معامل ارتباط للمتحولات التي أصبحت في المجموعة مع المتحولات الأخرى ونضيفها إلى تلك المجموعة ...الخ، ونتابع ذلك مادام ذلك لا يؤدي إلى انخفاض شديد في قيمة المعامل $B_{(j)}$. وبعدها ننتقل إلى تشكيل مجموعة جديدة بنفس الطريقة، وهكذا نتابع حتى نشكل m مجموعة غير متقاطعة. ثم نطبق عليها طريقة المجموعات لإنشاء محاور تمر من مراكز ثقلها، التي سنرمز لها بـ $c_1, c_2, \dots, c_m$.

2-6-2: كيفية تطبيق طريقة المجموعات:

لنفترض أن الحل النهائي مؤلف من عاملين F_1 و F_2 ، لذلك نقوم بإنشاء محورين إحداثيين مائلين $0T_1$ و $0T_2$ يتقاطعان في نفس المبدأ (0)، ويمر المحور الأول $0T_1$ من النقطة C_1 مركز ثقل المجموعة الأولى G_1 ويمر المحور الثاني $0T_2$ من النقطة C_2 مركز ثقل المجموعة الثانية G_2. وعندها نلاحظ أن مساقط المتحولات Z_j في كل مجموعة تزداد على المحور المار بمركز ثقلها، وتتناقص مساقطها على المحور الآخر المار بمركز المجموعة الأخرى (انظر الشكل (2-5)) .

وإذا أردنا أن ننسب هذه المتحولات إلى المحورين المائلين $0T_1$ و $0T_2$ ، فإن مصفوفة الصورة العاملية A تتغير وتتحول إلى مصفوفة أخرى تسمى الصورة العاملية المائلة ونرمز لها بـ P.

ويمكن الحصول على هذه المصفوفة P بواسطة تحويل المصفوفة A المتعامدة إلى المصفوفة المائلة P. ويستخدم لذلك التحويل مصفوفة خاصة تسمى مصفوفة التحويل T، سنتعرض لها لاحقاً .

وعندما ننتقل إلى المحاور المائلة $0T_1$ و $0T_2$ ، فإننا نصبح أمام محاور غير متعامدة (أي مرتبطة مع بعضها البعض). وهذا يعني أن معاملات الارتباط بين المحاور المائلة $0T_1$ و $0T_2$ أصبحت غير معدومة، وهذا يستدعي ظهور مصفوفة جديدة مؤلفة من معاملات الارتباط بين العوامل المائلة، والتي رمزنا لها بالرمز ϕ وهي تساوي:

$$\Phi = \begin{bmatrix} r_{T11} & r_{T12} & \cdots & r_{T1m} \\ r_{T_2T_1} & r_{T_2T_2} & \cdots & r_{T_2T_m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{T_mT_1} & r_{T_mT_2} & \cdots & r_{T_mT_m} \end{bmatrix} \qquad (133-2)$$

وهكذا نجد أن الحل العاملي المائل أصبح يتألف من مصفوفتين أساسيتين هما:

- A مصفوفة الصورة العاملية: وهي مؤلفة من الأمثال a_{jp} وهي مساقط المتحولات Z_j على العوامل العامة المائلة $T_1 0T_2$.
- Φ مصفوفة الارتباط للعوامل المائلة: وهي مؤلفة من معاملات ارتباط العوامل المائلة T_P مع بعضها البعض.

إن طريقة المجموعات تنطلق من تجميع المتحولات Z_j ضمن مجموعات متجانسة من حيث شدة الارتباط بين متحولاتها، أو من حيث طبيعة تلك المتحولات، أو من حيث تمركزها حول نقاط معينة، وتلعب مراكز ثقل المجموعات كمتحولات مركبة من المتحولات Z_j دوراً كبيراً في طريقة المجموعات، ولهذا فإننا سنقوم باستعراض مختصر لبعض خواص هذه المتحولات المركبة (المراكز)، لأنها تختلف عن المتحولات المعيارية Z_j من حيث تباناتها ومعاملات ارتباطها ، لذلك نفترض أنها ونعرف عليها المصطلحات التالية :

- أن عدد المجموعات المتقاربة التي يمكن تجميعها من الحل النهائي هو m مجموعة، ونرمز لها بـ G_P (حيث P: 1 2 3 …m) ، وإن كل مجموعة منها تضمن n_p متحولاً معيارياً Z_j، ونرمز لمركزها بالرمز C_P ، وبذلك يكون عدد العوامل مساوياً لعدد المجموعات m .

ولنفترض أن مجموع قيم المتحولات Z_j في المجموعة G_P يساوي T_P ونكتبه كما يلي:

$$T_P = \sum_{j=1}^{n_p} Z_j \qquad : Z_j \in G_j \qquad : \qquad (j: 1\ 2\ 3 \dots . np) \qquad (134-2)$$

وحيث أن (P: 1 2 3 …m)

- وأن مجموع معاملات ارتباط كل متحول Z_j مع جميع المتحولات في كل مجموعة G_P يساوي ω_{jp} ، ونحسبه كما يلي:

$$\omega_{jp} = \sum_{k=1}^{n_p} r_{jk} \qquad : Z_j \in G_p \qquad \begin{matrix}(p: 123 \dots m) \\ (j: 123 \dots n_p)\end{matrix} \qquad (135-2)$$

حيث أن ω_{jp} يشمل معاملات Z_j مع نفسها والمستبدلة بالتشاركية h_j^2،وحيث يتم وضع $r_{jj} = h_j^2$

- وأن مجموع معاملات الارتباط لكل متحولات المجموعة G_p مع كل متحولات المجموعة الأخرى G_q (بما فيها متحولات المجموعة مع نفسها) ، يساوي W_{pq} ونحسبه من العلاقة:

$$W_{pq} = \sum_{j=1}^{n_p} \omega_{jq} \qquad : \left(Z_j \in G_P \right) \qquad (p, q: 123 \dots m) \qquad (136-2)$$

ويمكن حساب هذا المجموع أيضاً من معاملات الارتباط العادية بين المتحولات Z_j و Z_k في المجوعتين G_p و G_q كما يلي:

$$W_{pq} = \sum_{j=1}^{n_p} \sum_{k=1}^{n_q} r_{jk} \qquad : \left(Z_j \in G_p \ , Z_k \in G_q \right) \qquad (137-2)$$

وحيث أن $(p, q: 1\ 2\ 3 \dots m)$

ويصبح هدفنا الآن حساب معاملات الارتباط بين العوامل المائلة المركبة T_p (وهي تشكل عناصر المصفوفة Φ) ، ثم حساب معاملات الارتباط بين تلك العوامل T_p مع المتحولات Z_j .(وهي تشكل عناصر مصفوفة الهيكل العاملي S) ويمكن حساب هذه المعاملات من خلال المجاميع المعرفة سابقاً، ولكن قبل ذلك نبحث عن كيفية حساب التباين بين العوامل المائلة T_p، وبما أن كل منها هو متحول مركب من عدة متحولات Z_j وحسب العلاقة (2–134) فإن تباينه يساوي :

$$\sigma_{T_p}^2 = n_p + 2 \sum r_{jk} \qquad : \left(Z_j, Z_k \in G_p \ , j < k \right) \qquad (138-2)$$

إن ظهور العدد n_p ناتج عن معاملات ارتباط المتحولات المعيارية Z_j مع نفسها ضمن المجموعة G_p، لأن كل معامل منها يساوي الواحد . وإذا استبدلنا المعاملات n_p والتي نرمز لها بـ r_{jj} بقيم التشاركيات h_j^2 ، وعندها نجد أن التباين $\sigma_{T_p}^2$ يمكن أن يكتب على الشكل التالي :

$$\sigma_{T_p}^2 = \sum_{j=1}^{n_p} \sum_{k=1}^{n_p} r_{jk} \qquad : \left(\ j, k \in G_p \ , r_{jj} = h_j^2 \right) \qquad (139-2)$$

وبمقارنة هذه العلاقة مع العلاقة (2–137) نستنتج أن :

$$\sigma_{T_p}^2 = W_{pp} = \sum_{j=1}^{n_p} \sum_{k=1}^{n_p} r_{jk} \qquad : \left(\ j, k \in G_p \ \right) \qquad (140-2)$$

والآن يمكننا أن نعرف معامل الارتباط بين المتحولين المركبين T_p و T_q كما يلي:

$$r_{T_pT_q} = \frac{\sum_{i=1}^{N}\left(T_{pi} * T_{qi}\right)}{N * \sigma_{T_p} * \sigma_{T_q}} \qquad (141-2)$$

حيث i هو دليل الجمع وهو مأخوذ من 1 حتى N .

إن الانحرافين المعياريين σ_{T_p} و σ_{T_q} يمكن حسابهما مباشرة من المعاملات الأصلية حسب العلاقة (2−139) وضمن كل مجموعة . أما البسط من (2−141) فيمكننا معالجة كما يلي:

$$\frac{1}{N}\sum_{i=1}^{N} T_{pi} * T_{qi} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum Z_{ji}\right)\left(\sum Z_{ki}\right) =$$

$$\sum\sum\left(\frac{1}{N}\sum_{i=1}^{N} Z_{ji} * Z_{ki} : j \in G_p, k \in G_q\right) = \qquad (142-2)$$

$$= \sum_{j=1}^{n_p}\sum_{k=1}^{n_q} r_{jk} : \left(j \in G_p, k \in G_q\right) = W_{pq}$$

وهكذا نجد من العلاقتين (2−140) و (2−142) أن العلاقة (2−141) تأخذ الشكل التالي:

$$r_{T_pT_q} = \frac{W_{pq}}{\sqrt{W_{pp}} * \sqrt{W_{qq}}} \qquad (143-2)$$

وهي تشكل عناصر مصفوفة المعاملات الارتباطية Φ بين العوامل T_p و T_q.

ولنقم الآن بحساب قيم معاملات الارتباط بين المتحولات Z_j والعوامل المائلة T_p (عناصر الهيكل العاملي المائلي S) . فنجد أن عنصر مصفوفة الهيكل S_{jp} يساوي معامل الارتباط بين Z_j والمتحول المركب T_p، أي يساوي r_{Z_jp}، وإن تباينه يحسب من العلاقة (2−140) ، ويمكن البرهان على أن عنصر مصفوفة الهيكل S_{jp} يساوي:

$$S_{jp} = \frac{\omega_{jp}}{\sqrt{Wpp}} \qquad (144-2)$$

وحتى نحصل على الحل النهائي المائل يجب أن نحصل على التوصيف الخطي للمتحولات Z_j بدلالة العوامل المائلة T_p وعلى مصفوفة معاملات الارتباط بين العوامل T_p والمتحولات Z_j (الهيكل العاملي S)، علماً بأن أمثال هذه المعاملات الخطية، والتي تشكل عناصر الصورة المائلة، هي عبارة عن إحداثيات النقاط الممثلة لأشعة المتحولات Z_j على جميع المحاور العاملية المائلة .

وعندما تكون المصفوفة الهيكلية S معلومة، وتكون مصفوفة الارتباطات البينية Φ معلومة، فإنه يمكننا حساب مصفوفة الصورة العاملية المائلة P من العلاقة المشهورة والمشابهة للعلاقة (1−55) التالية :

$$S = P * \Phi \qquad (145-2)$$

وإذا كانت Φ غير شاذة فإنه يمكننا حساب الصورة المائلة P من العلاقة :

$$P = S * \Phi^{-1} \qquad (146-2)$$

وبعد الحصول على الحل العاملي المائل P يمكننا الحصول على المصفوفة المحسوبة لمعاملات الارتباط R^* وللبواقي من العلاقة:

$$R^* = P * \Phi * P' \qquad (147-2)$$

وعندما تكون مصفوفتا الصورة العاملية P والبينية Φ معلومتين، فيمكننا من (2−145)، استبدال $P\Phi$ بـ S فيكون لدينا:

$$R^* = S * P' \qquad (148-2)$$

وبما أن R^* متناظرة فإن $R^{*\prime} = R^*$ ، وعندما نأخذ منقول الطرفين للعلاقة(2−148) نجد أن:

$$R^* = P * S' \qquad (149-2)$$

2−6−3: كيفية تحويل الحل المائل إلى حل متعامد (بعد طريقة المجموعات)

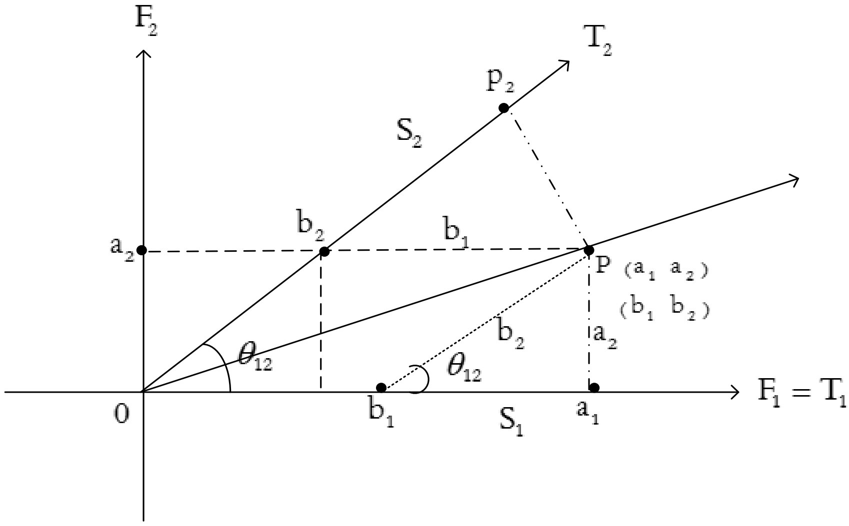
لقد لاحظنا أن طريقة المجموعات العاملية تعطينا حلول مائلة (بزاوية حادة أو منفرجة) تمر من مراكز ثقل المجموعات المتقاربة. ولتحويل تلك الحلول المائلة إلى حلول متعامدة نقوم بما يلي:

نفترض أنه لدينا محورين مائلين $0T_1$ و $0T_2$ ويصنعان بينهما زاوية قدرها θ_{12} ونريد تحويلهما إلى محورين متعامدين $0F_1$ و $0F_2$ وتحويل الصورة العاملية المائلة P إلى الصورة المتعامدة A . وحتى نستطيع فعل ذلك نرسم المحاور الإحداثية الجديدة والمتعامدة $0F_1$ و $0F_2$ كما يلي:

1− نعتبر مبدأ الإحداثيات (O) مبدءاً مشتركاً للنظامين الاحداثيين .

2− نرسم المحور $0F_1$ بحيث يكون منطبقاً على المحور المائل $0T_1$.

3− نرسم المحور $0F_2$ من النقطة 0 عمودياً على $0F_1$ كما في الشكل (2−6) .



الشكل (2−6) التمثيل المائل والمتعامد لـ Z_j

ولنفترض أن النقطة P_j تمثل رأس شعاع المتحول Z_j في R^2 واختصاراً سنرمز لها بـ P فقط .

وأن إحداثياتها في الاحداثيات المتعامدة $F_1\ 0F_2$ هي $P(a_1, a_2)$

وأن إحداثياتها في الاحداثيات المائلة $T_1\ 0T_2$ هي $P(b_1, b_2)$

وأن الزاوية بين المحورين المائلين $T_1\ 0T_2$ هي (θ_{12})

ومن هذا الشكل نجد أنه يمكننا التعبير عن الشعاع Z_j بدلالة إحداثيات النقطة P في كلا الاحداثيات كما يلي:

أ- في الاحداثيات المتعامدة نجد أن:

$$Z_j = a_1F_1 + a_2F_2 \qquad (150-2)$$

ب- في الاحداثيات المائلة نجد أن:

$$Z_j = b_1T_1 + b_2T_2 \qquad (151-2)$$

ومن الشكل أيضاً نجد أن الاحداثيات المتعامدة $P(a_1,a_2)$ مرتبطة مع الاحداثيات المائلة $P(b_1,b_2)$ للنقطة P بواسطة العلاقتين التاليتين :

$$\begin{aligned} a_1 &= b_1 + b_2cos\theta_{12} \\ a_2 &= 0 + b_2sin\theta_{12} \end{aligned} \qquad (152-2)$$

ومن جهة أخرى نعلم أن معامل الارتباط بين المحورين الواحدين المائلين $0T_1$ و $0T_2$ يساوي تجب الزاوية التي بينهما. أي أن:

$$r_{12} = \cos\theta_{12} \qquad (153-2)$$

ومنها نجد أن:

$$\sin\theta_{12} = \sqrt{1-r_{12}^2} \qquad (153\,a-2)$$

وهكذا نجد أن الاحداثيات المتعامدة للنقطة $P(a_1,a_2)$ بدلالة المائلة $P(b_1,b_2)$ (مع اختصار الرمز r_{12} إلى r)، تأخذ الشكل التالي :

$$\begin{aligned} a_1 &= b_1 + b_2r \\ a_2 &= 0 + b_2\sqrt{1-r^2} \end{aligned} \qquad (154-2)$$

ويمكن كتابتها على شكل مصفوفي كما يلي:

$$(a_1,a_2) *= (b_1,b_2) * \begin{bmatrix} 1 & 0 \\ r & \sqrt{1-r^2} \end{bmatrix} \qquad (155-2)$$

وبصورة عامة يمكننا تعميم ذلك على m عاملاً و n متحولاً فنجد أن :

$$A = P * T' \qquad (156-2)$$

حيث أن A_{n*m} :هي مصفوفة الصورة العاملية المتعامدة (a_{jp})

وأن P_{n*m} :هي مصفوفة الصورة العاملية المائلة (b_{jp})

وأن T'_{n*m} :هي منقول مصفوفة التحويل من الاحداثيات المائلة إلى الاحداثيات المتعامدة وهي مصفوفة مثلثية عليا وهي تساوي عندما $(m=2)$ ما يلي :

$$T = \begin{bmatrix} 1 & r \\ 0 & \sqrt{1-r^2} \end{bmatrix} \qquad (157-2)$$

وبما أنه لدينا من (2–146) أن الصورة المائلة P حسب (2–121) تساوي $P = S * \Phi^{-1}$ ، فإننا نجد أن الصورة المتعامدة A المعرفة في (2–156) تساوي :

$$A = S * \Phi^{-1} * T' \qquad (158-2)$$

ولكن المصفوفة Φ تحقق خاصية (جذر الوحدة التربيعي) $\Phi = T' * T$ وعندها نجد أن :

$$A = S * (T' * T)^{-1} * T' = S * \left(T^{-1} * T'^{-1}\right) * T' = S * T^{-1} * \left(T'^{-1} * T'\right)$$

$$A = S * T^{-1} \qquad (159-2)$$

حيث أن T هي مصفوفة مثلثية عليا مؤلفة من عناصر طريقة الجذر التربيعي، وهي العناصر الواردة في (2–157)، أما S فهي المصفوفة الهيكلية العاملية المائلة، وإن هذه العلاقة تساعدنا على الانتقال من الهيكلية المائلة S إلى الصورة العاملية المتعامدة A. ويمكن التعبير عنها بكتابة عبارات التحويل في العلاقة بدلالة المساقط S_{jp} على جملة من المحاور المائلة (بدلاً من الاحداثيات b_{jp} في الإحداثيات المائلة الأولية) .

مثال(2–7):

إن دراسة أجريت للتعرف على العوامل التي تؤثر على تحصيل تلاميذ المرحلة الابتدائية، واعتمدت 9 مؤشرات لمعرفة تأثيرها على تحصيلهم المدرسي، وجمعت البيانات عنهم ثم حسبت معاملات الارتباط الزوجية لهذه المؤشرات وقدرت التشاركيات h_j^2 ووضعتها على القطر الرئيسي فكانت المصفوفة الارتباطية المخففة كما يلي:

جدول (2–12) معاملات الارتباط لمؤشرات الدراسة:(مستقاة من Harman,P.165)

المؤشرات / المؤثرات	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
Z_1– معاني الكلمات	(0.81)	0.75	0.78	0.44	0.45	0.51	0.21	0.30	0.31
Z_2– تركيب الجمل	0.75	(0.69)	0.72	0.52	0.53	0.58	0.23	0.32	0.30
Z_3– الكلمات المز	0.78	0.72	(0.75)	0.47	0.48	0.54	0.28	0.37	0.37
Z_4– مسائل الحساب	0.44	0.52	0.47	(0.91)	0.82	0.82	0.33	0.33	0.31
Z_5– حساب الأعداد	0.45	0.53	0.48	0.82	(0.74)	0.74	0.37	0.36	0.36
Z_6– الأعداد الناقصة	0.51	0.58	0.54	0.82	0.74	(0.74)	0.35	0.38	0.38
Z_7– حالة الكفوف	0.21	0.23	0.28	0.33	0.37	0.35	(0.35)	0.45	0.52
Z_8– حالة الأحذية	0.30	0.32	0.37	0.33	0.36	0.38	0.45	(0.58)	0.67
Z_9– حالة الطاولة	0.31	0.30	0.37	0.31	0.36	0.38	0.52	0.67	(0.77)

ملاحظة: لقد تم استبدال العناصر القطرية بالتشاركيات h_j^2 المقدرة من العلاقة الثلاثية (1–71) لذلك وضعناها ضمن قوسين على القطر الرئيسي .

ولإيجاد العوامل المؤثرة على التحصيل الدراسي تم تقسيم هذه المؤشرات التسعة حسب شدة الارتباط بين عناصرها إلى 3 مجموعات، كل منها مؤلف من 3 مؤشرات ، ورمزنا لها كما يلي:

$$G_1(Z_1\ Z_2\ Z_3)\ ,\qquad G_2(Z_4\ Z_5\ Z_6\ )\ ,\qquad G_3(Z_7\ Z_8\ Z_9\ )$$

وعند حساب المجاميع T_p من البيانات الفعلية حسب (2−134) كانت تساوي:

$$T_1 = 187\ ,\qquad T_2 = 186\ ,\qquad T_3 = 168$$

وعند حساب المقادير ω_{jp} من مجموع المعاملات الارتباطية ضمن كل مجموعة حسب العلاقة (2−135) تم وضعها في الجدول (2−14) التالي .

فمثلاً نجد أن مجموعا معاملات ارتباط المتحول Z_1 في المجموعتين G_1 و G_2 يساويان :

$$\omega_{11} = 0.81 + 0.75 + 0.78 = 2.34$$
$$\omega_{12} = 0.44 + 0.45 + 0.51 = 1.40$$

جدول (2−14) قيم المقادير ω_{jp} محسوبة من (2−8)

المجموعات / المؤشرات	G_1	G_2	G_3
Z_j	ω_{j1}	ω_{j2}	ω_{j3}
Z_1	2.34	1.40	0.82
Z_2	2.16	1.63	0.85
Z_3	2.25	1.49	1.02
Z_4	1.43	2.55	0.97
Z_5	1.46	2.30	1.09
Z_6	1.63	2.30	1.11
Z_7	0.72	1.05	1.32
Z_8	0.99	1.07	1.72
Z_9	0.98	1.05	1.96

وعند حساب المجاميع W_{pq} تم وضعها في الجدول (2−15) التالي: وكمثال على ذلك نجد أن:

$$W_{11} = 2.34 + 2.16 + 2.25 = 6.75$$
$$W_{12} = 1.43 + 1.46 + 1.63 = 4.52$$

جدول (2–15): قيم المقادير W_{pq} محسوبة من (2–137)

المجموعة q / المجموعة p	G_1	G_2	G_3
G_1	6.75	4.52	2.96
G_2	4.52	7.15	3.17
G_3	2.96	3.17	4.98
$\sqrt{\omega_{pp}}$	2.60	2.67	2.23

ثم قمنا بحساب عناصر المصفوفة S_{jp} من العلاقة (2–144) وكمثال على ذلك نجد أن:

$$S_{11} = \frac{\omega_{11}}{\sqrt{\omega_{11}}} = \frac{2.34}{2.60} = 0.90$$
$$S_{12} = \frac{\omega_{12}}{\sqrt{\omega_{22}}} = \frac{1.40}{2.67} = 0.52$$

وهكذا نكون قد حصلنا على المصفوفة الهيكلية S المؤلفة من معاملات ارتباط المتحولات Z_j مع العوامل المائلة OT_1 , OT_2 , OT_3 ووضعناها في الجدول التالي:

جدول (2–16) المصفوفة الهيكلية S وعناصرها S_{jp}

المحاور / المتحولات	OT_1	OT_2	OT_3
Z_1	0.90	0.52	0.37
Z_2	0.83	0.61	0.38
Z_3	0.77	0.56	0.46
Z_4	0.55	0.96	0.43
Z_5	0.56	0.86	0.49
Z_6	0.64	0.86	0.50
Z_7	0.28	0.39	0.59
Z_8	0.38	0.40	0.76
Z_9	0.38	0.39	0.88

ولحساب عناصر الحل المائل P من العلاقة (2–146) قمنا بحساب عناصر المصفوفة Φ من العلاقة (2–133):

$$\Phi = \left[r_{T_pT_q}\right] = \left[\frac{W_{pq}}{\sqrt{W_{pp}}\sqrt{W_{qq}}}\right] \qquad (160-2)$$

فمثلاً نجد من الجدول (2−16) أن:

$$r_{T_1T_2} = \frac{W_{12}}{\sqrt{W_{11}}\sqrt{W_{22}}} = \frac{4.52}{(2.60)(2.67)} = 0.6511$$

$$r_{T_1T_3} = \frac{w_{13}}{\sqrt{W_{11}}\sqrt{W_{33}}} = \frac{2.69}{(2.60)(2.23)} = 0.4640$$

$$r_{T_2T_3} = \frac{W_{23}}{\sqrt{W_{22}}\sqrt{W_{33}}} = \frac{3.17}{(2.67)(2.23)} = 0.5324$$

وبذلك نجد أن المصفوفة Φ المتناظرة تساوي:

$$\Phi = \begin{bmatrix} 1 & 0.6511 & 0.4640 \\ 0.6511 & 1 & 0.5324 \\ 0.4640 & 0.5324 & 1 \end{bmatrix}$$

ومنها نجد أن Φ^{-1} تساوي:

$$\Phi^{-1} = \begin{bmatrix} 1.79582 & -1.01267 & -0.29411 \\ -1.01267 & 1.96663 & -0.57715 \\ -0.29411 & -0.57715 & 1.44374 \end{bmatrix}$$

ومنها يمكننا أن نحصل على الصورة العاملية المائلة P من الجداء التالي :

$$P = S * \Phi^{-1} = \begin{bmatrix} T_1 & T_2 & T_3 \\ 0.98 & -0.09 & -0.04 \\ 0.76 & 0.14 & -0.05 \\ 0.86 & -0.04 & 0.08 \\ -0.11 & 1.07 & -0.08 \\ -0.01 & 0.84 & 0.04 \\ 0.11 & 0.77 & 0.04 \\ -0.06 & 0.14 & 0.55 \\ 0.05 & -0.04 & 0.76 \\ 0.03 & -0.12 & 0.93 \end{bmatrix} = \begin{bmatrix} 0.98 & - & - \\ 0.76 & - & - \\ 0.86 & - & - \\ - & 1.07 & - \\ - & 0.84 & - \\ - & 0.77 & - \\ - & - & 0.55 \\ - & - & 0.76 \\ - & - & 0.93 \end{bmatrix}$$

وعند حذف الأمثال الصغيرة (التي أقل من 0.20) نحصل على الصورة العاملية المائلة، التي تعبر عن كل متحول Z_j بواسطة عامل واحد، ويتشكل منها 3 مجموعات تتضمن كل منها 3 متحولات وهي تكتب على الشكل التالي:

$$\begin{array}{ccc} \underline{G_1} & \underline{G_2} & \underline{G_3} \\ Z_1 = 0.98\, T_1 & Z_4 = 1.07\, T_2 & Z_7 = 0.55\, T_3 \\ Z_2 = 0.76\, T_1 & Z_5 = 0.84\, T_2 & Z_8 = 0.78\, T_3 \\ Z_3 = 0.86\, T_1 & Z_6 = 0.77\, T_2 & Z_9 = 0.93\, T_3 \end{array}$$

وهي تحقق شروط شكل الهيكل البسيط المعرف في الفقرة (2−1) وتفسر لنا العلاقات بين المتحولات Z_j حيث نجد أن متحولات كل مجموعة ذات طبيعة واحدة وترتبط بشدة مع بعضها .

وإذا أردنا أن نحول هذه الصورة المائلة إلى صورة متعامدة، نستخدم مصفوفة التحويل T من نظام الإحداثيات المائلة $F_1\,0F_2$ إلى نظام الإحداثيات المتعامدة $T_1, 0T_2$. لذلك نقوم بحساب عناصر المصفوفة T ثم نطبق عليها العلاقة:

$$A = S * T^{-1}$$

وقياساً على المصفوفة T المعرفة لعاملين بالعلاقة (2−157) نحسب عناصر المصفوفة T لتحويل ثلاثة عوامل مائلة إلى ثلاثة محاور متعامدة اعتماداً على المصفوفة Φ وذلك باستخدام (طريقة جذر الوحدة التربيعي) ثم نحسب T^{-1} فنحصل على أن:

$$T^{-1} = \begin{bmatrix} 1.000 & -0.8578 & -0.2448 \\ 0 & 1.375 & -0.4803 \\ 0 & 0 & 1.2013 \end{bmatrix}$$

وبإجراء الجداء السابق $(S * T^{-1})$ نحصل على الصور المتعامدة A التالية:

$$A = S * T^{-1} = \begin{array}{ccc} F_1 & F_2 & F_3 \end{array} \begin{bmatrix} 0.90 & -0.09 & -0.03 \\ 0.83 & 0.09 & -0.04 \\ 0.87 & -0.01 & 0.07 \\ 0.55 & 0.79 & -0.07 \\ 0.56 & 0.65 & 0.04 \\ \underline{0.63} & \underline{0.60} & 0.03 \\ 0.28 & 0.27 & 0.45 \\ 0.38 & 0.20 & 0.63 \\ 0.38 & 0.19 & 0.77 \end{bmatrix}$$

وهي تشبه الصورة المائلة P ولكنها أقل تعبيراً وأصعب تفسيراً لأن بعض الأمثال متشابكة. لذلك سنلجأ إلى الصورة المائلة في الحلول العاملية .

2−7 : الأشكال الممكنة للحلول العاملية :

لقد أشرنا عند عرض بعض الأمثلة العددية إلى أنه بعد الحصول على أمثال المكونات الأساسية المبينة في الجداول السابقة، نقوم بحذف الأمثال التي تقل قيمتها عن مقدار معين (مثل 0.10 أو 0.20)، وعندها سنحصل على الأمثال المعنوية المرافقة للمكونات الأساسية، وقد تؤدي هذه العملية إلى حذف بعض المكونات الأساسية كلياً، وبالتالي قد يأخذ الحل العام أحد الأشكال التالية :

1− الشكل العام للمكونات الأساسية: (طريقة Hotelling+ keley)

وفيه يتم التعبير عن المتحولات Z_j بدلالة n من العوامل العامة (وبدون عوامل التميز U_j) وهو يأخذ الشكل التالي (عدد العوامل = عدد المتحولات = n)

العوامل / المتحولات	F_1	F_2	F_3	F_4	F_5	F_6	
Z_1	x	X	x	x	x	x	بدون عوامل التميز U_j
Z_2	x	X	x	x	x	x	
Z_3	x	X	x	x	x	x	
Z_4	x	X	x	x	x	x	
Z_5	x	X	x	x	x	x	
Z_6	x	X	x	x	x	x	
Z_7	x	X	x	x	x	x	
Z_8	x	X	x	x	x	x	
Z_9	x	X	x	x	x	x	
حيث تشير علاقة x في الخلية (Pj) على وجود أمثال عددية للمتحول Z_j على العامل F_p							

الشكل (2–7) الشكل العام

2– شكل الهيكل البسيط Simple structure

وهو أهم الأشكال المستخدمة في عمليات تحليل العوامل، وفيه يتم التعبير عن المتحولات Z_j بدلالة عدد محدد m من العوامل العامة (m<n)، ويعتمد على النموذج المختصر (2–20) الذي يتضمن عوامل التميز U_j، ويستخدم هذا الشكل لتبسيط صورة الحل النهائي في الإحداثيات المتعامدة أو المائلة، وهو يشترط أن تحقق المصفوفة الهيكلية الشروط التالية (حسب Thurstone) :

أ– أن يتضمن كل سطر من مصفوفة الهيكل العاملي صفراً واحداً على الأقل، وهذا يعني أن كل شعاع من أشعة المتحولات Z_j يكون واقعاً على أحد المستويات الإحداثية، مما يجعل مسقطه على المحور المعامد مساوياً للصفر .

ب– أن يتضمن كل عمود من المصفوفة الهيكلية العاملية عدداً من الأصفار لا يقل عن m صفراً (m عدد العوامل العامة) .

ج– في كل زوج من أعمدة المصفوفة الهيكلية، يمكن أن نجد عدداً ن المتحولات تكون أمثالها الهيكلية مساوية للصفر في أحد هذين العمودين، وتكون ذات أهمية ومعنوية في العمود الآخر .

د– إذا كان عدد العوامل 4 أو أكثر $(m \geq 4)$ فإن عدداً من المتحولات في أي زوج من أعمدة المصفوفة الهيكلية يمكن أن يكون لها أمثال مساوية للصفر في كلا العمودين .

هـ– في أي زوج من أعمدة مصفوفة الهيكلية يمكن أن نجد عدداً قليلاً من المتحولات التي تكون أمثالها المتقابلة ذات أهمية معنوية في كلا العمودين .

وسنستخدم هذه الشروط عند إجراء عمليات التدوير في النظم الإحداثية المتعامدة أو المائلة، للحصول على الهيكل البسيط. الذي يؤدي إلى تخفيض تعقيدات الحل النهائي ويساعد على تفسير النتائج .

3- أهم الأشكال البسيطة : [Okon,Jan P.137] :

أ- الشكل الأحادي (طريقة بيرسون) : وفيه يتم التعبير عن كل المتحولات Z_j بدلالة عامل واحد من العوامل العامة مع عامل التميز U_j وهو يأخذ الشكل التالي:

العوامل / المتحولات	F_1	F_2	U_1	U_2	U_3	U_4	U_5	U_6
Z_1	x		x					
Z_2	x			X				
Z_3	x				X			
Z_4		x				X		
Z_5		x					x	
Z_6		x						X

الشكل (2-8)

ب- الشكل الثنائي (Dufactors): وفيه يتم التعبير عن كل من المتحولات Z_j بدلالة عامل واحد هو العامل الشامل F_1 مع عامل التميز U_j وهو يأخذ الشكل التالي :

العوامل / المتحولات	F_1	U_1	U_2	U_3	U_4	U_5	U_6
Z_1	x	X					
Z_2	x		X				
Z_3	x			X			
Z_4	x				x		
Z_5	x					x	
Z_6	x						X

الشكل (2-9)

ج- الشكل المزدوج befoctors (طريقة هوليزير): وفيه يتم التعبير عن كل متحول Z_j بدلالة عامل شامل F_1 وعامل أخر من العوامل $F_1, F_2 \dots . F_m$ مع العوامل Uj وهو يأخذ الشكل التالي:

العوامل / المتحولات	F_1	F_2	F_3	F_4	U_1	U_2	U_3	U_4	U_5	U_6
Z_1	x	x			X					
Z_2	x	x				x				
Z_3	x		x				x			
Z_4	x		x					x		
Z_5	x		x						x	
Z_6	x			x						X

الشكل (2−10)

د− الشكل المتعدد: (طريقة Thurstone): وفيه يتم التعبير عن المتحولات Z_j بدلالة عدد m من العوامل العامة $(m < n)$ مع وجود عوامل التميز U_j ويأخذ الشكل التالي (بعد حذف الأمثال الصغيرة منه وبدون نظام محدد)

العوامل / المتحولات	F_1	F_2	F_3	F_4	U_1	U_2	U_3	U_4	U_5	U_6
Z_1	X				X					
Z_2	x	x				x				
Z_3	x	x	X				x			
Z_4			X	x				x		
Z_5		x	X	x					x	
Z_6		x		x						X

الشكل (2−11)

هـ−الشكل المثلثي: وفيه يتم التعبير عن المتحول الأول Z_1 (أو غيره) بدلالة عامل واحد F_1 فقط، وعن المتحول Z_2 بدلالة عاملين F_2, F_1 ، وعن المتحول الثالث Z_3 بدلالة ثلاثة عوامل F_3, F_2, F_1 ...الخ وهو يأخذ مع عوامل التميز الشكل التالي:

العوامل / المتحولات	F_1	F_2	F_3	F_4	F_5	F_6	U_1	U_2	U_3	U_4	U_5	U_6
Z_1	x						x					
Z_2	x	x						x				
Z_3	x	x	x						x			
Z_4	x	x	x	x						X		
Z_5	x	x	x	x	X						x	
Z_6	x	x	x	x	X	X						x

الشكل (2−12)

ملاحظة: يمكن تغيير ترتيب مواقع المتحولات Z_j للحصول على هذا الشكل المثلثي .

4- الشكل القانوني للحل العاملي :

عندما نقوم بتحليل وحل إحدى مسائل التحليل العاملي بطريقتين مختلفتين أو أكثر. فإننا قد نحصل حلين مختلفين ظاهرياً من حيث أمثال المتحولات بدلالة العوامل . أو من حيث عدد العوامل المعتمدة m في كل منهما.لأن كل منهما يكون منسوباً إلى جملة إحداثيات خاصة به .

وهنا نتساءل فيما إذا كان هذان الحلان مختلفين أم متكافئين ؟

لذلك نقوم بتعريف الحلول المتكافئة بما يلي:

نقول عن أي حلين عاملين إنهما متكافئان، إذا تطابقت أمثالها وعواملها عند تحويلها إلى جملة إحداثيات ثابتة تسمى بالشكل القانوني . ولهذا كان من الضروري تحديد صيغة معينة نسميها الشكل القانوني للحل المدروس، وإن عملية تحويل أي حل إلى الشكل القانوني ماهي إلا وسيلة لتحويل ذلك الحل إلى جملة من الإحداثيات ثابتة رياضياً.

يتميز الشكل القانوني بأنه يحافظ على أن يأخذ كل عامل من العوامل المتتالية أكبر حصة ممكنة من التباين الإجمالي المحدد في فضاء العوامل العامة وبأن تكون عوامله مستقلة ومتعامدة. ومما سبق نتصور أنه يمكننا الحصول على صيغة الشكل القانوني لأي حل عاملي A من خلال تحويله إلى الشكل القانوني K بواسطة مصفوفة تحويل متعامدة ووحدوية T وفق علاقة التحويل التالية:

$$K = A * T \qquad (161-2)$$

حيث أن A هي مصفوفة الحل المدروس وتكون من المرتبة $n * m$

وأن T هي مصفوفة التحويل المتعامدة الوحدوية وتكون من المرتبة $m * m$

وأن K هي مصفوفة الصيغة القانونية للحل المدروس A وتكون من المرتبة $n * m$

وهكذا نجد أن جوهر عملية الحصول على صيغة الشكل القانوني يكمن في إيجاد مصفوفة التحويل T .

لذلك نضرب طرفي العلاقة (2−161) من اليسار بـ K' فنجد أن :

$$K'_{m*n} * K_{n*m} = K' * A * T = T'_{m*n} * A'_{m*n} * A_{n*m} * T_{m*m} \qquad (162-2)$$

وهنا نلاحظ أن المصفوفة الناتجة $K' * K$ هي مربعة من المرتبة $m * m$، وبما أن K' متعامدة (يجب أن تكون متعامدة) فإن الجداء $K' * K$ يكون من المرتبة $m * m$ ويساوي مصفوفة قطرية نرمز لها بـ Λ ونكتب : $K' * K = \Lambda_{m*m}$ وبذلك تأخذ العلاقة (2−162) الشكل التالي:

$$T' * (A' * A) * T = \Lambda \qquad (163-2)$$

وهنا نلاحظ أن الجداء $(A' * A)$ هو عبارة من مصفوفة متناظرة من المرتبة $m * m$ وعند مقارنة العلاقة (2−163) مع العلاقة (2−52) السابقة التالية:

$$A' * R * A = \Lambda \qquad (164-2)$$

نجد أنهما متشابهتان من حيث الشكل (حيث تحل $(A' * A)$ محل R)، وبناء على ذلك نستنتج أن العلاقة (2–163) ماهي إلا حالة خاصة من (النظرية الطيفية)، وهي تعنى أنه يمكن تحويل الجداء $(A' * A)$ إلى مصفوفة قطرية Λ بواسطة مصفوفة التحويل T ، وإن العناصر القطرية في المصفوفة الناتجة Λ هي القيم الذاتية لمصفوفة الجداء $(A' * A)$. وإن أعمدة مصفوفة التحويل T هي عبارة عن الأشعة الواحدية الذاتية المقابلة لتلك القيم الذاتية للجداء $(A' * A)$. وهكذا نجد أنه للحصول على مصفوفة تحويل أي حل ابتدائي A إلى الشكل القانوني K . يكفي أن نبحث عن القيم الذاتية والأشعة الذاتية للجداء $(A' * A)$ المقابلة لتلك القيم الذاتية ونشكل منها المصفوفة T.ثم نحسب K من العلاقة (2–161) فنحصل على الشكل القانوني المطلوب .

5– دراسة تكافؤ الحلول العاملية :

يمكننا دراسة تكافؤ أي حلين عامليين A وB ناتجين عن طريقتين مختلفتين، من خلال تحويلهما إلى الشكل القانوني، ويكونا متكافئين إذا تطابقت الأمثال والعوامل المتقابلة بهما ، أي إذا تحققت الشروط $(AT = K_1 = K_2 = BT)$، ومن جهة أخرى يمكننا دراسة العلاقة المباشرة بين هذين الحلين A وB، اللذين ينطلقان من نفس البيانات ويعتمدان على نفس المصفوفة الارتباطية. ونفترض أنهما متكافئان (رغم اختلاف الأمثال وعدد العوامل بينهما) وأنهما مرتبطان بعلاقة التكافؤ التالية :

$$R_h = A * A' = B * B' \qquad (165-2)$$

وبما أن كل من A وB يستند إلى جملة محاور إحداثية معينة، فإنه يمكننا نظرياً وعملياً أن نفترض أنه يوجد مصفوفة T تحول الحل A من الجملة الأولى إلى الحل B في الجملة الثانية، ونكتب ذلك على الشكل التالي :

$$A * T = B \qquad (166-2)$$

وبما أن A_{n*m} مصفوفة مستطيلة فليس لها مقلوب، لذلك نستفيد من خواص المصفوفة المركبة، $(A' * A)^{-1}(A' * A) = I$ ونضرب طرفي العلاقة (2–166) من اليسار بال $((A' * A)^{-1} * A')$ ، فنحصل على أن:

$$(A' * A)^{-1} * A' * A * T = (A' * A^{-1}) * A' * B$$

وبالاختصار نجد أن مصفوفة التحويل T تحسب من العلاقة :

$$T = (A' * A)^{-1} A' * B \qquad (167-2)$$

وبما أن العوامل متعامدة فإن مصفوفة الجداء $A' * A$ تشكل مصفوفة مربعة قطرية، وإن مقلوبها $(A' * A)^{-1}$ يكون موجوداً ويساوي مصفوفة قطرية أخرى وإن عناصرها تساوي مقاليب عناصر $(A' * A)$.

وهكذا نجد أنه انطلاقاً من الحلين A وB يمكننا أن نجد المصفوفة T لتحويل حل الطريقة الأولى A إلى حل الطريقة الثانية B . وذلك من خلال العلاقة (2–167) .

ويمكننا الاستفادة من العلاقة (2−167) لتحويل قيم العوامل في الحلين F و F^* حيث أن شرط التكافؤ يقتضي أن يكون :

$$A * F = B * F^* \qquad (168-2)$$

وبطريقة مشابهة نجد أن قيم العوامل F في الطريقة الأولى تحسب من خلال العلاقة:

$$F = (A' * A)^{-1} * A' * B * F^* = T * F^* \qquad (169-2)$$

وهكذا يمكننا بمساعدة المصفوفة T تحويل قيم العوامل في أحد الحلين إلى الحل الأخر . أي أن المصفوفة T تسمح لنا بتوضيح أو تفسير العلاقة بين أي حلين عاملين .

مثال (2−8): على التكافؤ : لنفترض أنه لدينا مسألة من ستة متحولات وبعد حلها حصلنا على الحلين:

A – الحل المتعامد الذي حصلنا عليه بطريقة العوامل الرئيسية .

B– الحل المائل حصلنا عليه بطريقة المجموعات لنفس المثال.

$$A = \begin{array}{c} \begin{matrix} F_1 & \quad F_2 \end{matrix} \\ \begin{bmatrix} 0.8619 & -0.2703 \\ 0.7448 & -0.3051 \\ 0.6815 & -0.1782 \\ 0.3477 & 0.7246 \\ 6.3785 & 0.5971 \\ 0.2339 & 0.4435 \end{bmatrix} \end{array} \quad B = \begin{array}{c} \begin{matrix} F_1 & \quad F_2 \end{matrix} \\ \begin{bmatrix} 0.7392 & -0.5190 \\ 0.6232 & -0.5067 \\ 0.6983 & -0.3735 \\ 0.5460 & 0.5875 \\ 0.5393 & 0.9588 \\ 0.3559 & 0.3537 \end{bmatrix} \end{array}$$

ومنها نجد أنه يجب أن يكون : $R_h = A * A' = B * B'$

ولإيجاد مصفوفة التحويل T بينهما نقوم بحساب وإجراء ما يلي:

$$A' * A = \begin{bmatrix} 2.0809 & 0.0000 \\ 0.0000 & 1.2762 \end{bmatrix}$$

$$(A' * A)^{-1} = \begin{bmatrix} 0.4806 & 0.0000 \\ 0.0000 & 0.7836 \end{bmatrix}$$

$$(A' * A)^{-1} * A' = \begin{bmatrix} 0.4142 & 0.3579 & 0.3275 & 0.1671 & 0.1819 & 0.1124 \\ -0.2118 & -0.2391 & -0.1396 & 0.5678 & 0.4679 & 0.3475 \end{bmatrix}$$

وبذلك نجد أن مصفوفة التحويل بينهما T وزاوية الدوران تساوي :

$$T = (A' * A)^{-1} * A' * B = \begin{bmatrix} 0.9546 & -0.2976 \\ 0.2976 & 0.9546 \end{bmatrix} = \begin{bmatrix} \cos(17.33) & -\sin(17.33) \\ \sin(17.33) & \cos(17.33) \end{bmatrix}$$

وللتحقق من ذلك نقوم بتحويل الحل A بواسطة T فنحصل على الحل B (تقريباً) كما يلي:

$$A * T = \begin{bmatrix} 0.7432 & -0.5145 \\ 0.6202 & -0.5129 \\ 0.6975 & -0.3729 \\ 0.5474 & 0.5882 \\ 0.5390 & 0.9574 \\ 0.3553 & 0.3538 \end{bmatrix} \simeq B$$

الفصل الثالث

قضايا التدوير Rotation

تمهيد:

إن الهدف الأساسي للتحليل العاملي هو التعبير عن عدد كبير من المتحولات Z_j بدلالة عدد محدود من العوامل الافتراضية F_p . وإن هذه العوامل يجب أن تعبر عن المتحولات المدروسة وتشكل توصيفاً مختصراً للعلاقات الموجودة بينها، وإنها قد تكون متعامدة (غير مرتبطة) أو مائلة (مرتبطة مع بعضها)، وإن الأمثال العددية المصاحبة لها تعبر عن معاملات ارتباطها مع المتحولات الأصلية، ومنها يمكننا حساب تشاركيات المتحولات وتباينات العوامل. ولكن تلك الأمثال والحلول المتشكلة منها نادراً ما تعطينا تفسيراً موضوعياً عن العلاقات بين المتحولات المدروسة ، لأنها تتغير من عينة لأخرى وتتأثر بإدخال المتحولات الجديدة التي تؤدي إلى إزاحة مراكز ثقل المجموعات، وإن هذا يؤدي إلى تغير وضع المحاور الإحداثية وبالتالي إلى تغير تلك الأمثال .

ولهذه الأسباب كان لابد من البحث عن أسس موضوعية للانتقال من وضعية معينة للإحداثيات إلى أخرى من أجل الحصول على حل عاملي أكثر تعبيراً وأسهل تفسيراً، وإن أهم المعايير المستخدمة في تحديد هذه الوضعية هو تحقيق مفهوم شكل الهيكل البسيط الذي عرفناه سابقاً . ويتم ذلك عن طريق تدوير المحاور الإحداثية المتعامدة أو المائلة حتى نحصل على وضعية الهيكل البسيط، ولذلك سنميز بين نوعين من التدوير هما:

- **التدوير المتعامد :** وهو تدوير في الإحداثيات المتعامدة، وفيه يتم تدوير جملة المحاور الإحداثية المتعامدة ككل حول المبدأ (0). مع الحفاظ على التعامد فيما بينها ، وهو يعطينا حلاً متعامداً آخر يحقق بعض مواصفات الهيكل البسيط.
- **التدوير المائل:** وفيه يتم تدوير كل من المحاور الإحداثية المتعامدة إلى محاور مائلة. أو تدوير كل من المحاور المائلة إلى محاور مائلة أخرى. وعندها تتغير الزوايا بين المحاور الجديدة ، وهو يعطينا حلاً مائلاً يحقق بعض أو كل مواصفات الهيكل البسيط .

3-1 التدوير المتعامد:

لقد لاحظنا أنه عند تحليل العوامل حسب النموذج المختصر (1-20)، كنا قد افترضنا أو اشترطنا أن تكون العوامل العاملة F_p مستقلة عن بعضها البعض (أي أن معاملات الارتباط بينها معدومة $r_{F_iF_j} = 0$) ، وهذا يعني أن جملة إحداثيات تلك العوامل متعامدة . لذلك يمكننا أن نرسم الأشعة

الممثلة للمتحولات Z_j في الفضاء R^m بدلالة مساقطها على المحاور العامة F_p المتعامدة . وإن أبسط حالة لتدوير المحاور الإحداثية هي حالة التدوير حول المبدأ (0) في المستوى $F_1 0 F_2$.

مثال (3−1): لنفترض أن أمثال الحل العاملي لـ 6 متحولات Z_j بدلالة عاملين F_1 و F_2 كانت كما يلي:

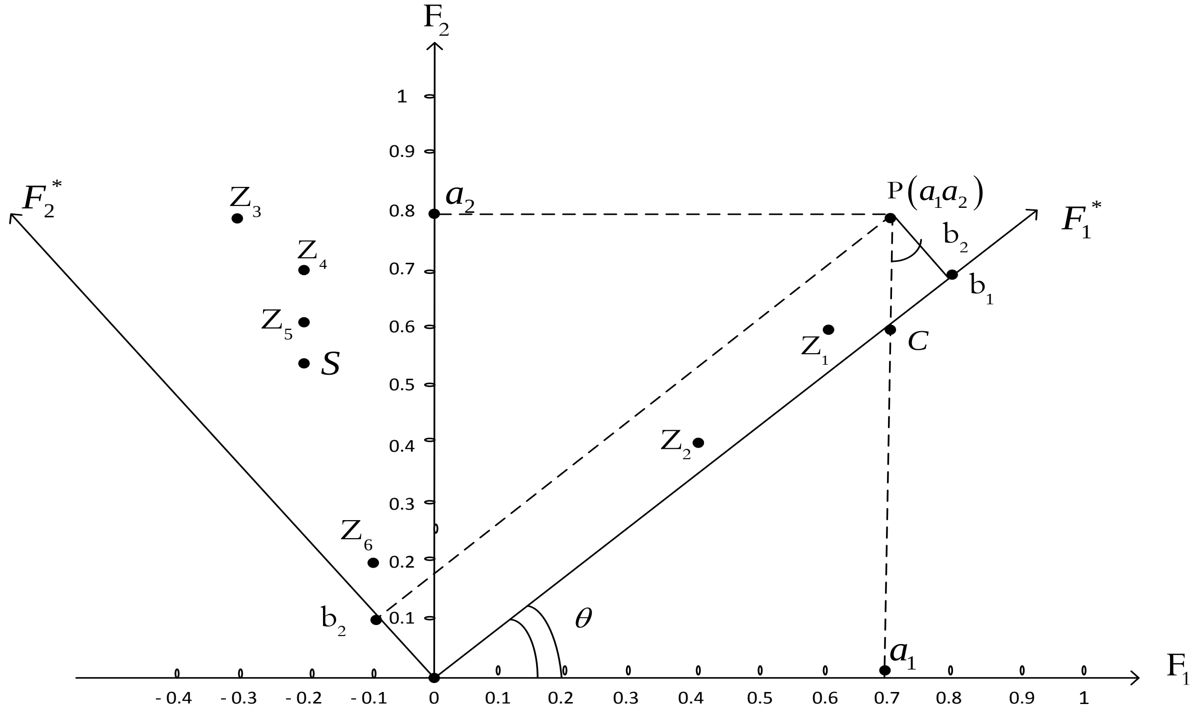
جدول (3−1): أمثال الحل العاملي للمثال الوارد في(*Uberla,p.174*) :

العوامل / المتحولات	F_1	F_2	التشاركيات h_j^2	الجداء $a_{j1}a_{j2}$
Z_1	0.60	0.60	0.72	0.36
Z_2	0.40	0.40	0.32	0.16
Z_3	−0.30	0.80	0.73	−0.24
Z_4	−0.20	0.70	0.53	−0.14
Z_5	−0.20	0.60	0.40	−0.12
Z_6	−0.10	0.20	0.05	−0.02
حصة F_p	0.7	2.05	2.75	0.00

وللتأكيد من تعامد المحورين $0F_1$ و $0F_2$ نأخذ جداء الأمثال المتقابلة فنجد من (العمود الأخير) أن:

$$\sum_{j=1}^{G} a_{j1}a_{j2} = 0 \qquad (1-3)$$

هذا يعني أن المحورين $0F_1$ و $0F_2$ متعامدان، لذلك نقوم برسم كل متحول Z_j على شكل شعاع ينتهي بنقطة P على المستوى $F_1 0 F_2$ وحسب إحداثياته (a_{j1}, a_{j2}) فنحصل على الشكل التالي:



الشكل (3−1)

وهنا نلاحظ أن عناصر هذا الحل مشتتة على المستوى $F_1 0 F_2$، لذلك سنحاول تدوير محاور الإحداثيات بعكس حركة عقارب الساعة بزاوية مقدارها θ ، بحيث نجعل جميع أو معظم النقاط تقع في الربع الأول الجديد وتصبح ذات إحداثيات موجبة ، فنحصل على المحورين الجديدين $0F_2^*$ و $0F_1^*$.

ولحساب الإحداثيات الجديدة لأية نقطة $P(b_1\ b_2)$ بدلالة الإحداثيات القديمة $(a_1 a_2)$ والزاوية θ نلاحظ من الشكل (3−1) أن الإحداثيات الجديدة تساوي:

$$b_1 = 0C + Cb_1$$
$$b_2 = CP\cos\theta$$

ومنها نحصل على العلاقتين التاليتين* (***Uberla,p.174***):

$$b_1 = a_1\cos\theta + a_2\sin\theta$$
$$b_2 = -a_1\sin\theta + a_2\, cos\,\theta \qquad (2-3)$$

والتي يمكن كتابتهما مصفوفياً على الشكل التالي :

$$[b1\ ,b2\ ] \ = [a_1\ ,a_2] * \begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix} \qquad (3-3)$$

ونرمز للمصفوفة الأخيرة بـ T ونسميها بمصفوفة التدوير بعكس حركة عقارب الساعة بزاوية مقدارها θ ونكتبها كمايلي** :

$$T = \begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix} \qquad (4-3)$$

---

* ملاحظة (1) : للبرهان على ذلك نلاحظ من الشكل (3−1) أن الإحداثيات الجديدة $(b_1\ b_2)$ تساوي:

$$b_1 = 0C + Cb_1 \qquad (a)$$
$$b_2 = CP\cos\theta \qquad (b)$$

ولحساب b_2 نجد من جهة أخرى أن: $CP = 0a_2 - a_1C = a_2 - a_1C$

ولكن $a\,,C$ ترتبط مع θ بالعلاقة : $\frac{a_1C}{0a_1} = tg\theta \quad \Rightarrow \quad a_1C = 0a_1\,tg\theta = a_1 tg\theta$

وبذلك نجد أن CP يساوي: $CP = a_2 - a_1\,tg\theta \qquad (C)$

نعوض ذلك في (b) فنحصل على أن: $b_2 = \left(a_2 - a_1\,tg\theta\right)\cos\theta = a_2\cos\theta - a_1\sin\theta$

ونكتبها حسب الترتيب لـ a_1 ثم a_2 فنحصل على أن: $b_2 = -a_1\sin\theta + a_2\cos\theta \qquad (d)$

أما بالنسبة لحساب b_1 فنجد أن $0C$ يرتبط مع θ بالعلاقة : $\frac{0a_1}{0C} = \cos\theta$

$$0C = \frac{0a_1}{\cos\theta} = \frac{a_1}{\cos\theta} \qquad (e)$$

وكذلك نجد أن Cb_1 يرتبط مع θ بالعلاقة: $\frac{Cb_1}{CP} = \sin\theta$ أي أن:

$$Cb_1 = CP\sin\theta = (a_2 - a_1 tg\theta)\sin\theta \qquad (f)$$

نعوض ذلك في (a) فنجد أن: $b_1 = \frac{a_1}{\cos\theta} + (a_2 - a_1 tg\theta)\sin\theta$

$$b_1 = \frac{a_1}{\cos\theta} + a_2\sin\theta - a_1\frac{\sin^2\theta}{\cos\theta}$$
$$b_1 = \frac{a_1}{\cos\theta} * [1 - \sin^2\theta] + a_2\sin\theta$$
$$b_1 = a_1\cos\theta + a_2\sin\theta \qquad (g)$$

وبذلك نصل على العلاقتين المستخدمتين كثيراً في عمليات التدوير المتعامد باتجاه عكس عقارب الساعة التاليتين:

$$b_1 = a_1\cos\theta + a_2\sin\theta$$
$$b_2 = -a_1\sin\theta + a_2\cos\theta \qquad (2-3)$$

** ملاحظة (2) : إذا كان التدوير مع حركة عقارب الساعة نستبدل θ بـ (θ −) ،وعندها فإن العلاقة(3−2) والمصفوفة T تأخذان الشكل التالي:

$$b_1 = a_1\cos\theta - a_2\sin\theta$$
$$b_2 = a_1\sin\theta + a_2\, cos\,\theta \quad \Rightarrow \quad T'' = \begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}$$

وإذا رمزنا لمصفوفة الحل النهائي في الإحداثيات $F_1 0 F_2$ بالرمز A ولمصفوفة الحل النهائي في الإحداثيات الجديدة $0F_2^*$ و $0F_1^*$ بالرمز B . فإننا نجد أن هاتين المصفوفتين مرتبطتان بواسطة T بالعلاقة التالية :

$$B = A * T \qquad (5-3)$$

ولتطبيق ذلك على الحل العاملي في الجدول (3−1) ، نفترض أن زاوية التدوير $\theta = 30^o$ ، فنجد أن $\cos(30) = 0.866 \;,\; \sin(30) = 0.50$، وإن مصفوفة التدوير T تأخذ الشكل التالي :

$$T = \begin{bmatrix} \cos(30) & -\sin(30) \\ \sin(30) & \cos(30) \end{bmatrix} = \begin{bmatrix} 0.866 & -0.500 \\ 0.500 & 0.866 \end{bmatrix}$$

وبذلك يمكننا حساب الإحداثيات الجديدة للحل النهائي السابق كما يلي:

$$B_r = \begin{array}{c} \begin{array}{cc} F_1 & F_2 \end{array} \\ \begin{bmatrix} 0.60 & 0.60 \\ 0.40 & 0.40 \\ -2.30 & 0.80 \\ -0.20 & 0.70 \\ -0.20 & 0.60 \\ -0.10 & 0.20 \end{bmatrix} \end{array} \begin{bmatrix} 0.866 & -0.500 \\ 0.500 & 0.866 \end{bmatrix} = \begin{array}{c} \begin{array}{cc} F_1^* & F_2^* \end{array} \\ \begin{bmatrix} 0.8196 & 0.2196 \\ 0.5464 & 0.1464 \\ 0.1402 & 0.8428 \\ 0.1768 & 0.7062 \\ 0.1268 & 0.6196 \\ \underline{0.0134} & \underline{0.2232} \end{bmatrix} \\ \begin{array}{cc} 1.0375 & 1.7129 \end{array} \end{array}$$

وهكذا نستنتج أن عملية التدوير أعطتنا صورة جديدة للحل النهائي تتميز بما يلي:

1− إن جميع أمثال المتحولات Z_j أصبحت موجبة .

2− إن أمثال Z_1 و Z_2 ازدادت على $0F_1^*$ وتناقصت على $0F_2^*$ وذلك لأن المساقط تغيرت على هذين المحورين .

3− إن أمثال المتحولات Z_3 و Z_4 و Z_5 و Z_6 قد تغيرت فأصبحت أمثالها على $0F_2^*$ أكبر مما كانت عليه كما إن أمثالها على $0F_1^*$ أصبحت صغيرة .

4− إن التشاركية الاجمالية لهذين العاملين بقيت ثابتة (2.75)، ولكن حصة العامل الأول $0F_1^*$ ازدادت إلى (1.0375) ونقصت حصة العامل الثاني $0F_2^*$ وأصبحت (1.7129) .

ملاحظة: إذا أردنا أن نجري عملية التدوير السابقة بحيث يمر المحور الجديد $0F_1^*$ من النقطة التي تمثل المتحول الأول Z_1 والتي إحداثياتها (0.60، 0.60) علينا أن نقوم بحساب زاوية التدوير θ من العلاقة:

$$tg\theta = \frac{a_2}{a_1} = \frac{0.60}{0.60} = 1$$

أي أن زاوية التدوير التي تجعل $0F_1^*$ يمر من Z_1 تساوي $\theta = 45^o$ ، ومنها نجد أن مصفوفة التدوير الجديدة T تساوي:

$$T = \begin{bmatrix} \cos(45) & -\sin(45) \\ \sin(45) & \cos(45) \end{bmatrix} = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}$$

وإن صورة الحل الجديد الذي يمر من Z_1 تحسب من العلاقة :

$$B_2 = \begin{bmatrix} 0.60 & 0.60 \\ 0.40 & 0.40 \\ -0.30 & 0.80 \\ -0.20 & 0.70 \\ -0.20 & 0.60 \\ -0.10 & 0.20 \end{bmatrix} \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} = \begin{bmatrix} 0.8485 & 0 \\ 0.5657 & 0 \\ 0.3536 & 0.7778 \\ 0.3536 & 0.6364 \\ 0.2828 & 0.5657 \\ \underline{0.0707} & \underline{0.2121} \end{bmatrix}$$
$$\begin{matrix} 1.375 & 1.376 \end{matrix}$$

وهو يختلف عن الحل السابق ويتميز عنه بما يلي:

1- إن أمثال المتحولين Z_1 و Z_2 ازدادت على المحور $0F_1^*$ وانعدمت على $0F_2^*$ المتعامد مع $0F_1^*$ (لأن OF1 يمر من النقطتين Z_1 و Z_2 ، اللتين تقعان على استقامة واحدة) .

2- إن أمثال المتحولات Z_3 و Z_4 و Z_5 و Z_6 ازدادت قليلاً على $0F_1^*$ وتراجعت قليلاً على $0F_2^*$ بسبب تغير موقع $0F_2^*$ بالنسبة للنقاط التي تمثل هذه المتحولات .

3- إن قيمة التشاركية الإجمالية بقيت ثابتة (2.75) ولكن حصة العامل الأول F_1^* أصبحت (1.375) وحصة العامل الثاني F_2^* أصبحت (1.376) .

4- إن الشكل الجديد أصبح قريباً من شكل الهيكل البسيط لأنه أصبح يحقق بعض شروطه (راجع شروط الهيكل البسيط) .

والسؤال الآن هو: هل يمكن تحويل الحل المفروض إلى شكل أفضل من الشكل السابق .

والجواب هو: نعم يمكن ذلك بتدوير الحل المفروض بزاوية θ , بحيث يمر المحور الثاني $0F_2^*$ بين تجمع النقاط التي تمثل المتحولات Z_3 و Z_4 و Z_5 و Z_6 ، وإن أفضل وضع لـ $0F_2^*$ هو الوضع الذي يمر من مركز ثقل ذلك التجمع S والذي يتميز بأن إحداثياته تحسب من متوسطات إحداثيات هذه المتحولات على المحورين القديمين، أي أن إحداثيات S تساوي :

$$S\left(\frac{1}{4}\sum_{j=3}^{6} a_{j1}\ ,\frac{1}{4}\sum_{j=3}^{6} a_{j2}\right) = S(-0.20\,, 0.575)$$

ولحساب الزاوية التي يصنعها $0S$ مع $0F_1$ نحسب ظلها من العلاقة :

$$tg\theta = -\frac{0.575}{0.20} = -2.875$$

ومن جدول التوابع المثلثية نجد أن الزاوية التي يصنعها $0S$ مع $0F_1$ تساوي: $\theta_1 = 109.189^o$

وبذلك تكون الزاوية التي يصنعها $0S$ مع $0F_2$ تساوي: $\theta_2 = 109.189 - 90 = 19.189^o$

وهي زاوية التدوير المناسبة للمحاور، حتى يمر المحور الجديد $0F_2^*$ من مركز ثقل تجمع المتحولات Z_3 و Z_4 و Z_5 و Z_6، ومنه نجد أن:

$$\cos(19.189) = 0.9444 \qquad \sin(19.189) = 0.3287$$

وإن مصفوفة التحويل T أصبحت تساوي:

$$T = \begin{bmatrix} 0.9444 & -0.3287 \\ 0.3287 & 0.9444 \end{bmatrix}$$

ولإيجاد الحل الجديد في المحاور الجديدة نحسب الجداء $B = A * T$ فنحصل على أن:

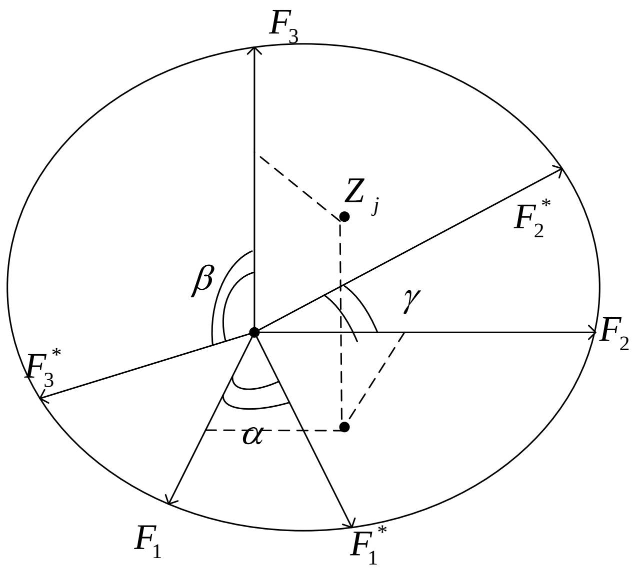
$$B_3 = \begin{array}{c} \begin{array}{cc} F_1 & F_2 \end{array} \\ \begin{bmatrix} 0.60 & 0.60 \\ 0.40 & 0.40 \\ -0.30 & 0.80 \\ -0.20 & 0.70 \\ -0.20 & 0.60 \\ -0.10 & 0.20 \end{bmatrix} \end{array} \begin{bmatrix} 0.9444 & -0.3287 \\ 0.3287 & 0.9444 \end{bmatrix} = \begin{array}{c} \begin{array}{cc} F_1^* & F_2^* \end{array} \\ \begin{bmatrix} 0.76386 & 0.36942 \\ 0.50924 & 0.24628 \\ 0.02036 & 0.85413 \\ 0.04121 & 0.72682 \\ 0.00834 & 0.63238 \\ -0.0287 & 0.22175 \end{bmatrix} \end{array}$$

وهذا يعني أن أمثال (مساقط) المتحولات Z_3 وZ_4 وZ_5 وZ_6 قد ازدادت على المحور الجديد $0F_2^*$ وتناقصت على $0F_1^*$. بينما تناقصت أمثال Z_1 وZ_2 قليلاً على $0F_1^*$ وتزايدت على $0F_2^*$. واحتفظت التشاركية الاجمالية بقيمتها (2.75) .

ملاحظة: حتى تصبح الصورة الأخيرة أكثر تعبيراً وأسهل تفسيراً ، يجب أن نجعل المحور الأول $0F_1^*$ يمر من مركز ثقل التجمع الأول المؤلف من النقطتين Z_2, Z_1 ، وأن نجعل المحور الثاني يمر من مركز ثقل التجمع الثاني المؤلف من Z_3 وZ_4 وZ_5 وZ_6، وعندها نجد أن ذلك يزيد بآن واحد قيم مساقط (أمثال) متحولات كل تجمع على المحور المار به، وينقص من مساقطه على المحور المقابل له .

ولكن تحقيق هذه الرغبة يخل بشروط تعامد الإحداثيات ويجعلنا ننتقل من الإحداثيات المتعامدة إلى الإحداثيات المائلة، لأن المحوين الجديدين $0F_1^*$ و$0F_2^*$ يصبحان غير متعامدين. وهذا أمر يحتاج إلى معالجة خاصة وسنتعرض لها لاحقاً .

أما عندما يكون الحل النهائي ممثلاً بثلاثة عوامل F_1 وF_2 وF_3 ، فإن عمليات التدوير تجري في فضاء العوامل العامة R^3 وفي كل مستو على حدة ، كما في الشكل (3−2) وكما يلي :



الشكل (3−2): عمليات التدوير في R^3

1- نجري التحويل الأول في المستوى $F_1\ 0F_2$ بتدوير الإحداثيات بزاوية $\propto$ حول المحور المتعامد $0F_3$ (كما يدور قرص CD حول محوره) وذلك باستخدام المصفوفة T_{12} التالية :

$$T_{12} = \begin{bmatrix} \cos\propto & -\sin\propto & 0 \\ \sin\propto & \cos\propto & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (6-3)$$

2- نجري التحويل الثاني في المستوى $F_1\ 0F_3$ بتدوير الإحداثيات بزاوية β حول المحور المتعامد $0F_2$ (كما تدور عجلة السيارة حول محورها) وذلك باستخدام المصفوفة T_{13} التالية :

$$T_{13} = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \qquad (7-3)$$

3- ثم نجري التحويل الثالث في المستوى $F_2\ 0F_3$ بتدوير الإحداثيات بزاوية γ حول المحور المتعامد $0F_1$ (كما تدور المروحة في السيارة حول محورها)، وذلك باستخدام مصفوفة التحويل T_{23} التالية:

$$T_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix} \qquad (8-3)$$

ويمكننا إجراء هذه التحويلات الثلاثة دفعة واحدة باستخدام المصفوفة الناتجة عن جداء هذه المصفوفات ونرمز لها بـ T والتي تساوي:

$$T = T_{12} * T_{13} * T_{23} \qquad (9-3)$$

وهكذا نحصل على المصفوفة الإجمالية T لتدوير الإحداثيات حول محاورها المتعامدة الثلاثة . وتكون مرتبتها 3*3، وإذا رمزنا للحل الجديد في الإحداثيات الجديدة بالرمز B فإنه يساوي:

$$B_{n*3} = A_{n*3} * T_{3*3} = A * T_{12} * T_{13} * T_{23} \qquad (10-3)$$

وإذا رمزنا لعناصر المصفوفة T بالرمز λ_{ij} فإنه يمكننا كتابة الحل النهائي B على الشكل التالي:

$$B_{n*3} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & b_{n3} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{bmatrix} \qquad (11-3)$$

ومنها يمكننا أن نحصل على عناصر الحل النهائي B وأن نكتب الصورة النهائية بدلالة المحاور الجديدة F^* على الشكل التالي:

$$Z^* = BF^* \qquad (12-3)$$

كما يمكننا كتابة وحساب أي متحول Z_j بدلالة الأمثال الجديدة من (3-12) على الشكل التالي:

$$Z_j^* = b_{j1}F_1^* + b_{j2}F_2^* + b_{j3}F_3^* \qquad : (j: 123 \dots n): \text{حيث} \qquad (12\,a-3)$$

ومن جهة أخرى نجد أن العلاقة (3-11) تساعدنا على حساب الأمثال الجديدة b_{j1} و b_{j2} و b_{j3} لكل متحول Z_j في الإحداثيات الجديدة $0F_1^*$ و $0F_2^*$ و $0F_3^*$ كما يلي:

$$b_{j1} = a_{j1}\lambda_{11} + a_{j2}\lambda_{21} + a_{j3}\lambda_{31}$$
$$b_{j2} = a_{j1}\lambda_{12} + a_{j2}\lambda_{22} + a_{j3}\lambda_{32} \qquad (13-3)$$
$$b_{j3} = a_{j1}\lambda_{13} + a_{j2}\lambda_{23} + a_{j3}\lambda_{33}$$

حيث: $(j: 123 \ldots n)$

ولإظهار بعض خواص المصفوفة الإجمالية T_{3*3} نكتبها على الشكل التالي:

$$T = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{bmatrix} \qquad (14-3)$$

وإن أهم خواصها عندما تكون الإحداثيات متعامدة هي:

1- إنها مصفوفة متعامدة، أي أنها تتألف من أشعة عمودية متعامدة، وهذا يعني أن الجداء السلمي لكل عمودين منها يساوي الصفر .

2- إن طول كل شعاع عمودي فيها يساوي الواحد (أشعة واحدية)، أي أن مجموع مربعات عناصر أي عمود منها يساوي الواحد .

إن هذه الخواص تعني أن عناصر الأعمدة في T يجب أن تحقق الشروط التالية:

أ- شروط التعامد: إن يكون الجداء السلمي لأي عمودين فيها مساوياً للصفر:

$$< \lambda_1 , \lambda_2 >= \lambda_{11}\lambda_{12} + \lambda_{21}\lambda_{22} + \lambda_{31}\lambda_{32} = 0$$
$$< \lambda_1 , \lambda_3 >= \lambda_{11}\lambda_{13} + \lambda_{21}\lambda_{23} + \lambda_{31}\lambda_{33} = 0 \qquad (15-3)$$
$$< \lambda_2 , \lambda_3 >= \lambda_{21}\lambda_{13} + \lambda_{21}\lambda_{23} + \lambda_{23}\lambda_{33} = 0$$

ب- شروط الواحدية : إن يكون طول أو مربع طول كل عمود فيها مساوياً للواحد:

$$\lambda_{11}^2 + \lambda_{21}^2 + \lambda_{31}^2 = 1$$
$$\lambda_{12}^2 + \lambda_{22}^2 + \lambda_{32}^2 = 1 \qquad (16-3)$$
$$\lambda_{13}^2 + \lambda_{23}^2 + \lambda_{33}^2 = 1$$

وهذه الشروط تشكل 6 معادلات وتتضمن 9 مجاهيل هي λ_{ij} ، أي أنها تترك لنا 3 درجات حرية (3-6-9) للتدوير في الفضاء الثلاثي R^3 ، وهي عمليات التدوير التي أشرنا إليها في المستويات الثلاثة حول المحاور المعامدة لها كما في الشكل (3-2)السابق.

3-2 طرائق التدوير في الإحداثيات المتعامدة [Harman, P.317] :

لقد لاحظنا أن المشكلة الرئيسية في عمليات التدوير، هي كيفية حساب الزاوية المناسبة للتدوير، بحيث نحصل على حل يحقق شروط الهيكل البسيط المذكورة سابقاً . ولمعالجة هذه القضية لاحظنا من الأمثلة السابقة أن أفضل وضع لتمثيل نقطة ما Z_j (أو مجموعة من النقاط) هو أن يمر أحد المحاور فيها (أو في مركزها)، وكذلك وجدنا أنه عند تدوير أحد المحاور ليقترب من تلك النقطة فإن إحداثياتها على ذلك المحور ستزيد ولكن إحداثياتها على المحاور الأخرى ستتناقص. وإن الجداء السلمي للإحداثيات الجديدة لأي نقطتين مختلفتين يبدأ بالتناقص، واعتماداً على هذه الملاحظة توصل (فيرجيوسن) إلى تشكيل مقياس

اقتصادي للتدوير مؤلف من مجموع جداءات إحداثيات مجموعة النقاط المجتمعة، ولكن بما أن هذه المجاميع قد تأخذ قيماً سالبة أو موجبة، فقد اقترح أن يتم استبدال المجموع العادي بمجموع مربعات جداء إحداثيات تلك النقاط مأخوذاً على جميع العوامل m وعلى جميع المتحولات n . وهو يأخذ الشكل التالي:

$$\sum_{j=1}^{n}\sum_{p<q=1}^{m}\left(b_{jp}*b_{jq}\right)^{2}\Rightarrow \min \qquad (17-3)$$

وهذا يتألف من $\frac{m(m-1)}{2}$ مجموعاً مؤلفاً من n زوجاً من الاحداثيات الجديدة للمتحولات Z_j .

ولمعالجة هذه القضية ننطلق من العلاقة التي تربط بين الحل قبل التدوير A (الحل القبلي) والحل بعد التدوير B (الحل البعدي) بواسطة مصفوفة التحويل T في الإحداثيات المتعامدة كما رأينا سابقاً وهي:

$$B=A*T \qquad (18-3)$$

أي أن المصفوفة T تحول المصفوفة A إلى المصفوفة B، مع شرط أن تبقي قيمة التشاركية h_j^2 الخاصة بكل متحول Z_j ثابتة (= مربع طوله) ، وهذا يعني رياضياً أن التشاركيات الفردية قبل وبعد التدوير يجب أن تحقق العلاقة:

$$\sum_{p=1}^{m}b_{jp}^{2}=\sum_{p=1}^{m}a_{jp}^{2}=h_j^2=const \qquad (j:123\quad n) \qquad (19-3)$$

وإذا أخذنا مربعات هذه التشاركيات فإنه يجب أن يبقى ثابتاً أيضاً $\left(\sum b_{jp}^{2}\right)^{2}=const$ وإن مفكوكة يجب أن يساوي مقداراً ثابتاً، أي أن:

$$\left(\sum b_{jp}^{2}\right)^{2}=\sum_{p=1}^{m}b_{jp}^{4}+2*\sum_{p<q=1}^{m}b_{jp}^{2}b_{jq}^{2}=const \qquad (20-3)$$

(21–3) وذلك من أجل كل متحول Z_j (أي أن $j:123\quad n$)

نقوم الآن بأخذ مجموع الطرف الأيمن على جميع المتحولات Z_j فنحصل على التركيب الذي يجب أن يساوي مقداراً ثابتاً، كما يلي :

$$\sum_{j=1}^{n}\sum_{p=1}^{m}b_{jp}^{4}+2*\sum_{j=1}^{n}\sum_{p<q=1}^{m}\left(b_{jp}^{2}b_{jq}^{2}\right)=const \qquad (22-3)$$

وهنا نلاحظ أن مجموع الحدين في (3–22) يساوي مقداراً ثابتاً بشكل دائم لأنه أصبح مرتبطاً بالتشاركية الإجمالية . أي أنه إذا ازداد أحدهما يؤدي إلى تناقص الآخر، وهكذا نجد أن عملية تحويل المصفوفة A إلى المصفوفة B وإيجاد زاوية التدوير، يمكن أن تستند إلى مدخلين مترابطين وذلك حسب الحد الذي يتم استخدامه في حل هذه المسألة. وبنتيجة المقارنة الرياضية بين هذين الحدين مع المقياس(3–17) ، نجد أنه إذا استخدمنا الحد الثاني فإن الأمر يقتضي جعله أصغر ما يمكن، أما إذا استخدمنا الحد الأول فإن الأمر يقتضي جعله أكبر ما يمكن.

وانطلاقاً من ذلك فإن (فيرجيوسن) نفسه اقترح أن يتم تعظيم الحد الأول ورمز له بـ Q، وأصبح هدفه عند التدوير هو تعظيم المقدار التابع التالي:

$$Q = \sum_{j=1}^{n}\sum_{p=1}^{m} b_{jp}^{4} \Rightarrow \max \qquad (23-3)$$

وهنا نلاحظ أن المقدار Q يرتبط بقيم الأمثال العاملية الجديدة، التي بدورها تتأثر بمكان توضع المحاور الإحداثية بعد التدوير . ولهذا يجب تدوير المحاور بحيث يأخذ المقدار Q أكبر قيمة ممكنة .

ومن الناحية النظرية نعلم أن الحد الأعلى لـ Q يساوي n، وهو عبارة عن قيمته ما تكون عند درجة التعقيد لكل من المتحولات مساوية للواحد (كل متحول Z_j يمثل بعامل واحد F_p) .

أي أن Q يبلغ حده الأقصى عندما تصبح درجة تعقيد كل من المتحولات أصغر ما يمكن. وعندها يمكننا أي نقول أن مجموعة المتحولات موزعة على شكل يحقق بعض شروط شكل الهيكل البسيط .

3-2-1: طريقة (كوارتي ماكس Quartimax)

وسميت بذلك لأنها تهدف إلى تعظيم مجموع الأسس الرابعة لأمثال المتحولات Z_j على العوامل الجديدة F_p^*، وتعتمد هذه الطريقة بشكل أساسي على تعظيم المقدار Q، ولإيجاد زاوية التدوير في أحد المستويات، وليكن المستوى F_p0F_q المعرف بالمحورين $0F_p$ و $0F_q$ فقط، نرمز لها بـ φ_{pq} ونلاحظ أن المقدار Q يأخذ في ذلك المستوي الشكل المختصر التالي :

$$Q_{pq} = \sum_{j=1}^{n}\left(b_{jp}^{4} + b_{jq}^{4}\right) \qquad (24-3)$$

ومعلوم إن الإحداثيين الجديدين b_{jp} و b_{jq} يعطيان بدلالة الإحداثيين القديمين a_{jp} و a_{jq} والتوابع المثلثية لزاوية الدوران φ_{pq} (أو اختصاراً φ) بالعلاقتين التاليتين (انظر العلاقة (3-2)) :

$$\begin{aligned} b_{jp} &= a_{jp}\cos\varphi + a_{jp}\sin\varphi \\ b_{jq} &= -a_{jq}\sin\varphi + a_{jq}\cos\varphi \end{aligned} \qquad (25-3)$$

وبالتعويض في العلاقة (3-24) نحصل على المقدار Q التابع للزاوية φ فقط كما يلي:

$$Q_{pq}(\varphi) = \sum_{j=1}^{n}\left[\left(a_{jp}\cos\varphi + a_{jp}\sin\varphi\right)^{4} + \left(-a_{jq}\sin\varphi + a_{jq}\cos\varphi\right)^{4}\right] \qquad (26-3)$$

والآن يصبح هدفنا إيجاد تلك الزاوية φ_{pq} (اي φ) لتدوير أي محورين كالمحورين $0F_p$ و $0F_q$ والتي تجعل قيمة التابع $Q_{pq}(\varphi)$ أكبر ما يمكن، وعندما نعطي p وq قيمها الممكنة التالية:

$$P: 1\ 2\ 3 \ldots\ldots (m-1) \qquad q: p+1\,, p+2\,, \ldots\ldots m$$

فإن المصفوفة النهائية للإحداثيات الجديدة B ستساوي جداء المصفوفة الأساسية A في مصفوفات التحويل المتتالية ولجميع الأزواج الممكنة والتي تبلغ $\frac{m(m-1)}{2}$ زوجاً أو تدويراً كما يلي:

$$B = A * T_{12} * T_{13} \ldots\ldots T_{pq} \ldots\ldots T_{(m-1)m} \qquad (27-3)$$

وعندما ننجز جميع هذه التدويرات نكون قد أنجزنا ما يسمى (دورة كاملة circle)، وهنا نلاحظ أنه في نهاية كل دورة كاملة ستصبح قيمة المؤشر Q أكبر مما كانت عليه في الدورة السابقة وتقترب من حدها الأعلى n، ومن جهة أخرى ولتحديد قيمة الزاوية φ لأي عملية تدويرية T_{pq} في المستوى $F_p\ 0F_q$ ، والتي تجعل قيمة التابع $Q_{pq}(\varphi)$ أكبر ما يمكن نقوم بما يلي:

أ- نشتق العلاقة (3−26) بالنسبة للزاوية φ .

ب- نضع هذا المشتق مساوياً للصفر فنحصل على : $\frac{dQ(\varphi)}{d\varphi}=0$.

ج- نقوم بحل معادلة المشتق بالسنبة لـ φ . وبعد حسابات مطولة نحصل على أن :

$$tg(4\varphi)=\frac{2*\sum_{j=1}^{n}\left(2*a_{jp}*a_{jq}\right)\left(a_{jp}^2-a_{jq}^2\right)}{\sum_{j=1}^{n}\left[\left(a_{jp}^2-a_{jq}^2\right)^2-\left(2a_{jp}*a_{jq}\right)^2\right]}=\frac{v}{\delta} \qquad (28-3)$$

حيث v هي القيمة العددية للبسط ، و δ هي القيمة العددية للمقام .

ويمكننا أن نكتب المعادلة (3−28) كما يلي: $tg(4\varphi)=\frac{\sin(4\varphi)}{\cos(4\varphi)}=\frac{v}{\delta}$

أي أن معادلة المشتق الأول تساوي :

$$\frac{dQ}{d\varphi}=v*\cos(4\varphi)-\delta*\sin(4\varphi)=0 \qquad (29-3)$$

وبما أن دورة كل من $\sin(4\varphi)$ و $\cos(4\varphi)$ تساوي $\frac{\pi}{2}$ ، لذلك يجب علينا أن نبحث عن حلول للزاوية φ في المجال $[0\,,90]$. وتدل التجارب أن أفضل الخيارات لحلول φ هي الواقعة في المجال $[-45\,,45]$. وعند تحديد قيمة φ التي تجعل Q أكبر ما يمكن، يجب أن نتحقق من شرط سالبية المشتق الثاني (عند القيمة المحددة لـ φ) أي يجب أن يتحقق (بعد الاختصار على 4) الشرط التالي :

$$\frac{1}{4}\frac{d^2Q}{d^2\varphi}=-\delta\cos(4\varphi)-v\sin(4\varphi)<0 \qquad (30-3)$$

ولتطوير هذا الشرط نحسب $\cos(4\varphi)$ من (3−29) ونعوضها في (3−30)، وبعد الاصلاح نجد أن الشرط (3−30) يصبح كما يلي:

$$-\frac{\delta^2+v^2}{v}\sin(4\varphi)<0 \qquad (31-3)$$

وبما أن (δ^2+v^2) مقدار موجب فإن الشرط الأخير يصبح كما يلي :

$$\frac{\sin(4\varphi)}{v}>0 \qquad \Leftrightarrow\ \frac{1}{v}tg(4\varphi)*\cos(4\varphi)>0 \qquad (32-3)$$

وهذا يعني أن إشارة $\sin(4\varphi)$ يجب أن تكون مثل إشارة v، وبما أن كل منهما يمكن أن يكون موجباً أو سالباً، فإنه سيكون لدينا أربعة إمكانيات (بدائل)، وكل إمكانية تقابل الربع الذي تقع فيه الزاوية (4φ) وذلك حسب توافق إشارات v و δ و $tg(4\varphi)$ وكذلك $\cos(4\varphi)$ وهي المعروضة في الجدول التالي:

جدول (3−2): كيفية تحديد حساب زاوية التدوير:

القطاع الذي يقع فيه زاوية التدوير φ	الربع الذي تقع فيه الزاوية (4φ)	إشارة $\cos(4\varphi)$	الإشارات في العلاقة (3−32)		
			اشارة $tg(4\varphi)$	إشارة المقام δ	إشارة البسط v
$0^\circ < \varphi < 22.5^\circ$	الأول $0^\circ < 4\varphi < 90^\circ$	+	+	+	+
$22.5^\circ < \varphi < 45^\circ$	الثاني $90^\circ < 4\varphi < 180^\circ$	−	−	−	+
$-45^\circ < \varphi < -22.5^\circ$	الثالث $-180^\circ < 4\varphi < -90^\circ$	−	+	−	−
$-22.5^\circ < \varphi < 0^\circ$	الرابع $-90^\circ < 4\varphi < 0^\circ$	+	−	+	−

وبناء على إشارتي كل من البسط v والمقام δ يتم تحديد الربع التي تقع فيه الزاوية (4φ) ضمنه من العلاقة: $4\varphi = arctg\left(\frac{v}{\delta}\right)$، وبعدها يتم حساب φ وتعويضها في مصفوفة التحويل T، ثم نقوم بضرب المصفوفة A من اليمين بـ T فنحصل على الحل الجديد $B = A * T$ ، وهو سيكون الحل النهائي بعد التدوير الأول في المستوي T_{pq} .

ملاحظة: حتى نتأكد من حصولنا على أكبر قيمة للتابع $Q_{pq}(\varphi)$، علينا أن نتأكد من تحقق الشرط (3−32)

مثال (3−2): لنأخذ الحل الأولي لثمانية متحولات Z_j بدلالة عاملين F_1 و F_2 ، المبين في العمودين 2 و 3 من الجدول (3−3) ، ولإجراء عملية التدوير على هذا الحل نقوم بما يلي:

نحسب قيمتي البسط والمقام الواردتين في العلاقة (3−28) من خلال حساب عناصر كل منهما على بيانات المثال التالي، ولذلك نعد الجدول التالي:

جدول (3−3): الحسابات اللازمة لحساب $tg(4\varphi)$:

المتحولات	الحل الأولي A		المربعات		الجداء	الفروقات	الحل النهائي B		الجداء × الفروقات
	a_{j1}	a_{j2}	a_{j1}^2	a_{j2}^2	$2\ a_{j1}^2 * a_{j2}^2$	$a_{j1}^2 - a_{j2}^2$	b_{j1}	b_{j2}	
Z_1	0.830	−0.396	0.6889	0.1568	−0.6574	0.5321	0.899	0.196	−0.3498
Z_2	0.818	−0.469	0.6691	0.2200	−0.7673	0.4491	0.934	0.131	−0.3446
Z_3	0.777	−0.470	0.6037	0.2209	−0.7304	0.3828	0.902	0.105	−0.2796
Z_4	0.798	−0.401	0.6368	0.1608	−0.6400	0.4760	0.876	0.172	−0.3046
Z_5	0.786	0.500	0.6178	0.2500	0.7860	0.3678	0.315	0.877	0.2891
Z_6	0.672	0.458	0.4516	0.2098	0.6156	0.2418	0.250	0.774	0.1489
Z_7	0.594	0.444	0.3528	0.1971	0.5275	0.1557	0.197	0.715	0.0821
Z_8	0.647	0.333	0.4186	0.1109	0.4309	0.3077	0.307	0.660	0.1326
المجموع	4.4394	1.5263	2.5776	0.3053	3.4247	1.1706	3.5563	2.4112	−0.6260

ومن نتائج هذه الحسابات نقوم بحساب البسيط v والمقام δ الواردين في العلاقة (3−28)، وبعد إجراء جداء عناصر عمودي الجداء والفروقات (الموجودين في الجدول (3−3)) نجد أن قيمة v تساوي:

$$v = 2 * \sum_{j=1}^{n} \left(2a_{j1} * a_{j2}\right)\left(a_{j1}^2 - a_{j2}^2\right) = 2(-0.6260) = -1.2520$$

ومن السطر الأخير المتضمن مجموع المربعات نجد أن:

$$\delta = \sum \left[\left(a_{j1}^2 - a_{j2}^2 \right)^2 - \left(2a_{j1} * a_{j2} \right)^2 \right] = (1.1706) - 3.4247 = -2.2541$$

ثم نعوض ذلك في العلاقة (3−28) فنجد أن:

$$tg(4\varphi) = \frac{v}{\delta} = \frac{-1.2520}{-2.2541} = 0.5554$$

وبما أن البسط $v < 0$. وإن المقام $\delta < 0$ ، فإن الزاوية (4φ) حسب معطيات السطر الثالث من الجدول (3−2) تقع في الربع الثالث، أي أنها محصورة في المجال: $-180 < 4\varphi < -90^{\circ}$، وهذا يعني أن φ محصورة في المجال $-45^{\circ} < \varphi < -22.5^{\circ}$.

وبالعودة إلى جداول التوابع المثلثية أو باستخدام الآلات الحاسبة نجد أن الزاوية (4φ) تساوي :

$$4\varphi = arctg\ (0.5554) - 180 = 29.05 - 180 = -150.95^{\circ}$$

وهكذا نجد أن زاوية الدوران المطلوبة φ تساوي: $\varphi = -\frac{150.95}{4} = -37.7375$.

وحتى نحافظ على أن يكون إتجاه التدوير بعكس حركة عقارب الساعة في المستوى $F_1\,0F_2$ ، أي يكون ملائماً للمصفوفة (3−4). فإننا نترك القيمة السالبة للزاوية φ (للدلالة على اتجاه حركة الدوران المعاكسة)، ونقوم بحساب عناصر مصفوفة التحويل (3−4). وعندها نجد أن:

$$\cos\varphi = \cos(-37.74) = 0.7908 \qquad sin\,\varphi = sin(-37.74) = -.6120$$

وبذلك نجد أن مصفوفة التحويل T تأخذ الشكل التالي:

$$T = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} = \begin{bmatrix} 0.7908 & 0.6120 \\ -0.6120 & 0.7908 \end{bmatrix}$$

ولحساب عناصر الحل النهائي B نقوم بتطبيق العلاقة: $B = A * T$، ومنها حصلنا على الحل B ووضعناه في العمودين الأخيرين من الجدول (3−3)، وهو يحافظ على ثبات قيمة التشاركية الاجمالية والمساوية في الحلين (5.9657)، ويجعل قيمة المؤشر Q أكبر ما يمكن($Q = 4.0991$)، مقابل القيمة 2.883 للحل الأولي .

3−2−2: طريقة فري ماكس *Varimax*

لقد لاحظنا من الفقرة السابقة أن طريقة (كوارتي ماكس) كانت تهدف إلى تبسيط التعبير في كل سطر من أسطر المصفوفة A (أي تبسيط التعبير عن كل متحول Z_j)، ولكن طريقة (فري ماكس) التي وضعها *(Kaiser)* تهدف إلى تلبية شروط الهيكل البسيط من خلال الضغط على الأعمدة (أي من خلال الضغط على العوامل F_p وعلى تبايناتها .

ولقد عرف *(Kaiser)* ببساطة تباين العامل F_p من خلال تباين مربعات أمثاله، فإذا كان عمود أمثال العامل F_p على المتحولات $Z_1, Z_2, Z_3, \ldots\ldots Z_n$ في الحل النهائي كما يلي:

$$b_{1p}, b_{2p}, b_{3p}, \ldots\ldots b_{np} \qquad (33-3)$$

فإن قيم هذه الأمثال كمعاملات ارتباط تحقق المتراجحة : $-1 \leq b_{jp} \leq +1$، وإن مربع كل منها يقع في المجال $]0,1[$ ،وهي تشكل أبسط شكل لها عندما يأخذ بعضها قيماً قريبة من الواحد ويأخذ بعضها الآخر قيماً قريبة من الصفر. (أي عندما تأخذ b_{jp}^2 قيماً متطرفة)، وهذا يعني أن تباينها (أي تباين b_{jp}^2) في هذه الحالة سيكون أكبر ما يمكن، وإذا رمزنا لتباين مربعات الأمثال العمودية b_{jp} للعامل F_p مأخوذاً على جميع المتحولات بالرمز S_p^2 . نجد أنه يساوي :

$$S_p^2 = \frac{1}{n}\sum_{j=1}^{n}\left(b_{jp}^2 - \bar{b}_{jp}^2\right)^2 \qquad (p: 123 \ldots . m) \qquad \mathbf{(34-3)}$$

وإن مفكوكه يساوي :

$$S_p^2 = \frac{1}{n}\sum_{j=1}^{n} b_{jp}^4 - \frac{1}{n^2}\left(\sum_{j=1}^{N} b_{jp}^2\right)^2 \qquad (p: 123 \ldots . m) \qquad \mathbf{(35-3)}$$

وحتى تشمل هذه العلاقة جميع العوامل الداخلة في الصورة العاملية، نأخذ مجموع المقدار S_p^2 على جميع العوامل F_p، فنحصل على المقدار التالي:

$$S^2 = \sum_{p=1}^{m} S_p^2 = \frac{1}{n}\sum_{p=1}^{m}\sum_{j=1}^{n} b_{jp}^4 - \frac{1}{n^2}\sum_{p=1}^{m}\left(\sum_{j=1}^{n} b_{jp}^2\right)^2 \qquad \mathbf{(36-3)}$$

وهكذا توصل (Kaiser) إلى أن عناصر الحل النهائي b_{jp} التي تعطينا الهيكل البسيط ، يجب أن تجعل المقدار S^2 أكبر ما يمكن .

ولكن التجارب العملية أظهرت أن لهذا المعيار عيباً كبيراً، وهو أنه يتيح للمتحولات ذات التشاركيات الكبيرة أن تؤثر على حساب زاوية التدوير φ أكثر من المتحولات ذات التشاركيات الصغيرة، أي تترك لها وزن أكبر في تحديد الحل النهائي.

ولهذا قام (Kaiser) نفسه بتحسين المعيار السابق وتعديله بتقسيم عناصر الحل النهائي b_{jp} على جذور التشاركيات المقابلة لها (h_j) للمتحولات Z_j، وسمى الناتج بالأمثال المميزة للمتحول Z_j (لأنها تجعل طول شعاع كل سطر مساوياً للواحد) . ثم قام بتعريف تابع جديد للتعبير عن بساطة الحل النهائي بالعلاقة التالية (استخرجه من S^2 بتقسيم كل b_{jp} على h_j ثم ضربه بـ n^2) .

$$V = n\sum_{p=1}^{m}\sum_{j=1}^{n}\left(\frac{b_{jp}}{h_j}\right)^4 - \sum_{p=1}^{m}\left(\sum_{j=1}^{n}\frac{b_{jp}^2}{h_j}\right)^2 \Rightarrow \max \qquad \mathbf{(37-3)}$$

ثم أطلق على هذا المؤشر المعياري اسم (فري ماكس)، ولتحديد زاوية التدوير φ_{pq} في المستوى $F_p\,0F_q$ (لعاملين فقط) نستخدم العلاقة (3−28) ونكتب عناصر الحل النهائي b_{jp} و b_{jq} بدلالة عناصر الحل الأولي a_{jp} و a_{jq} كما يلي :

$$\begin{aligned} b_{jp} &= a_{jp}\cos\varphi_{pq} + a_{jq}\sin\varphi_{pq} \\ b_{jq} &= -a_{jp}\sin\varphi_{pq} + a_{jq}\cos\varphi_{pq} \end{aligned} \qquad (38-3)$$

ويصبح هدفنا الآن إيجاد زاوية التدوير φ_{pq} في المستوى $F_p\, 0F_q$، ثم البحث عنها في جميع المستويات الأخرى في فضاء العوامل العامة R^m، و للحصول على جميع عناصر الحل النهائي B نطبق العلاقة:

$$B = A\, T_{12} * T_{13} \dots T_{pq} \dots . T_{(m-1)m} \qquad (39-3)$$

ولإيجاد هذه الزاوية φ_{pq} التي تجعل التابع V أكبر ما يمكن نجري ما يلي :

نقوم بمعيرة عناصر الحل الأولي a_{jp} وعناصر الحل النهائي b_{jp} ونرمز لهذه القيم الممعيرة والواقعة في المستوى $F_p\, 0F_q$ كما يلي:

$$x_j = \frac{a_{jp}}{h_j} \quad , \quad y_j = \frac{a_{jq}}{h_j} \qquad \left(\text{للحل الأولي}\right) \qquad (40-3)$$

$$X_j = \frac{b_{jp}}{h_j} \quad , \quad Y_j = \frac{b_{jq}}{h_j} \qquad \left(\text{للحل النهائي}\right) \qquad (41-3)$$

وهكذا نجد إنهما يرتبطان عبر المصفوفة T بالعلاقة:

$$\left(X_j\, , Y_j\right) = \left(x_j\, , y_j\right) \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

وهي تعطينا أن العلاقة (3–38) تأخذ الشكل التالي:

$$\begin{aligned} X_j &= \frac{b_{jp}}{h_j} = x_j \cos\varphi + y_j \sin\varphi \\ Y_j &= \frac{b_{jq}}{h_j} = -x_j \sin\varphi + y_j \cos\varphi \end{aligned} \qquad (42-3)$$

وبتعويض هذه العناصر في العلاقة (3–37) نحصل على تابع للزاوية φ فقط ونرمز له $V(\varphi)$.

ولإيجاد القيمة العظمى لهذا التابع نأخذ مشتقه بالنسبة لـ φ ونضعه مساوياً للصفر فنحصل على المعادلة الآتية:

$$\frac{dV(\varphi)}{d\varphi} = 0 \qquad (43-3)$$

ومن هذه المعادلة وبعد إجراء حسابات معقدة نحصل على أن:

$$tg(4\varphi) = \frac{D - 2\dfrac{A * B}{n}}{C - \dfrac{(A^2 - B^2)}{n}} = \frac{v'}{\delta'} \qquad (44-3)$$

حيث أن الرموز في (3–44) تعني وتساوي ما يلي:

$$\begin{aligned} u_j &= x_j^2 - y_j^2 & A &= \sum_{j=1}^{n} u_j \\ v_j &= 2x_j * y_j & B &= \sum_{j=1}^{n} v_j \\ C &= \sum_{j=1}^{n} \left(u_j^2 - v_j^2\right), & D &= 2\sum_{j=1}^{n} u_j * v_j \end{aligned} \qquad (45-3)$$

ويتم تحديد قيمة الزاوية φ من العلاقة (3_44) ضمن المجال $]45 , 45-[$ ، بحيث تجعل قيمة المقدار V في (3−37) أكبر ما يمكن ، وذلك اعتماداً على الجدول (3−2) الوارد في طريقة (كوارتي ماكس) وحسب قيمتي وإشارتي البسط v' والمقام δ' (كما فعلنا سابقاً) .

وأخيراً نشير إلى أن قيمة المشتق الثاني للتابع $V(\varphi)$ يجب أن تحقق الشرط التالي: $\frac{d^2V}{d^2\varphi} < 0$

مثال (3−3):

لنأخذ الحل الأولي الوارد في المثال السابق (3−2) لثمانية متحولات. ولنجري عليه الحسابات اللازمة والمبينة في الجدول التالي ($m = 2$ لعاملين فقط) :

جدول (3−4) بيانات المثال السابق والحسابات اللازمة عليها:

المتحولات	أمثال الحل الأولي		جذر التشاركية	الأمثال الممعيرة للحل الأولي		حسابات المقادير الوسيطة			
	a_{j1}	a_{j2}	h_j	x_j	y_j	u_j	v_j	$u_j^2 - v_j^2$	$2u_jv_j$
Z_1	0.830	−0.396	0.9196	0.9026	−0.4306	0.6293	−0.7773	−0.2082	−0.9783
Z_2	0.818	−0.469	0.9429	0.8675	−0.4974	0.5041	−0.8630	−0.4896	−0.8718
Z_3	0.777	−0.470	0.9081	0.8556	−0.5176	0.4641	−0.8857	−0.5691	−0.8221
Z_4	0.798	−0.401	0.9831	0.8935	−0.4490	0.5967	−08024	−0.2878	−0.9576
Z_5	0.786	0.500	0.9316	0.8437	0.5367	0.4238	0.9056	−0.6405	0.7676
Z_6	0.672	0.458	0.8132	0.8264	0.5632	0.3657	0.9309	−0.7328	0.6809
Z_7	0.594	0.444	0.7416	0.8010	0.5987	0.2832	0.9591	−0.8397	0.5432
Z_8	0.647	0.333	0.7277	0.8891	0.4576	0.5811	0.8137	−0.3244	0.9457
المجموع	----	-----	-----	-----	------	3.8490	0.2809	-4.0921	−0.6930

ومن هذا الجدول نجد أن:

$$A = \sum_{j=1}^{8} u_j = 3.8490 \qquad B = \sum_{j=1}^{8} v_j = 0.2809$$

$$C = \sum_{j=1}^{8} \left(u_j^2 - v_j^2\right) = -4.0921 \quad D = 2\sum_{j=1}^{8} u_j * v_j = -0.6930$$

وبتعويض ذلك في العلاقة (3−44) نجد أن:

$$tg(4\varphi) = \frac{-0.6930 - \frac{2}{8} * 3.8490 * 0.2809}{-4.0921 - \frac{1}{8}[(3.8490)^2 - (0.2809)^2]} =$$

$$tg(4\varphi) = \frac{-0.9633}{-5.9341} = 0.1623 \qquad \Longrightarrow \quad arctg\,(0.1623) = 19.22^{\circ}$$

وهنا نلاحظ أولاً أن إشارة البسط سالبة وإشارة المقام سالبة. أي أن الزاوية (4φ) تقع في المجال: $-180 < 4\varphi \le -90$ ، وهذا يعني أن زاوية التدوير تقع في المجال: $-45 < \varphi \le -22.5$ ، ومن جداول التوابع المثلثية أو من الآلات الحاسبة نجد أن:

$$4\varphi = arctg(0.1623) - 180 = 19.22 - 180 = -170.78^{\circ}$$

ومنها نجد أن قيمة زاوية التدوير $\varphi = -\frac{170.78}{4} = -42.70^{\circ}$ ،

ومنها نحسب عناصر مصفوفة التحويل فنجد أن :

$$\sin\varphi = -0.6782 \qquad \cos\varphi = 0.7349$$

وهكذا نجد أن مصفوفة التدوير T تأخذ الشكل التالي :

$$T = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} = \begin{bmatrix} 0.7349 & 0.6782 \\ -0.6782 & 0.7349 \end{bmatrix}$$

وبضرب عناصر الحل الأولي الممعير x من اليمين بـ T نحصل على الحل النهائي الممعير X كما يلي:

$$X = x * T$$

ثم نقوم بالانتقال إلى حساب عناصر الحل النهائي B (قبل المعيرة) بضرب عناصر الحل النهائي الممعير X بقيم جذور التشاركيات h_j ، أي نحسبها من العلاقة $b_{jp} = \left(X_j * h_j\right)$ ، ونضع النتائج في الجدول التالي: (انظر العمودين الأخيرين b_{j1} و b_{j2}).

جدول (3-5): الحسابات اللازمة لإيجاد الحل النهائي:

المتحولات	الحل النهائي بالأمثال الممعيرة		مربعات العناصر		الحل النهائي b_{jp} $b_{jp} = X_j * h_j$		شكل الهيكل البسيط	
	X_j	Y_j	X_j^2	Y_j^2	b_{j1}	b_{j2}	b_{j1}	b_{j2}
Z_1	0.9554	0.2957	0.9128	0.0874	0.879	0.272	0.879	0.000
Z_2	0.9749	0.2228	0.9504	0.0496	0.919	0.210	0.919	0.000
Z_3	0.9798	0.1999	0.9600	0.0400	0.890	0.182	0.890	0.000
Z_4	0.9611	0.2760	0.9237	0.0762	0.858	0.246	0.858	0.000
Z_5	0.2560	0.9666	0.0665	0.9343	0.238	0.900	0.000	0.900
Z_6	0.2254	0.9744	0.0508	0.9495	0.183	0.792	0.000	0.792
Z_7	0.1826	0.9832	0.0333	0.9667	0.135	0.729	0.000	0.720
Z_8	0.3431	0.9393	0.1177	0.8823	0.250	0.684	0.000	0.684
المجموع	----	---	4.0142	3.9860	----	----	----	----
مجموع المربعات	----	---	3.5331	3.5049	3.316	2.648	----	----

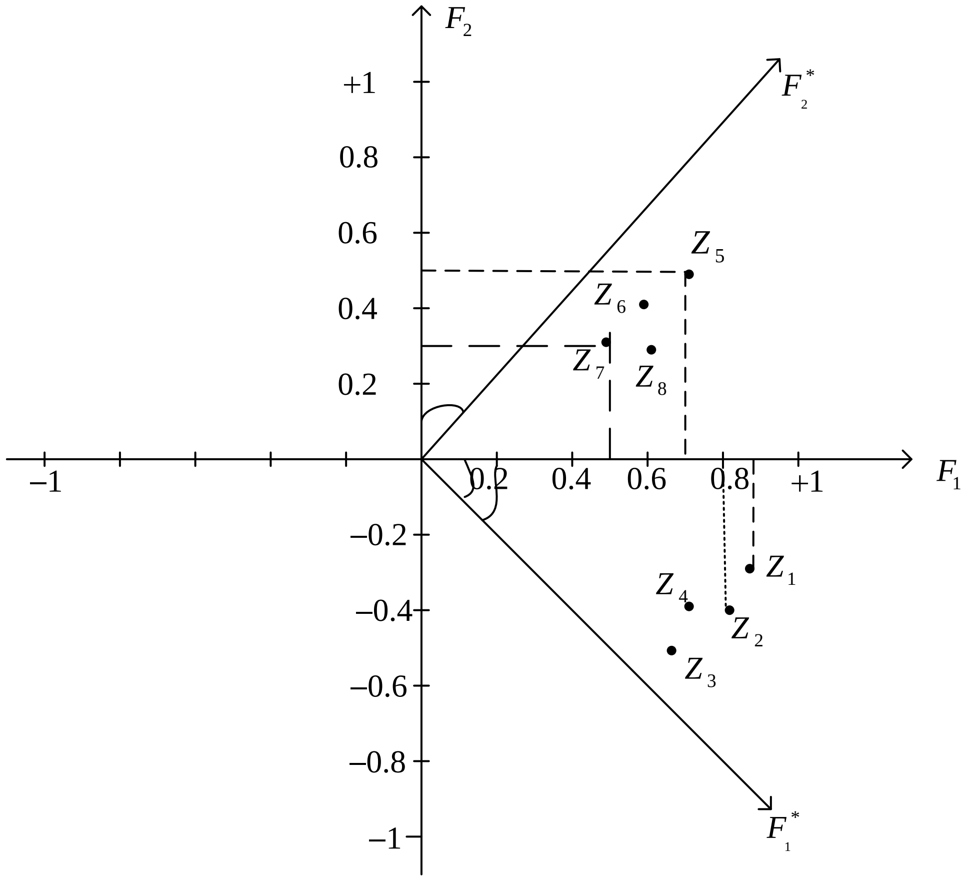
وهكذا نجد أن عناصر الحل النهائي قد أصبحت موجبة وأخذت قيماً أكثر تعبيراً وأسهل تفسيراً، لأن أمثال المتحولات الأربعة الأولى قد ازدادت وأخذت قيم معنوية في العامل الأول F_1^* وأمثال غير معنوية في العامل الثاني F_2^* ، وإن أمثال المتحولات الأربعة الأخرى قد تناقصت في F_1^* وتزايدت في F_2^* ، وهذا

يعني أن العامل الأول يتألف من المتحولات الأربعة الأولى وتشكل العامل الأول F_1^* ، وإن المتحولات الأربعة الأخرى تشكل العامل الثاني F_2^*، ولقد قمنا بحساب المربعات X_j^2 و X_j^2 لاستخدامها في حساب قيمة التابع V، حيث وجدنا أن قيم عناصر هذا الحل تعطينا القيمة العظمى للتابع V وتساوي :

$$V = 8(3.5331 + 3.5049) - [(4.0141)^2 + (3.9860)^2] = 24.3012$$

مقابل قيمة V المقابلة للحل الأولي A والتي كانت تساوي : $V = 0.4078$

ويمكن توضيح عملية التدوير السابقة على الشكل التالي:



الشكل (3−3): التدوير المتعامد للمثال السابق

ومن الشكل نلاحظ أن الحل الأخير أصبح يتألف من عاملين:

الأول: ويضم Z_4, Z_3, Z_2, Z_1 والثاني: ويضم المتحولات Z_8, Z_7, Z_6, Z_5

وأخيراً نستنتج أن الحل النهائي المدور الجديد أصبح أقرب ما يكون إلى شكل الهيكل البسيط ، ويكفي أن نعتبر الأمثال التي تقل عن (0.30) مهملة أو معدومة. فنحصل على العمودين الأخيرين اللذين يحققان شروط الهيكل البسيط حيث يوجد صفر في كل سطر، ويوجد أكثر من صفر في كل عمود .

مثال(3−4): لنأخذ الحل العاملي المتعامد للمثال (1−2) السابق ولنعالجه حسب طريقة vary max للحصول على هيكل بسيط، ولذلك نجري الحسابات التالية:

جدول (3−6): حسابات التدوير حسب varimax

المتحولات	X_j	Y_j	جذر التشاركيات h_j	الأمثال الممعيرة		المقادير المطلوبة			
				x_j	y_j	u_j	v_j	$u_j^2-v_j^2$	$2u_j\ v_j$
Z_1	0.5810	0.8064	0.9939	0.5846	0.8113	−0.3165	0.9486	−0.7997	−0.6005
Z_2	0.7671	−0.5448	0.9409	0.8153	−0.5790	0.3295	−0.9441	−0.7828	−0.6222
Z_3	0.6724	0.7260	0.9895	0.6795	0.7337	−0.0766	0.9971	−0.9883	−0.1528
Z_4	0.9324	−0.1043	0.9382	0.9938	−0.1112	0.9753	−0.2210	0.9024	−0.4311
Z_5	0.7911	−0.5582	0.9682	0.8171	−0.5765	0.3353	−0.9421	−0.7751	−0.6318
المجموع					-----	1.2470	−0.1615	−2.4435	−2.4384

وهكذا نجد من العلاقة (3−44) أن:

$$tg(4\varphi)=\frac{-2.4384-\frac{2}{5}*(1.2470)(-0.1615)}{-2.4453-\frac{(1.2470)^2-(-0.1615)^2}{5}}=\frac{-2.3578438}{-2.751085}=0.857059$$

ومن الجدول (3−2) نجد أن الزاوية (4φ) تقع في الربع الثالث، ومن جداول التوابع المثلثية نجد أن:

$$(4\varphi)=arctg(0.8570)-180=40.5985-180=-139.4015$$

أي أن زاوية التدوير المناسبة تساوي:

$$\varphi=-\frac{139.4015}{4}=-34.85$$

وبذلك نجد أن عناصر مصفوفة التحويل T تساوي :

$$T=\begin{bmatrix}\cos(-34.85) & -\sin(-34.85)\\ \sin(-34.85) & \cos(-34.85)\end{bmatrix}=\begin{bmatrix}0.82065 & 0.57143\\ -0.57143 & 0.82065\end{bmatrix}$$

ومنها نجد أن الإحداثيات الجديدة لصورة الحل السابق تساوي ما يلي:

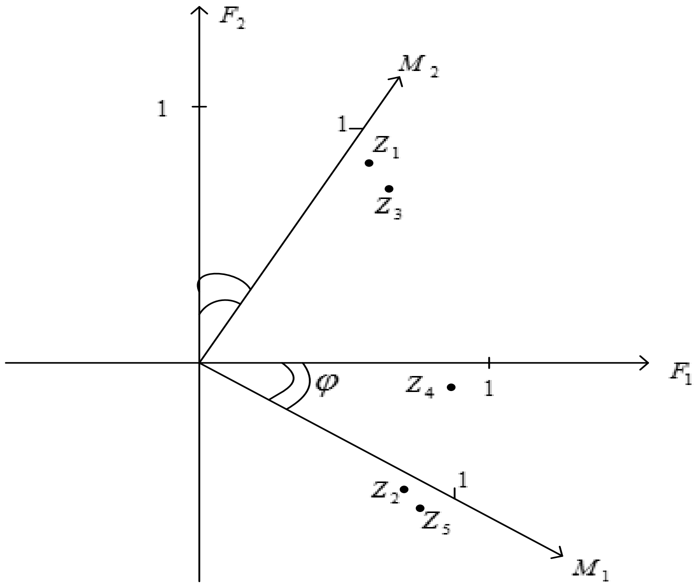
$$B=A*T=\begin{array}{c}\begin{matrix}F_1 & F_2\end{matrix}\\ \begin{bmatrix}0.5810 & 0.8064\\ 0.7671 & -0.5448\\ 0.6724 & 0.7260\\ 0.9324 & -0.1043\\ 0.7911 & -0.5582\end{bmatrix}\end{array}*\begin{bmatrix}0.82065 & 0.57143\\ -0.57143 & 0.82065\end{bmatrix}=\begin{array}{c}\begin{matrix}M_1 & M_2\end{matrix}\\ \begin{bmatrix}0.0160 & 0.9988\\ 0.9408 & -0.0088\\ 0.1370 & 0.9801\\ 0.8248 & 0.4471\\ 0.9682 & -0.0060\end{bmatrix}\end{array}$$

ومما سبق نلاحظ أن الحل الجديد B الذي حصلنا عليه بعد تدوير المحاور الإحداثية بمقدار(34.85−) درجة، يتميز بأنه يحقق معظم شروط الهيكل البسيط، ولإظهار ذلك يكفي أن نعتبر الأمثال التي تقل عن (0.20) مهملة أو مساوية للصفر فنجد أن ذلك الحل يساوي :

$$B=\begin{bmatrix}0.000 & 0.9938\\ 0.9408 & 0.000\\ 0.000 & 0.9801\\ 0.8248 & 0.4471\\ 0.9682 & 0.000\end{bmatrix}$$

وهو يتضمن في كل سطر صفراً واحداً على الأقل (ماعدا سطر Z_4) ويتضمن صفرين في كل عمود، وهذه أهم خواص الهيكل البسيط .

وأخيراً يمكننا أن نرسم الحل المدور الجديد مع المحاور الإحداثية الجديدة M_1oM_2 كما يلي:



الشكل (3−4): الحل بعد التدوير (قارنه مع الشكل (1−2))

وهو يشير إلى وجود عاملين رئيسيين:

الأول: يتألف من $Z_5\ ,Z_2$ والثاني: يتألف من $Z_3\ ,Z_1$ ، أما المتحول Z_4 فله موقع خاص بينهما، ولكنه أقرب منه إلى العامل الأول، لذلك يمكن إدخاله ضمن العامل الأول .

3−3 : التدوير المائل وطرائقه: [Harman P.336]

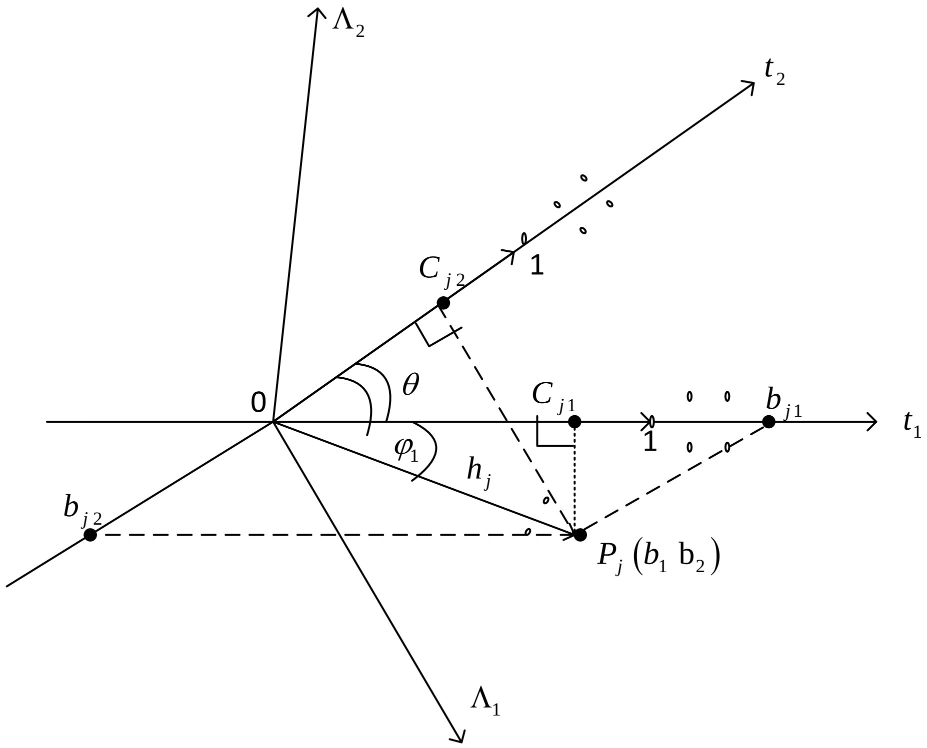
1−3−3: تمهيد:

يستخدم التدوير المائل لتحسين الصورة العاملية للمسألة المطروحة. ويهدف إلى تحويل الحل المتعامد (أو المائل) إلى حل مائل (آخر) محمول على محاور مائلة تمر من مراكز ثقل المجموعات التي تشكلها المتحولات Z_j في الفضاء R^m (كما في الحالة التي رأيناها في طريقة المجموعات) .

ولتوضيح معاني الحل المائل نفترض أن الشعاعين t_1 و t_2 واحديان ويشكلان في المستوى محورين مائلين بينهما بزاوية θ (حادة أو منفرجة). ولتسهيل عملية تمثيل المتحولات Z_j على هذه المحاور نفترض أن كل متحول Z_j يتمثل على هذا المستوى بنقطة P_j، أو بشعاع oP_j طوله h_j $\left(h_j \leq 1\right)$، فتكون الإحداثيات المائلة (الموازية للمحاور) للنقطة P_j في هذا النظام المائل هي $P\left(b_{j1}\ ,\ b_{j2}\right)$، وهي تمثل أمثال (أثقال) ذلك المتحول على المحورين المائلين ot_1 و ot_2 كما في الشكل (3−5)، أي يمكننا أن نكتب أي متحول Z_j بدلالة t_1 و t_2 على شكل تركيب خطي كما يلي:

$$Z_j = b_{j1}t_1 + b_{j2}t_2 \qquad (46-3)$$

ومن الشكل (3−5) نلاحظ أن القطعة المستقيمة b_{j1} (أمثال Z_j على ot_1) أكبر من الواحد الصحيح $(b_{j1} > 1)$ ، أي أن الأمثال b_{jk} في الإحداثيات المائلة t_1ot_2 يمكن أن تأخذ قيماً أكبر من الواحد، وذلك على عكس الاحداثيات المتعامدة التي تكون دائماً أصغر من الواحد $(a_{jp} < 1)$، لأنها عبارة عن المساقط العمودية للشعاع oP_j الذي طوله h_j $(h_j \leq 1)$ على المحورين ot_2, ot_1 ، ولأن الأمثال العمودية a_{jp} تمثل معاملات الارتباط بين المتحولات Z_j والعوامل F_p .



الشكل (3−5) التدوير المائل

لقد أشرنا إلى أن طول أي شعاع Z_j في الفضاء R^m يساوي جذر التشاركية h_j^2 لذلك الشعاع وتساوي :

$$\ell_j = \|Z_j\| = h_j = oP_j \qquad (47-3)$$

وإذا رمزنا للزاوية بين الشعاع oP_j (الممثل لـZ_j) والمحور t_1 بالرمز φ_j ولمسقطه العمودي على t_1 بالرمز C_{j1} ولمسقطه العمودي على t_2 بالرمز C_{j2}، فإننا نجد من الشكل (3−5) أن:

$$\cos\varphi_j = \frac{C_{j1}}{h_j} \qquad (48-3)$$

$$C_{j1} = h_j \cos\varphi_j \qquad (49-3)$$

ومن جهة أخرى نعلم أنه إذا كانت المتحولات Z_j ممثلة على شكل أشعة في R^m، فإن معاملات الارتباط بينهما تساوي جداؤهما السلمي، وهكذا نجد أن معامل الارتباط بين Z_j و t_1 يساوي جداؤهما السلمي ويساوي :

$$r_{jt_1} = \ell_j * \ell_{t_j} * \cos\varphi_j \qquad (50-3)$$

وبما أن طول الشعاع t_1 يساوي الواحد $\left(\ell_{t_j} = 1\right)$ وأن $\left(\ell_j = h_j\right)$ فإن معامل الارتباط r_{jt_1} يساوي:

$$r_{jt_1} = h_j \cos\varphi_j \qquad (51-3)$$

ومن العلاقتين (3−51) و (3−49) نستخلص مباشرة أن :

$$r_{jt_1} = C_{j1} \qquad (52-3)$$

وبطريقة مشابهة يمكننا أن نبرهن على أن:

المساقط العمودية للشعاع oP_j (الممثل لـ Z_j) على المحورين المائلين ot_1 و ot_2 ، ماهي إلا معاملات ارتباط Z_j مع المحورين ot_1 و ot_2 على الترتيب . وهي تختلف عن الأمثال (b_1 , b_2) التي تعبر عن الإحداثيات المائلة للمتحول Z_j والشكل (3−5) يوضح ذلك .

ومع أن كلتا الحالتين تشكلان مساقط oP_j على المحورين ot_1 و ot_2 ، ولكن الأمثال $\left(b_{j1} , b_{j2}\right)$ تنشأ من إسقاط oP_j على t_1 بشكل يوازي t_2 ، ومن أسقاط oP_j على t_2 بشكل يوازي t_1، ويمكن أن تكون موجبة أو سالبة، والشيء الأهم هو أن قيمتها يمكن أن تكون أكبر من الواحد (وذلك حسب موقع المتحول Z_j في المستوى t_1ot_2) .

أما المساقط العمودية $\left(C_{j1} , C_{j2}\right)$ لـ Z_j على المحورين المائلين ot_1 و ot_2 فهي تساوي معاملات الارتباط بين Z_j والمحورين ot_1 و ot_2 ، وإن هذه المساقط يمكن أن تكون موجبة أو سالبة ولكن قيمها المطلقة لا يمكن أن تتجاوز الواحد. (حسب خواص معامل الارتباط) .

وإذا اقتربت الزاوية بين t_1 و t_2 إلى الزاوية القائمة $(90°)$ فإن كلا النوعين من المساقط يتطابقان.

مما سبق يمكننا أن نستخلص هندسياً من الشكل (3−5) ورياضياً من العلاقات السابقة أنه يوجد لدينا مصفوفتان أو صورتان عامليتان للمحاور المائلة هما:

أ− مصفوفة الصورة العاملية المائلة ونرمز لها بـ P: وهي عبارة عن مصفوفة الأمثال b_{jp} للمتحولات Z_j على المحاور المائلة $t_1 t_2 t_3 \ldots\ldots t_m$ ويمكن أن تكون قيمها المطلقة أكبر من الواحد .

ب− مصفوفة الهيكل العاملي ونرمز لها بـ S : وهي عبارة عن مصفوفة معاملات الارتباط بين المتحولات Z_j والعوامل المائلة $t_1, t_2, t_3, \ldots\ldots, t_m$، وتكون قيمها المطلقة أصغر من الواحد .

أما في الإحداثيات المتعامدة فإن هاتين المصفوفتين تتطابقان .

وبطريقة مشابهة لما فعلناه في الفصل الأول عند استخراج العلاقات (1−49) و(1−55) و(1−56) واستناداً إلى العلاقة (3−5) ، يمكننا أن نستخرج علاقات مشابهة وبرموز مناسبة للإحداثيات المائلة، حيث نجد أن:

− المصفوفة الهيكلية S المؤلفة من معاملات الارتباط بين المتحولات Zj والعوامل المائلة $t_1,t_2,t_3,\ldots,t_m$ ، والتي تساوي بدلالة الصورة العاملية الأولية A وبدلالة مصفوفة التحويل T ما يلي:

$$S = A * T \qquad (54-3)$$

− وهي تساوي من جهة أخرى وقياساً على العلاقة (1−55) ما يلي :

$$S = P * \emptyset \qquad (55-3)$$

- حيث أن : $\emptyset$ هي مصفوفة المعاملات الارتباطية بين العوامل المائلة نفسها والتي تساوي:

$$\emptyset = T' * T \quad : \quad (\emptyset \neq I) \qquad (56-3)$$

وهكذا نجد من (3−55) أن :

$$P = S * \emptyset^{-1} \qquad (57-3)$$

ومن العلاقة (3−56) نجد أن : $\emptyset^{-1} = T^{-1} * (T')^{-1}$ وبالاستفادة من العلاقة (3−54) نجد أن العلاقة (3−57) تأخذ الشكل التالي:

$$P = S * \emptyset^{-1} = (A * T)T^{-1} * (T')^{-1} = A * (T')^{-1} \qquad (58-3)$$

أي أنه يمكننا حساب مصفوفة الصورة المائلة P من المصفوفة A مباشرة من العلاقة :

$$P = A * (T')^{-1} \qquad (59-3)$$

إن التعبير عن نقاط المتحولات في نظام الإحداثيات المائلة t_1ot_2 ، يسمى بالحل العاملي الأول (الأولاني) ، ويسمى المحوران ot_1 و ot_2 وما شابههما بالمحاور الأولى .

وبعد حساب وتحديد أمثال هذه النقاط على ot_1 و ot_2 نقوم بتحسين ذلك الحل والبحث عن حل آخر يسمى بالحل الثاني منسوب لإحداثيات جديدة متعامدة تخالفياً مع الأولى t_1ot_2 ويتم إنشاؤها كما يلي:

نقوم بإنشاء محورين إحداثيين جديدين $o\wedge_1$ و $o\wedge_2$ ، بحيث يكون $o\wedge_1$ متعامداً مع ot_2 ويكون $o\wedge_2$ متعامداً مع ot_1 ، ونسميهما بالمحاور الثانية ، وهما المحوران المرسومان على الشكل (3−5) السابق .

ومن الشكل (3−5) يتضح لنا أن مساقط النقاط الأربعة القريبة من ot_1 على المحور $o\wedge_2$ تصبح مساوية للصفر تقريباً .

وكذلك نجد أن مساقط النقاط الأربعة القريبة من ot_2 على المحور الثاني $o\wedge_1$ هي أيضاً تصبح مساوية للصفر تقريباً .

وبذلك نكون قد حصلنا على الحل الثاني الذي يحقق شروط الهيكل البسيط . وهكذا نجد أنه توجد علاقة بين هذين النوعين من الحل ويمكننا الانتقال من أحدهما إلى الآخر .

ولكن في المسائل العملية غالباً ما يلجؤون مباشرة إلى الحل المائل الثاني (بدلالة المحاور الثانية $\wedge_2 o \wedge_1$) لأن المحاور الثانية تشكل اتجاهات مناسبة لإنشاء نظام إحداثيات عند التدوير ويحقق لنا معظم شروط الهيكل البسيط .

وإن العقبة الأولى في التدوير المائل هي ضرورة التفريق بين الصورة العاملية المائلة P والهيكل العاملي المائل S وفهم معاني كل منهما .

أما العقبة الثانية تتلخص في فهم الفروقات بين الحلين المائلين الأول والثاني حيث أن

الحل المائل الأول يكون منسوباً للإحداثيات المائلة الأولى t_1ot_2 .

أما الحل المائل الثاني فيكون منسوباً للإحداثيات المائلة الثانية $\wedge_2 o \wedge_1$ المتعامدة تخالفياً مع t_1ot_2 .

3-3-2: خوارزمية التدوير للحصول على الحل الأول المائل :

لقد لاحظنا مما سبق أن الحلول العاملية قد تكون متعامدة أو مائلة، وسنقدم فيما يلي خوارزمية للانتقال من اي حل متعامد إلى حل مائل، يحقق بدرجة ما شروط الهيكل البسيط، وتتألف من الخطوات التالية:

1- الحصول على حل ابتدائي متعامد باستخدام أي من الطرق التي تعطينا حلول متعامدة مثل:

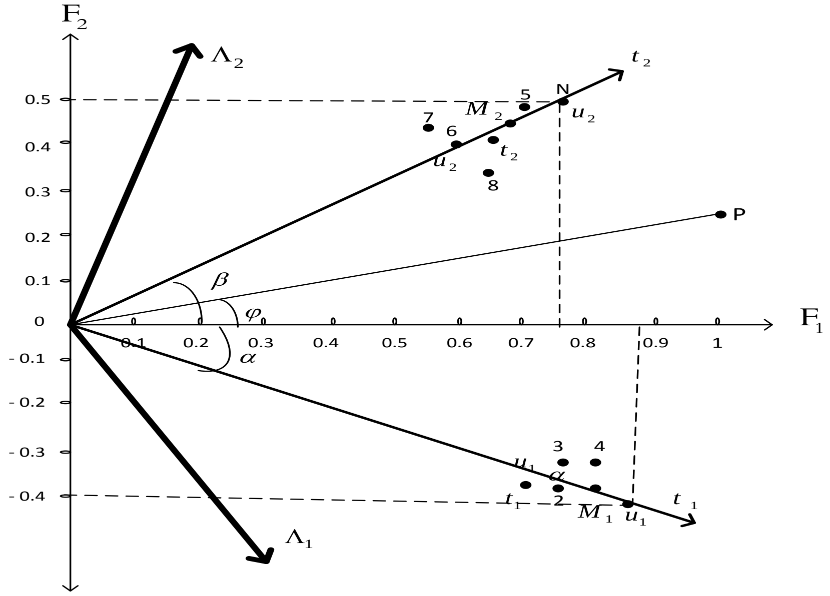
طريقة العوامل الرئيسية – طريقة البواقي الصغرى – طريقة الامكانية العظمى، وسيكون معرفاً على m عاملاً متعامداً .

2- **إيجاد الصورة المخففة *Redution Pattern***: أي إيجاد المحاور الإحداثية المائلة، التي تمر من مراكز ثقل المجموعات التي تشكلها المتحولات Z_j المترابطة بشدة مع بعضها البعض. ونرمز لتلك المجموعات بالرمز G_p ولعدد عناصرها بالرمز n_p ولعددها بالرمز n_1 $(n_1 < n)$، فنحصل على n_1 محوراً مائلاً يمثل n_1 متحولاً مركباً بدلاً من n متحولاً في فضاء العوامل R^m، فمن الشكل السابق (3–5) نلاحظ أن المتحولات الثمانية تشكل مجموعتين من المتحولات المترابطة بشدة، الأولى: G_1 وتتألف من المتحولات الأربعة الأولى وتقع في الربع الرابع، والثانية :G_2 وتتألف من المتحولات الأربعة الأخرى وتقع في الربع الأول، ولإيجاد المحاور التي تمر من مراكز ثقل هاتين المجموعتين نأخذ متوسط أو مجموع المتحولات Z_j في كل مجموعة G_p، ونشكل لكل منها متحولاً مركباً نرمز له بـ v فيكون لدينا متحولان مركبان جديدان هما v_1 , v_2 ويعرفان بالعلاقتين التاليتين :

$$v_1 = Z_1 + Z_2 + Z_3 + Z_4 \qquad \text{(للمجموعة الأولى)}$$

$$v_2 = Z_5 + Z_6 + Z_7 + Z_8 \qquad \text{(للمجموعة الثانية)}$$

والشكل البياني (3–6) يبين لنا أن الشعاعين oP_1 و oP_2 يشكلان المتحولين المركبين v_1 و v_2 اللذين يمران من مركزي ثقل المجموعتين G_1 و G_2 ، ورمزنا لهما بـ M_1 و M_2 ، .



الشكل (3–6)

ولتحويل هذين الشعاعين إلى محاور إحداثية مائلة، علينا أن نأخذ الشكل المعياري لهما. ولذلك علينا أن نحسب التباين والانحراف المعياري لها ثم نعرف عليها متحولين معياريين كما يلي:

$$U_1 = \frac{v_1}{Sv_1}$$

$$U_2 = \frac{v_2}{Sv_2}$$

واعتماداً على العلاقة (2−139) نجد أن تباين كل متحول مركب v_p يساوي مجموع معاملات الارتباط بين المتحولات Z_j المنتمية للمجموعة G_p أي أن:

$$S_{v_p}^2 = \sum_{j=1}^{n_p}\sum_{k=1}^{n_p} r_{jk} \qquad : \quad (j\,,k\ \in G_p) \qquad (60-3)$$

وفي مثالنا هذا نجد أنه يمكننا من الجدول (2−11) أن نحسب تباين المتحول المركب الأول v_1 ، حيث يساوي مجموع عناصر المصفوفة الجزئية لمعاملات الارتباط المقابلة للمتحولات Z_1, Z_2, Z_3, Z_4 مع بعضها البعض (علماً بأنه يشمل حساب معامل كل متحول مع نفسه والذي يساوي الواحد): وهو يساوي: $S_{v_1}^2 = 14.03626$

وإن انحرافه المعياري يساوي : $S_{v_1} = 3.7465$

وبطريقة مشابهة نجد أن تباين المتحول المركب v_2 يساوي :

$$S_{v_2}^2 = 11.639697$$

وإن انحرافه المعياري يساوي : $S_{v_2} = 3.4117$

والآن لنقم بالتعبير خطياً عن المتحولين المركبين v_1 و v_2 بدلالة العوامل العامة الأصلية F_1 و F_2 ، حيث نعلم مما سبق أن الأمثال الخطية لهذين المركبين على F_1 و F_2 هي عبارة عن معاملات ارتباطهما مع F_1 و F_2 لذلك علينا حساب معاملات ارتباط v_1 و v_2 مع F_1 و F_2 .

وبما ان العوامل العامة F_1 و F_2 غير مرتبطة ومعطية بشكلها المعياري، فإن معامل ارتباط المتحول المركب U_p مع العامل F يساوي ما يلي:

$$r_{U_pF} = \frac{\sum_{j=1}^{n_p} r_{jF}}{S_{v_p}} \qquad : (j \in G_p) \qquad (61-3)$$

وذلك لأن معاملات ارتباط المتحولات المختلفة مع العوامل F، ماهي إلا أمثال المتحولات الداخلة في المجموعة G_p الناجمة عن الصورة العاملية الأصلية (المتعامدة). وفي مثالنا هذا نجد من الجدول (2−11) أن معامل ارتباط المتحول المركب الأول U_1 مع العامل الأول F_1 يساوي (مجموع المعاملات في عمود F_1 المقابلة للمتحولات Z_1, Z_2, Z_3, Z_4 فقط) مقسوماً على انحرافه المعياري: أي أن:

$$r_{U_1F_1} = \frac{0.856 + 0.848 + 0.808 + 0.831}{3.7465} = 0.8923$$

وهكذا نجد أن معامل ارتباط U_1 مع F_2 يساوي :

$$r_{U_1F_2} = \frac{-0324 - 0.412 - 0.409 - 0.342}{3.4117} = -0.3969$$

ومنهما نجد أنه يمكننا التعبير بشكل معياري عن المركب U_1 بدلالة F_1 و F_2 من خلال العلاقة المعيارية التالية:

$$U_1 = \frac{v_1}{Sv_1} = 0.8923F_1 - 0.3969F_2 \qquad (62-3)$$

وبطريقة مشابهة يمكننا التعبير بشكل معياري عن المركب U_2 بدلالة F_1 و F_2 من خلال العلاقة المعيارية التالية:

$$U_2 = \frac{v_2}{Sv_2} = 0.7495F_1 + 0.5639F_2 \qquad (63-3)$$

وهنا نلاحظ أن أمثال $\mathbf{U_1}$ و U_2 على $\mathbf{F_1}$ و $\mathbf{F_2}$ (انظر الشكل (3−6)) هي مساقطهما على $\mathbf{0F_1}$ و $\boldsymbol{O}\mathbf{F_2}$.

إن هاتين العلاقتين تشكلان ما يسمى بالصورة المخففة للحل الأصلي وهي تتصف بنفس الصفات التي تتمتع بها الصورة العاملية العادية. ويمكن كتابتهما مصفوفياً كما يلي:

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0.8923 & -0.3969 \\ 0.7495 & 0.5639 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \qquad \Longrightarrow U = B * F \qquad (64-3)$$

3− **حساب مصفوفة التحويل:** بما أن اتجاهات المحاور المائلة للحل الأولي تتطابق مع الاتجاهات الموافقة لها للمتحولين المركبين $\boldsymbol{U}_1$ و U_2 ، فإنه يمكننا تحديد وحساب مصفوفة التحويل T بمساعدة الصورة العاملية المخففة B . لذلك نقوم بتحويل أمثال الصورة المخففة التي تمثل إحداثيات نهايات الأشعة المركبة $U1$ و U_2 إلى اشعة واحدية، فنحصل على الشعاعين العاملين المائلين t_1 و t_2 .

ويتم ذلك بتقسيم هذه الأمثال على طول الشعاع الذي يمثلها. فمثلاً نجد بالنسبة للشعاع $\boldsymbol{U1}$ أن طوله يحسب من العلاقة :

$$\|U_1\| = \sqrt{(0.8923)^2 + (-0.3969)^2} = 0.9766$$

وبتقسيم أمثال U_1 في (3−62) على هذا الطول نحصل على الاتجاهات الواحدية للعامل المائل t_1 بدلالة العاملين F_1 و F_2 وتساوي:

$$t_{11} = \frac{0.8923}{0.9766} = 0.91377 \quad , \qquad t_{21} = \frac{-0.3969}{0.9766} = -0.4064$$

وبطريقة مشابهة نحصل على اتجاهات العامل المائل t_2 بدلالة العاملين F_1 و F_2 كما يلي:

$$\|U_2\| = \sqrt{(0.7495)^2 + (0.5639)^2} = 0.93794$$

$$t_{12} = \frac{0.7495}{0.93794} = 0.7991 \quad , \qquad t_{22} = \frac{0.5639}{0.93794} = 0.6012$$

وهكذا نكون قد حصلنا على العوامل المائلة t_1 و t_2 بدلالة العوامل الأصلية F_1 و F_2 . ويمكن أن نكتبهما كما يلي:

$$\begin{matrix} t_1 \\ t_2 \end{matrix} = \begin{matrix} 0.9137F_1 & -0.4064F_2 \\ 0.7991F_1 & 0.6012F_2 \end{matrix}$$

ويمكن كتابتها بدلالة منقول مصفوفة التحويل T' على الشكل التالي :

$$U = T' * F = \begin{bmatrix} 0.9137 & -0.4064 \\ 0.7991 & 0.6012 \end{bmatrix} * \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \qquad (65-3)$$

ومنها نحصل على أن مصفوفة التحويل T من المحاور $F_1, 0F_2$ إلى المحاور $t_1, 0t_2$ المائلة هي:

$$T = \begin{array}{cc} t_1 & t_2 \\ \end{array} \qquad \begin{array}{cc} t_1 & t_2 \end{array}$$
$$T = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix} = \begin{bmatrix} 0.9137 & 0.7991 \\ -0.4064 & 0.6012 \end{bmatrix} \qquad (66-3)$$

ويشترط في مصفوفة التحويل T أن تكون ذات أشعة واحدية ، أن يكون مجموع مربعات عناصر اي عمود فيها مساوياً للواحد. وهكذا يمكننا الحصول على الصورة العاملية المائلة P من العلاقة (3−59) وعلى الصورة العاملية المخففة من العلاقة (3−65) ، أو من أي علاقات خطية مشابهة لهما إذا كان عدد المجموعات المتشكلة أكثر من اثنين .

4− حساب معاملات الارتباط بين العوامل المائلة t_1 و t_2 :

بعد حساب أمثال اتجاهات المحاور المائلة t_1 و t_2 ... يمكننا حساب معاملات الارتباط بين اي عاملين جديدين واحديين ، حيث نعلم أن الجداء السلمي لأي شعاعين واحديين يساوي قيمة معامل الارتباط بينهما (أو يساوي *cos* الزاوية بينهما)، وهكذا نجد أن معامل الارتباط بين العاملين t_1 و t_2 يساوي الجداء السلمي لهما ويساوي:

$$r_{t_1t_2} = (0.9137)(0.7991) + (-0.4064)(0.612) = 0.4858 = \cos\theta_{12}$$

وبنفس الطريقة نحسب معاملات ارتباط كل من العوامل مع نفسها فنجد أنها تساوي الواحد ، ومما سبق يمكننا حساب الزاوية التي بينهما فنجد أنها تساوي:

$$\theta_{12} = arc\cos(0.4858) \ = 61°$$

وبصورة عامة عندما يكون لدينا أكثر من مجموعتين فإن الجداء السلمي لأي شعاعين t_p و t_q من مصفوفة التحويل يساوي معامل الارتباط بينهما $r_{t_pt_q}$، وعندها فإن مصفوفة معاملات الارتباط بين العوامل الجديدة t_p و t_q والتي نرمز لها بـ $\emptyset$ تساوي :

$$\emptyset = T' * T \qquad (67-3)$$

5− **إيجاد مصفوفة الهيكل العاملي المائل *S***: وهو عبارة عن مصفوفة معاملات الارتباط بين *المتحولات* الأصلية Z_j والعوامل المائلة الجديدة t_1 و t_2 .

ولإيجاد هذه المعاملات لنفترض أن المحور المائل t_1 يصنع مع العامل العام F_1 زاوية قدرها ∝، وأن المحور المائل t_2 يصنع مع العامل العام F_1 أيضاً زاوية قدرها β، ولنأخذ أي نقطة P_j تمثل احد أشعة المتحولات Z_j . ولنفترض أن الزاوية التي يصنعها الشعاع oP_j مع F_1 هي φ_j، فنجد من الشكل (3−7) أن مسقط ذلك الشعاع على t_1 يساوي :

$$D(OM) = D(OP) * \cos\big(\varphi + (-\propto)\big) \qquad \text{(تعتبر هنا ∝ سالبة)} \qquad (68-3)$$

وبما أنه لدينا :

$$D(OP) = \|Z_j\| = h_j \qquad (69-3)$$

$$D(OM) = \|OM\| = r_{jt_1} \qquad (70-3)$$

فإننا نجد أن:

$$r_{jt_1} = h_j \cos(\varphi - \propto) \qquad (71-3)$$

ثم نقوم بحساب مفكوك $\cos(\varphi - \propto)$ فنحصل على أن :

$$r_{jt_1} = h_j(\cos\varphi \cos\propto + \sin\varphi \sin\propto) \qquad (72-3)$$

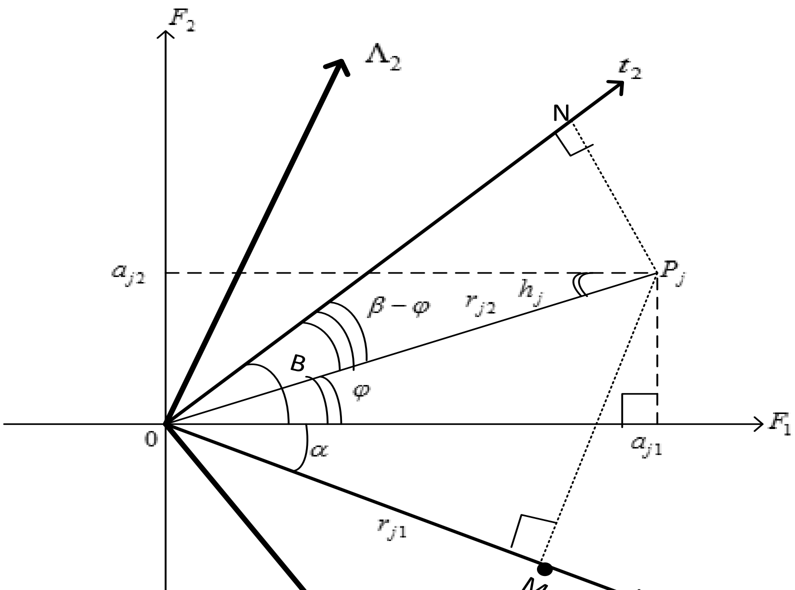
$$r_{jt_1} = (h_j \cos\varphi)\cos\propto + (h_j \sin\varphi)\sin\propto \qquad (73-3)$$

ونلاحظ من الشكل (3−7) أن كل $h_j \cos\varphi$ يساوي a_{j1} مسقط oP_j على oF_1 للمتحول Z_j وأن كل $h_j \sin\varphi$ يساوي a_{j2} مسقط oP_j على oF_2 للمتحول Z_j، وأخيراً نجد أن:

$$r_{jt_1} = a_{j1}\cos\propto + a_{j2}\sin\propto \qquad (74-3)$$

وبطريقة مشابهة نجد أن المسقط $D(ON)$ للشعاع oP_j على oT_2 يساوي معامل الارتباط بينهما ويساوي $r_{jt_2} = h_j\cos(\beta - \varphi)$ ومنها نجد أن:

$$r_{jt_2} = a_{j1}\, cos\,\beta + a_{j2}\sin\beta \qquad (75-3)$$



الشكل (3−7)

وهكذا نجد أنه يمكننا توحيد العلاقتين (3−74) (3−75) وكتابتهما مصفوفياً على الشكل التالي:

$$(r_{jt_1}, r_{jt_2}) = (a_{j1}, a_{j2})\begin{bmatrix}\cos\propto & \cos\beta \\ \sin\propto & \sin\beta\end{bmatrix} \qquad (75a-3)$$

ويمكن تعميم هذه العمليات على أكثر من عاملين، فنحصل على صيغة لتحويل الصورة العاملية المتعامدة الأصلية A إلى الهيكل العاملي المائل S من خلال الشكل التالي:

$$S = A * T \qquad (76-3)$$

حيث أن عناصر أعمدة المصفوفة T هي التوابع المثلثية $(\cos\theta)$ لاتجاهات المحاور المائلة بالنسبة للمحاور المتعامدة والمعرفة بالعلاقة $(3-75a)$.

مثال (3−5): لنأخذ الصورة العاملية للمثال السابق (3−3) لثمانية متحولات، والمطلوب العمل لإيجاد الهيكل العاملي الذي يمثل معاملات ارتباط للمتحولات Z_j مع العوامل المائلة t_1 و t_2 . نقوم بضرب الصورة العاملية A من اليمين بمصفوفة التحويل T المحسوبة من (3−66) فنجد أن:

$$S = A * T = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Z_8 \end{array} \begin{array}{c} \begin{array}{cc} F_1 & F_2 \end{array} \\ \begin{bmatrix} 0.856 & -0.324 \\ 0.848 & -0.412 \\ 0.808 & 0.409 \\ 0.831 & 0.342 \\ 0.750 & 0.571 \\ 0.631 & 0.492 \\ 0.569 & 0.510 \\ 0.607 & 0.351 \end{bmatrix} \end{array} \begin{bmatrix} 0.9137 & 0.7991 \\ -0.4064 & 0.6012 \end{bmatrix} = \begin{array}{c} \begin{array}{cc} r_{jt_1} & r_{jt_2} \end{array} \\ \begin{bmatrix} 0.914 & 0.489 \\ 0.912 & 0.430 \\ 0.904 & 0.400 \\ 0.808 & 0.458 \\ 0.453 & 0.943 \\ 0.377 & 0.800 \\ 0.313 & 0.761 \\ 0.412 & 0.696 \end{bmatrix} \end{array}$$

ولإيجاد مصفوفة الهيكل العاملي المخففة التي تمثل معاملات ارتباط المتحولات المركبة u_1 و u_2 مع العاملين المائلين t_1 و t_2 ، نقوم بضرب الصورة العاملية B المعرفة في (3−64) من اليمين بمصفوفة التحويل T فنجد أن :

$$S_h = B * T = \begin{array}{c} u_1 \\ u_2 \end{array} \begin{bmatrix} 0.8923 & -0.3969 \\ 0.7496 & 0.5639 \end{bmatrix} * \begin{bmatrix} 0.9137 & 0.7991 \\ -0.4064 & 0.6012 \end{bmatrix} = \begin{bmatrix} 0.977 & 0.473 \\ 0.456 & 0.938 \end{bmatrix}$$

ولإيجاد الصورة العاملية المائلة P بدلالة العاملين t_1 و t_2 ، نضرب الصورة A الأصلية من اليمسن بمقلوب منقول المصفوفة T فنجد أن:

$$P = A * (T')^{-1} = \begin{bmatrix} 0.885 & 0.059 \\ 0.960 & -0.035 \end{bmatrix}$$
$$\cdots\cdots\cdots \quad \cdots\cdots\cdots$$

وهي العناصر الموجودة في الجدول (3−7) التالي (انظر العمودين الأخيرين) :

وكذلك يمكننا حساب الصورة العاملية المخففة للمتحولين U_1 و U_2 بدلالة العاملين t_1 و t_2 من العلاقة:

$$P_h = B * (T')^{-1} = \begin{bmatrix} 0.977 & -0.001 \\ 0.000 & 0.938 \end{bmatrix}$$

وهكذا نكون قد حصلنا على الحل العاملي الأول وهو يتألف من جزأين هما: الهيكل العاملي المائل S والصورة العاملية المائلة P، وكذلك الهيكل العاملي المخفف S_h والصورة العاملية المخففة P_h ونضعها في جدول موحد كما يلي:

جدول (3−7) نتائج الحسابات للحصول على الحل النهائي المائل : P , P_h و S , S_h

المتحولات الأصلية	معاملات الهيكل العاملي المائل S		الصورة العاملية المائلة P	
	r_{jT_1}	r_{jT_2}	الصورة على t_1	الصورة على t_2
Z_1	0.914	0.489	0.885	0.059
Z_2	0.942	0.430	0.960	−0.035
Z_3	0.904	0.400	0.929	−0.053
Z_4	0.898	0.458	0.881	0.028
Z_5	0453	0.943	−0.007	0.946
Z_6	0.377	0.800	−0.015	0.807
Z_7	0.313	0.761	−0.074	0.797
Z_8	0.412	0.696	0.097	0.649
المتحولات المخففة	الهيكل العاملي المخفف S_h		الصورة العاملية المخففة P_h	
U_1	0.977	0.474	0.977	−0.001
U_2	0.456	0.938	0.000	0.938

ومن هذا الجدول نلاحظ أن أمثال المتحولات الأربعة الأولى في العمود t_1 قد تحسنت عما كانت عليه. كما أن أمثال المتحولات الأربعة الأخرى في العمود t_2 قد تحسنت أيضاً عما كانت عليه ، بينما انخفضت الأمثال الأخرى في كلا العمودين وأصبحت معدومة تقريباً .

6− **إيجاد الصورة العاملية المائلة P** : إن الصورة العاملية المخففة هي التعبير الخطي عن المتحولات المركبة U_1 و U_2 ، بدلالة العوامل المائلة t_1 و t_2 . وإن أمثال هذه الصورة هي إحداثيات نهايات أشعة المتحولات Z_j في الإحداثيات المائلة. ويمكن تحديد قيمها مباشرة من الصورة العاملية الأصلية A وذلك باستخدام العلاقة: $P = A * (T')^{-1}$ كما فعلنا في المثال السابق .

ولكن يمكننا حساب هذه الأمثال بعد الحصول على الهيكل العاملي المائل S وباستخدام العلاقة التي تربط الصورة المائلة بالهيكل العاملي والمعرفة بالعلاقة: $P = S * \emptyset^{-1}$.

ولهذا نكتب الصورة العاملية المائلة لمثالنا الحالي كما يلي :

$$Z_j = b_{j1}t_1 + b_{j2}t_2 \qquad (j: 123 \dots 8) \qquad (77-3)$$

وعلينا أن نحدد الأمثال b_{j1} و b_{j2} المقابلة لجميع المتحولات Z_j، لذلك نضرب طرفي العلاقة (3−77) بـ t_1 (ثم بـ t_2) ونأخذ مجموع الطرفين على جميع القيم N، ونقسم الناتج على N فنحصل على جملتين من المعادلات الخطية بأمثال من معاملات الارتباط هما:

$$r_{jt_1} = b_{j1} + b_{j2}r_{t_1t_2} \qquad (j: 123 \dots 8) \qquad (78-3)$$

$$r_{jt_2} = b_{j1}r_{t_2t_1} + b_{j2} \qquad (j: 123 \dots 8) \qquad (79-3)$$

فإذا كان الهيكل العاملي S معلوماً فإن المعاملات r_{jt_1} , r_{jt_2} تكون معلومة، كما تكون المعاملات $r_{t_2t_1}$, $r_{t_1t_2}$ بين العوامل نفسها معلومة، وهي في مثالنا الحالي تساوي ما يلي:

$$r_{t_1t_2} = r_{t_2t_1} = 0.4858$$

وذلك لأن جداء الشعاعين t_1 و t_2 يساوي:

$$\langle t_1, t_2 \rangle = 0.9137 * 0.7991 - 0.4064 * 0.6012 = 0.4858$$

وبذلك نجد أن مصفوفة أمثال جملة المعادلات لتحديد الشعاع المجهول b يكون لها الشكل التالي:

$$\emptyset = \begin{bmatrix} 1 & 0.4858 \\ 0.4858 & 1 \end{bmatrix}$$

وهي تبقى ثابتة لجميع المتحولات Z_j ، وتصلح لحساب قيم الأمثال b المقابلة لها من العلاقة $P = S * \emptyset^{-1}$.

وبصورة عامة واعتماداً على (3−56) نجد أن العلاقة التي تربط بين عناصر الصورة العاملية P والهيكل العاملي S في فضاء العوامل العامة هي العلاقة المستنبطة من العلاقة (1−55) وهي:

$$S = P * \emptyset \qquad (80 - 3)$$

ومنها يمكننا حساب عناصر الهيكل العاملي S إذا كانت عناصر المصفوفة $\emptyset$ وعناصر الصورة العاملية P معلومة .

ومن العلاقة السابقة يمكننا أن نحسب الصورة P بدلالة S من العلاقة المعكوسة كما يلي :

$$P = S * \emptyset^{-1} \qquad (81 - 3)$$

فإذا كانت عناصر الهيكل العاملي S معلومة، فإنه يمكننا حساب عناصر الصورة العاملية المائلة P ولكن لدينا من جهة أخرى أن:

$$\emptyset = T' * T \qquad \Rightarrow \quad \emptyset^{-1} = (T' * T)^{-1} = T^{-1} * T'^{-1}$$
$$S = A * T \qquad (82 - 3)$$

نعوض ذلك في (3−81) فنجد أن :

$$P = A * T * T^{-1} * T'^{-1}$$
$$P = A * (T')^{-1} \qquad (83 - 3)$$

وعليها يمكن الاعتماد في حساب الصورة العاملية المائلة P كما فعلنا في المثال السابق .

7− **حساب حصص العوامل المائلة من التشاركية الاجمالية:** نشير بداية إلى أن قيمة التشاركية الاجمالية للعوامل المائلة بعد التحويل تبقى ثابتة ومساوية لقيمة التشاركية الاجمالية للعوامل الأصلية. وهي في مثالنا تساوي (5.995) (انظر الجدول (3−3)) .

ولكن بعد أن نحصل على الصورة العاملية المائلة يظهر لنا نوعان من الحصص التشاركية منفردة ومختلطة، ويمكننا حساب هذه الحصص من التباين الكلي (أو من إجمالي التشاركية) لكل عامل من العوامل المائلة .

فحسب العلاقة (3−77) المعممة على m عاملاً نجد أن تشاركية المتحول Z_j، في الاحداثيات المائلة تساوي:

$$h_j^2 = b_{j1}^2 + b_{j2}^2 + \cdots + b_{jm}^2 + 2\sum_{p<q=1}^{m} b_{jp} * b_{jq} * r_{t_p t_q} \qquad (84-3)$$

$$(P: 123 \dots m - D)$$

$$h_j^2 = \sum_{P=1}^{m} b_{jp}^2 + 2\sum_{p<q=1}^{m} b_{jp} * b_{jq} * r_{t_p t_q}$$

وهنا نلاحظ أن هذه التشاركية مؤلفة من قسمين هما:

- مجموع مربعات الأمثال b_{jp}^2 وتسمى الحصص المنفردة للعوامل من التباين الكلي للمتحول Z_j .
- مجموع الجداءات $\left(2r_{t_p t_q} * b_{jp} * b_{jq}\right)$ وهي تمثل الحصص المختلطة للعوامل من التباين الكلي للمتحول Z_j، وهذا يعني أن حصة كل عامل تتحدد ليس فقط من الحصة المنفردة بل من العلاقة المتبادلة مع العوامل الأخرى .

وإذا أخذنا المجموع الكلي عمودياً لتشاركيات جميع المتحولات Z_j وفرزناها حسب العوامل t_p، فإننا نحصل على الحصة الاجمالية المنفردة للعامل t_p من العلاقة:

$$V_p = \sum_{j=1}^{n} b_{jp}^2 \qquad (P = 123 \dots .m) \qquad (85-3)$$

وعلى الحصة الاجمالية المختلطة من العلاقة :

$$V_{pq} = 2r_{T_p T_q} * \sum_{j=1}^{n} b_{jp} * b_{jq} \qquad (P < q) \qquad (86-3)$$

وهما تشكلان مصفوفة مثلثية يتألف قطرها الرئيسي من التشاركيات الاجمالية المنفردة ، وتتألف عناصرها غير القطرية من قيم التشاركيات المختلطة .

ولتوضيح ذلك نأخذ المتحول الأول Z_1 (الوزن) من مثالنا الحالي فنجد أنه يساوي بدلالة t_1 و t_2 كما يلي: (حسب الاحداثيات المائلة) : $Z_1 = 0.885t_1 + 0.059t_2$

ولكن معامل ارتباط Z_1 مع t_1 يساوي: $r_{1t_1} = 0.914$

وإن معامل ارتباط Z_1 مع t_2 يساوي: $r_{1t_2} = 0.489$

وإن حصة العامل t_1 من تباين المتحول Z_1 تساوي: $(0.885)^2 = 0.783$

وإن حصة العامل t_2 من تباين المتحول Z_1 تساوي: $(0.059)^2 = 0.003$

وإن حصتهما المختلطة تساوي: $2(0.885)(0.059)(0.4858) = 0.051$

وإن التشاركية الإجمالية لـ Z_1 تساوي: $h_1^2 = 0.783 + 0.003 + 0.051 = 0.837$

وبعبارة أخرى نقول أن 78.3% من تباين Z_1 (الوزن) يعود إلى العامل t_1 (القامة) وفقط 0.3% يعود إلى العامل t_2 (السمعة). وإن 5.1% من تباينه يعود إلى العاملين t_1 و t_2 معاً، أي أن هذه العوامل تفسر 83.7% من تباين Z_1 . ويبقى 16.3% من تباين Z_1 تعود إلى عوامل أخرى .

كما نجد أن قيم هذه التشاركيات في مثالنا الحالي تساوي:

$V_1 = 3.365$ و $V_2 = 2.611$ ، أما $V_{12} = -0.021$

8- الحصول على الحل الثاني المائل المناسب للشكل البسيط :

لقد لاحظنا أن الصورة العاملية في الحل الأول المائل تقترب كثيراً من الشكل البسيط لأن بعض عناصر ذلك الحل تصبح معدومة أو شبه معدومة (سالبة أو موجبة)، ولكن (*Thurstone*) قدم مخرجاً آخر يضمن ظهور عدد من الأصفار في الحل العاملي وذلك بإنشاء محاور جديدة متعامدة تخالفياً مع المحاور الأولى و تسمى المحاور الثانية ويتلخص ذلك في مثالنا السابق كما يلي:

- ننشأ من مبدأ الاحداثيات محوراً عمودياً على t_1 ونسميه $O\,\Lambda_2$.
- ننشأ من مبدأ الاحداثيات محوراً آخر عمودياً على t_2 ونسميه $O\,\Lambda_1$.

ومن الشكل (3−7) نلاحظ أن مساقط المتحولات الأربعة الأولى القريبة من المحور t_1 على المحور الجديد Λ_2 تساوي أو تقترب من الصفر، كما نلاحظ أن مساقط المتحولات الأربعة الأخرى القريبة من المحور t_2 على المحور الجديد Λ_1 تساوي أو تقترب من الصفر .

وهكذا سنحصل على صورة عاملية مائلة تحقق معظم شروط الهيكل البسيط. ويمكن الحصول عليها مباشرة من تدوير الصورة العاملية الأصلية *A* كما يلي:

لنفترض أنه لدينا حل عاملي أولي متعامد *A*، وأنه لدينا مصفوفة مجهولة Λ تساعدنا مباشرة على تحويل الحل *A* إلى هيكل مائل V معرف على محوري الإحداثيات الجديدة $\Lambda_1 o \Lambda_2$، فإن العلاقة (3−54) تأخذ الشكل التالي:

$$\mathrm{V} = A * \Lambda \qquad (87-3)$$

علماً بأن أعمدة المصفوفة Λ (كما أعمدة *T* في (2−13)) تتضمن الاتجاهات المثلثية(*COS*) للمحاور المائلة Λ_p بالنسبة لجملة المحاور القائمة F_1 , F_2 . وإذا علمنا عناصر مصفوفة التحويل *T* يمكننا حساب المصفوفة المطلوبة Λ بدون العودة إلى الشكل البياني .

إن عناصر أعمدة المصفوفة *T* ماهي إلا اتجاهات المحورين t_1 , t_2 المارين من مركزي المجموعتين G_1 , G_2، أما عناصر أعمدة المصفوفة Λ فيمكن النظر إليها وكأنها اتجاهات أشعة إضافية في جملة الاحداثيات الأصلية .

وبذلك يمكن اعتبار جملة الاحداثيات الجديدة $\Lambda_1 o \Lambda_2$ وكأنها توسع للصورة العاملية *A* (كما الصورة المخففة السابقة) . ولكن بما أن نسب الاتجاهات المثلثية في المصفوفة *T* موزعة على الأعمدة، فإنه

عندما نضربها بمصفوفة أخرى فإننا نقوم بعملية تحويل ، وباستخدام العلاقة التحويلية (3−87) السابقة وتطبيقها على منقول مصفوفة التحويل T نحصل على مصفوفة D تساوي:

$$D = T' * \Lambda \qquad (88-3)$$

وبما أن عناصر المصفوفة D تنتج من الجداءات السلمية لأعمدة T و Λ، فإن عناصرها هي معاملات الارتباط بين الأشعة t_p و Λ_p (P: 123 … m) ، كما إن المصفوفة D هي مصفوفة قطرية لأن عناصرها غير القطرية تكون معدومة بسبب تعامد الشعاعين t_1 مع V_2 على $o\Lambda_2$ وكذلك t_2 مع V_1 على $o\Lambda_1$.

وبذلك نجد أن المصفوفة Λ تساوي :

$$\Lambda = (T')^{-1} * D \qquad (89-3)$$

ومن جهة أخرى لدينا من العلاقة (3−82) أن: $\emptyset = T' * T$ فإنه يكون لدينا :

$$\emptyset^{-1} = T^{-1} * (T')^{-1} \qquad (90-3)$$

لنضرب طرفي العلاقة (3−90) من اليسار بـ T فنحصل على أن:

$$T * \emptyset^{-1} = T * T^{-1} * (T')^{-1} = (T')^{-1} \qquad (90\,a-3)$$

وهكذا نجد بعد تعويض $(T')^{-1}$ في (3−89) أن المصفوفة Λ تحسب من العلاقة التالية :

$$\Lambda = T * \emptyset^{-1} * D \qquad (91-3)$$

حيث D هي مصفوفة معاملات ارتباط t_p مع Λ_p، ومن جهة أخرى نلاحظ أن العلاقة (3−89) تشير إلينا أنه يمكن حساب عناصر المصفوفة Λ من خلال معيرة أعمدة المصفوفة $(T')^{-1}$ ووضعها أعمدة لـ Λ . وبعد الحصول على مصفوفة التحويل Λ يمكننا حسب العلاقة (3−87) الحصول على مصفوفة الهيكل العاملي الجديد V، وذلك بضرب مصفوفة الصورة العاملية الأصلية A من اليمين بالمصفوفة Λ فنحصل على أن:

$$\left(\text{الهيكل}\right) \qquad V = A * \Lambda \qquad (92-3)$$

والجدير بالذكر أن المصفوفة Λ تشبه المصفوفة T وتحقق علاقة مشابهة للعلاقة (3−67) وهي أن:

$$\Lambda' * \Lambda = \psi \qquad (93-3)$$

وللوصول إلى الحل العاملي الثاني، يجب علينا الحصول على الصورة العاملية المائلة له W بدلالة المحاور الاحداثية Λ_1, Λ_2 وهذا يتم من خلال تطبيق إحدى العلاقتين التاليتين :

$$\left(\text{الصورة}\right) \qquad W = V * \psi^{-1} = V * \Lambda^{-1} * \Lambda'^{-1} \qquad (94-3)$$

وبتعويض V من العلاقة (3−92) نحصل أخيراً على الصورة العاملية المائلة W بدلالة Λ_1, Λ_2 :

$$W = A * (\Lambda')^{-1} \qquad (95-3)$$

مثال (3−5): كتطبيق على ذلك نأخذ مثالنا السابق المؤلف من ثمانية متحولات، ونقوم أولاً بحساب عناصر المصفوفة Λ من العلاقة (3−91)، ولذلك نقوم بحساب المصفوفة $(T')^{-1}$ من العلاقة (3−90a) وذلك اعتماداً على المصفوفتين T و θ فنجد مما سبق أن:

$$(T')^{-1} = T * \emptyset^{-1} = \begin{bmatrix} 0.9137 & 0.7991 \\ -0.4064 & 0.6012 \end{bmatrix} * \begin{bmatrix} 1 & 0.4858 \\ 0.4858 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0.6879 & 0.4649 \\ -0.9142 & 1.0452 \end{bmatrix}$$

أي أن

$$(T')^{-1} = \begin{bmatrix} 0.6879 & 0.4649 \\ -0.9142 & 1.0452 \end{bmatrix}$$

وبإجراء عملية معيرة على أعمدة هذه المصفوفة (بتحويلها إلى أشعة واحدية) وذلك بتقسيم عناصر كل عمود فيها على طوله:

فنجد أن طول العمود الأول: $\ell_1 = \sqrt{0.6879^2 + 0.9142^2} = 1.144$

وأن طول العمود الثاني : $\ell_2 = \sqrt{0.4649^2 + 1.0452^2} = 1.1439$

وهكذا نجد أن عناصر المصفوفة Λ تساوي :

$$\Lambda = \begin{bmatrix} \frac{0.6879}{1.144} & \frac{0.4649}{1.1439} \\ \frac{-0.9142}{1.144} & \frac{1.0452}{1.1439} \end{bmatrix} = \begin{bmatrix} 0.6012 & 0.4064 \\ -0.7991 & 0.9137 \end{bmatrix}$$

وهي مصفوفة التحويل Λ إلى المحاور الإحداثية المائلة الجديدة $\Lambda_1 o \Lambda_2$ ، وللحصول على مصفوفة الهيكل العاملي الجديد V ، نضرب الصورة العاملية الأصلية A من اليمين بمصفوفة التحويل Λ فنحصل على ما يلي:

$$\mathrm{V} = \mathrm{A} * \Lambda = \begin{bmatrix} 0.856 & -0.324 \\ 0.846 & -0.412 \\ 0.808 & -0.409 \\ 0.813 & -0.342 \\ 0.750 & 0.571 \\ 0.631 & 0.492 \\ 0.569 & 0.510 \\ 0.607 & 0.351 \end{bmatrix} * \begin{bmatrix} 0.6012 & 0.4064 \\ -0.7991 & 0.9137 \end{bmatrix} = \begin{bmatrix} 0.774 & 0.052 \\ 0.839 & -0.032 \\ 0.813 & -0.045 \\ 0.773 & 0.025 \\ -0.005 & 0.827 \\ -0.014 & 0.706 \\ -0.065 & 0.607 \\ 0.084 & 0.567 \end{bmatrix}$$

وللحصول على الهيكل المخفف نطبق العلاقة:

$$\mathrm{S}_h = B * \Lambda = \begin{bmatrix} 0.8923 & -0.3969 \\ 0.7496 & 0.5639 \end{bmatrix} \begin{bmatrix} 0.6012 & 0.4064 \\ -0.7991 & 0.9137 \end{bmatrix} = \begin{bmatrix} 0.854 & -0.000 \\ -0.000 & 0.820 \end{bmatrix}$$

وللحصول على الصورة العاملية الجديدة W نطبق مباشرة العلاقة (3−94) فنجد أن:

$$\mathrm{W} = V * \psi^{-1}$$

لذلك نقوم بحساب ψ^{-1} كما يلي:

$$\psi^{-1} = (\Lambda' * \Lambda)^{-1}$$

$$\Lambda' * \Lambda = \begin{bmatrix} 0.6012 & -0.7991 \\ 0.4064 & 0.9137 \end{bmatrix} * \begin{bmatrix} 0.6012 & 0.4064 \\ -0.7991 & 0.9137 \end{bmatrix} = \begin{bmatrix} 1 & -0.4858 \\ -0.4858 & 1 \end{bmatrix}$$

$$\psi^{-1} = (\Lambda' * \Lambda)^{-1} = \begin{bmatrix} 1.3089 & 0.6359 \\ 0.6359 & 1.3089 \end{bmatrix}$$

وبذلك نجد أن الصورة العاملية في الإحداثيات المائلة W تساوي :

$$W = V * \psi^{-1} = \begin{bmatrix} 0.774 & 0.052 \\ 0.839 & -0.032 \\ 0.813 & -0.045 \\ 0.773 & 0.025 \\ -0.005 & 0.827 \\ -0.014 & 0.706 \\ -0.065 & 0.607 \\ 0.085 & 0.567 \end{bmatrix} \begin{bmatrix} 1.3089 & 0.6359 \\ 0.6359 & 1.3089 \end{bmatrix} = \begin{bmatrix} 1.046 & 0.559 \\ 1.079 & 0.492 \\ 1.034 & 0.458 \\ 1.027 & 0.524 \\ 0.513 & 1.079 \\ 0.431 & 0.915 \\ 0.358 & 0.871 \\ 0.471 & 0.796 \end{bmatrix}$$

وكان يمكن حساب الصورة العاملية الجديدة W من الصورة العاملية الأصلية A مباشرة من العلاقة التالية $W = A(\Lambda')^{-1}$ ، فنحصل على نفس الجواب السابق:

وللحصول على الصورة العاملية المخففة الجديدة نطبق العلاقة:

$$Ph = B(\Lambda')^{-1} = \begin{bmatrix} 0.8923 & -0.3969 \\ 0.7495 & 0.5639 \end{bmatrix} * \begin{bmatrix} 1.045 & 0.91423 \\ -0.46495 & 0.6878 \end{bmatrix} = \begin{bmatrix} 1.118 & 0.542 \\ 0.522 & 1.073 \end{bmatrix}$$

وأخيراً يمكننا حساب المصفوفة D من العلاقة:

$$D = T' * \Lambda = \begin{bmatrix} 0.9137 & 0.7991 \\ -0.4064 & 0.6012 \end{bmatrix} \begin{bmatrix} 0.6012 & 0.4064 \\ -0.7991 & 0.9137 \end{bmatrix} = \begin{bmatrix} 0.8741 & 0 \\ 0 & 0.8741 \end{bmatrix}$$

أي أن معامل الارتباط بين زوجي المحاور للعوامل Λ, t يساوي 0.8741

9- العلاقات بين الحلين المائلين الأول والثاني:

لقد أشرنا أيضاً إلى أن أي حل مائل يتألف من جزأين :

- مصفوفة الصورة العاملية P: وتتألف من أمثال المتحولات Z_j المرافقة للعوامل العامة F_p أو للعوامل t_p
- مصفوفة الهيكل العاملي S: ويتألف من معاملات الارتباط بين المتحولات المدروسة Z_j والعوامل العامة F_p أو العوامل t_p .

وحسب الرموز المستخدمة في الحلين الأول والثاني وجدنا أن:

$$P = A(T')^{-1} \qquad (96-3)$$

ومنها نجد أن الصورة العاملية الأصلية A تساوي:

$$A = P * T' \qquad (97-3)$$

ومن جهة أخرى وجدنا أن الهيكل العاملي V للحل الثاني يساوي :

$$V = A * \Lambda \qquad (98-3)$$

أي أن :

$$A = V * \Lambda^{-1} \qquad (99-3)$$

وبإجراء المطابقة بين العلاقتين (3-97) و(3-99) نجد أن:

$$P * T' = V * \Lambda^{-1} \qquad (100-3)$$

ومن جهة أخرى نجد من (3-89) أن:

$$\Lambda^{-1} = D^{-1} * T' \qquad (101-3)$$

أي أن:

$$P * T' = V * D^{-1} * T' \qquad (102-3)$$

وإذا ضربنا من اليمين طرفي العلاقة بـ $(T')^{-1}$ نجد أن العلاقة بين V , P تساوي:

$$P = V * D^{-1} \qquad (103-3)$$

وبضربها من اليمين بـ D نحصل على أن:

$$V = P * D \qquad (104-3)$$

وهكذا نجد أنه يمكننا من (3–103) حساب مصفوفة الصورة للحل الأول P من مصفوفة الهيكل العاملي V للحل الثاني ومن مقلوب مصفوفة معاملات الارتباط D بين عوامل الحل الأول t_p وعوامل الحل الثاني V . كما يمكن حساب V بدلالة P وD من العلاقة (3–104) .

وبطريقة مشابهة يمكننا التوصل على علاقتين مشابهتين بين مصفوفة الهيكل للحل الأول S مع الصورة العاملية للحل الثاني W فيكون لدينا ما يلي:

$$W = S * D^{-1} \qquad (105-3)$$

$$S = W * D \qquad (106-3)$$

3–3–3: طرائق الحصول على الحلول المائلة:

إن عملية الحصول على الحلول المائلة تجري ضمن شروط معقدة وتتطلب حسابات مطولة أكثر مما تتطلبه الحلول المتعامدة التي تبدو وكأنها حالة خاصة من الحلول المائلة .

وحتى نحصل على الحل المائل المناسب للهيكل البسيط يعتمد على نفس المعايير المستخدمة في الحلول المتعامدة، مع إلغاء شرط تعامد العوامل العامة، وهذا ما يجعل المعايير المتكافئة في الحلول المتعامدة (مثل Q , K) غير متكافئة في الحلول المائلة ، لذلك لا يمكن النظر للطرائق المائلة وكأنها تعميم للطرق المتعامدة. أما الطرائق المائلة فهي :

1– **طريقة** (Oblimax) :

تعتمد هذه الطريقة على إيجاد مصفوفة التحويل Λ، التي تقوم بتحويل مصفوفة الحل العاملي الأصلي A إلى مصفوفة الحل المائل النهائي للهيكلية V. أي التي تحقق العلاقة:

$$V = A * \Lambda \qquad (107-3)$$

حيث أن عناصر المصفوفة النهائية للهيكل العاملي V تحقق المعيار (K– max) التالي:

$$K = \frac{\sum_{j=1}^{n} \sum_{p=1}^{m} v_{jp}^4}{\left(\sum \sum v_{jp}^2\right)^2} \qquad (108-3)$$

وهذه العلاقة هي عبارة عن كتابة جديدة للعلاقتين (3–24) و(3–20) الواردتين في الفقرة السابقة، ولكن باستخدام رموز جديدة لتفريقها عن رموز الحلول المتعامدة .

وهنا يجب أن نتذكر أنه في الإحداثيات المائلة أن مصطلح الأمثال (الأثقال) المستخدم لعناصر الهيكل العاملي V ، يعني أنها تساوي معاملات الارتباط بين المتحولات Z_j والمجاور الإحداثية Λ_p، أي أن:

$$v_{jp} = r_{Z_j \Lambda_p} \qquad (109-3)$$

ويسمى الحل المائل V ، الذي يعظم قيمة المعيار K في $(3-108)$ بالحل المائل (Oblimax) ، ويتم حساب المساقط والمعاملات v_{jp} على المحاور الثانية Λ_p من خلال التدوير المتتالي في المستوى P، بحيث تأخذ القيمة الخاصة به K_p أكبر قيمة ممكنة. حيث أن قيمة K_p في المستوى P تساوي :

$$\mathrm{K}_p = \frac{\sum v_{jp}^4}{\left(\sum v_{jp}^2\right)^2} \qquad \Longrightarrow Max \qquad (110-3)$$

وعند إجراء عمليات التكرار والمعاودة فإن عناصر الهيكل العاملي v_{jp} تتغير، وإن قيمة التابع K تكبر بالتدريج حتى تأخذ أكبر قيمة لها .

وهنا نشير إلى أن العلاقة $(3-110)$ تعبر عن تأثير تعظيم v_{jp} على محور واحد للإحداثيات، أما العلاقة $(3-\ 108)$ فإنها تعطينا التأثير الاجمالي لجميع عمليات التعظيم المتتالية على الحل النهائي في V. وتتم معالجة ذلك على الحواسيب بواسطة برامج خاصة بذلك (ولذلك لم نتعرض للمعالجة الرياضية لهذه المسائل في هذا الكتاب) .

2- طريقة (Quartimin) :

وهي تعتمد على المعيار (N– min) في الحلول المتعامدة، ولكن العوامل بعد التدوير يمكن أن تكون غير متعامدة، وهنا علينا أن نختار عناصر المصفوفة V للهيكل العاملي، بحيث تجعل قيمة المعيار (N– min) أصغر ما يمكن حيث أن:

$$\mathrm{N} = \sum_{p<q=1}^{m}\sum_{j=1}^{n} v_{jp}^2 * v_{jq}^2 \qquad \Longrightarrow Min \qquad (111-3)$$

وعندها فإن التحويل $\mathrm{V} = \mathrm{A} * \Lambda$ يعطينا الحل (Quartimin) ويتم الحصول عليه بطريقة المعاودة على أعمدة المصفوفة Λ عموداً تلو الاخر .

3- **طريقة** (Oblimin):

لقد استنبط (Kaiser) هذه الطريقة من طريقة Varimax بعد أن ألغى شرط تعامد العوامل فيها وسماها بطريقة Kovarimin، وتوصل إلى معيار جديد هو تصغير قيمة التابع C^* التالي:

$$C^* = \sum_{p<q=1}^{m}\left(n\sum_{j=1}^{n} v_{jp}^2 v_{jq}^2 - \sum_{j=1}^{n} v_{jp}^2 * \sum_{j=1}^{n} v_{jq}^2\right) \qquad (112-3)$$

وهو يساوي التباين المشترك بين مربعات عناصر مصفوفة الهيكل V، وأخيراً اقترح (Kaiser) استبدال العلاقة السابقة بعلاقة تتضمن الأمثال المميرة لعناصر C^* وهي :

$$C = \sum_{p<q=1}^{m}\left[n\sum_{j=1}^{n}\left(\frac{v_{jp}^2}{h_j^2}\right)\left(\frac{v_{jq}^2}{h_j^2}\right) - \sum_{j=1}^{n}\frac{v_{jp}^2}{h_j^2} * \sum_{j=1}^{n}\frac{v_{jq}^2}{h_j^2}\right] \qquad (113-3)$$

وهنا تبدأ المعاودة بمعيرة المصفوفة العاملة الأصلية V حسب الأسطر، أي يتم معيرة أشعة المتحولات، أي إنه بعد الحصول على المصفوفة الجديدة V نحسب أطوال الأشعة السطرية في

المصفوفة V ونقوم بضرب عناصر أسطر المصفوفة V بأطوالها أو بقيم الجذور التربيعية لتشاركاتها فنحصل الصورة المطلوبة P ، وهناك شكل آخر لهذا المعيار هو : $B^* = N + \frac{C^*}{n}$

4- **طريقة** (Oblimin) المباشرة :

تعتمد هذه الطريقة على معالجة أمثال الحل المائل، فلقد انطلق (جينري وكيمبون) من أنه للحصول على حل ذي هيكل بسيط، لابد أن نحصل على أصغر قيمة لتابع مؤلف من عناصر الحل الأول للصورة العاملية ، فبدلاً من العلاقات السابقة استنبطا علاقة جديدة تحتوي على عناصر الهيكل وتشكل تابعاً لطريقة Oblimin المباشرة وهي:

$$F(P) = \sum_{p<q=1}^{m} \left(\sum_{j=1}^{n} b_{jp}^2 * b_{jq}^2 - \frac{\sigma}{n} \sum_{j=1}^{n} b_{jp}^2 * \sum_{j=1}^{n} b_{jq}^2 \right) \qquad (114-3)$$

حيث P هي مصفوفة الصورة العاملية الأولى: $P = [b_{jp}]$

ويمكن معيرة هذه الأمثال كما في العلاقة (3-113) السابقة، وللحصول على الحل المباشر (Oblimin) ، يجب علينا تصغير التابع F(p) .

وأخيراً نرفق جدولاً لمقارنة مفاهيم الطرق المتبقية في التدوير المتعامد وفي التدوير المائل مع الرموز والشروط المتقابلة لكل منهما .

جدول (3-8): مفاهيم طرائق التدوير ورموزها:

رموز الطرائق المائلة	رموز الطرائق المتعامدة	المفاهيم
العوامل غير متعامدة t_p p: 123 … m	العوامل متعامدة F_p p: 123 … m	العوامل الأصلية
$A = (a_{jp})\ (j: 123 \dots n)$	$A = (a_{jp})\ (j: 123 \dots n)$	مصفوفة الصورة العاملية (مصفوفة الأمثال)
$\Lambda = (\lambda_{pq})\ (p, q: 123 \dots m)$	$T = (t_{pq})\ (p, q: 123 \dots m)$	مصفوفة التحويل
$V = A * \Lambda = (v_{jp})$ V مصفوفة الهيكل العاملي الثاني (من الحل الثاني): $P = (b_{jp})$ P مصفوفة الصورة العاملية الأولى (من الحل الأول):	$B = A * T = (b_{jp})$	الحل النهائي
للحل الأول t_p للحل الثاني Λ_p	M_p	العوامل النهائية
K = Max N = Min C = Min F = Min B = Min D = Min	Q = Max M = Max N = Max K = Max V = Max	طرائق التدوير للحصول على الحل النهائي

المصدر: Harry. H. Harman (1972). Modern Factor Analysis, university of Chicago press Trans. Moscow Statistics P.337

الفصل الرابع

طرائق حساب العوامل العامة F_p بدلالة المتحولات Z_j

4-1 تمهيد:

تستخدم في التحليل العاملي عدة طرائق لتحديد وحساب قيم العوامل العامة F_p، وحساب أمثالها بدلالة المتحولات المشاهدة Z_j . وتعتمد هذه الطرائق على نوع النموذج المستخدم في التحليل العاملي .

فإذا كان النموذج المستخدم هو نموذج المكونات الأساسية (2-19)، الذي يتضمن n عاملاً عاماً F_p (بقدر عدد المتحولات Z_j) والذي يأخذ الشكل التالي [Harman,P.365] :

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jn}F_n \qquad (1-4)$$

حيث أن : $j: 123 \dots n$

فعندها يمكن كتابة المعادلات(4-1) بشكل مصفوفي كما يلي :

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} * \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix} \qquad (2-4)$$

أو على الشكل التالي:

$$Z = A * F \qquad (a2-4)$$

حيث أن A: هي مصفوفة الأمثال $[a_{jp}]$ وهي مصفوفة مربعة (في هذه الحالة) . وإذا كانت A غير شاذة (محددها لا يساوي الصفر) فإن مقلوبتها A^{-1} تكون موجودة، نضرب طرفي العلاقة $(a2-4)$ من اليسار بـ A^{-1} فنحصل على العوامل F_p كما يلي:

$$F = A^{-1} * Z = B * Z \qquad (3-4)$$

حيث رمزنا للمصفوفة A^{-1} بالرمز $B = (b_{jp})$ ، وإذا كانت A غير شاذة فإن المصفوفة B تكون معروفة. وعندها يمكن كتابة علاقة العوامل بدلالة المتحولات مصفوفياً كما يلي:

$$\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \dots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} * \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix} \qquad (4-4)$$

وبذلك يمكننا استخلاص العلاقات التي تعطينا العوامل $F_1F_2 \dots F_n$ من الجداء (4-4) كما يلي:

$$\begin{aligned} F_1 &= b_{11}Z_1 + b_{12}Z_2 + b_{13}Z_3 \dots b_{1n}Z_n \\ F_2 &= b_{21}Z_1 + b_{22}Z_2 + b_{23}Z_3 \dots b_{2n}Z_n \\ &\vdots \\ F_n &= b_{n1}Z_1 + b_{n2}Z_2 + b_{n3}Z_3 .. b_{nn}Z_n \end{aligned} \qquad (5-4)$$

وتستخدم هذه العلاقات لحساب القيم المختلفة للعوامل F_p، والتي نرمز لها بالرمز f_{pk} والمقابلة لكل جملة من قيم المتحولات $Z_j(\mathfrak{z}_{1k}\ \mathfrak{z}_{2k}\ \mathfrak{z}_{3k} \dots \mathfrak{z}_{nk})$، وإذا كانت A معلومة وغير شاذة فإن المعادلات السابقة تعطينا حلاً وحيداً لقيم العوامل f_{pk}، ولا يتعلق بأي تقديرات ممكنة لها(لأن عدد المعادلات المستقلة يساوي عدد المتحولات). وإن مجموع تباينات هذه العوامل تساوي تماماً التباين الاجمالي للمتحولات Z_j، أي أن مجموع التشاركيات h_j^2 لهذه العوامل يساوي n (بعدد المتحولات Z_j)، وذلك لأنه يفترض أن يكون تأثير عوامل التميز معدوماً، لأنها غير واردة في النموذج .

وأخيراً يمكننا أن نكتب أي عامل F_p بدلات المتحولات Z_j بدلالة علاقة ارتباطية (انحدارية) كما يلي:

$$\tilde{F}_p = b_{p1}\mathrm{Z}_1 + b_{p2}\mathrm{Z}_2 + b_{p3}\mathrm{Z}_3 \dots b_{pn}\mathrm{Z}_n \qquad (6-4)$$

حيث أن : $P{:}\,123 \dots n$ ،

ومنها يمكننا حساب قيم العوامل F_p المقابلة لقيم المتحولات Z_j . علماً بأن العوامل F_p و Z_j مأخوذة بصيغها المعيارية، لذلك فإن متوسطاتها تساوي الصفر وانحرافاتها المعيارية تساوي الواحد، وبما أنه يفترض أن تكون العوامل F_p غير مرتبطة مع بعضها البعض فإنها تحقق العلاقة :

$$F * F' = I \qquad (7-4)$$

وإذا قمنا بترتيب هذه العوامل F_p تنازلياً حسب تبايناتها أو حسب حصصها من التباين الاجمالي، فإن العامل ذا التباين الأكبر يسمى بالعامل الأول ونرمز له بـ F_1، والعامل ذا التباين الذي يليه يسمى بالعامل الثاني F_2، والذي يليه بالثالث F_3 ...الخ .

أما إذا كنا نريد التعامل مع النموذج المختصر(2−20)، فإننا ننطلق من أن تباينات كل من العوامل الأخيرة ستكون صغيرة أو مهملة . لذلك يمكننا السماح بمقدار معين من الخطأ وافتراض أن نموذج التحليل العاملي يتضمن m عاملاً عاماً فقط $(m < n)$ وهي العوامل الأولى من النموذج السابق، وهي التي تسمى بالعوامل الرئيسية. وعندها فإن نموذج التحليل العاملي يأخذ شكل العوامل الرئيسية (بدون حدود التميز) التالي :

$$\mathrm{Z}_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m \qquad (8-4)$$

حيث أن : $j{:}\,1,2,3, \dots n$

وإذا رمزنا لمصفوفة الأمثال بـ A ولعمود Z_j بـ Z ولعمود F_p بـ F يمكننا كتابة هذا النموذج بشكل مصفوفي كما يلي:

$$Z_{n*1} = A_{n*m} * F_{m1} \qquad (9-4)$$

وبما أن المصفوفة A في هذا النموذج ليست مربعة فإن مقلوبها غير معروف، وللتخلص من هذه المشكلة نضرب طرفي العلاقة من اليسار بمنقولها A' فنجد أن :

$$A' * Z = A' * A * F \qquad (10-4)$$

وبما أن A متعامدة فإن الجداء $(A' * A)$ يعطينا مصفوفة مربعة و قطرية، نرمز لها بالرمز $\Lambda = A' * A$ ، فإذا كانت Λ غير شاذة فإننا نقوم بضرب طرفي (4–10) بالمقلوب $\Lambda^{-1} = (A' * A)^{-1}$ ، فنحصل على أن:

$$(A' * A)^{-1} * A' * Z = (A^{-1} * A) * (A' * A) * F$$

والتي يمكننا كتابتها كما يلي:

$$F = \Lambda^{-1} * A' * Z \qquad (11-4)$$

وبما أن المصفوفة $\Lambda_m = A' * A_n$ هي مصفوفة قطرية ، لأن أعمدة A متعامدة بسبب تعامد العوامل F_p، فإن عناصر Λ القطرية هي القيم الذاتية لـ A وتساوي مجموع مربعات عناصر أعمدة A الموافقة لـF_p، وهي تمثل حصة تباين العامل الموافق F_p من التباين الاجمالي S^2 للمتحولات Z_j .

وإذا رمزنا لعناصر المصفوفة Λ القطرية بـ λ_p ، فإن مقلوبها Λ^{-1} هو أيضاً مصفوفة قطرية، وإن عناصرها مؤلفة من مقاليب العناصر القطرية وتساوي $\frac{1}{\lambda_p}$، وأخيراً يمكننا أن نكتب العلاقة (4–11) تفصيلياً كمايلي :

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_m \end{bmatrix} = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\lambda_3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\lambda_m} \end{bmatrix} * \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ a_{13} & a_{23} & \cdots & a_{n3} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ \vdots \\ Z_n \end{bmatrix} \qquad (12-4)$$

أو على الشكل التالي :

$$F_{m*1} = \Lambda^{-1}_{m*m} \quad , \quad A'_{m*n} \quad , \quad Z_{n*1} \qquad (a12-4)$$

وعند إجراء الضرب نجد أنه يمكننا أن نكتب كل عامل F_p على شكل علاقة جبرية بدلالة المتحولات Z_j والأمثال a_{jp} والقيم الذاتية λ_p على الشكل التالي:

$$\tilde{F}_p = \frac{a_{1p}}{\lambda_p} Z_1 + \frac{a_{2p}}{\lambda_p} Z_2 + \frac{a_{3p}}{\lambda_p} Z_3 + \cdots \frac{a_{np}}{\lambda_p} Z_n \qquad (13-4)$$

حيث أن: $P: 1,2,3, \dots m$، وهي تشكل m معادلة تمثل جميع العوامل F_p ، وهي شكل آخر للعلاقة السابقة (4–6) التالية:

$$\tilde{F}_p = b_{1p} Z_1 + b_{2p} Z_2 + b_{3p} Z_3 + \cdots + b_{np} Z_n \qquad (a\,6-4)$$

ملاحظة هامة:

نلاحظ من مطابقة العلاقتين (4–6) و(4–13) ، أن الأمثال b_{jp} تساوي : $b_{jp} = \frac{a_{jp}}{\lambda_p}$.

أي أنه يمكننا حساب أمثال المعادلة الخطية (4−6) التي تعطينا F_p بدلالة Z_j من العلاقة (4−13) بطريقة بسيطة جداً وهي: أن نأخذ الأمثال العمودية المقابلة للعامل F_p مع المتحولات Z_j في المصفوفة A ونقسمها على القيمة الذاتية λ_p المقابلة للعامل F_p. أي نطبق العلاقة:

$$b_{jp} = \frac{a_{jp}}{\lambda_p} \qquad (b\,6-4)$$

ولذلك علينا أن أولاً أن نوجد المصفوفة A ثم نحسب المصفوفة القطرية $\Lambda = A' * A$ فنحصل على القيم الذاتية لـ A وهي $\lambda_1, \lambda_2, \dots \lambda_m$ ثم نطبق العلاقة (4−13) مباشرة .

وكتطبيق على حساب مجموع مربعات عناصر عمود A المقابل لـF_p، وإيجاد قيمة كل λ نأخذ المثال التالي:

مثال (4−1): ليكن لدينا الحل العاملي A_{6*2} لستة متحولات على عاملين F_1 و F_2 كالتالي :

$$A_{6*2} = \begin{array}{c} \begin{array}{cc} F_1 & F_2 \end{array} \\ \begin{bmatrix} 0.839 & 0.345 \\ 0.775 & 0.419 \\ 0.772 & 0.315 \\ 0.428 & -0.731 \\ 0.486 & -0.658 \\ \underline{0.371} & \underline{-0.606} \end{bmatrix} \\ \begin{array}{cc} \lambda_1 & \lambda_2 \end{array} \end{array} \quad \Longrightarrow \quad \begin{aligned} \lambda_1 &= \sum_{j=1}^{6} a_{j1}^2 = 2.45656 \\ \lambda_2 &= \sum_{j=1}^{6} a_{j2}^2 = 1.72580 \\ \frac{1}{\lambda_1} &= 0.40707 \\ \frac{1}{\lambda_2} &= 0.57944 \end{aligned}$$

ومنه نحسب مجموع مربعات عناصر كل عمود ، فنحصل مباشرة على القيمتين الذاتيتين λ_1 و λ_2 وعلى مقلوبيهما المبينتين على اليمين أعلاه .

ويمكن حسابهما بطريقة أخرى ، حيث نجد أن A' تساوي:

$$A'_{2*6} = \begin{bmatrix} 0.839 & 0.775 & 0.772 & 0.428 & 0.486 & 0.371 \\ 0.345 & 0.419 & 0.315 & -0.731 & -0.658 & -0.606 \end{bmatrix}$$

ثم نقوم بحساب الجداء $A' * A$ فنجد أن:

$$A' * A = \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 2.45656 & 0 \\ 0 & 1.72580 \end{bmatrix}$$

فنحصل على نفس القيمتين لـ λ_p التي حسبناهما أعلاه من مجموع مربعات عناصر أعمدة A، ولحساب Λ^{-1} نجد أن:

$$\Lambda^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} = \begin{bmatrix} 0.40707 & 0 \\ 0 & 0.57944 \end{bmatrix}$$

ومنهما نجد أن عناصر الجداء $\Lambda^{-1} * A'$ تساوي:

$$\Lambda^{-1} * A' = \begin{bmatrix} 0.342 & 0.315 & 0.391 & 0.174 & 0.198 & 0.151 \\ 0.199 & 0.243 & 0.183 & -0.424 & -0.382 & -0.351 \end{bmatrix}$$

وبوضع نتائج هذه الحسابات في العلاقة (4−12) نجد أن:

$$F = \Lambda^{-1} * A' * Z$$

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0.342 & 0.315 & 0.391 & 0.174 & 0.198 & 0.151 \\ 0.199 & 0.243 & 0.183 & -0.424 & -0.382 & -0.351 \end{bmatrix} * \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \end{bmatrix}$$

ومنها نحصل على المعادلتين المعياريتين اللتين تعطيانا العاملين F_1 و F_2 بدلالة المتحولات المعيارية $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6$، حيث نجد من الجداء المصفوفي أن:

$$\widetilde{F_1} = 0.342Z_1 + 0.315Z_2 + 0.391Z_3 + 0.174Z_4 + 0.198Z_5 + 0.151Z_6$$
$$\widetilde{F_2} = 0.199Z_1 + 0.343Z_2 + 0.183Z_3 - 0.424Z_4 - 0.382Z_5 - 0.351Z_6$$

ملاحظة هامة (1): كان يمكن تطبيق (4−13) واستخلاص هاتين المعادلتين مباشرة من مصفوفة الصورة العاملية A ، وذلك بتقسيم عناصر عمودي F_1 و F_2 فيها على القيم الذاتية المقابلة لها λ_1 و λ_2 (أي على حصتيهما من التباين) على الترتيب، ثم نضع العناصر الناتجة أمثالاً للمتحولات Z_j المقابلة لها فنحصل تماماً على المعادلتين السابقتين.

ومن هاتين المعادلتين يمكننا أن نحصل على تقديرات عددية لقيم F_1 و F_2 المقابلة لأي جملة من قيم المتحولات المعيارية $(Z_1, Z_2, Z_3, Z_4, Z_5, Z_6)$، فمثلاً إذا كانت الجملة الأولى من قيم المتحولات Z_j هي: (0.10, 0.05, 0.10 ,0.80, 0.70, 0.50)، فإننا بعد التعويض في المعادلتين السابقتين نجد أن $\widetilde{F_1}$ و $\widetilde{F_2}$ يأخذان القيمتين التالتين:

$$\tilde{f}_{11} = 0.342(0.10) + 0.315(0.05) + 0.391(0.10) + 0.174(0.80) + 0.198(0.70) + 0.151(0.50) = 0.442$$
$$\tilde{f}_{12} = 0.199(0.10) + 0.343(0.05) + 0.183(0.10) - 0.424(0.80) - 0.382(0.70) - 0.351(0.50) = -0.732$$

وعند تعويض جميع جمل القيم المشاهدة للمتحولات المدروسة Z_j في العلاقة (4−13)، فإننا نحصل على عمودين للقيم العددية للعاملين F_1 و F_2، ويتضمنان N قيمة في كل منهما (لكل منهما قيمة مقابل كل حالة)، وللتأكد من تعامد هذين العاملين نقوم بحساب المصفوفة الارتباطية لها، والتي يجب أن تشكل مصفوفة قطرية واحدية I أو مصفوفة قريبة منها. أي أن ϕ يجب أن تساوي :

$$\phi \approx \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (14-4)$$

ملاحظة (2): يمكن تحويل المعادلتين المعياريتين (4−13) إلى معادلتين أصليتين بدلالة قيم X الأصلية وذلك باستبدال كل Z_j بما تساويه من العلاقة $Z_j = \frac{X_j - \bar{x}_j}{\sigma_j}$، فنحصل في النهاية على تقديرات أصلية لـ F_p من العلاقة (4−13) تساوي :

$$F_p^* = \frac{a_{1p}}{\lambda_p}\left(\frac{X_1 - \bar{x}_1}{\sigma_1}\right) + \frac{a_{2p}}{\lambda_p}\left(\frac{X_2 - \bar{x}_2}{\sigma_2}\right) + \cdots \frac{a_{np}}{\lambda_p}\left(\frac{X_n - \bar{x}_n}{\sigma_n}\right) \qquad (15-4)$$

وبعد الاصلاح نجد أنها تساوي:

$$F_p^* = \frac{a_{1p}X_1}{\lambda_p * \sigma_1} + \frac{a_{2p}X_2}{\lambda_p * \sigma_2} + \cdots \frac{a_{np}X_n}{\lambda_p * \sigma_n} - C_p \qquad (16-4)$$

حيث أن الثابت C_p يساوي:

$$C_p = \frac{a_{1p}\bar{x}_1}{\lambda_p * \sigma_1} + \frac{a_{2p}\bar{x}_2}{\lambda_p * \sigma_2} + \cdots \frac{a_{np}\bar{x}_n}{\lambda_p * \sigma_n}$$

2-4 طرائق حساب أمثال العوامل F_p :

هناك عدة طرائق لحساب هذه الأمثال واهمها هي:

4-2-1: طريقة الانحدار المتعدد MultiRegression

تعتمد هذه الطريقة على مفاهيم الانحدار المتعدد، وعلى النموذج العاملي المختصر (2-20)، وتنطلق من وجود علاقة خطية لأي عامل F_p كتابع مجهول مع n متحولاً معيارياً ومعلوماً Z_j ، وتكتب على الشكل التالي:

$$F_p = \beta_{p_1}Z_1 + \beta_{p_2}Z_2 + \cdots \beta_{p_n}Z_n \qquad (17-4)$$

حيث أن : $P: 123 \ldots m$

وحسب نظريات تحليل الانحدار المتعدد فإن الأمثال المعيارية β_{p_j} تحسب من جملة المعادلات الخطية التالية:

$$\begin{aligned} &\beta_{p_1} + r_{12}\beta_{p2} + r_{13}\beta_{p3} + \cdots + r_{1n}\beta_{pn} = S_{1p} \\ &r_{21}\beta_{p1} + \beta_{p_2} + r_{13}\beta_{p3} + \cdots + r_{2n}\beta_{pn} = S_{2p} \\ &\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \\ &r_{n1}\beta_{p1} + r_{n2}\beta_{p_2} + r_{n3}\beta_{p3} + \cdots + \beta_{pn} = S_{np} \end{aligned} \qquad (18-4)$$

حيث أن: r_{jk} هو معامل الارتباط بين أي متحولين Z_j و Z_k وأن $(r_{jj} = 1)$ ، علماً بأن هذه المعاملات معلومة أو يمكن حسابها مباشرة من قيم المتحولات Z_j، ولكن المعاملات $S_{jp} = r_{Z_j p}$ وهي معاملات الارتباط بين المتحولات Z_j والعوامل F_p في المصفوفة الهيكلية S ، وهي معاملات مجهولة لأن قيم العوامل F_p مازالت مجهولة .

وهنا نلاحظ أن المعاملات المرافقة للمقادير المجهولة β_{p_j} تشكل المصفوفة المتناظرة لمعاملات الارتباط الميدانية r_{jk}، ومنها يمكننا حساب أو تقدير قيم الأمثال β_{p_j} من خلال العلاقات التالية (حسب طريقة المحددات) :

$$\beta_{p_j} = \frac{-|\Delta_{jp}|}{|R|} \qquad \begin{pmatrix} P: 1,2,3 \ldots m \\ i: 1,2,3 \ldots \ldots n \end{pmatrix} \qquad (19-4)$$

حيث أن $|R|$: هي قيمة محدد المصفوفة الارتباطية للمتحولات Z_j .

وأن $|\Delta_{jp}|$: هي قيمة محدد المتمم الجبري للعنصر S_{jp} في المصفوفة Δ الموسعة من R بتطويقها بالعناصر S_{jp} كما يلي :

$$\Delta=\begin{bmatrix} 1 & S_{1p} & S_{2p} & \dots & S_{np} \\ S_{1p} & 1 & r_{12} & \dots & r_{1n} \\ S_{2p} & r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ S_{np} & r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix} \;,\; R=\begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & \dots & 1 \end{bmatrix}$$

ومن جهة أخرى يمكننا أن نعبر عن المحددات $|\Delta_{jp}|$ بدلالة العناصر S_{jp} والمتممات الجبرية $|R_{kj}|$ المقابلة لها من المصفوفة الارتباطية R، وهكذا يمكننا كتابة (4−19) كما يلي:

$$\beta_{p_j}=\frac{S_{1p}|R_{1j}|+S_{2p}|R_{2j}|+S_{3p}|R_{3j}|+\dots+S_{np}|R_{nj}|}{|R|} \qquad (20-4)$$

حيث أن: $|R_{kj}|$ هو المتمم الجبري للعنصر r_{kj} في المصفوفة الارتباطية R .

وبتعويض (4−20) في العلاقة (4−17) وكتابتها بشكل مصفوفي نحصل على معادلة لحساب قيم العامل F_p بدلالة المتحولات Z_j وتكتب مصفوفياً كما يلي:

$$\tilde{F}_p=S'_p*R^{-1}*Z \qquad P:1,2,3\dots m \qquad (21-4)$$

حيث أن: S_p هو الشعاع العمودي $S_p(S_{1p},S_{2p},S_{3p},\dots,S_{np})$ المأخوذ من العمود P في المصفوفة الهيكلية S للارتباط بين العوامل F_p والمتحولات Z_j .

وبتوحيد العلاقات (4−21) في علاقة مشتركة، نحصل على الصيغة المصفوفية التالية:

$$\tilde{F}=S'*R^{-1}*Z \qquad (22-4)$$

وبما أن مصفوفة الهيكل S بشكل عام ترتبط مع الصورة A بالعلاقة $S=A*\phi$ وأن ϕ متناظرة، فإن العلاقة (4−22) تأخذ الشكل التالي:

$$\tilde{F}=\phi*A'*R^{-1}*Z \qquad (23-4)$$

حيث أن ϕ: هي مصفوفة المعاملات الارتباطية بين العوامل F_p .

ولكن عندما تكون العوامل F_p غير مرتبطة (متعامدة) فإن المصفوفة ϕ تصبح مصفوفة واحدية $(\phi=I)$ وعندها فإن العلاقة (4−23) تأخذ الشكل التالي:

$$\tilde{F}=A'*R^{-1}*Z \qquad (24-4)$$

وبعد الحصول على هذه العلاقة الانحدارية لكل عامل F_p نقوم بحساب معامل الارتباط المتعدد R_p^2 من العلاقة التالية (أو غيرها) .

$$R_p^2=\beta_{p1}S_{1p}+\beta_{p2}S_{2p}+\dots+\beta_{pn}S_{np} \qquad (25-4)$$

وكلما كانت قيمة R_p^2 قريبة من الواحد كان التمثيل جيداً .

وأخيراً يمكننا الحصول على معادلات تقدير العوامل F_p في أي إحداثيات أخرى بواسطة التدوير . لذلك نفترض أن هناك علاقة بين الصورة A الأصلية في الإحداثيات الأولى F_p وبين الصورة المدورة B في الاحداثيات الثانية M_p ، وإن مصفوفة التحويل بينهما هي T . فعندها يجب أن يكون لدينا :

$$B=A*T \qquad (26-4)$$

ومن جهة أخرى يمكننا أن نكتب المتحولات الأصلية Z_j في الاحداثيات الجديدة M_p كما يلي:

$$Z = B * M \qquad (27-4)$$

حيث رمزنا للإحداثيات الجديدة بـ $M_1, M_2, \dots M_k$، وبما أن B غير مربعة (بصورة عامة) فإننا نضرب من اليسار الطرفين بـ B' فنحصل على مصفوفة مربعة $(B' * B)$، والعلاقة (4−27) تصبح كما يلي:

$$B' * Z = B' * B * M \qquad (28-4)$$

ثم نضرب الطرفين بـ $(B' * B)^{-1}$ من اليسار فنحصل على أن :

$$M = (B' * B)^{-1} * B' * Z \qquad (29-4)$$

ومنها نحصل على معادلات انحدار العوامل M_p في الاحداثيات الجديدة كما يلي:

$$\widetilde{M}_p = b_{1p}Z_1 + b_{2p}Z_2 + \cdots + b_{np}Z_n \qquad (30-4)$$

حيث أن: $P: 1,2,3, \dots, m$

ومنها يمكننا حساب القيم العددية للعوامل M_p وذلك بتعويض القيم العددية لـ Z_j في (4−30)، فنحصل على N قيمة لكل من العوامل M_p .

وكتطبيق على ذلك نأخذ المثال (4−3) ذي خمسة المتحولات، والذي تم تدوير صورته العاملية A من الاحداثيات المتعامدة F_p إلى الإحداثيات M_p فأصبحت صورته المائلة B كما يلي:

$$B = \begin{bmatrix} 0.0160 & 0.9938 \\ 0.9408 & -0.0088 \\ 0.1370 & 0.9801 \\ 0.8248 & 0.4471 \\ 0.9682 & -0.0060 \end{bmatrix}$$

ومنها نجد أن B' تساوي :

$$B' = \begin{bmatrix} 0.0160 & 0.9408 & 0.1370 & 0.8248 & 0.9682 \\ 0.9938 & -0.0088 & 0.9801 & 0.4471 & -0.0060 \end{bmatrix}$$

وأن :

$$B' * B = \begin{bmatrix} 2.2218 & 0.5048 \\ 0.5048 & 2.1481 \end{bmatrix}$$

$$(B' * B)^{-1} = \begin{bmatrix} 0.4161 & -0.0978 \\ -0.0978 & 0.4885 \end{bmatrix}$$

$$(B' * B)^{-1} * B' = \begin{bmatrix} -0.0905 & 0.3923 & -0.0388 & 0.2995 & 0.4035 \\ 0.4839 & -0.0963 & 0.4654 & 0.1378 & -0.0976 \end{bmatrix}$$

وهكذا نجد أن العلاقة (4−30) تعطينا المعادلتين الانحداريتين للعوامل الجديدة M_p كما يلي:

$$M_1 = -0.0905Z_1 + 0.3923Z_2 - 0.0388Z_3 + 0.2995Z_4 + 0.4035Z_5$$
$$M_2 = 0.4839Z_1 - 0.0963Z_2 + 0.4654Z_3 + 0.1378Z_4 - 0.0976Z_5$$

وبتعويض القيم المعيارية للمتحولات Z_j نحصل على القيم العددية للعوامل M_p في الإحداثية الجديدة M_p

4−2−2: طريقة (Bartlitt)

تعتمد هذه الطريقة على إيجاد تقدير قيم العوامل F_p بواسطة تصغير قيم عوامل التميز U، ولتقديم هذه الطريقة نعتمد على النموذج المختصر التالي:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + d_jU_j \qquad (31-4)$$

ومنها يمكننا حساب عوامل التميز فنجد أن:

$$U_j = \frac{\left(Z_j - a_{j1}F_1 - a_{j2}F_2\right)}{d_j} \qquad (32-4)$$

ثم نشكل تابع مجموع مربعاتها على جميع المتحولات Z_j فنجد أن:

$$U\left(F_1, F_2\right) = \sum_{j=1}^{n} U_j^2 = \frac{\sum^n \left(Z_j - a_{j1}F_1 - a_{j2}F_2\right)^2}{d_j^2} \qquad (33-4)$$

ولإيجاد أصغر قيمة للتابع $U[F_1, F_2]$ نقوم باشتقاقه بالنسبة لـ F_2 و F_1 ونضع المشتقين مساويين للصفر، فنجد أن:

$$\begin{aligned} \frac{\partial U}{\partial F_1} &= 2\sum \frac{1}{d_j^2}\left(Z_j - a_{j1}F_1 - a_{j2}F_2\right) * \left(-a_{j1}\right) = 0 \\ \frac{\partial U}{\partial F_2} &= 2\sum_{j=1}^{n} \frac{1}{d_j^2}\left(Z_j - a_{j1}F_1 - a_{j2}F_2\right) * \left(-a_{j2}\right) = 0 \end{aligned} \qquad (34-4)$$

وبعد الاصلاح يمكن كتابة هاتين العلاقتين على الشكل التالي:

$$\begin{aligned} \sum^{n} \left(\frac{a_{j1}^2}{d_j^2}\right)\tilde{F}_1 + \sum^{n} \left(\frac{a_{j1}a_{j2}}{d_j^2}\right)\tilde{F}_2 &= \sum^{n} \frac{a_{j1}}{d_j^2} Z_j \\ \sum \left(\frac{a_{j1}a_{j2}}{d_j^2}\right)\tilde{F}_1 + \sum^{n} \left(\frac{a_{j1}^2}{d_j^2}\right)\tilde{F}_2 &= \sum^{n} \frac{a_{j2}}{d_j^2} Z_j \end{aligned} \qquad (35-4)$$

ومنهما يمكننا حساب العاملين $\tilde{F}_1$ و $\tilde{F}_2$ بدلالة المتحولات Z_j .

ويمكن تعميم هذه الطريقة على m عاملاً. وعندها فإن العلاقتين (4–35) تأخذان الشكل المصفوفي التالي :

$$(A' * D^{-2} * A) * F = A' * D^{-2} * Z \qquad (36-4)$$

وبفرض أن المصفوفة $(A' * D^{-2} * A)$ غير شاذة فإن مقلوبها يكون موجوداً، وبضرب طرفي (4–36) بذلك المقلوب نحصل على المعادلة التي تعطينا العوامل F وهي:

$$\tilde{F} = (A' * D^{-2} * A)^{-1} * A' * D^{-2} * Z \qquad (37-4)$$

ولتطبيق هذه الطريقة علينا إجراء ما يلي:

1- الحصول على الصورة A ثم إيجاد المنقول A' وحساب عناصر المصفوفة D^2.
2- حساب D^{-2} ثم حساب $A' * D^{-2}$.
3- حساب الجداء $(A' * D^{-2} * A)$ ثم حساب المقلوب $(A' * D^{-2} * A)^{-1}$.
4- حساب الجداء النهائي: $(A' * D^{-2} * A)^{-1} * A * D^{-2}$.
5- حساب قيم F من العلاقة (4–37) وذلك بتعويض قيم Z فيها .

4–2–3: الطريقة المتسارعة (طريقة Lederman, Anderson, Rubin)

وهي تعتمد على النموذج (2–20) وعلى العلاقة العامة التي تربط بين المصفوفة الارتباطية R والمصفوفة العاملية الموسعة M المعرفة بالعلاقة (1–66) التالية :

$$R = M * \begin{bmatrix} \phi & 0 \\ 0 & I \end{bmatrix} M' \qquad (38-4)$$

حيث أن M هي المصفوفة الموسعة $M = [A|, D]$ وتساوي :

$$M = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} & d_1 & 0 & 0 & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & d_2 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & 0 & d_n \end{bmatrix} = (A, D) \quad (39-4)$$

وهكذا يمكننا أن نكتب العلاقة (4−31) كما يلي:

$$R = (A, D) \begin{bmatrix} \phi & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A' \\ D \end{bmatrix} = [A * \phi \,, D] \begin{bmatrix} A' \\ D \end{bmatrix} = A * \phi * A' + D^2 \qquad (40-4)$$

وتهدف هذه العلاقة إلى تبسيط العلاقة الانحدارية (4−23) واختصار الحسابات في تقدير قيم العوامل العامة F :

وسنناقش الحالة العامة التي يمكن أن يكون فيها نظام الاحداثيات مائلاً ويتضمن mعاملاً مائلاً كما يلي:

لنضرب الآن طرفي (4−40) من اليسار بـ $A' * D^{-2}$ فنحصل على أن:

$$A' * D^{-2} * R = A' * D^{-2} * A * \phi * A' + A' * D^{-2} * D^2$$

$$A' * D^{-2} * R = [A' * D^{-2} * A * \phi + I] * A' \qquad (41-4)$$

وإذا رمزنا للجداء داخل القوس بـ K فيكون لدينا:

$$K = A'_{m*n} * D^{-2}_{n*n} * A_{n*m} * \phi_{m*m} \qquad (42-4)$$

وهو عبارة عن مصفوفة من المرتبة m*m .

وعندها يمكننا أن نكتب العلاقة (4−41) السابقة كما يلي:

$$A' * D^{-2} * R = [K + I] * A' \qquad (43-4)$$

ولنضرب الآن طرفي العلاقة (3−43) من اليسار بـ $(K + I)^{-1}$ فنجد أن:

$$(K + I)^{-1} * A' * D^{-2} * R = A'$$

ثم نضربها من اليمين بالمقلوب R^{-1} ونبدل مواقع الطرفين فنحصل على أن:

$$A' * R^{-1} = (I + K)^{-1} * A' * D^{-2} * I \qquad (44-4)$$

ومن جهة أخرى لدينا من العلاقة (4−23) أن :

$$F = \phi * A' * R^{-1} * Z \qquad (45-4)$$

لذلك نعوض $A' * R^{-1}$ في العلاقة (4−45) فنحصل على أن:

$$\tilde{F} = \phi * (I + K)^{-1} * A' * D^{-2} * Z \qquad (46-4)$$

ويمكن كتابة هذه العلاقة بين قيم أعمدة $\tilde{F}$ وقيم أعمدة Z، فنحصل على قيم هذه الأعمدة ونكتب ذلك كما يلي:

$$\tilde{\mathfrak{f}} = \phi * (I + K)^{-1} * A' * D^{-2} * \mathfrak{z} \qquad (47-4)$$

ولتبسيط العلاقة (4−45) نضرب طرفيها بـ $[\phi(I + K)^{-1}]^{-1}$ والذي يساوي $(I + K) * \phi^{-1}$ فنحصل على أن:

$$(I + K) * \phi^{-1} * F = A'_{m*n} * D^{-2} * Z \qquad (48-4)$$

إن نتيجة العلاقة السابقة (4–48) تعطينا عمودين من مرتبة (m*1) في كل طرف وبإجراء المطابقة بينهما نحصل على m معادلة خطية .

والطرف الأيمن يتضمن مصفوفات بسيطة. أما الطرف الأيسر فيكون أكثر تعقيداً. لذلك نقوم بتبسيط الطرف الأيسر بإعادة الرمز K إلى ما يساويه من (4–42) فنجد أن:

$$(I+K)*\phi^{-1}*F=(I+A'*D^{-2}*A*\phi)\phi^{-1}*F$$
$$=(\phi^{-1}+A'*D^{-2}*A)*F=A'*D^{-2}*Z$$

وبوضع $J=A'*D^{-2}*A$ ثم $L=\phi^{-1}+J$ يكون لدينا:

$$L*F=A'*D^{-2}*Z \qquad (49-4)$$

ومنها نستنتج أن F تقدر من العلاقة العامة التالية:

$$\tilde{F}=L^{-1}*A'*D^{-2}*Z \qquad (50-4)$$

أما عندما تكون العوامل F_p غير مرتبطة (متعامدة) فإن $\phi=\phi^{-1}=I$، وعندها فإن العلاقة (4–50) تأخذ الشكل التالي:

$$\tilde{F}=(I+J)^{-1}*A'*D^{-2}*Z \qquad (51-4)$$

وعند استخدام العلاقة (4–50) في المسائل العملية ينصح اتباع الخطوات التالية :

1- نحسب مصفوفة الصورة الأصلية A .
2- نقسم كل عناصر السطر j من A على d_j^2 فنحصل على $D^{-2}*A$.
3- نضرب من اليسار $D^{-2}*A$ في المنقول A' فنحصل على $J=A'*D^{-2}*A$.
4- نحسب المصفوفة ϕ ثم ϕ^{-1} ونضعهما مع المصفوفة J فنحصل على المصفوفة L .
5- نحسب المقلوب L^{-1} .
6- نحسب $A'*D^{-2}$ وذلك بتقسيم كل عناصر السطر j من A' على d_j^2 .
7- نحسب الجداء $(L^{-1}*A'*D^{-2}*\mathfrak{z})$ فنحصل على عمود القيم $\tilde{\mathfrak{f}}$.

أما إذا كانت العوامل غير مرتبطة فتكون ϕ واحدية، فعندها نقوم بحساب L من $L=I+J$ ، ثم نقوم بحساب L^{-1} ونضربها من اليسار بـ $A'*D^{-2}$ فنحصل على معادلة قيم F كما في العلاقة (4–51).

4–2–4: طريقة المتحولات المثالية (طريقة *Harman*)

هي عبارة عن أسلوب آخر لتقدير العوامل يعتمد على التعبير عن العوامل بواسطة مساقط متحولات (مثالية) Z_j^* في فضاء العوامل العامة. وبذلك يمكننا كتابة العلاقة $Z=A*F$ بدلالة القيم على الشكل التالي:

$$\mathfrak{z}^*=A*\mathfrak{f} \qquad (52-4)$$

حيث $\mathfrak{z}^*$ هو شعاع عمود مؤلف من مساقط المتحولات على محاور فضاء العوامل العامة ونرمز له بـ $\mathfrak{z}_1^*,\mathfrak{z}_2^*,\dots,\mathfrak{z}_n^*$، وبناء على ذلك يمكننا التعبير عن قيم العوامل $\mathfrak{f}$ بواسطة علاقات خطية بدلالة متحولات إفتراضية Z_j^* .

لنضرب الآن من اليسار العلاقة (4−52) بالمصفوفة A' فنجد أن :

$$A' * \mathfrak{z}^* = A' * A * \mathfrak{f} \qquad (53 - 4)$$

وبما أننا رمزنا للجداء $A' * A$ بالرمز Λ_m للدلالة على المصفوفة القطرية المتضمنة m قيمة ذاتية هي $\lambda_1 , \lambda_2 , \dots , \lambda_m$، فإننا نكتب الجداء $A' * A$ كما يلي :

$$A' * A = \Lambda_m = Diag\,(\lambda_1 , \lambda_2 , \dots , \lambda_m) \qquad (54 - 4)$$

وإذا ضربنا طرفي العلاقة (4−53) من اليمين بـ $(A' * A)^{-1}$ نحصل على أن $\tilde{\mathfrak{f}}$ تساوي :

$$\tilde{\mathfrak{f}} = (A' * A)^{-1} * A' * \mathfrak{z}^* = \Lambda^{-1} * A' * \mathfrak{z}^* \qquad (55 - 4)$$

وهي علاقة شبيهة بالعلاقة الأولى (4−11) للعوامل الرئيسية. والفرق الوحيد بينهما هو استخدام المتحولات (المثالية) بدلاً عن المتحولات المشاهدة .

وهذه القيم – كما هو متوقع– تختلف عن قيم العاملين الناتجين عن طريق الانحدار المتعدد أو عن الطريقة المتسارعة ولكنها قريبة منها. وإن سبب ذلك التقارب هو أن قيم متحولات التميز U تكون صغيرة، وهذا يؤدي بدوره إلى تقارب المتحولات المثالية مع المتحولات المشاهدة .

وأخيراً نشير إلى أن طريقة الانحدار المتعدد هي الأكثر استخداماً وكفاءة رغم تعقيدات الحسابات فيها. وقد تكون تقديراتها متحيزة أو تكون قيم العوامل فيها مترابطة. أما طريقة Bartlitt فتتميز بأن تقديراتها تكون غير متحيزة وغير مترابطة (إلا في حالة التعامد). أما تقديرات طريقة Anderson فتكون معيارية (ذات متوسطات =0 ، وتباين =1) ، وقد تكون مترابطة بينها في حالة التعامد أو مترابطة مع عوامل التميز U .

المراجع العلمية:

1- **Frank, wood, (2009). Priacipal Component Anatysis, Jolliffe 2.edit North Scott.**

2- **Harman, H. Harry; Modern Factor Analysis, secd. Edit. Trans. Russian . Statistica, Moscow. 1972.**

3- **Harris, H. R. A Primer of Multivariate Statistics, LEA. New york third Edit 2001.**

4- **Harman, Harry. (1976) Modern Factor analysis, 3Ed. Chicaqo press (QA278,5,H38) 1976 .**

5- **Okon, Jan. Factor Analysis, Trans, Russian, statistica Moscow 1974.**

6- **Uberla,J. Factor Analysis, Trans. Russian, Statistica. Moscow 1980.**

7- ابو فايد، أحمد: التحليل العاملي مفاهيمه وتطبيقاته في SPSS ، جامعة الأزهر، غزة، 2016 .

أسس
التحليل الإحصائي متعدد المتغيرات

الجزء الخامس
قضايا التمييز والتصنيف

7-6: طرائق التحليل العنقودي الهرمي التجميعي (الربط المنفرد- الربط التام- الربط المتوسط)

7-7: التحليل العنقودي غير الهرمي وطرائقه .

7-7-1: معايير العنقدة .

7-7-2: طريقة K متوسطا .

7-7-3: طرائق التجزئة السريعة .

7-7-4: طريقة النمذجة المختلطة .

7-7-5: طريقة النمذجة غير الخطية .

7-7-6: طريقة K متوسطا الغامضة .

المراجع المعتمدة:

الفصل الأول
المفاهيم العامة للتحليل التمييزي

1-1 : تمهيد:

إن كل إنسان يمارس التحليل التمييزي يومياً وبصورة لاشعورية . فعندما نسير في الشارع نميز فوراً بين الناس حسب الجنس ونفرزهم إلى ذكور وإناث، كما نميز بينهم حسب العمر ونفرزهم دون أن نسألهم إلى كبار وشباب وأطفال . والطبيب عندما يفحص المرضى يميز بينهم بسرعة حسب حالاتهم الصحية (مريض أو سليم)، وذلك قبل إجراء الدراسة الطبية اللازمة، وأحياناً يعالجهم دون إجراء التحاليل لهم، والمتجول في الغابة يميز بين أنواع الأشجار بسرعة، والشرطي الذي يقف على بوابة المطار الدولي يميز في لحظة خاطفة بين المسافرين حسب جنسياتهم ويفرزهم إلى أوروبيين وأفارقة وهنود وصينيين ... الخ .

وهكذا نجد أن عمليات التمييز كثيرة جداً وهي تهدف إلى تحديد المجموعة التي ينتمي إليها أي عنصر من عناصر المجتمع المدروس، أو تحديد انتماء أي عنصر جديد إلى إحدى مجموعات المجتمع المدروس، ولكن عملية التمييز بين العناصر وتحديد المجموعات التي تنتمي إليها ليست بهذه البساطة والسهولة .

فمثلاً لا نستطيع ببساطة التمييز بين الأخيار والأشرار أو بين المثقفين وغير المثقفين أو بين الأذكياء والأغبياء ...الخ، وذلك لأن التمييز في مثل هذه الحالات يحتاج إلى دراسة معمقة وتحديد المؤشرات التي تساعدنا في عملية التمييز .

وهنا نطرح الأسئلة التالية :

1- كيف استطعنا فرز هؤلاء الناس وتحديد المجموعات التي ينتمون إليها وبهذه السرعة الفائقة ؟
2- ماهي المتحولات الكمية أو النوعية التي اعتمدنا عليها في عمليات الفرز وتحديد الانتماء ؟
3- ماهي الصيغ أو التراكيب الرياضية التي استخدمناها وطبقناها على تلك المتحولات لاتخاذ القرار المناسب لفرز هذه العناصر وتحديد المجموعات التي تنتمي إليها ؟

إن الإجابة على هذه الأسئلة هي: إننا استخدمنا بعض الصفات أو بعض الخواص الملازمة لهذه العناصر وقمنا بتركيبها ضمن صيغ معينة (فورية)، ومنها حصلنا على مؤشر معين ساعدنا على فرز تلك العناصر وعلى تحديد المجموعات التي تنتسب إليها .

ولكن بعض عمليات الفرز والتمييز قد لا تكون صحيحة بسبب التشابه بين خواص العناصر، أو بسبب الالتباس بينها . ففي مثال شرطي المطار الدولي فقد يقع في خطأ ما. فالشخص الذي اعتبره اوروبياً قد يكون كندياً والإفريقي قد يكون أمريكياً والهندي قد يكون باكستانياً والصيني قد يكون يابانياً ...الخ .

وهكذا نجد أن عمليات التمييز بين العناصر ليست أمراً سهلاً، وهي تحتاج إلى دراسات معمقة وإلى أدوات رياضية واحصائية متقدمة، يأتي في مقدمة هذه الأدوات ما يسمى التابع التمييزي، الذي يستخدم لفرز عناصر العينة المدروسة إلى المجموعات المناسبة لها (مع احتمال الخطأ في ذلك) .

وقبل أن ندخل في استعراض التحليل التمييزي نشير إلى بعض المصطلحات المتشابهة والمتداخلة، وذلك حتى يتم التفريق بينهما عند معالجة المسائل المختلفة وهذه المصطلحات هي [حسب [4] ص437]

1- **التمييز:** وهو يشترط وجود مجتمع مقسم إلى مجموعتين أو أكثر (محددتين مسبقاً)، وأن يتم سحب عينة عشوائية من كل منها، ويتلخص هدفه ومهمته في دراسة خواص هذه العينات، ثم استخلاص قاعدة معينة مبنية على بيانات العينة، تسمح لنا بفرز عناصر المجتمع إلى تلك المجموعات وتحديد الحدود الفاصلة بينها، وبحيث تمكننا من تحديد المجموعة التي ينتمي إليها أي عنصر جديد، دون أن يكون لدينا أي فكرة مسبقة عن المجموعة ينتمي إليها ذلك العنصر .
2- **التصنيف:** وهو يتناول توزيع عناصر المجتمع أو عينة منه إلى مجموعات محددة، وبحيث يتم تصنيف عناصر تلك العينة (أو المجتمع) إلى مجموعات متجانسة داخلياً ومختلفة عن بعضها البعض أكثر ما يمكن، وينتج عنه مجموعات محددة وذات مواصفات معينة، تساعدنا على التنبؤ بانتماء إي عنصر جديد من المجتمع إلى إحدى المجموعات المحددة.
أي أن التمييز يشترط مسبقاً وجود المجموعات (المجتمعات) المحددة، ولكن التصنيف يهدف إلى تشكيل هذه المجموعات المختلفة في المجتمع .
3- **التجزئة:** وهي تتناول تجزئة عناصر عينة معينة (أو عناصر المجتمع الكلي)، وتقسيمها إلى مجموعات جزئية، بغض النظر عن وجود حدود طبيعية لتلك التجزئة أم لا، مثل توزيع الطلاب عشوائياً على الشعب الدراسية بدون أية شروط مسبقة .

وهنا نشير إلى أن التحليل التمييزي يتناول عمليتين أساسيتين هما:

التمييز: وهو القيام بفرز عناصر المجتمع إلى مجموعات محددة ومنفصلة وتحديد الحدود الفاصلة بينها .

التصنيف: وهو القيام بالتنبؤ بانتماء أي عنصر جديد إلى إحدى المجموعات وفق قاعدة معينة .

1-2 تعريف التحليل التمييزي:

إن التحليل التمييزي هو عبارة عن تقنية إحصائية رياضية قوية، تستخدم لتوصيف عناصر المجتمع المدروس، والموزعة على مجموعات محددة ومنفصلة ومتكاملة (أكثر أو يساوي من مجموعتين)، وتحديد الحدود الفاصلة بينها، واستخلاص قاعدة معينة لتحديد انتماء أي عنصر إليها، ونرمز لهذه المجموعات بـ $G_1 , G_2 , G_3 \dots G_g$، حيث أن: g هو عدد تلك المجموعات في المجتمع المدروس، وأن $(g \geq 2)$.

كما يستخدم التحليل التمييزي للتنبؤ بانتماء أي عنصر جديد إلى إحدى المجموعات السابقة أو تصنيفه في احداها حسب قاعدة معينة .

لذلك يقسم التحليل التمييزي إلى قسمين أساسيين هما:

- التحليل التمييزي الوصفي: وهو الذي يتناول توصيف العناصر ضمن المجموعات ووضع قواعد الانتماء إليها .
- التحليل التمييزي التنبؤي: وهو الذي يتناول تصنيف العناصر إلى مجموعات ويتنبأ بانتماء أي عنصر جديد إلى إحدى تلك المجموعات .

كما يمكن تقسيم التحليل التمييزي إلى نوعين آخرين هما:

- التحليل التمييزي البسيط: وهو الذي يعتمد على متحول واحد X لتوصيف عناصر المجتمع وتصنيفها ضمن مجموعاته المتعددة $(g \geq 2)$.
- التحليل التمييزي المتعدد: وهو الذي يعتمد على عدة متحولات مستقلة (متحولين أو أكثر) لتوصيف عناصر المجتمع وتصنيفها ضمن مجموعاته المتعددة $(g \geq 2)$. ونرمز لهذه المتحولات بالرموز $X_1\ X_2\ ...\ X_p$ حيث أن p هو عدد المتحولات المستقلة .

وكذلك يمكن تقسيم التحليل التمييزي إلى نوعين آخرين هما:

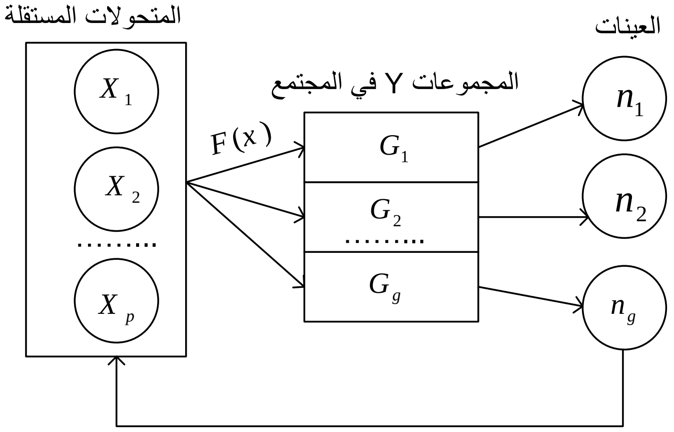
- التحليل التمييزي الخطي: وهو الذي يعتمد على النماذج الخطية للفصل بين مجموعات المجتمع المدروس .
- التحليل التمييزي غير الخطي: هو الذي يعتمد على النماذج غير الخطية للفصل بين مجموعات المجتمع المدروس .

وأخيراً يمكن تقسيم التحليل التمييزي إلى نوعين آخرين هما:

- التحليل التمييزي الكمي: وهو الذي يعتمد على المتحولات الكمية (المقاسة) في توصيف عناصر المجتمع، (كالدخل والوزن والعمر ...الخ) .
- التحليل التمييزي النوعي: وهو الذي يعتمد على المتحولات النوعية (غير المقاسة) في توصيف عناصر المجتمع (كالجنس والتعليم والمكان والحالة الاجتماعية ...الخ) .

وسنركز في منشورنا هذا على التحليل التمييزي الكمي الخطي الذي يهدف إلى دراسة التأثير المشترك لعدد متحولات كمية مستقلة هي $X_1\ , X_2\ ...\ X_p$ على تابع نوعي Y مؤلف من عدة مجموعات (أو فئات) هي: $G_1\ , G_2\ ...\ G_g$ وتحديد دور كل منها في تصنيف عناصر المجتمع إلى هذه المجموعات ، وذلك من خلال بيانات العينات العشوائية المسحوبة من تلك المجموعات، ثم العمل على إنشاء تركيب خطي (أو غير خطي) $F_j(x)$ لتلك المتحولات المستقلة، واستخدامه في التمييز بين تلك المجموعات التي يتألف من المجتمع المدروس .

ويمكن تمثيل التأثير المشترك للمتحولات المستقلة X على المجموعات التابعةY بيانياً على الشكل التالي:



الشكل (1−1) آلية عمل التحليل التمييزي

إن الشكل (1−1) يبين أن جملة المتحولات المستقلة $\left(X_1\ X_2\ \dots X_p\right)$ تؤثر على كل مجموعة من المجموعات التي يتألف منها التابع Y. وذلك من خلال التراكيب الخطية (أو غير الخطية)، المعرفة عليها $F_1(\mathrm{x}),\ F_2(x), \dots F_g(x)$، وأن كل عنصر من عناصر هذه المجموعات يعطينا جملة من القياسات لهذه المتحولات وعند تعويضها في التركيب الخطي المقابل لتلك المجموعة نحصل على قيمة محددة للتابع $F_j(x)$ تميز ذلك العنصر عن غيره في المجموعة المذكورة .

إن تطبيق التحليل التمييزي يتطلب أن نقوم بسحب عينة عشوائية من كل مجموعة من مجموعات المجتمع Y بحجوم متساوية أو مختلفة $n_1, n_2, \dots n_g$، فتتشكل لدينا عينات طبقية، وعينة كلية حجمها $n = \sum_{j=1}^{g} n_j$. ويمكن تفريغ بيانات هذه العينات في جداول مناسبة وحساب متوسطاتها وتبايناتها كما يلي:

جدول (1−1) بيانات العينات المسحوبة من المجموعات (حالة ثلاث مجموعات):

بيانات عينة المجموعة j من n_j عنصراً فيها للقيم x_{ijk}	التباينات	المتوسطات	بيانات عينة المجموعة الثانية من n_2 عنصراً 1 2 … n_2	التباينات	المتوسطات	بيانات عينة المجموعة الأولى من n_1 عنصراً 1 2 3 … n_1	البيانات المتحولات
نرمز لقيم المتحولات بـ x_{ijk} حيث يرمز i للمتحولات X_i ويرمز j للمجموعة G_j ويرمز k للمشاهدة k من n_j وحيث أن: $i: 1\ 2\ 3 \dots p$ $j: 1\ 2\ 3 \dots . g$ $k: 1\ 2\ 3 \dots\ n_j$	σ_{12}^2	$\bar{x}_{12}$	$x_{121}, x_{122} \dots x_{12n_2}$	σ_{11}^2	$\bar{x}_{11}$	$x_{111}, x_{112}, x_{113} \dots x_{11n_1}$	X_1
	σ_{22}^2	$\bar{x}_{22}$	$x_{221}, x_{222} \dots x_{22n_2}$	σ_{21}^2	$\bar{x}_{21}$	$x_{211}, x_{212}, x_{213} \dots x_{21n_1}$	X_2
	σ_{32}^2	$\bar{x}_{32}$	$x_{321}, x_{322} \dots x_{32n_2}$	σ_{31}^2	$\bar{x}_{31}$	$x_{311}, x_{312}, x_{313} \dots x_{31n_1}$	X_3
	σ_{42}^2	$\bar{x}_{42}$	$x_{421}, x_{422} \dots x_{42n_2}$	σ_{41}^2	$\bar{x}_{41}$	$x_{411}, x_{412}, x_{413} \dots x_{41n_1}$	X_4
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	σ_{p2}^2	$\bar{x}_{p2}$	$x_{p21}, x_{p22} \dots x_{p2n_2}$	σ_{p1}^2	$\bar{x}_{p1}$	$x_{p11}, x_{p12}, x_{p13} \dots x_{p1n}$	X_p

وأخيراً نشير إلى أننا سنركز في منشورنا هذا على التحليل التمييزي الخطي المتعدد الذي يعتمد على المتحولات المستقلة الكمية، ونعرفه كما يلي:

التحليل التمييزي الخطي: هو الذي يستخدم نماذج خطية في تركيب المتحولات المستقلة الكمية $X_1\ X_2\ \dots X_p$ لاستخدامها في التمييز بين عناصر المجموعات . وهذه النماذج يكون لها الشكل التالي :

$$\mathrm{F}(x) = B_0 + B_1X_1 + B_2X_2 + \cdots + B_pX_p + \varepsilon \qquad (1-1)$$

حيث أن $B_0\ B_1\ B_2\ \dots\ B_p$ هي الأمثال التمييزية (المعاملات) في التابع $\mathrm{F}(x)$.

وأن ε هو متحول الخطأ العشوائي في النموذج (1-1) . وأن $\mathrm{F}(x)$ هو تابع كامن يصاغ بواسطة تركيب خطي للمتحولات X المؤثرة في التابع Yويسمى بالتابع التمييزي الفاصل بين المجموعات (Discriminant Function) .

إن الهدف الأساسي للتحليل التمييزي هو اختبار فيما إذا كان تصنيف عناصر المجموعات في التابع Y، يعتمد على متحول واحد على الأقل من المتحولات المستقلة $X_1\ X_2\ \dots\ X_p$، ويمكن وضع فرضية العدم والفرضية البديلة كما يلي:

H_0: إن تصنيف عناصر المجموعات في التابع Y لايعتمد على أي من المتحولات المستقلة $X_1\ X_2X_3\ \dots X_p$.

H_1: إن تصنيف عناصر المجموعات في التابع Y يعتمد على واحد على الأقل من المتحولات المستقلة $X_1\ X_2\ \dots X_p$.

ويمكن كتابة هاتين الفرضيتين كما يلي :

$$\begin{aligned} &H_0: \quad B_i = 0 \qquad i : 1\ 2\ 3 \dots p \quad \text{من أجل} \\ &H_1: \quad B_i \neq 0 \qquad \text{من أجل قيمة واحدة عن الأقل} \end{aligned} \qquad (2-1)$$

1-3 الافتراضات والشروط لتطبيق التحليل التمييزي:

يشترط عند تطبيق التحليل التميزي على مجموعات مجتمع ما باستخدام متحولات $X_1\ X_2X_3\ \dots\ X_p$ مايلي :

1- أن تكون المجموعات في المجتمع المدروس محددة ومتكاملة وغير متقاطعة فيما بينها، وأن يكون عددها $(g \geq 2)$، وأن تكون حجومها متقاربة أو غير مختلفة كثيراً . ويمكن إنشاء المجموعات على أسس ديموغرافية أو جغرافية أو إدارية أو غيرها .

2- أن تكون المشاهدات (أو القياسات) مستقلة عن بعضها البعض . أي ان تكون العينات الطبقية مسحوبة بطريقة عشوائية من المجموعات . وأن يكون حجم العينة الكلية n أكبر بعشرين مرة من عدد المتحولات المستقلة p [Steven 1996] وأن لا يقل حجمها في كل الأحوال عن $n = 30$ عنصراً.

3- أن تكون المتحولات $X_1\ X_2X_3\ \dots\ X_p$ مستقلة عن بعضها البعض . وأن لا يكون عددها كبيراً، وإذا كان عددها كبيراً فيجب اختزالها حسب معايير محددة . وفي كل الأحوال يجب أن لا يزيد عددها p عن 5% من حجم العينة الكلية n، وعندما يكون $p = 1$ فإننا نكون أمام التحليل التمييزي البسيط .

4- أن تكون المتحولات المستقلة (الكمية) $X_1\ X_2\ \dots X_p$ خاضعة للتوزيع الطبيعي المتعدد، أو أن يكون كل منها خاضعاً للتوزيع الطبيعي العام . وإذا كان هذا الشرط غير محقق نقوم بإجراء بعض

التحويلات اللازمة أو نستغني عن بعض المتحولات، ولكن إذا كان حجم العينة كبيراً فيطبق عليها قانون الأعداد الكبيرة وتعتبر طبيعية تقاربياً .

5- أن تكون مصفوفات التباين والتباين المشترك للمتحولات المستقلة X داخل كل مجموعة، متشابهة (متماثلة Similarity)، أي أن تكون تلك المصفوفات متساوية أو متجانسة على الأقل .

6- أن تكون مقادير الأخطاء (البواقي) موزعة عشوائياً. أي أن يكون توقعها مساوياً للصفر .

ملاحظة: يمكن التساهل في الشروط الأخيرة عندما يكون حجم العينة كبيراً .

وللتأكد من تحقق هذه الشروط يجب إجراء الاختبارات اللازمة لذلك . وسنعرضها لاحقاً في فقرة خاصة .

4-1 أهداف وخطوات التحليل التمييزي :

يهدف التحليل التمييزي إلى تحقيق عدة أغراض أهمها:

1- التحري عن الاختلافات الموجودة بين المجموعات التي يتألف منها المجتمع المدروس .
2- تحديد الطريقة الأكثر توفيراً في التفريق أو الفصل أو التمييز بين المجموعات .
3- حذف المتحولات المرتبطة بضعف مع المجموعة المحددة .
4- تحديد المتحولات المستقلة التي تميز كل مجموعة وأي المتحولات التي لا تميزها .
5- تصنيف العناصر أو الحالات ضمن المجموعات حسب قاعدة معينة .
6- اختبار النظرية القائمة من خلال مقارنة فيما إذا كانت الحالات المصنفة سابقاً مشابهة للحالات المصنفة لاحقاً .

أما خطوات التحليل التمييزي فتتلخص بما يلي :

1- تحديد نوع التحليل التمييزي المناسب الذي يوصلنا إلى النتائج الاحصائية المطلوبة، فهل التحليل الوصفي هو المناسب أم التنبؤي؟ ، أو البسيط أم المتعدد، أو الخطي أم غير الخطي، أو الكمي أم النوعي أم اللوجيستي ؟.
2- تحديد المتحولات المستقلة المناسبة للتحليل التمييزي المختار واللازمة لتحقيق أهداف البحث وجمع البيانات عنها .
3- تحديد عدد المجموعات في المجتمع التي ستستخدم في التحليل $(g \geq 2)$ وسحب العينات الطبقية منها وجمع البيانات اللازمة منها .
4- اختبار البيانات والتأكد من أنها تحقق الافتراضات والشروط المفروضة عليها .
5- إجراء التحليل التمييزي حسب خطواته العملية والحصول على النتائج المطلوبة .
6- تفسير النتائج والعمل على الاستفادة منها .

الفصل الثاني

التحليل التمييزي البسيط (حسب متحول واحد ومجموعتين)

2-1 تمهيد:

يعتبر التحليل التمييزي البسيط أساساً لأنواع التمييز الأخرى، لأنه يمكن أن يتم تحويل معظم الأنواع الأخرى إلى التحليل التمييزي البسيط . ولاستعراض هذا النوع البسيط نأخذ مجتمع القروض المصرفية الذي يقدمها أحد المصارف إلى المواطنين ونعرضها كما يلي:

لنفرض أن مدير المصرف لاحظ أن الكثير من القروض الممنوحة متعثرة السداد (أكثر من ستة أشهر)، فطلب من المسؤولين عنها تحديد أسباب التعثر لمعالجة هذه الظاهرة فقالوا له: توجد أسباب كثيرة لذلك ومنها (الدخل والعمر والملاءة والجنس والمهنة والخبرة والمنافسة والالتزام والأخلاق ...الخ)، ولكن المدير طلب منهم تحديد مؤشر واحد أو مؤشرين لاستخدامهما في قرار الموافقة على منح القروض، فاستقر أمرهم الأول على اعتبار (الدخل الشهري للمقترض X) هو أهم الأسباب المؤدية إلى تعثر القروض، لأنه إذا كان دخل المقترض ضعيفاً، فإن قدرته على تسديد الأقساط تكون ضعيفة، وبالتالي فإن قرضه يكون عرضة للتعثر .

وللتأكد من ذلك قام فريق العمل بسحب عينة عشوائية من القروض المتعثرة بحجم $n_1 = 8$، وسحب عينة عشوائية من القروض غير المتعثرة بحجم $n_2 = 7$، ثم قام بتصنيف عناصر هاتين العينتين حسب دخل المقترض في كل منها ثم معالجتها، فكانت بيانات ومتوسطات وتباينات دخول أصحاب هذه القروض كما يلي:

جدول (2-1): بيانات عينيتي القروض المتعثرة وغير المتعثرة . (من *Lesson2: DA SPSS* بتصرف)

	مجموعة القروض المتعثرة G_1								مجموعة القروض غير المتعثرة G_2						
	عناصر عينة القروض المتعثرة : $n_1 = 8$								عناصر عينة القروض غير المتعثرة : $n_2 = 7$						
الرقم	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7
دخل صاحب القرض	56	56	45	56	29	56	56	30	72	72	56	72	64	64	56
المتوسط والتباين	$\bar{x}_1 = 48$ $S_1^2 = (12.036)^2$								$\bar{x}_2 = 65.14$ $S_2^2 = (7.192)^2$						

وكان المتوسط العام المثقل للدخول في هاتين العينتين (المتعثرين وغير المتعثرين) يساوي :

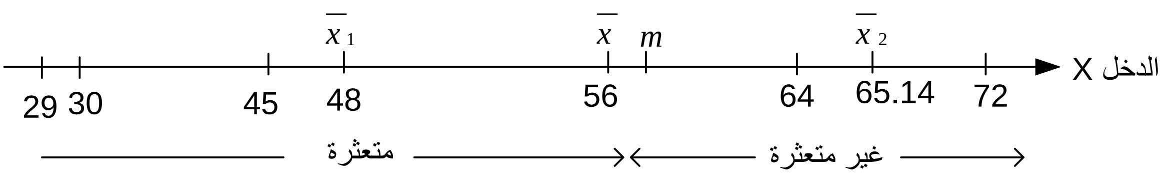
$$\bar{x} = \frac{8(48) + 7(65.14)}{15} = 56$$

وإن التباين العام والانحراف المعياري يساويان :

$$S^2 = (12.153)^2 \quad , \quad S = 12.153$$

ويمكن حساب القيمة الوسطى للمتوسطين $\bar{x}_1$ و $\bar{x}_2$ من العلاقة : $m = \frac{\bar{x}_1 + \bar{x}_2}{2} = 56.57$

إن هذه المتوسطات أو المراكز ترسم على المحور المخصص للدخل X على الشكل التالي:



الشكل رقم (2−1): محور توزيع القروض حسب الدخل

وهنا تساؤل المسؤولون: كيف سنفرز القروض حسب الدخل فقط إلى قروض متعثرة وقروض غيرمتعثرة ؟ وماهو المعيار الذي يمكن اعتماده لتحقيق ذلك الفرز بأقل خطأ ممكن ؟

إن الجواب السهل والبسيط على ذلك هو أن نستفيد من المتوسط العام $\bar{x}$، ونعتبره الحد الفاصل بين هاتين المجموعتين ونقول:

إذا كان $X \leq \bar{x}$ فإن القرض يكون متعثراً

أما إذا كان $X > \bar{x}$ فإن القرض يكون غير متعثر .

وفي مثالنا هذا نلاحظ أن $\bar{x} = 56$، وهذا يعني أن جميع القروض المتعثرة $(n = 8)$ ستصنف على أنها متعثرة وهذا تصنيف صحيح .

ولكننا سنقع في خطأ ملحوظ عند تصنيف القروض غير المتعثرة، حيث سنعتبر القرضين (3) و (7) من القروض غير المتعثرة قروضاً متعثرة (وهذا تصنيف خاطئ). وهكذا نجد إن هذا المعيار (اعتبار المتوسط العام $\bar{x}$ هو الحد الفاصل) هو معيار معقول ولكنه يحتمل الخطأ، فعند دراسة الطلبات الجديدة للزبائن فإنه حسب هذا المعيار سنعتبر كل من كان دخله $X \leq 56$ من أصحاب القروض المتعثرة، وهذا أمر مشكوك فيه لأن بيانات العينة الثانية تفيدنا أن نسبة لا بأس بها من أصحاب الدخول $X \leq 56$ هم من أصحاب القروض غير المتعثرة (وليس المتعثرة) .

إن الوقوع في هذا الخطأ يحدث لعدة أسباب أهمها:

1− إذا كانت الدراسة لا تستند إلى خصائص المجتمع ولا تعتمد على التوزيعات التكرارية الاحتمالية للدخول في كل مجموعة في المجتمع .

2− إذا كانت العينات المسحوبة غير عشوائية وكان حجمها الكلي صغيراً $(n < 30)$.

3− إذا كانت حجوم العينات الطبيعية لا تتناسب مع حجوم المجموعات . أي إذا كانت نسبة حجم كل منها لا يتوافق مع نسبة حجم المجموعة في المجتمع .

فمثلاً: إذا كانت نسبة القروض المتعثرة 30% أي $(P_1 = 0.30)$ من القروض الكلية .

وكانت نسبة القروض غير المتعثرة 70% أي $(P_2 = 0.70)$ من القروض الكلية .

فإن نسبة حجمي العينتين يجب أن يتوافق مع هاتين النسبتين. أي يجب أن يكون :

$$\frac{n_1}{n} = P_1 \qquad \frac{n_2}{n} = P_2 \qquad (1-2)$$

4- إذا كانت حجوم العينات لا تأخذ بعين للاعتبار الانحرافات المعيارية داخل كل مجموعة، لأن حجوم العينات يجب أن تتناسب طرداً مع الانحرافات المعيارية الداخلية لتلك المجموعات أي يجب أن يكون :

$$n_1 = \frac{N_1}{N}\sqrt{S_{11}} \qquad n_1 = \frac{N_2}{N}\sqrt{S_{22}} \qquad (2-2)$$

حيث أن N_1 و N_2 هما عدد العناصر في المجموعتين وأن $(N_1 + N_2)$ = حجم المجتمع N.

5- إذا كانت حجوم العينات واحتمالات الخطأ لا تأخذ بعين الاعتبار تكلفة التصنيفات الخاطئة Misclassification cost . لأن حجوم العينات تتناسب عكساً مع تكلفة التصنيفات الخاطئة .

ولهذا يجب اتباع طرائق رياضية دقيقة لتحديد الحدود الفاصلة بين المجموعات، وأهم هذه الطرائق هي:

- طريقة التكلفة المتوقعة للتصنيفات الخاطئة ECM .
- طريقة بايز (Bayse) أو الاحتمالات الشرطية اللاحقة .
- طريقة الجوار الأقرب أو مسافة Mahalanobis .
- طريقة معيار فيشر Fishir's certion .
- طريقة تحليل التباين الأحادي أو المتعدد ANOVA أو MANOVA .

وسنتعرض لهذه الطرائق باختصار يتناسب مع التحليل التمييزي البسيط كما يلي :

2-2 طريقة التكلفة المتوقعة للتصنيفات الخاطئة ECM: (Expected Cost of Misclassification)

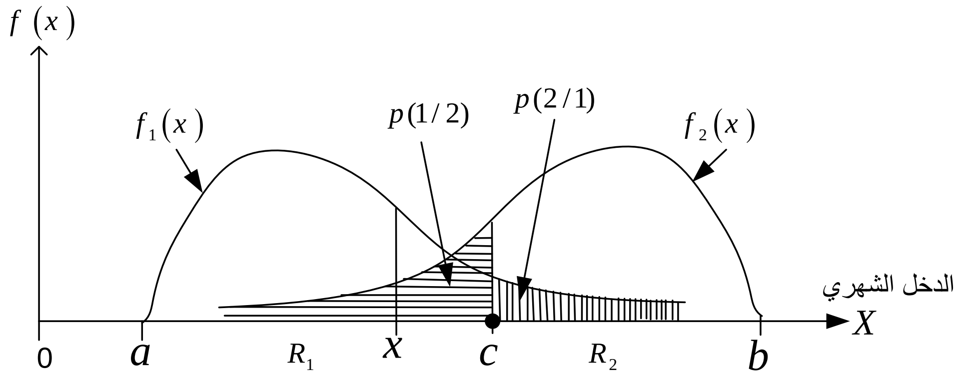
لنتابع مثالنا السابق حول القروض، ولنرمز للمجموعتين كما يلي:[Johnson,Wichern, P. 485 بتصرف وإضافة]

نرمز لمجموعة القروض المتعثرة بالرمز G_1 ولنسبتها في المجتمع P_1 .

نرمز لمجموعة القروض غير المتعثرة بالرمز G_2 ولنسبتها في المجتمع P_2 .

إن P_1 و P_2 يسميان بالاحتمالات السابقة . لأنهما يعطيانا احتمال انتماء أي عنصر جديد إلى إحدى هاتين المجموعتين ويحققان العلاقة $P_1 + P_2 = 1$.

ولنفترض الآن أننا جمعنا البيانات اللازمة عن جميع دخول المقترضين X في كل مجموعة على حدة، وحسبنا تكراراتها النسبية واستخلصنا منها التوزيع الاحتمالي للدخول في G_1 ورمزنا له بـ $f_1(x)$. والتوزيع الاحتمالي للدخول في G_2 ورمزنا له بـ $f_2(x)$. والمعرفين على مجال مشترك لهما هو $[a\ ,b]$ وكان لهما الشكل التالي :



الشكل (2–2): التوزيعات التكرارية في G_1 و G_2

ونعلم أنه من خواص هذين التوزيعين إنهما يحققان الخاصتين التاليتين

$$f_1(x) \geq 0 \qquad \int_a^b f_1(x)dx = 1 \qquad \textbf{(3-2)}$$

$$f_2(x) \geq 0 \qquad \int_a^b f_2(x)dx = 1 \qquad \textbf{(4-2)}$$

والمطلوب تحديد النقطة الفاصلة C، التي تفرز القروض بأقل خطأ ممكن إلى مجموعتين: مجموعة المتعثرة G_1 على اليسار، ومجموعة غير المتعثرة G_2 على اليمين، والتي تقسم المجال $[a,b]$ إلى مجالين غير متقاطعين $[a,c]$ و $]c\ ,d]$ ، ونرمز لهما اختصاراً كما يلي:

بحيث يكون المجال $R_1 =]a\ ,\ c]$ مقابلاً للمجموعة المتعثرة G_1 .

ويكون المجال $R_2 =]c\ ,b]$ مقابلاً للمجموعة غير المتعثرة G_2 .

ولتحديد انتماء كل قرض إلى إحدى هاتين المجموعتين نقوم بحساب الاحتمالات الشرطية المقابلة لهما، فنجد أن احتمال أن يتم تصنيف القرض حسب الدخل لـ X خطأ في المجموعة الثانية G_2 وهو أصلاً من المجموعة الأولى G_1 يساوي:

$$P(2/1) = P[x \in R_2/G_1] = \int_{R_2} f_1(x)dx \qquad (5-2)$$

وهو يمثل بالمساحة التي يشكلها $f_1(x)$ فوق المجال R_2 وهي المساحة المظللة بخطوط عمودية . وهو يعبر عن احتمال أن ينتمي عنصر المجموعة G_1 خطأ إلى المجموعة G_2 ، أي هو احتمال أن يتم تصنيف عنصر المجموعة المتعثرة G_1 خطأ في المجموعة غير المتعثرة G_2 .

أما احتمال أن يتم تصنيف عنصر المجموعة G_2 خطأ في المجموعة G_1 فهو يساوي :

$$P(1/2) = P[x \in R_1/G_2] = \int_{R_1} f_2(x)dx \qquad (6-2)$$

وهو يمثل بالمساحة المظللة بخطوط أفقية والتي يشكلها التوزيع $f_2(x)$ فوق المجال R_1 .

ويطلق على الاحتمالين الشرطيين $P(1/2)$ و $P(1/2)$ وما شابههما مصطلح الاحتمالات اللاحقة، وهي تختلف عن الاحتمالات السابقة P_2 و P_1 التي ذكرناها سابقاً .

وبناءً على ذلك يمكننا حساب الاحتمالات المختلفة للتصنيف الصحيح وللتصنيف الخاطئ إلى كل من هاتين المجموعتين كما يلي:

إن احتمال أن يكون تصنيف العنصر x إلى المجموعة G_1 صحيحاً يساوي الجداء التالي =

= (احتمال أن يكون أصله من G_1)*(احتمال تصنيفه في G_1) .

ولنرمز لذلك الاحتمال بالرمز $P(1,1)$ فيكون لدينا :

$$P(1,1) = P[x \in R_1/G_1] * P(G_1) = P(1/1) * P_1 \qquad (7-2)$$

وإن احتمال أن يكون تصنيف العنصر x إلى المجموعة G_1 خاطئاً يساوي الجداء التالي =

(احتمال أن يكون أصله من G_2)*(احتمال تصنيفه خطأ في G_1) .

ولنرمز لذلك الاحتمال بالرمز $P(1,2)$ ويكون لدينا :

$$P(1,2) = P[x \in R_1/G_2] * P(G_2) = P(1/2) * P_2 \qquad (8-2)$$

أما احتمال أن يكون تصنيف العنصر x إلى المجموعة G_2 صحيحاً فيساوي الجداء التالي =

(احتمال أن يكون أصله من G_2)*(احتمال تصنيفه في G_2) .

ولنرمز لذلك الاحتمال بالرمز $P(2,2)$ ويكون لدينا :

$$P(2,2) = P[x \in R_2/G_2] * P(G_2) = P(2/2) * P_2 \qquad (9-2)$$

وإن احتمال أن يكون تصنيف العنصر x إلى المجموعة G_2 خطأ يساوي الجداء التالي =

(احتمال أن يكون أصله من G_1)*(احتمال تصنيفه خطأ في G_2) .

ولنرمز لذلك الاحتمال بالرمز $P(2,1)$ ويكون لدينا :

$$P(2,1) = P[x \in R_2/G_1] * P(G_1) = P(2/1) * P_1 \qquad (10-2)$$

ويمكن وضع هذه الاحتمالات الشرطية في جدول رباعي كما يلي:

جدول (2–2): الاحتمالات المتقاطعة للتصنيف في المجموعتين G_1 و G_2 :

الاحتمالات السابقة	المجموعات	G_1	G_2
P_1	G_1	$P(1/1) * P_1$	$P(2/1) * P_1$
P_2	G_2	$P(1/2) * P_2$	$P(2/2) * P_2$

ويصبح هدفنا الآن أن نجعل مجموع الاحتمالين المقابلين للحالتين الخاطئتين أصغر ما يمكن، وهو الذي يساوي:

$$P(2/1) * P_1 + P(1/2) * P_2 = TPM \qquad (11-2)$$

ويسمى هذا الاحتمال بالاحتمال الاجمالي للتصنيف الخاطئ، ويرمز له بـ TPM من الكلمات (Total Probability Misclassification)، وبناءً على العلاقتين (2–5) و(2–6) يمكن كتابة TPM على الشكل التالي:

$$TPM = P_1 * \int_{R_2} f_1(x)dx + P_2 * \int_{R_1} f_2(x)dx \qquad (12\,a-2)$$

وعلينا الآن أن نجعل هذا الاحتمال الاجمالي TPM أصغر ما يمكن .
ولكن ذلك الأمر يتعلق بالتكلفة المترتبة على التصنيفات الخاطئة في كلا المجموعتين G_1 و G_2 .
فإذا افترضنا أن تكلفة التصنيف الصحيح في كل مجموعة تساوي الصفر(لأن ذلك لا يسبب ضرراً فيها). وإن تكلفة التصنيف الخاطئ في المجموعة الأولى تساوي $C(1/2)$. وإن تكلفة التصنيف الخاطئ في المجموعة الثانية تساوي $C(2/1)$.
فيمكننا إنشاء جدول مقابل للجدول (2−3) يتضمن تقاطع هذه التكاليف كما يلي:

جدول (2−3): تكاليف التصنيف الخاطئ :

المجموعات	G_1	G_2
G_1	0	$C(2/1)$
G_2	$C(1/2)$	0

ولحساب القيمة المتوقعة (المتوسط) لهذه التكاليف، نقوم بضربها بالاحتمالات الشرطية للتصنيف الخاطئ، المقابلة من الجدول (2−2) السابق، فنحصل على ما تسمى القيمة المتوقعة لتكلفة التصنيف الخاطئ (Expected Cost of misclassification)، والتي يرمز لها بـ ECM وهي تساوي حسب الجدولين (2−2) و (2−3) ما يلي:

$$\mathrm{ECM} = \mathrm{C}(2/1) * P(2/1) * P_1 + C(1/2) * P(1/2) * P_2 \qquad (12-2)$$

ويصبح هدفنا الآن أن نجعل القيمة المتوقعة ECM أصغر ما يمكن .
وبناءً على العلاقتين (2−5) و (2−6) نجد أن:

$$\mathrm{ECM} = \mathrm{C}(2/1) * P_1 \int_{R_2} f_1(x)dx + \mathrm{C}(1/2) * P_2 \int_{R_1} f_2(x)dx \qquad (13-2)$$

وبما أنه لدينا $R = R_1 \cup R_2$ فإن التكامل الشامل لـ $f_1(x)$ حسب (2−3) على المجال الكلي R يساوي الواحد ويساوي ما يلي :

$$1 = \int_{R} f_1(x)dx = \int_{R_1} f_1(x)dx + \int_{R_2} f_1(x)dx \qquad (14-2)$$

ومنها نجد أن تكامل $f_1(x)$ على R_2 فقط يساوي :

$$\int_{R_2} f_1(x)dx = 1 - \int_{R_1} f_1(x)dx \qquad (15-2)$$

وبالتعويض في (2−13) نحصل على أن :

$$\mathrm{ECM} = \mathrm{C}(2/1) * P_1 \left[1 - \int_{R_1} f_1(x)dx\right] + \mathrm{C}(1/2) * P_2 \int_{R_1} f_2(x)dx$$

ومنها نحصل على أن:

$$\mathrm{ECM} = \int_{R_1} [\mathrm{C}(1/2) * P_2 f_2(x) - \mathrm{C}(2/1) * P_1 f_1(x)]dx + \mathrm{C}(2/1) * P_1 \qquad (16-2)$$

وبما أن قيم P_1 و P_2 وكذلك $C(1/2)$ و $C(2/1)$ هي أعداد موجبة، وأن جميع قيم التوزيعين $f_1(x) \geq 0$ و $f_2(x) \geq 0$ من أجل جميع قيم x في المجال R_1، فإن هذين التوزيعين هما المقداران الوحيدان في ECM، اللذان يرتبطان بـ x .

وبما أن قيمة الحد الأخير $C(2/1) * P_1$ في العلاقة موجبة ومعلومة، فإن ECM تأخذ أصغر قيمة لها إذا كانت نتيجة التكامل في $(2-16)$ سالبة أو معدومة .

وهكذا نجد أن ECM تأخذ أصغر قيمة ممكنة إذا كان المجال R_1 يتضمن قيماً معينة لـ x تجعل التكامل في $(2-16)$ سالباً أو معدوماً. (أي تجعل المساحة أو الحجم الذي يحدده ذلك التكامل على R_1 سالباً أو معدوماً) . وهذا يعني أن يحقق التكامل المذكور العلاقة التالية:

$$\int_{R_1} [C(1/2) * P_2 f_2(x) - C(2/1) * P_1 f_1(x)]\, dx \leq 0 \qquad (17-2)$$

وهذا لا يمكن أن يحدث إلا إذا كان المقدار المكامل (الذي تحت التكامل)سالباً أو معدوماً، أي أن ECM تأخذ أصغر قيمة ممكنة لها عندما وفقط عندما تتحقق المتراجحة التالية:

$$C(1/2) * P_2 f_2(x) - C(2/1) * P_1 f_1(x) \leq 0 \qquad (18-2)$$

وذلك من أجل قيم معينة لـ x، أي أنه باستثناء قيم x التي تجعل المقدار الأيسر موجباً. فإن المجال R_1 يجب أن يتضمن مجموعة من النقاط x التي تحقق المتراجحة $(2-18)$ والتي نكتبها على الشكل التالي:

$$C(1/2) * P_2 f_2(x) \leq C(2/1) * P_1 f_1(x) \qquad (19-2)$$

أي أن المجال R_1 يتحدد من خلال العلاقة $(2-19)$ والتي يمكن كتابتها على الشكل التالي :

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \qquad (20-2)$$

كما أن المجال R_2 يجب أن يتضمن مجموعة من النقاط x التي تحقق المتراجحة المكملة التالية :

$$R_2: \frac{f_1(x)}{f_2(x)} < \left[\frac{C(1/2)}{C(2/1)}\right] * \left[\frac{P_2}{P_1}\right] \qquad (21-2)$$

ومما سبق نستنتج القاعدة التالية:

إن تحديد المنطقتين R_1 و R_2 في R، التي تجعلان ECM أصغر ما يمكن، يتم بواسطة تحديد قيم x بدلالة النسب، التي تحقق المتراجحة التالية:

$$R_1: \begin{bmatrix} \text{نسبة الكثافة} \\ \text{الاحتمالية} \end{bmatrix} \geq \left[\text{نسبة التكلفة}\right] * \begin{bmatrix} \text{معكوس نسبة} \\ \text{الاحتمالات السابقة} \end{bmatrix} \qquad (20-2)'$$

$$R_2: \begin{bmatrix} \text{نسبة الكثافة} \\ \text{الاحتمالية} \end{bmatrix} < \left[\text{نسبة التكلفة}\right] * \begin{bmatrix} \text{معكوس نسبة} \\ \text{الاحتمالات السابقة} \end{bmatrix} \qquad (21-2)'$$

وذلك لأن استخدام النسب أسهل من البحث عن القيم الأصلية لمركباتها .

حالات خاصة: مما سبق يمكن أن نستخلص بعض الحالات الخاصة لهذه القاعدة وهي:

أ– حالة تساوي الاحتمالات السابقة: أي عندما يكون $\frac{P_2}{P_1}=1$ فإن القاعدة تصبح كما يلي:

$$R_1:\frac{f_1(x)}{f_2(x)}\geq\left[\frac{\mathrm{C}(1/2)}{\mathrm{C}(2/1)}\right]\quad,\quad R_2:\frac{f_1(x)}{f_2(x)}<\left[\frac{\mathrm{C}(1/2)}{\mathrm{C}(2/1)}\right]\qquad(22-2)$$

ب–حالة تساوي تكلفة التصنيف الخاطئ، أي عندما يكون $C(1\,/\,2)=C(2\,/\,1)$ فإن :

$$R_1:\frac{f_1(x)}{f_2(x)}\geq\frac{P_2}{P_1}\quad,\quad R_2:\frac{f_1(x)}{f_2(x)}<\frac{P_2}{P_1}\qquad(23-2)$$

ج– حالة تساوي الاحتمالات السابقة وتساوي التكلفة (وكذلك في حالة $1=\frac{\mathrm{C}(2/1)}{\mathrm{C}(1/2)}*\frac{\mathrm{P}_2}{\mathrm{P}_1}$) فإن:

$$R_1:\frac{f_1(x)}{f_2(x)}\geq 1\quad,\quad R_2:\frac{f_1(x)}{f_2(x)}<1\qquad(24-2)$$

ويتم استخدام هذه الحالات الخاصة في التطبيقات العملية في عدة حالات كما يلي :

إذا كانت الاحتمالات السابقة P_1 وP_2 غير معروفة فإنها تفترض على أنها متساوية، وعندها فإن عملية تصغير ECM تنحصر في مقارنة نسبة التوزيعين الاحتماليين مع نسبة تكلفة التصنيف الخاطئ (الحالة أ).

وعندما تكون تكلفة التصنيف الخاطئ غير محددة فإنها تفترض على أنها متساوية، وعندها يتم مقارنة نسبة التوزيعين الاحتماليين مع نسبة الاحتماليين السابقين $\frac{P_2}{P_1}$ (مع ملاحظة أن نسبة الاحتماليين متعاكسة مع نسبة التوزيعين) (الحالة ب) .

د– وأخيراً عندما تكون نسبتا الاحتماليين والتكلفتين متساويتين أو تساويان الواحد، أو عندما تكون أحدهما تساوي مقلوب الأخرى، فإن منطقة التصنيف R_1 تحدد ببساطة بواسطة مقارنة قيم التوزيعين مع بعضهما، وفي هذه الحالة نجد أنه إذا كانت x_0 مشاهدة جديدة وكانت:

$$R_1:\left[\left(f_1(x_0)\right)\geq f_2(x_0)\ \text{أي إذا كان}\right]\qquad\frac{f_1(x_0)}{f_2(x_0)}\geq 1\qquad(25-2)$$

فإننا ننسب x_0 إلى المجموعة G_1.

وبالمقابل إذا كانت :

$$R_2:\left[\left(f_1(x_0)\right)<f_2(x_0)\ \text{أي إذا كان}\right]\qquad\frac{f_1(x_0)}{f_2(x_0)}<1\qquad(26-2)$$

فإننا ننسب x_0 إلى المجموعة G_2 .

وكثيراً ما يتم استخدام الحالة الأخيرة في التطبيقات العملية لأنها تسهل عمليات تصنيف العناصر والتمييز بين المجموعات، ولكنها تجعلنا نفقد الفرق بين التمييز والتصنيف (لأننا لم نعد نهتم بالحد الفاصل).

مثال (2-1): لنفترض أن التحليل الاحصائي لجميع القروض الممنوحة في المصرف حسب الدخل X أعطانا النتائج التالية:

1- إن نسبة القروض المتعثرة P_1 تساوي 30% من إجمالي القروض .

2- إن نسبة القروض غير المتعثرة P_2 تساوي 70% من إجمالي القروض .

3- إن الدخول الشهرية لأصحاب القروض تقاس بالآلاف وتتحول في المجال [75 ,25]، وإن المدرج التكراري للدخول X في مجموعة القروض المتعثرة G_1 يخضع لتوزيع بيتا $\beta_{2,4}(x)$ المعرف بالآلاف على المجال المحدد [75 ,25] والذي يعطى بالعلاقة التالية :

$$f_1(x) = \beta_{2,4}(x) = C_1(x-25)^{2-1}(75-x)^{4-1} \qquad (a)$$

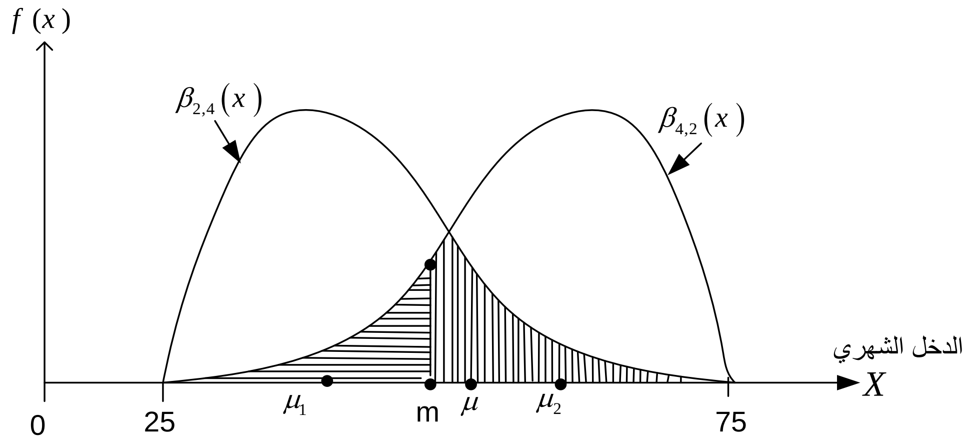
حيث: C_1 ثابت التوزيع الاحتمالي الذي يجعل المساحة تحته تساوي الواحد .

4- إن المدرج التكراري للدخول X في مجموعة القروض غير المتعثرة G_2 يخضع لتوزيع بيتا $\beta_{4,2}(x)$ المعرف بالآلاف على نفس المجال المحدد [75 ,25] والذي يعطى بالعلاقة التالية :

$$f_2(x) = \beta_{4,2}(x) = C_2(x-25)^{4-1}(75-x)^{2-1} \qquad (b)$$

حيث: C_2 ثابت التوزيع الاحتمالي الذي يجعل المساحة تحته تساوي الواحد .

5- يرسم هذان التوزيعان المنحنيين التاليين :



الشكل (2-3) منحنيا التوزيعين $\beta_{24}(x)$ و $\beta_{42}(x)$

ومن هذين التوزيعين يمكننا حساب توقعي وتبايني الدخول في المجموعتين μ_1 و μ_2، σ_1^2 و σ_2^2 من العلاقات التالية:

$$\mu_1 = \frac{a*q+b*p}{p+q} = \frac{25*4+75*2}{4+2} = 41.667$$

$$\sigma_1^2 = \frac{pq(b-a)^2}{(p+q)^2(p+q+1)} = 79.25 \qquad (C)$$

$$\mu_2 = \frac{a*q+b*p}{p+q} = \frac{25*2+75*4}{4+2} = 58.333$$

$$\sigma_2^2 = \frac{pq(b-a)^2}{(p+q)^2(p+q+1)} = 79.25 \qquad (e)$$

وهنا نلاحظ أن التباينين متساويان $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 79.25$، وأن المتوسط العام للدخل : $\mu = 53.333$

ولكن القيمة الوسطى لمتوسطيهما تساوي:

$$m = \frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{2}(41.667 + 58.333) = \frac{100}{2} = 50 \qquad (f)$$

فهل تصلح هذه القيمة لتكون حداً فاصلاً بين عناصر المجموعتين G_1 و G_2 ؟

6- الآن لنفترض أن تكلفة التصنيف الخاطئ لعناصر G_1 في G_2 تساوي $\mathrm{C}(2/1) = 20$، وإن تكلفة التصنيف الخاطئ لعناصر G_2 في G_1 تساوي $\mathrm{C}(1/2) = 50$، وتكلفة التصنيف الصحيح معدومة .

7- مما سبق نلاحظ أن الاحتمالين السابقين P_1 و P_2 معلومان ويساويان نسبتي كل مجموعة في المجتمع، أي أن: $P_2 = 0.70$ و $P_1 = 0.30$.

وهكذا نجد أنه لتحديد مجال التمييز والتصنيف R_1 للمجموعة الأولى G_1 على المحور OX نطبق العلاقة (2-20) كما يلي :

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \frac{\mathrm{C}(1/2)}{\mathrm{C}(2/1)} * \frac{P_2}{P_1} \geq \frac{50}{20} * \frac{0.70}{0.30} = 5.8333 \qquad (e)$$

وإن تحديد مجال التمييز R_2 للمجموعة الثانية G_2 يتم كما يلي:

$$R_2 : \frac{f_1(x)}{f_2(x)} < \frac{\mathrm{C}(1/2)}{\mathrm{C}(2/1)} * \frac{P_2}{P_1} < \frac{50}{20} * \frac{0.70}{0.30} = 5.8333 \qquad (g)$$

وبتعويض $f_1(x)$ و $f_2(x)$ في العلاقة (e) نجد أن R_1 تحسب كما يلي:

المجال R_1 يتحدد كمايلي :

$$R_1 : \frac{C_1(x-25)^1(75-x)^3}{C_2(x-25)^3(75-x)} \geq 5.8333 \qquad (h)$$

وبما أن التوزيعين $B_{2,4}(x)$ و $B_{4,2}(x)$ متناظران، فإن $(C_1 = C_2)$، وبعد إجراء الاختصارات الممكنة يكون لدينا :

$$R_1 : \frac{(75-x)^2}{(x-25)^2} \geq 5.8333$$

وللحصول على الحد الفاصل نأخذ المساواة ونجذر الطرفين فنجد أن:

$$\frac{75-x}{x-25} = \pm\sqrt{5.8333} = \pm 2.415 \qquad (I)$$

$$75 - x = \pm 2.415(x-25)$$

فإذا أخذنا الاشارة الموجبة نجد أن :

$$75 - x = 2.415x - 60.381$$

$$135.381 = 3.415x$$

$$x = \frac{135.381}{3.415} = 39.64 \qquad (k)$$

وهذه القيمة تمثل النقطة x_0 التي تمثل الحد الفاصل بين المجموعتين G_1 و G_2 .

أي أن القيمة $x = 39.64$ هي قيمة الدخل التي تفصل بين القروض المتعثرة G_1 والقروض غير المتعثرة G_2 والتي تجعل ECM أصغر ما يمكن .

وذلك ضمن شروط وخواص هذه المسألة، وهي تعطينا المجالين R_1 و R_2 المقابلين لـ G_1 و G_2 كما يلي:

$$R_1 = [25\ ,39.64] \qquad\qquad R_2 =]39.64\ ,75]$$

أما إذا أخذنا الإشارة السالبة للعلاقة (I) نجد أن:

$$(75 - x) = -2.415(x - 25) = -2.415x + 60.281$$

$$14.719 = -1.415x \implies x = -10.40 \qquad \text{(حل مرفوض)}$$

حالات خاصة:

1- لو إننا افترضنا أن تكلفتي التصنيف الخاطئ في المجموعتين متساويتان. فإن ذلك يعني أن $C(1/2) = C(2/1)$ وهذا يجعل القاعدة (2-20) تأخذ الشكل التالي:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \frac{P_2}{P_1} =$$

$$R_1 : \frac{C_1(x-25)(75-x)^3}{C_2(x-25)^3(75-x)} \geq \frac{0.70}{0.30} = 2.3333$$

وبعد الاختصار (ولأن $C_1 = C_2$) نحصل على أن:

$$\frac{(75-x)^2}{(x-25)^2} \geq 2.3333$$

وللحصول على الحد الفاصل نأخذ المساواة ونجذر الطرفين فنحصل على أن :

$$\frac{(75-x)}{(x-25)} = \pm 1.5275 \qquad (k)$$

وإذا أخذنا الإشارة الموجبة نحصل على أن:

$$75 - x = 1.5275(x - 25) = 1.5275x - 38.188$$
$$113.1875 = 2.5275x$$
$$x = 44.78$$

وهي تعطينا أن: $R_1 = [25\ ,44.78]\quad ,\quad R_2 =]44.78\ ,75]$

أما الإشارة السالبة لـ (k) فتعطينا حلاً مرفوضاً . (تأكد من ذلك بنفسك)

2- لو افترضنا أن $C(1/2) = C(2/1)$ وكانت نسبة القروض المتعثرة P_1 مساوية لنسبة القروض غير المتعثرة $(P_1 = P_2)$ فإن ذلك يجعل القاعدة (2-20) تأخذ الشكل التالي :

$$\frac{f_1(x)}{f_2(x)} \geq 1$$

$$\frac{C_1(x-25)(75-x)^3}{C_2(x-25)^3(75-x)} \geq 1$$

وبعد الاختصارات (ولأن $C_1 = C_2$) نحصل على أن:

$$\frac{(75-x)^2}{(x-25)^2} \geq 1$$

وللحصول على الحد الفاصل نأخذ المساواة ونجذر الطرفين فنحصل على أن:

$$\frac{75-x}{x-25} = \pm 1$$

لنأخذ الإشارة الموجبة (لأن السالبة تعطينا حلاً مرفوضاً) .

فنجد أن
$$75 - x = x - 25$$
$$100 = 2x \implies x = 50$$

وهكذا نحصل على أن الحد الفاصل بين المجموعتين G_1 و G_2 في هذه الحالة يساوي $x_0 = 50$ ، وهو يساوي القيمة الوسطى لمتوسطي المجموعتين μ_1 و μ_2 ، وهو يعطينا أن :

$$R_1 = [25\ ,50] \quad , \quad R_2 =]50\ ,75]:$$

وهو يقسم المجال $[75\ ,\ 25]$ إلى قسمين متساويين.

ملاحظة: إن نتائج هذه المسألة تختلف عن نتائج المثال العددي الأول المعروض في الجدول (2–1) لأننا أخذنا جميع عناصر المجتمع وحسبنا التوزيعين الاحتماليين للدخل في المجموعتين، ولم نكتف بعينة صغيرة منهما .

3–2 طريقة بايز Bayese (بتكاليف متساوية):

يمكن تصنيف القروض حسب دخول أصحابها حسب طريقة أخرى تسمى طريقة (بايز) أو طريقة الاحتمالات اللاحقة . وتعتمد هذه الطريقة على حساب احتمال تقاطعات الحوادث، فإذا كان لدينا حادثان B, A من جملة Ω ، فإن احتمال تقاطعهما يحسب من العلاقة التالية:

$$P(A \cap B) = P(A) * P(B/A) = P(B) * P(A/B) \qquad (27-2)$$

ومنها نحسب الاحتمالات اللاحقة $P(B/A)$ فنجد أنها تساوي:

$$P(B/A) = \frac{P(B) * P(A/B)}{P(A)} \qquad (28-2)$$

ولتطبيق ذلك في التحليل التمييزي نتجاهل تكلفة التصنيف الخاطئ أو نفترضها متساوية، ونفترض أنه لدينا عنصراً عشوائياً مثل x_0 من دخول مجتمع القروض . فهو يمكن أن يكون واقعاً في أحد المجالين R_1 أو R_2 من R، ويمكن كتابة ذلك كحادث كما يلي:

$$x_0 = R \cap x_0 = [R_1 \cup R_2] \cap x_0 = [R_1 \cap x_0] \cup [R_2 \cap x_0] \qquad (29-2)$$

وهكذا نجد أن الاحتمال الكلي لوقوع x_0 في أحد المجالين R_1 و R_2، يساوي حسب (2–27) و(2–29) ما يلي:

$$P(x_0) = P(R_1) * P(x_0/R_1) + P(R_2) * P(x_0/R_2) \qquad (30-2)$$

وبما أن: $P(R_1) = P_1$، و$P(R_2) = P_2$، وأن: $P(x_0/R_1) = f_1(x_0)$، وأن: $P(x_0/R_2) = f_2(x_0)$ فإن العلاقة (2–30) تأخذ الشكل التالي:

$$P(x_0) = P_1 f_1(x_0) + P_2 * f_2(x_0) \qquad (31-2)$$

ومن جهة أخرى نجد أن احتمال أن يقع x_0 في R_1 يساوي حسب العلاقة (2–27) ما يلي:

$$\boldsymbol{P(x_0 \cap R_1) = P(x_0) * P(R_1/x_0) = P(R_1) * P(x_0/R_1) = P_1 f_1(x_0)} \qquad \boldsymbol{(32-2)}$$

ومنها نجد أن الاحتمال اللاحق $P(R_1/x_0)$ يساوي:

$$P(R_1/x_0) = \frac{P_1 f_1(x_0)}{P(x_0)} = \frac{P_1 * f_1(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} \qquad (33-2)$$

وهو احتمال أن ينتمي العنصر المختار x_0 إلى المجال R_1، ويكون القرض المقابل له منتمياً لمجموعة القروض المتعثرة G_1 .

وبطريقة مشابهة نجد أن:

$$P(R_2/x_0) = \frac{P_2 * f_2(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} \qquad (34-2)$$

وهو احتمال أن ينتمي العنصر المختار x_0 إلى المجال R_2، وبذلك يكون القرض المقابل له منتمياً لمجموعة القروض غير المتعثرة G_2 .

ملاحظة: هنا يجب أن يكون:

$$P(R_1/x_0) + P(R_2/x_0) = 1 \qquad (35-2)$$

ويتم استخدام هذين الاحتمالين اللاحقين في تصنيف القروض إلى إحدى المجموعتين G_1 وG_2 حسب أكبر قيمة لهذين الاحتمالين كما يلي:

قاعدة (بايز) : إذا كان $P(R_1/x_0) > P(R_2/x_0)$، فإننا نصنف القرض المقابل في G_1

$(36-2)$ أما إذا كان $P(R_2/x_0) > P(R_1/x_0)$، فإننا نصنف القرض المقابل في G_2

مثال (2-2): لنفترض في مثالنا السابق إن دخل أحد الزبائن كان $x_0 = 40$، فكيف سنصنف القرض الذي سيحصل عليه هذا الزبون؟

نعلم من البيانات السابقة أن الاحتمالات السابقة تساوي:

$$P_1 = 0.30 \qquad\qquad P_2 = 0.70$$

ولحساب قيمتي $f_1(x_0)$ و$f_2(x_0)$ نعوض $x_0 = 40$ في قانوني التوزيع بيتا (a) و(b) فنجد أن: (علماً بأن $C_1 = C_2 = C$) :

$$f_1(x_0) = B_{2,4}(x_0) = C(40-25)(75-40)^3 = C * 643125$$
$$f_2(x_0) = B_{4,2}(x_0) = C(40-25)^3(75-40) = C * 118125$$

وبذلك نجد أن الاحتمالين اللاحقين يساويان:

$$P(R_1/x_0) = \frac{P_1 f_1(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} = \frac{0.3 * C * 643125}{0.3 * C * 643125 + 0.7 * C * 118125} = 0.70$$
$$P(R_2/x_0) = \frac{P_2 f_2(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} = \frac{0.7 * C * 118125}{0.3 * C * 643125 + 0.7 * C * 118125} = 0.30$$

وبمقارنة هذين الاحتمالين نجد أن:

$P(R_1/x_0) > P(R_2/x_0)$, لذلك فإننا نصنف القرض المقابل لذلك الدخل ضمن مجموعة القروض المتعثرة G_1. وهذا أمر يتوافق مع طريقة القيمة المتوقعة للتكاليف ECM، حيث وجدنا أنه عندما تكون التكاليف متساوية فإن الحد الفاصل التي يميز بين المجموعتين هو $\bar{x} = 44.78$ ، وبما أن $x_0 = 44.78 < 40$ فإن القرض يقع ضمن المجال R_1 المقابل للقروض المتعثرة .

ولنفترض الآن أن دخل زبون آخر كان $x_0 = 55$ فكيف سنصنف القرض الذي سيحصل عليه ذلك الزبون؟

للإجابة على ذلك السؤال نلاحظ أولاً أن قيمتي كل من الاحتمالين السابقين هي نفسها وتساوي:

$$P_1 = 0.30\,, P_2 = 0.70$$

ثم نقوم بحساب $f_1(x_0)$ و $f_2(x_0)$ فنجد أن: (لأن $C_1 = C_2 = C$):

$$f_1(x_0) = B_{24}(x_0) = C(55-25)(75-55)^3 = C * 240000$$
$$f_2(x_0) = B_{42}(x_0) = C(55-25)^3(75-55) = C * 540000$$

وبذلك نجد أن :

$$P(R_1/x_0) = \frac{0.3 * C * 240000}{0.3 * C * 240000 + 0.7 * C * 540000} = 0.16$$
$$P(R_2/x_0) = \frac{0.7 * C * 240000}{0.3 * C * 240000 + 0.7 * C * 540000} = 0.84$$

وبما أن $P(R_2/x_0) > P(R_1/x_0)$ فإننا نصنف القرض المقابل للدخل $x = 55$ ضمن مجموعة القروض غير المتعثرة G_2 .

وهذا يتفق مع طريقة القيمة المتوقعة للتكاليف ECM عندما تكون التكاليف متساوية، حيث نجد أن $x_0 > 44.78$، أي أن x_0 يقع في المجال R_2 المقابل لمجموعة القروض غير المتعثرة G_2 .

4-2 طريقة المسافة الأقرب إلى المركزين (لمتحول واحد ومجموعتين):

بعد دراسة خصائص المجموعات في المجتمع يمكننا تصنيف أي عنصر جديد مثل x_0 إلى إحدى المجموعتين بالاعتماد على قربه من مركزها . لذلك نقوم بحساب مربعي المسافة على المحور OX بين تلك النقطة ومركزي المجموعتين الأولى والثانية فنحصل على أن:

$$D_1^2 = (x_0 - \mu_1)^2 \qquad D_2^2 = (x_0 - \mu_2)^2 \qquad (37-2)$$

ثم نقارن بينهما ونتخذ القرار كما يلي:

إذا كانت : $D_1^2 < D_2^2$ نصنف العنصر x_0 في المجال R_1 والقرض في G_1 .

أما إذا كانت : $D_2^2 < D_1^2$ نصنف العنصر x_0 في المجال R_2 والقرض في G_2 . $(38-2)$

مثال (2−3): لنفترض أن دخل أحد المقترضين في المثال السابق كان $x_0 = 60$، ولتصنيفه في إحدى المجموعتين نحسب مربعي المسافة عن المركزين μ_1 و μ_2، فنجد أن :

$$D_1^2 = (60 - \mu_1)^2 = (60 - 41.666)^2 = 336.11$$
$$D_2^2 = (60 - \mu_2)^2 = (60 - 58.333)^2 = 2.772$$

ومن المقارنة نجد أن: $D_2^2 < D_1^2$، لذلك نصنف x_0 في المجال R_2 والقرض المقابل له في G_2، أي في مجموعة القروض غير المتعثرة . وهذا يتفق مع الطرائق السابقة .

2−5 طريقة تحليل التباين (لمتحول واحد ومجموعتين):

تعتمد هذه الطريقة على تحليل التباين الكلي للمتحول X في المجتمع المدروس. لذلك نعرف مجموع مربعات انحرافات قيم X عن المتوسط العام μ . ثم نقوم بتحليله إلى مجموع مربعات الانحرافات عن متوسطي المجموعتين μ_1 و μ_2 (داخل كل مجموعة)، وإلى مجموع مربعات انحرافات متوسطي المجموعتين μ_1 و μ_2 عن المتوسط العام μ كمايلي:

لذلك نرمز بـ x_{ij} للعنصر i الذي ينتمي إلى المجموعة G_j، ولنرمز لمجموع مربعات الانحرافات عن المتوسط العام μ بـ SST من (Sum Square Total) فيكون لدينا :

$$SST = \sum_{i=1}^{n_j} \sum_{j=1}^{2} \left(x_{ij} - \mu\right)^2 \qquad (39-2)$$

نطرح ونضيف متوسط المجموعة μ_j من داخل القوس ونعالجه كمايلي:

$$\begin{aligned} SST &= \sum_{i=1}^{n_j} \sum_{j=1}^{2} \left(x_{ij} - \mu_j + \mu_j - \mu\right)^2 \\ &= \sum_{i=1}^{n_j} \sum_{j=1}^{2} \left[\left(x_{ij} - \mu_j\right)^2 + \left(\mu_j - \mu\right)^2 - 2\left(x_{ij} - \mu_j\right)\left(\mu_j - \mu\right)\right] \\ &= \sum_{i=1}^{n_j} \sum_{j=1}^{2} \left(x_{ij} - \mu_j\right)^2 + \sum_{j=1}^{2} \sum_{i=1}^{n_i} \left(\mu_j - \mu\right)^2 - \qquad (\mathbf{40-2}) \\ &\qquad -2 \sum_{j=1}^{2} \left(\mu_j - \mu\right) \sum_{i=1}^{n_i} \left(x_{ij} - \mu_j\right) \end{aligned}$$

نلاحظ أن الحد الأخير معدوم لأن $\sum_{i=1}^{n_i}\left(x_{ij} - \mu_j\right) = 0$ في كل مجموعة j، وبعد إجراء بعض الاصلاحات نحصل على أن:

$$SST = \sum_{j=1}^{2} \sum_{i=1}^{n_i} \left(x_{ij} - \mu_j\right)^2 + \sum_{j=1}^{2} n_i\left(\mu_j - \mu\right)^2 + 0 \qquad (41-2)$$

ومنها نلاحظ أن الحد الأول هو عبارة عن مجموع مربعات انحرافات القيم x_{ij} عن متوسطات المجموعات μ_j وذلك داخل كل مجموعة. ولهذا يسمى مجموع المربعات داخل المجموعات ويرمز له بالرمز SSW من (Sum Square Within) .

أما الحد الثاني فهو عبارة عن مجموع مربعات انحرافات المتوسطين μ_1 و μ_2 عن المتوسط العام μ، لذلك يسمى مجموع المربعات بين المجموعات [Sum Square Between Groups] ويرمز له بالرمز SSB . وهكذا نجد أن SST يساوي :

$$SST = SSW + SSB \qquad (42-2)$$

نقسم الطرفين في (2−42) على SST فنحصل على العلاقة الهامة التالية :

$$\frac{SSW}{SST} + \frac{SSB}{SST} = 1 \qquad (43-2)$$

والسؤال الآن كيف يمكن توظيف ذلك في تصنيف العناصر X إلى المجموعتين المذكورتين ؟

والجواب الطبيعي هو أن نجعل النسبة الداخلية $\frac{SSW}{SST}$ أصغر ما يمكن، أو أن نجعل النسبة البديلة $\frac{SSB}{SST}$ أكبر ما يمكن، وهو الأفضل .

وبكلام آخر علينا أن نصنف العناصر X داخل المجموعات بحيث تكون متجانسة داخلياً، أو بحيث نجعل مربعي الفرق بين المتوسطين μ_1 و μ_2 عن المتوسط العام μ أكبر ما يمكن. ولكن تحقيق هذا الهدف ليس بالأمر السهل ويحتاج إلى معالجات معقدة وأدوات جديدة (يمكن البحث عنها في المراجع المختصة) .

2−6 طريقة فيشر (Fisher): أو طريقة المسافة النسبية بين المركزين μ_1 و μ_2 :

انطلاقاً من طريقة تحليل التباين التي تتطلب أن نجعل النسبة $\frac{SSB}{SST}$ أكبر ما يمكن، قدم (فيشر) معياراً جديداً مشابهاً لتلك النسبة، وهو أن نجعل مربع المسافة بين متوسطي المجموعتين μ_1 و μ_2 مقسوماً على التباين الكلي σ_x^2 أكبر ما يمكن . أي أن نجعل المعيار التالي أكبر ما يمكن .

$$\frac{(\mu_1 - \mu_2)^2}{\sigma_x^2} = \frac{D^2}{\sigma_x^2} \quad \Longrightarrow Max \qquad (44-2)$$

واعتماداً على هذه الطريقة يمكننا أن نزيح الحد الفاصل بين المجموعتين m إلى اليمين أو اليسار، بحيث نجعل المقدار (2−44) أكبر ما يمكن. لأن ذلك الانزياح سيغير من انتماء العناصر X إلى المجموعتين، وبالتالي سيغير من قيمتي μ_1 و μ_2 ومن الفرق $(\mu_1 - \mu_2)$ بينهما .

مثال (2−4): لنأخذ معطيات مثال القروض السابق، ولنفترض إننا جعلنا الحد الفاصل بين الدخل في المجموعتين مساوياً لـ $x_0 = 50$ ، فعندها تتغير عناصر المجموعتين وتصبح كما يلي:

$$G_1' = \{45 \quad 29 \quad 30\} \qquad : \quad n_1 = 3$$
$$G_2' = \{56\ 56\ 56\ 56\ 56\ 56\ 56\ 64\ 64\ 72\ 72\ 72\} \ : \ n_2 = 12$$

وإن متوسطي هاتين المجموعتين يصبحان مساويان لـ :

$$\tilde{\mu}_1 = \bar{x}_1 = \frac{104}{3} = 34.6667$$

$$\tilde{\mu}_2 = \bar{x}_2 = \frac{624}{12} = 52$$

علماً بأن التباين العام لـ X يساوي $\sigma_x^2 = 13.153$ ويبقى ثابتاً .

وبذلك نجد أن معيار (فيشر) لمربع المسافة النسبية بينهما يساوي :

$$F = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\sigma_x^2} = \frac{(52 - 34.666)^2}{13.153} = \frac{300.47}{13.153} = 22.844$$

وهي أكبر مما كانت عليه في التصنيف الفعلي الأول، لأن هذه النسبة حسب التصنيف الأول والذي كان حده الفاصل $x_0 = 56$ كانت تساوي:

$$F = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\sigma_x^2} = \frac{(65.14 - 48)^2}{13.153} = \frac{293.78}{13.153} = 22.33$$

أما إذا أخذنا الحد الفاصل حسب القيمة المتوقعة للتكلفة ECM والمساوي لـ $x_0 = 39.64$ فإننا نحصل على مجموعتين جديدتين هما:

$$G_1'' = \{ 29 , 30\} \qquad : \quad n_2 = 2$$

$$G_2'' = \{45, 56, 56 , 56, 56, 56, 56, 56, 64, 64, 72, 72, 72\} \quad : \; n_2 = 13$$

وإن متوسطيهما يساويان :

$$\tilde{\mu}_1 = \bar{x}_1 = \frac{59}{2} = 29.5$$

$$\tilde{\mu}_2 = \bar{x}_2 = \frac{781}{13} = 60.077$$

وهكذا نجد أن النسبة السابقة تصبح مساوية لـ :

$$F = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\sigma_x^2} = \frac{(60.077 - 29.5)^2}{13.153} = \frac{934.95}{13.153} = 71.08$$

وهي أكبر مما كانت عليه في الحالتين المذكورتين سابقاً . وهذا يعطينا تصنيفاً للقروض ضمن مجموعتين متجانستين داخلياً ومتباينتين خارجياً، وهما المجموعتان G_1'' و G_2'' المذكورتين أعلاه، وهذا ما يجعل مربع المسافة بين مركزي هاتين المجموعتين أكبر ما يمكن .

الفصل الثالث

التحليل التمييزي الخطي (بمتحولين ومجموعتين)

3-1 : تمهيد:

لنأخذ المثال الوارد في التمهيد (2-1) والذي يتضمن القروض المصرفية الممنوحة لزبائن أحد المصارف. ولنفترض الآن أن هذه القروض مصنفة إلى مجموعتين: المتعثرة G_1 وغير المتعثرة G_2، ولنفترض إننا نريد أن نميز بينهما باستخدام متحولين هما:

- الدخل الشهري لصاحب القرض ونرمز له بـ X_1 (بالآلاف) .
- العمر الذي يبلغه صاحب القرض عند المنح ونرمز له بـ X_2 (بالسنوات)

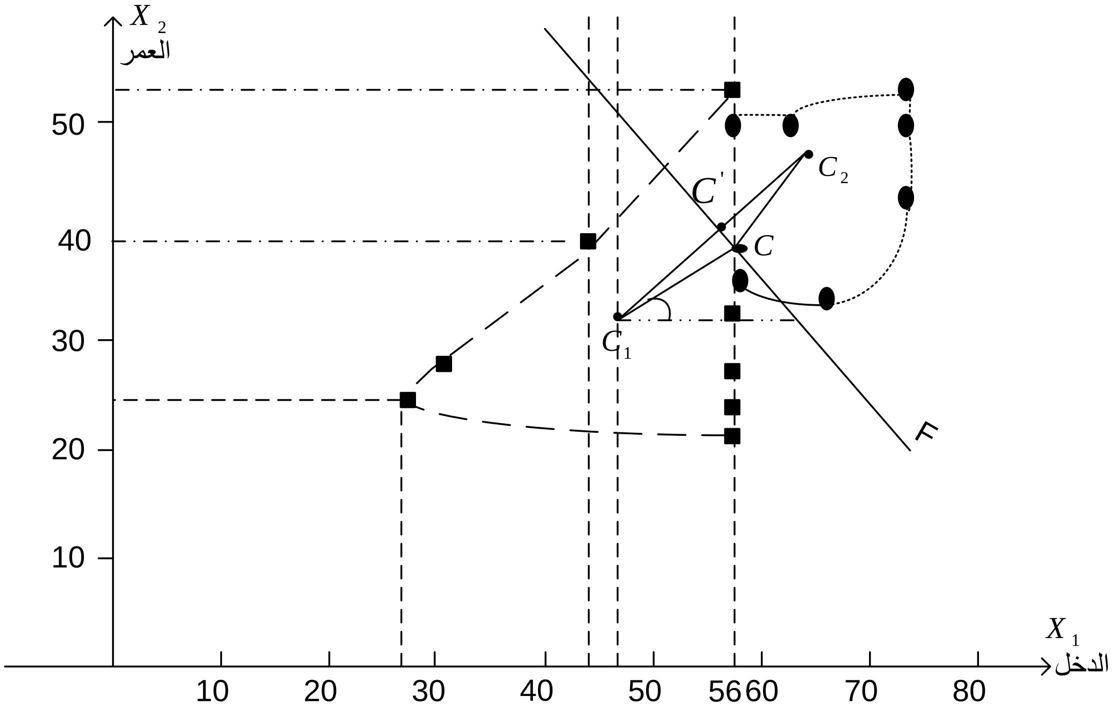
ولنفترض كالسابق أننا سحبنا عينتين عشوائيتين من هاتين المجموعتين بحجمين: $n_1 = 8$ و $n_2 = 7$ ، فحصلنا منهما على البيانات المبينة في الجدول التالي:

جدول (3-1): بيانات عينتي القروض . (فرضية)

مجموعة القروض غير المتعثرة G_2								
عناصر عينة القروض غير المتعثرة $n_2 = 7$								المتوسط
الرقم	1	2	3	4	5	6	7	
الدخل X_1	72	72	56	72	64	64	56	65.14
العمر X_2	54	50	48	44	47	35	36	44.86
إشارة F	+	+	+	+	+	+	−	+
D_1^2	8.95	…	…	…	…	….	0.51	
D_2^2	1.57	…	…	…	…	…	31.8	
التصنيف حسب المسافة الاصغر	2	2	2	2	2	2	1	2

مجموعة القروض المتعثرة G_1									
عناصر عينة القروض المتعثرة $n_1 = 8$									المتوسط
الرقم	1	2	3	4	5	6	7	8	
الدخل X_1	56	56	45	56	29	56	56	30	48
العمر X_2	52	23	40	34	26	25	27	29	32
إشارة F	+	−	−	−	−	−	−	−	−
D_1^2	4.6	1.29	6.88	….	…	….	…	2.33	
D_2^2	2.6	10.14	8.30	…	…	…	…	28.84	
التصنيف حسب المسافة الاصغر	2	1	1	1	1	1	1	1	1

وعند رسم هذه القروض كنقاط هندسية إحداثياتها (X_1 , X_2) في المستوي جعلنا المحور الأفقي للدخل X_1 والعمودي للعمر X_2 فحصلنا على الشكل البياني التالي:



الشكل (3−1) مواقع عناصر المجموعتين: ■عناصر G_1 ، ● عناصر G_2 ومراكزهما

ومن البيانات السابقة نرسم مركزي هاتين المجموعتين ونحسب إحداثياتهما فنجد أن:

C_1 مركز المجموعة الأولى G_1 وإحداثياته $C_1(48\,,32)$.

C_2 مركز المجموعة الثانية G_2 وإحداثياته $C_2(65.14\,,44.86)$.

كما يمكننا حساب ورسم المركز العام لهذه القروض. ونرمز له بـ C ونحسب إحداثياته من متوسطي الدخل والعمر في العينتين معاً فنحصل على أن $C(56\,,38)$، كما يمكننا حساب القيمة الوسطى لمركزي المجموعتين من متوسطي إحداثيات المركزين السابقين فنحصل على القيمة الوسطى للمركزين ونرمز له بـ C' وتكون إحداثياته :

$$C'\left(\frac{48+65.14}{2}\,,\frac{32+44.86}{2}\right)=C'(56.57\,,38.43)$$

ملاحظة: إن القيمة الوسطى C' تقع على القطعة المستقيمة $C_1\,,C_2$ وتقع في منتصفها. ولكن المركز العام C يقع خارج القطعة المستقيمة $C_1\,,C_2$ كما موضح على الشكل (3−1) .

والسؤال الآن كيف يمكننا أن نصيغ علاقة رياضية للتمييز بين هاتين المجموعتين؟

الجواب يمكن أن يكون بعدة طرائق رياضية هي:

3−2 : طريقة المستقيم الفاصل بين عناصر المجموعتين، ويسمى تابع التمييز القانوني Canonical Discriminant Function

تتلخص هذه الطريقة بإنشاء مستقيم يمر من نقطة القيمة الوسطى للمركزين $C'(56.57\,,38.43)$ (أو من أي نقطة مناسبة أخرى)، ويتعامد (أو يتقاطع) مع القطعة المستقيمة الواصلة بين مركزي المجموعتين

$C_1(48\,,32)$ و $C_2(65.14\,,44.86)$، فنحصل على مستقيم يفرز عناصر العينة إلى مجموعتين جديدتين منفصلتين: المجموعة الأولى تقع تحته(على يساره)، والمجموعة الثانية تقع فوقه(على يمينه) .

ولإيجاد معادلة هذا المستقيم الفاصل بين هاتين المجموعتين نقوم بإيجاد ميل القطعة المستقيمة $\overline{C_1\,C_2}$ الواصلة بين مركزي المجموعتين من العلاقة :

$$m = tg\theta = \frac{44.86 - 32}{65.14 - 48} = \frac{12.86}{17.14} = 0.75029$$

ثم نقوم بحساب ميل المستقيم المتعامد معها من العلاقة:

$$m' = -\frac{1}{m} = \frac{-1}{0.75020} = -1.3328$$

ثم نقوم بحساب معادلة المستقيم المعامد، والذي يمر من نقطة القيمة الوسطى للمركزين $C'(56.57\,,38.43)$ وميله يساوي $m' = -1.3328$ ، من العلاقة :

$$(X_2 - 38.43) = -1.3328(X_1 - 56.57)$$

وبعد الإصلاح نحصل على معادلة المستقيم التالي:

$$X_2 + 1.3328X_1 - 113.8265 = 0 \qquad (1-3)$$

وهنا نلاحظ أن هذا المستقيم يقسم نقاط المستوى إلى نصفين: نصف موجب يقع فوقه ونصف سالب يقع تحته وجزء معدوم يقع عليه . كما أن هذا المستقيم يفرز عناصر العينة إلى مجموعتين منفصلتين:

المجموعة الأولى: هي النقاط التي تقع تحته (في النصف السالب) وتضم معظم نقاط القروض المتعثرة في G_1 [ماعدا القرض الأول (52 ، 56)] .

المجموعة الثانية: وهي النقاط التي تقع فوقه (في النصف الموجب) ويضم معظم نقاط القروض غير المتعثرة في G_2 (ماعدا القرض الأخير (36 ، 56) .

أي أنه يمكننا استخدام هذا المستقيم (أو أي مستقيم أفضل منه) كأداة لفرز القروض إلى متعثرة وغير متعثرة باحتمال خطأ معقول .

ولذلك نفترض أن معادلة المستقيم تساوي تابعاً F ونكتبها كما يلي:

$$F = X_2 + 1.3328X_1 - 113.8265 \qquad (2-3)$$

ولاستخدام هذه المعادلة في تصنيف عناصر العينة أو في تصنيف أي عنصر جديد من المجتمع. نقوم بتعويض إحداثيات صاحب القرض($X_1\,,X_2$) في المعادلة $(3-2)$ ونحسب قيمة F ثم نتخذ القرار (في هذا المثال) حسب إشارة قيمة F كما يلي:

إذا كانت $F \leq 0$ فإن القرض ينتمي إلى المجموعة المتعثرة G_1 .

أما إذا كانت $F > 0$ فإن القرض ينتمي إلى المجموعة غير المتعثرة G_2 .

وإذا طبقنا ذلك على عناصر العينتين المسحوبتين من ($G_1\,,G_2$)، نحصل على السطرF في الجدول $(3-1)$. وإذا قارنا التصنيف الجديد مع التصنيف الفعلي أو الإداري السابق نحصل على الجدول التالي:

جدول (3−2): نتائج التصنيف الجديد مع التصنيف الأصلي:

	التصنيف الجديد			
التصنيف الفعلي الإداري	المجموعات	G_1	G_2	المجموع
	G_1	7	1	8
	G_2	1	6	7

وهنا نلاحظ أن نسبة التصنيف الصحيح تساوي نسبة مجموع عدد العناصر القطرية على حجم العينة الكلية n، أي:

$$P = \frac{7+6}{15} = 0.8667 = (86.67\ \%)$$

وإن نسبة التصنيف الخاطئ تساوي نسبة مجموع العناصر غير القطرية على حجم العينة الكلية n،، أي:

$$q = \frac{1+1}{15} = 0.1333 = (13.33\ \%)$$

وهذا يعني أنه يمكننا الاعتماد على المستقيم السابق لتصنيف عناصر المجتمع مع خطأ معقول .

ملاحظة: إن هذا المستقيم الفاصل بين هاتين المجموعتين ليس هو الفاصل المثالي بينهما، بل هو أحد المستقيمات الفاصلة بينهما، لأنه يمكن أن ننشأ مستقيمات كثيرة أخرى تفصل بين هاتين المجموعتين وبطريقة أفضل. فمثلاً يمكننا أن نحرك هذا المستقيم يميناً أو يساراً، أو إلى الأعلى أو إلى الأسفل، لنحصل على لانهاية من المستقيمات الفاصلة بين هاتين المجموعتين. كما يمكننا أن نغير ميلان هذا المستقيم m' إلى الأكبر أو إلى الأصغر، وندوره حول نقطة القيمة الوسطى نحو اليمين أو نحو اليسار بزاوية θ، لنحصل على لانهاية أخرى من المستقيمات الفاصلة بين هاتين المجموعتين .

لذلك كان لابد لنا من أن نضع أسس رياضية دقيقة لإنشاء ذلك المستقيم الفاصل بين هاتين المجموعتين بحيث نجعل الاحتمال الاجمالي للتصنيف الخاطئ TPM أصغر ما يمكن، أو نجعل القيمة المتوقعة لتكاليف التصنيف الخاطئ ECM أصغر ما يمكن. وهذا ما سنتعرض إليه لاحقاً .

وأخيراً نشير إلى أنه عندما تكون المجموعات متقاطعة ومتشابكة، فإنها تكون غير قابلة للفصل بواسطة خط مستقيم، وتسمى مثل هذه المجموعات بالمجموعات غير القابلة للفصل خطياً، وفي هذه الحالة نلجأ إلى استخدام منحنيات مناسبة للفصل بينها، وبذلك ينتقل التحليل إلى التحليل التمييزي غير الخطي، وهو موضوع خارج إطار هذا الكتاب.

3−3 : طريقة الجوار الأقرب (أصغر المسافات عن مراكز المجموعات):

تعتمد هذه الطريقة على حساب مربع المسافة بين كل عنصر x_0 ومركزي المجموعتين (G_1 , G_2)، واللذين يساويان حسب الجدول (3−1) ما يلي:

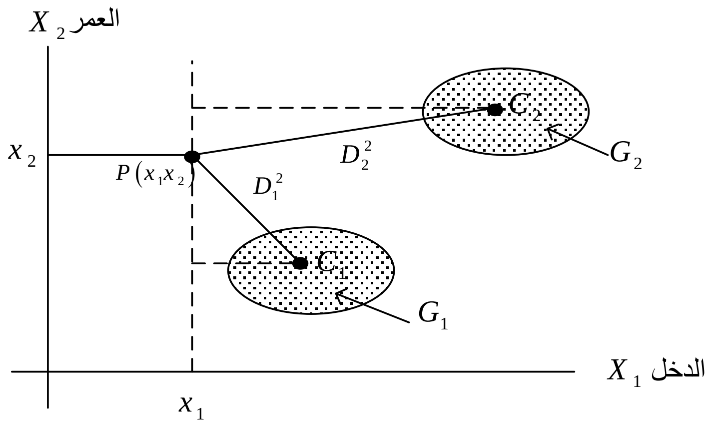
جدول (3−3): إحداثيات مركزي المجموعتين بالرموز والأعداد:

	G_1	G_2
X_1	$\bar{x}_{11}$	$\bar{x}_{12}$
X_2	$\bar{x}_{21}$	$\bar{x}_{22}$

=

	G_1	G_2
الدخل X_1	48	65.14
العمر X_2	32	44.86

لنأخذ الآن نقطة ممثلة لأحد القروض الجديدة مثل النقطة P والتي إحداثياتها يساويان $P(x_1\,,x_2\,)$، ولنرمز لمربعي المسافة الهندسية منها حتى مركزي المجموعتين C_1 و C_2 بالرمزين D_1^2 و D_2^2 على الترتيب ونرسمها على الشكل البياني كما يلي:



الشكل (3−2): المسافات عن مركزي المجموعتين

ولتحديد المجموعة التي ينتمي إليها ذلك القرض الجديد نحسب مربعي المسافتين D_1^2 و D_2^2 ثم نقارن بينهما ونتخذ القرار حول الانتماء كما يلي:

إذا كانت $D_1^2 \leq D_2^2$ فإننا ننسب النقطة P إلى المجموعة G_1 .

أما إذا كانت $D_1^2 > D_2^2$ فإننا ننسب النقطة P إلى المجموعة G_2 .

ولكن هذه المقارنة لا تجوز إلا إذا كانت القيم العددية لهاتين المسافتين مقاسة بوحدات قياس موحدة، وبما أن X_1 − في مثالنا− يقاس بآلاف الليرات و X_2 يقاس بالسنوات، فلا يجوز مقارنة هذين المتحولين بطريقة مباشرة .

وللتخلص من هذه المشكلة نقوم بتحويل هذين المتحولين إلى متحولين معياريين، وذلك بإجراء التحويل المعياري $Z = \frac{X-\bar{X}}{\sigma}$، على كل متحول وفي كل مجموعة على حدة، فنحصل على قيم مجردة من الوحدات . ولتوضيح ذلك نحسب الانحرافات المعيارية لكل متحول وفي كل مجموعة على حدة، فنحصل على الانحرافات المعيارية التالية:

جدول (3−4): الانحرافات المعيارية لعناصر العينيتن:

المجموعات / المتحولات	G_1	G_2
X_1	S_{11}	S_{12}
X_2	S_{21}	S_{22}

=

المجموعات / المتحولات	G_1	G_2
X_1	12.036	7.198
X_2	9.769	7.081

وبذلك نجد أن مربع المسافة المعيارية أو الإحصائية بين النقطة$P(X_1\,,X_2)$ ومركز المجموعة الأولى $C_1(48\,,32)$ يحسب من العلاقة:

$$D_1^2=\left(\frac{X_1-48}{12.036}\right)^2+\left(\frac{X_2-32}{9.769}\right)^2 \qquad (3-3)$$

وإن مربع المسافة المعيارية أو الإحصائية بين النقطة$P(X_1\,,X_2)$ ومركز المجموعة الثانية $C_2(65.14\,,44.86)$ يحسب من العلاقة:

$$D_2^2=\left(\frac{X_1-65.14}{7.198}\right)^2+\left(\frac{X_2-44.86}{7.081}\right)^2 \qquad (4-3)$$

ومنهما نحصل على القيمتين D_1^2 و D_2^2 كقيم عددية مجردة من وحدات القياس . لذلك يمكن مقارنتهما وتحديد المجموعة التي تنتمي إليها النقطة P ، وهي المجموعة التي تقابل أصغر المربعين D_1^2 أو D_2^2 .

فمثلاً نجد أن مربعي المسافة للقرض الأول من G_1 وهو الذي يقابل النقطة $P(56\,,52)$ يساويان :

$$D_1^2=\left(\frac{56-48}{12.036}\right)^2+\left(\frac{52-32}{9.769}\right)^2=0.44418+4.1914=4.6332$$

$$D_2^2=\left(\frac{56-65.14}{7.198}\right)^2+\left(\frac{52-44.86}{7.081}\right)^2=1.6124+1.0167=2.6291$$

وعند المقارنة نجد أن:$D_2^2<D_1^2$ وهذا يعني أن النقطة Pهي أقرب إلى مركز المجموعةG_2، لذلك سنعتبر ذلك القرض قرضاً غير متعثر وينتمي إلى المجموعة G_2، مع أنه ينتمي إلى G_1(**انظر الجدول (1−3)**).

وهكذا قمنا بحساب مربعات المسافات المقابلة لبعض عناصر العينة (عينة القروض) في المثال السابق ووضعناها في سطرين ملحقين بالجدول (3−1)، وبناء على مقارنة D_1^2 و D_2^2 في هذين السطرين تم تحديد انتماء كل عنصر من عناصر عينة البحث، ووضعناها في السطر الأخير، فحصلنا على نفس النتائج التي حصلنا عليها بطريقة المستقيم الفاصل المبينة في الجدول (3−2).

ويمكن الاستفادة من هذه الطريقة في تحديد انتماء أي قرض جديد والتنبؤ بمصيره قبل اتخاذ قرار الموافقة على منحه .

ملاحظة: لنأخذ المجموعتين الجديدتين G_1' و G_2'، اللتين أصبحتا بعد التصنيف الجديد حسب طريقة الجوار الأقرب تضمان العناصر التالية:

$$G_1':\{7_{(1)},2,3,4,5,6,7,8\}$$

$$G_2':\{\,1,2,3,4,5,6,\ 8_{(1)}\,\}$$

وإذا حسبنا متوسطات X_1 و X_2 في هاتين المجموعتين الجديدتين نحصل على المتوسطات والتباينات التالية:

الانحرافات المعيارية الجديدة				المتوسطات الجديدة للمجموعات		
	G_1	G_2			G_1	G_2
X_1	12.036	7.198		X_1	48	56
X_2	6	6.283		X_2	30	47.14

أي أن إحداثيات مركز المجموعة الجديدة G_1' أصبح $C_1'(48, 30)$.

وأن إحداثيات مركز المجموعة الجديدة G_2' أصبح $C_2'(56, 47.14)$.

وإن مربع المسافة الهندسية بين هذ المركزين يساوي:

$$D_1'^2 = (48 - 56)^2 + (30 - 47.11)^2 = 357.78$$

بينما كان مربع المسافة بينهما حسب التصنيف الفعلي (الإداري) يساوي مربع البعدين المركزين الفعليين اللذين إحداثياتها $C_1(48, 32)$ $C_2(56, 44.86)$ ويساوي:

$$D_1^2 = (48 - 56)^2 + (32 - 44.86)^2 = 211.38$$

وهكذا نلاحظ أن التصنيف الجديد أدى إلى جعل مربع المسافة بين مركزي المجموعتين أكبر مما كان عليه قبل ذلك. وأصبح (357.78) بدلاً من (211.38) . وهذا هو المعيار الذي استنبطه (فيشر) للتمييز بين عناصر المجموعتين G_1 و G_2.

كما نلاحظ أن التصنيف السابق جعل مجموع مربعات الانحرافات داخل كل مجموعة SSW يصبح أصغر أو يساوي مما كان عليه، وبالمقابل جعل مجموع مربعات الانحرافات بين هاتين المجموعتين SSB يصبح أكبر مما كان عليه .

وهذا يعني أن المجموعتين أصبحتا أكثر تجانساً داخلياً وأكثر اختلافاً خارجياً، وهو ما نسعى إليه من خلال التحليل التمييزي بين المجموعات .

3-4 طريقة القيمة المتوقعة للتكاليف ECM (لمتحولين طبيعيين (X_1, X_2) ومجموعتين (G_1, G_2)): [Johnson,Wichern. P.485 بتصرف] .

قبل البحث في التحليل التمييزي لهذين المتحولين نرمز لهما بـ X ولتوقعهما بـ $\boldsymbol{\mu}$ ونكتبهما على شكل شعاعين عمودين كما يلي:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad (5-3)$$

وبذلك نجد أن التوقع الرياضي لهما في المجتمع ككل يساوي:

$$\boldsymbol{E}(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \boldsymbol{\mu} \qquad (6-3)$$

وعندها فإن مصفوفة التباين المشترك لهما تساوي:

$$\boldsymbol{V} = cov\,(X_1\,, X_2) = E(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'$$
$$= E\begin{bmatrix} \mathrm{X}_1-\mathrm{\mu}_1 \\ \mathrm{X}_2-\mathrm{\mu}_2 \end{bmatrix} * [(X_1-\mu_1)\,,(X_2-\mu_2)]$$
$$\boldsymbol{V} = E\begin{bmatrix} (\mathrm{X}_1-\mathrm{\mu}_1)^2\,, & (\mathrm{X}_1-\mathrm{\mu}_1)(X_2-\mu_2) \\ (X_2-\mu_2)(\mathrm{X}_1-\mathrm{\mu}_1)\,, & (\mathrm{X}_2-\mathrm{\mu}_2)^2 \end{bmatrix} = \begin{bmatrix} E(\mathrm{X}_1-\mathrm{\mu}_1)^2\,, & E(\mathrm{X}_1-\mathrm{\mu}_1)(X_2-\mu_2) \\ E(X_2-\mu_2)(\mathrm{X}_1-\mathrm{\mu}_1)\,, & E(\mathrm{X}_2-\mathrm{\mu}_2)^2 \end{bmatrix}$$
$$\boldsymbol{V} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad (7-3)$$

حيث أن: σ_{11} هو تباين X_1 و σ_{22} هو تباين X_2

وأن: $\sigma_{12} = \sigma_{21}$ هو التباين المشترك للمتحولين $(X_1\,, X_2)$. وأن V هي مصفوفة مربعة متناظرة .

ومن العلاقة (3−7) يمكننا أيضاً حساب المصفوفة الارتباطية المتناظرة للمتحولين $(X_1\,, X_2)$ والتي تعتمد على أن معامل الارتباط بينهما ρ_{12} يحسب من العلاقة: $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}*\sqrt{\sigma_{22}}}$ وبذلك يكون لدينا:

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}*\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}*\sqrt{\sigma_{22}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}*\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}*\sqrt{\sigma_{22}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \qquad (8-3)$$

حيث أن: $\rho_{12} = \rho_{21}$ هو معامل الارتباط بين المتحولين $(X_1\,, X_2)$

ومن المصفوفة (3−7) يمكننا استخلاص مصفوفة الانحرافات المعيارية الخاصة بهذين المتحولين، والتي سنرمز لها (تجاوزاً) بالرمز $\sigma^{\frac{1}{2}}$. وهي المصفوفة القطرية المؤلفة من جذور العناصر القطرية في المصفوفة V ونكتبها كما يلي:

$$\boldsymbol{\sigma}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} \qquad (9-3)$$

وبسهولة يمكننا أن نستخلص أن V تساوي:

$$\boldsymbol{V} = \boldsymbol{\sigma}^{\frac{1}{2}} * \boldsymbol{\rho} * \boldsymbol{\sigma}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} * \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix}$$
$$= \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} * \begin{bmatrix} \sqrt{\sigma_{11}} & \rho_{12}\sqrt{\sigma_{22}} \\ \rho_{21}\sqrt{\sigma_{11}} & \sqrt{\sigma_{22}} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \,, & \rho_{12}\sqrt{\sigma_{11}}*\sqrt{\sigma_{22}} \\ \rho_{21}\sqrt{\sigma_{22}}*\sqrt{\sigma_{11}} \,, & \sigma_{22} \end{bmatrix}$$

وبتبديل ρ_{12} و ρ_{21} من العلاقة: $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\,\sqrt{\sigma_{22}}}$ نحصل على أن:

$$\boldsymbol{V} = \boldsymbol{\sigma}^{\frac{1}{2}} * \boldsymbol{\rho} * \boldsymbol{\sigma}^{\frac{1}{2}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad (10-3)$$

وإذا ضربنا طرفي العلاقة (3−10) من اليمين ثم من اليسار ب. $\boldsymbol{\sigma}^{-\frac{1}{2}}$ نحصل على أن:

$$\boldsymbol{\rho} = \boldsymbol{\sigma}^{-\frac{1}{2}} * \boldsymbol{V} * \boldsymbol{\sigma}^{-\frac{1}{2}} \qquad (11-3)$$

حيث أن: $\boldsymbol{\sigma}^{-\frac{1}{2}}$ هي مقلوب مصفوفة الانحرافات المعيارية $\boldsymbol{\sigma}^{\frac{1}{2}}$، وهي مصفوفة قطرية تكون عناصرها مساوية لمقاليب عناصر المصفوفة $\sigma^{\frac{1}{2}}$ أي أن:

$$\boldsymbol{\sigma}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} \end{bmatrix} \qquad (12-3)$$

ويمكن تعميم ذلك على عدة متحولات $X_1, X_2, X_3, \dots, X_p$ وعلى عدة مجموعات $G_1, G_2, \dots, G_p$ كما سنرى لاحقاً.

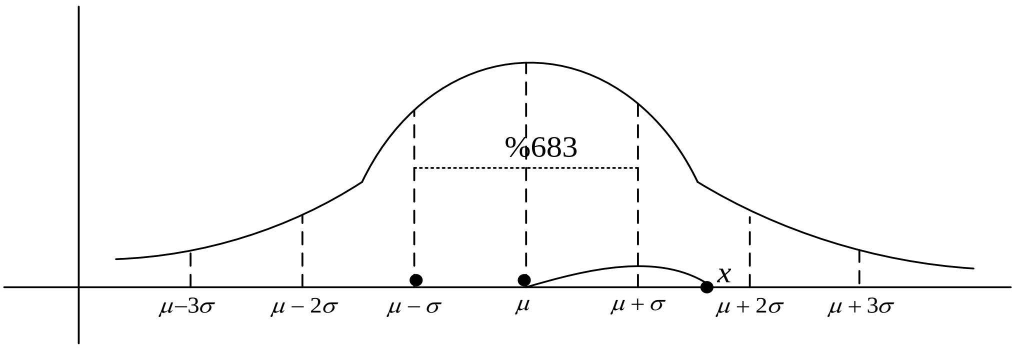
ولتطبيق طريقة القيمة المتوقعة للتكاليف على هذه الحالة (لمتحولين طبيعيين X_1 و X_2 وعلى المجموعتين G_1 و G_2) نفترض أن التوزيع الطبيعي لـ X_1 هو $N(\mu_1 \ \sigma_1^2)$. وهو يعطي بالعلاقة :

$$f(x_1) = \frac{1}{\sqrt{2\pi} * \sigma_1} * e^{-\frac{1}{2}\left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2} \ : \ -\infty < X_1 < +\infty \qquad (13-3)$$

وإن التوزيع الطبيعي لـ X_2 هو $N(\mu_2 \ \sigma_2^2)$ ، وهو يعطي بالعلاقة :

$$f(x_2) = \frac{1}{\sqrt{2\pi} * \sigma_2} * e^{-\frac{1}{2}\left(\frac{X_2 - \mu_2}{\sigma_2}\right)^2} \ : \ -\infty < X_2 < +\infty \qquad (14-3)$$

وإن كل منها يرسم الشكل البياني التالي :



الشكل (3−3): المنحني العام للتوزيع الطبيعي

وهنا نلاحظ أنه يمكن كتابة الحد الأسي في كل منهما على الشكل التالي :

$$D^2 = \left(\frac{X - \mu}{\sigma}\right)^2 = (X - \mu) * (\sigma^2)^{-1} * (X - \mu) \qquad (15-3)$$

وهو يمثل مربع المسافة الإحصائية بين النقطة X والمتوسط μ مقاسة بواحدات الانحراف المعياريσ.

ويمكننا تعميم هذه العلاقة على أي شعاع X مؤلف من P متحولاً وذات متوسط عام $\boldsymbol{\mu}$ ومصفوفة تباين مشترك $\boldsymbol{V}$ كما يلي:

$$\Delta^2 = (\boldsymbol{X} - \boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X} - \boldsymbol{\mu}) \qquad (16-3)$$

حيثΔ : هو مربع المسافة الإحصائية من X حتى μ في الفضاء R^p .

وبناء على ذلك يمكننا تعريف التوزيع المشترك للمتحولات $X_1, X_2, \dots, X_p$ باستبدال مربع المسافة D^2 في (3–15) بمربع المسافة Δ^2 ، فنحصل على العلاقة التالية :

$$f\left(X_1, X_2, \dots, X_p\right) = \frac{1}{(2\pi)^{\frac{P}{2}}|\boldsymbol{V}|^{\frac{1}{2}}} e^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X}-\boldsymbol{\mu}) \qquad (16\,a-3)$$

ولإيجاد التوزيع المشترك للمتحولين (X_1, X_2) نستبدل كل $\boldsymbol{X}$ بالشعاع $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ والتوقع $\boldsymbol{\mu}$ بالشعاع $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ ، وكما نستبدل مقلوب مصفوفة التباين $\boldsymbol{V}^{-1}$ بما تساويه والذي نحسبه من $\boldsymbol{V}$ كما يلي:

بما أن V تساوي:

$$\boldsymbol{V} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad |V| = \sigma_{11}\,\sigma_{22} - \sigma_{12}\,\sigma_{21} \qquad (17-3)$$

فإن مقلوبها V^{-1} يساوي:

$$\boldsymbol{V}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - {\sigma_{12}}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} \qquad (18-3)$$

ونكتب التوزيع المشترك للمتحولين (X_1, X_2) قياساً على العلاقة $(16\,a-3)$ كما يلي:

$$f(X) = f(X_1, X_2) = \frac{1}{(2\pi)|V|^{\frac{1}{2}}} e^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X}-\boldsymbol{\mu}) \qquad (19-3)$$

وإن هذا التوزيع المشترك يعرف على المجموعة الأولى G_1 بالعلاقة :

$$f_1(\boldsymbol{X}) = \frac{1}{(2\pi)|V_1|^{\frac{1}{2}}} e^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu_1})' * \boldsymbol{V_1}^{-1} * (\boldsymbol{X}-\boldsymbol{\mu_1}) \qquad (20-3)$$

ويعرف على المجموعة الأولى G_2 بالعلاقة :

$$f_2(\boldsymbol{X}) = \frac{1}{\sqrt{2\pi}|V_2|^{\frac{1}{2}}} e^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu_2})' * \boldsymbol{V_2}^{-1} * (\boldsymbol{X}-\boldsymbol{\mu_2}) \qquad (21-3)$$

وهنا سنميز بين حالتين هما:

- **الحالة الأولى:** عندما يكون لدينا $\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V}$ افتراضياً أو فعلياً .

وهذا يعني إننا نفترض أن مصفوفتي التباين المشترك متساويتان $(\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V})$، وإن تكلفة التصنيف الصحيح معدومة $C(2/2) = C(1/1) = 0$، وأن الاحتمال السابق للانتماء إلى المجموعة G_1 يساوي P_1، وإن الاحتمال السابق للانتماء إلى المجموعة G_2 يساوي P_2 .

وبذلك نجد أن قاعدة القيمة المتوقعة للتكاليف (2–21)، (2–22) تأخذ الشكل التالي:

$$R_1: \frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} \geq \left[\frac{C(1/2)}{C(2/2)}\right] * \left[\frac{P_2}{P_1}\right] \qquad (22-3)$$

$$R_2: \frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} < \left[\frac{C(1/2)}{C(2/1)}\right] * \left[\frac{P_2}{P_1}\right] \qquad (23-3)$$

ولذلك نحسب النسبة $\frac{f_1(x)}{f_2(x)}$ للتوزيعين الطبيعيين فنجد أنها عندما تكون $(V_1 = V_2 = V)$ تساوي:

$$\frac{f_1(X)}{f_2(X)}=\frac{\dfrac{1}{(2\pi)|V|^{\frac{1}{2}}}e^{-\frac{1}{2}}[(\boldsymbol{X}-\boldsymbol{\mu_1})'*\boldsymbol{V^{-1}}*(\boldsymbol{X}-\boldsymbol{\mu_2})}{\dfrac{1}{(2\pi)|V|^{\frac{1}{2}}}e^{-\frac{1}{2}}[(\boldsymbol{X}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}}*(\boldsymbol{X}-\boldsymbol{\mu_2})} \qquad (\mathbf{24-3})$$

نختصر على العدد $\frac{1}{(2\pi)|V|^{\frac{1}{2}}}$ ونرفع المقام إلى البسط فنجد أن:

$$\frac{f_1(X)}{f_2(X)}=\boldsymbol{e^{-\frac{1}{2}}}[(\boldsymbol{X}-\boldsymbol{\mu_1})'*\boldsymbol{V^{-1}}(\boldsymbol{X}-\boldsymbol{\mu_1})]+\frac{\mathbf{1}}{\mathbf{2}}[(\boldsymbol{X}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}}(\boldsymbol{X}-\boldsymbol{\mu_2})]=\boldsymbol{e^{Z}} \qquad (\mathbf{25-3})$$

لنأخذ الأس العام ونرمز له بـ $\boldsymbol{Z}$ ونعالجه كما يلي:

$$\boldsymbol{Z}=-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{X}-\boldsymbol{\mu_1})'*\boldsymbol{V^{-1}}(\boldsymbol{X}-\boldsymbol{\mu_1})+\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{X}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}}(\boldsymbol{X}-\boldsymbol{\mu_2})=$$

$$=-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{X}-\boldsymbol{\mu_1})'*\boldsymbol{V^{-1}X}+\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{X}-\boldsymbol{\mu_1})'*\boldsymbol{V^{-1}\mu_1}+\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{X}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{X}-\boldsymbol{\mu_2})'\boldsymbol{V^{-1}\mu_2}=$$

$$=-\frac{\mathbf{1}}{\mathbf{2}}[(\boldsymbol{X}-\boldsymbol{\mu_1})'-(\boldsymbol{X}-\boldsymbol{\mu_2})']\boldsymbol{V^{-1}X}+\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{X'}*\boldsymbol{V^{-1}\mu_1}-\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\mu_1'V^{-1}\mu_1}-\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{X'}*\boldsymbol{V^{-1}\mu_2}+\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\mu_2'V^{-1}\mu_2}=$$

$$=\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}+\frac{\mathbf{1}}{\mathbf{2}}[\boldsymbol{\mu_1'V^{-1}X}]'-\frac{\mathbf{1}}{\mathbf{2}}[\boldsymbol{\mu_2'V^{-1}X}]'-\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\mu_1'V^{-1}\mu_1}+\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\mu_2'V^{-1}\mu_2}=$$

$$=\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}+\frac{\mathbf{1}}{\mathbf{2}}[(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}]'-\frac{\mathbf{1}}{\mathbf{2}}[(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}}(\boldsymbol{\mu_1}+\boldsymbol{\mu_2})$$

$$\boldsymbol{Z}=(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})*\boldsymbol{V^{-1}}(\boldsymbol{\mu_1}+\boldsymbol{\mu_2}) \qquad (\mathbf{27-3})$$

وذلك لأن الحد الثاني يساوي: عدد حقيقي $=[(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}]'=[(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}$

ويدمج مع الحد الأول .

وبذلك نجد أن القاعدة (3–22) للقيمة المتوقعة للتكاليف نأخذ الشكل التالي:

$$R_1:\frac{f_1(X)}{f_2(X)}=e^{(\mu_1-\mu_2)'*V^{-1}X-\frac{1}{2}(\mu_1-\mu_2)'*V^{-1}(\mu_1+\mu_2)}\geq\left[\left[\frac{C(1/2)}{C(2/1)}\right]*\left[\frac{P_2}{P_1}\right]\right] \qquad (\mathbf{28-3})$$

ثم نأخذ اللوغاريتم الطبيعي للطرفين فنجد أن القاعدة (3–22) تأخذ الشكل التالي:

$$R_1:(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}}(\boldsymbol{\mu_1}+\boldsymbol{\mu_2})\geq \ell n\left[\left[\frac{C(1/2)}{C(2/1)}\right]*\left[\frac{P_2}{P_1}\right]\right] \qquad (\mathbf{29-3})$$

$$R_2:(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}X}-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})'*\boldsymbol{V^{-1}}(\boldsymbol{\mu_1}+\boldsymbol{\mu_2})< ln\left[\left[\frac{C(1/2)}{C(2/1)}\right]*\left[\frac{P_2}{P_1}\right]\right] \qquad (\mathbf{30-3})$$

وإذا لم تكن المعالم $\boldsymbol{\mu_1}$ و$\boldsymbol{\mu_2}$ وV معلومة نقوم بتقديرها من العينتين المسحوبتين من المجموعتين G_1 وG_2 بواسطة المؤشرات المقابلة لها $\boldsymbol{\bar{X}_1}$ و$\boldsymbol{\bar{X}_2}$ وبمصفوفة التباين المدمجة *Pooled* والمركبة من المصفوفتين $\boldsymbol{S_1}$ و$\boldsymbol{S_2}$ كما يلي:

$$\boldsymbol{S_p}=\frac{(n_1-1)\boldsymbol{S_1}+(n_2-1)\boldsymbol{S_2}}{n_1+n_2-2} \qquad (\mathbf{31-3})$$

وبالتالي نحصل على قاعدة القيمة المتوقعة للتكاليف في العينات من العلاقة التقديرية التالية:

$$R_1:(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})'*\boldsymbol{S_p^{-1}}*\boldsymbol{X}-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\bar{X}_1}-\boldsymbol{\bar{X}_2})'*\boldsymbol{S_p^{-1}}(\boldsymbol{\bar{X}_1}+\boldsymbol{\bar{X}_2})\geq \ell n\left[\frac{C(1/2)}{C(2/1)}*\frac{P_2}{P_1}\right] \qquad (\mathbf{32-3})$$

$$R_2: (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X - \frac{1}{2}(\bar{X}_1 - \bar{X}_2)' * S_p^{-1}(\bar{X}_1 + \bar{X}_2) < ln\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] \qquad (33-3)$$

فإذا أخذنا أية نقطة مثل x_0 وكانت تحقق العلاقة الأولى (3-23) فإننا ننسب x_0 إلى G_1 وإذا كان العكس ننسبها إلى G_2 .

ملاحظة1: إن الحد الأول في العلاقة (3-32) ماهو إلا معادلة تابع خطي لـ X لذلك سنرمز له بـ Y، فيكون لدينا :

$$Y = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X = \ell' * X \qquad (34-3)$$

أما الحد الثاني فهو عبارة عن عدد حقيقي ناتج عن جداءات الأطراف التي فيه .

ملاحظة2: يمكن أن نرمز للطرف الأيسر في العلاقة (3-32) بـ W كما فعل أندرسون *(Anderson)* فيكون لدينا :

$$W(X) = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X - \frac{1}{2}(\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * (\bar{X}_1 + X_2) =$$

$$W(X) = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * \left[X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)\right] \qquad (35-3)$$

وهي معادلة مستقيم ميله $\left[(\bar{X}_1 - \bar{X}_2)' S_p^{-1}\right]$ ويمر من النقطة الوسطى $\tilde{m} = \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$

وهذه العلاقة تسمى تابع التصنيف لأندرسون *(Anderson)* وهي تشير إلى أن المستقيم ***W(X)*** يمر من نقطة القيمة الوسطى *m*. وبذلك فإن قاعدة التصنيف (3-32) و(3-33) تأخذ الشكل التالي:

$$R_1 : W(X) \geq \ell n\left[\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right]\right] \qquad (36-3)$$

$$R_2 : W(X) < ln\left[\left[\frac{C(2/1)}{C(1/2)} * \frac{P_2}{P_1}\right]\right] \qquad (36\,a-3)$$

أما إذا كان الطرف الأيمن في (3-32): $\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] = 1$ فإن $\ell n\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] = 0$

وعندها تصبح قاعدة التصنيف كما يلي: إذا كان x_0 أي عنصر من المجتمع فإننا نصنفه كما يلي :

(3-37) إذا كان $W(x_0) \geq 0$ فإننا نصنف x_0 في المجموعة G_1

(3-37a) وإذا كان $W(x_0) < 0$ فإننا نصنف x_0 في المجموعة G_2

وهذا يتفق مع قاعدة المستقيم الفاصل بين المجموعتين، أي أنه إذا كان لدينا مجتمعان طبيعيان وكان لهما نفس مصفوفة التباين المشترك، فإن قاعدة المستقيم الفاصل بينهما تكافئ قاعدة القيمة المتوقعة للتكلفة ECM، وذلك عندما تكون فيها الاحتمالات المسبقة متساوية وتكون تكاليف التصنيف الخاطئ متساوية .

وبصورة عامة إذا كانت المتحولات طبيعية ومصفوفات التباينات متساوية، فإن مؤشر التصنيف W(X) المعرف في (3−35) يمكن أن يحسب لكل مشاهدة x_0، ثم يتم العمل على تصنيف المشاهدات بمقارنة قيم W(X)مع قيمة اللوغاريتم$\ell n\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right]$ ويتم التصنيف كما في العلاقتين(3−36) و(3− 36a).

مثال 3−1: لنفترض أنه لدينا مجموعتين G_1 و G_2 ونريد تمييزهما بمتحولين طبيعيين X_1 و X_2، لذلك سحبنا منهما عينتين بحجمين $n_1 = n_2 = 3$ فحصلنا منهما على البيانات التالية:

$$\boldsymbol{XG_1} = \begin{bmatrix} 2 & 4 & 3 \\ 12 & 10 & 2 \end{bmatrix} \quad , \quad \boldsymbol{\bar{X}_1} = \begin{bmatrix} 3 \\ 10 \end{bmatrix} \quad , \quad \boldsymbol{S_1} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\boldsymbol{XG_2} = \begin{bmatrix} 5 & 3 & 4 \\ 7 & 9 & 5 \end{bmatrix} \quad , \quad \boldsymbol{\bar{X}_2} = \begin{bmatrix} 4 \\ 7 \end{bmatrix} \quad , \quad \boldsymbol{S_2} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

وبما أن مصفوفتي التباينين $\boldsymbol{S_1} = \boldsymbol{S_2}$ ، فإننا نحسب مصفوفة التباين المدمجة منهما بواسطة العلاقة (3−31) فنجد أن:

$$\boldsymbol{S_p} = \frac{1}{3+3-2}\left[2\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} + 2\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\right] = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} = \boldsymbol{\tilde{V}}$$

ثم نحسب مقلوبها $\boldsymbol{S_p^{-1}}$ فنجد أن:

$$\boldsymbol{S_p^{-1}} = \frac{1}{4-1}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$$

وبفرض أن الاحتمالين المسبقين للمجموعتين متساويان $(P_1 = P_2)$ وأن تكاليف التصنيف الخاطئ متساوية $(C(1/2) = C(2/1))$، فعندها سيكون لدينا $\ell n\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] = 0$ وهو الطرف الأيمن للعلاقة (3−32)، ثم نقوم بحساب معادلة المؤشر *W(X)* من العلاقة (3−35) فنجد أن:

$$\boldsymbol{W(X)} = (\boldsymbol{\bar{X}_1} - \boldsymbol{\bar{X}_2})' * \boldsymbol{S_p^{-1}} * \left[\boldsymbol{X} - \frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{\bar{X}_1} + \boldsymbol{\bar{X}_2})\right]$$

$$\boldsymbol{W(X)} = \left(\begin{bmatrix} 3 \\ 10 \end{bmatrix} - \begin{bmatrix} 4 \\ 7 \end{bmatrix}\right)' * \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} * \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \frac{1}{2}\left(\begin{bmatrix} 3 \\ 10 \end{bmatrix} + \begin{bmatrix} 4 \\ 7 \end{bmatrix}\right)\right)$$

$$\boldsymbol{W(X)} = \left(\frac{1}{3}[-1\ ,3]\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} X_1 - \frac{7}{2} \\ X_2 - \frac{17}{2} \end{bmatrix}\right)$$

$$\boldsymbol{W(X)} = \frac{1}{3}[-1\ ,2]\begin{bmatrix} X_1 - \frac{7}{2} \\ X_2 - \frac{17}{2} \end{bmatrix} = \frac{1}{3}\left[\left(-X_1 + \frac{7}{2}\right) + (2X_2 - 17)\right]$$

$$\boldsymbol{W(X)} = -\frac{1}{3}X_1 + \frac{2}{3}X_2 - \frac{27}{6}$$: وهي معادلة التابع التمييزي الفاصل بين المجموعتين

فإذا أخذنا المشاهدة الأولى $\begin{bmatrix} 2 \\ 12 \end{bmatrix}$ من G_1 نجد أن قيمة *W(X)* المقابلة لها تساوي:

$$W_1 = -\frac{1}{3}2 + \frac{2}{3}12 - \frac{27}{6} = 8 - \frac{31}{6} > 0$$

أي أن المشاهدة $\begin{bmatrix}2\\12\end{bmatrix}$ ستصنف في المجموعة الأولى G_1 وهي أصلاً من G_1 وهو تصنيف صحيح .

وإذا أخذنا المشاهدة الثانية $\begin{bmatrix}4\\10\end{bmatrix}$ من G_1 نجد أن قيمة W(X) المقابلة لها تساوي:

$$W_2 = -\frac{1}{3}4 + \frac{2}{3}10 - \frac{27}{6} = \frac{20}{3} - \frac{35}{6} > 0$$

أي أن المشاهدة $\begin{bmatrix}4\\10\end{bmatrix}$ ستصنف أيضاً في G_1 وهو تصنيف صحيح أيضاً .

وإذا أخذنا المشاهدة $\begin{bmatrix}3\\8\end{bmatrix}$ من G_1 نجد أن قيمة W(X) المقابلة لها تساوي :

$$W_3 = -\frac{1}{3}3 + \frac{2}{3}8 - \frac{27}{6} = \frac{16}{3} - \frac{33}{6} < 0$$

أي أن المشاهدة $\begin{bmatrix}3\\8\end{bmatrix}$ ستصنف في المجموعة G_2 وهي أصلاً من G_1 وهذا تصنيف خاطئ .

وإذا أخذنا المشاهدة الأولى $\begin{bmatrix}5\\7\end{bmatrix}$ من G_2 نجد أن *W(X)* تساوي:

$$W_4 = -\frac{1}{3}5 + \frac{2}{3}7 - \frac{27}{6} = \frac{14}{3} - \frac{37}{6} < 0$$

أي أن المشاهدة $\begin{bmatrix}5\\7\end{bmatrix}$ ستصنف في المجموعة G_2 وهي أصلاً من G_2 وهو تصنيف صحيح .

وإذا أخذنا المشاهدة الثانية $\begin{bmatrix}3\\9\end{bmatrix}$ من G_2 نجد أن *W(X)* تساوي:

$$W_5 = -\frac{1}{3}3 + \frac{2}{3}9 - \frac{27}{6} = 5 - \frac{37}{6} > 0$$

أي أن المشاهدة $\begin{bmatrix}3\\9\end{bmatrix}$ ستصنف في المجموعة G_1 وهي أصلاً من G_2 وهو تصنيف خاطئ .

وإذا أخذنا المشاهدة الثالثة $\begin{bmatrix}4\\5\end{bmatrix}$ من G_2 نجد أن *W(X)* تساوي:

$$W_6 = -\frac{1}{3}4 + \frac{2}{3}5 - \frac{27}{6} = \frac{10}{3} - \frac{35}{6} < 0$$

أي أن المشاهدة $\begin{bmatrix}3\\9\end{bmatrix}$ من G_2 ستصنف في المجموعة G_2 وهي أصلاً من G_2 وهو تصنيف صحيح .

وبذلك نحصل على جدول تقاطع التصنيف الجديد مع التوزيع الأصلي التالي :

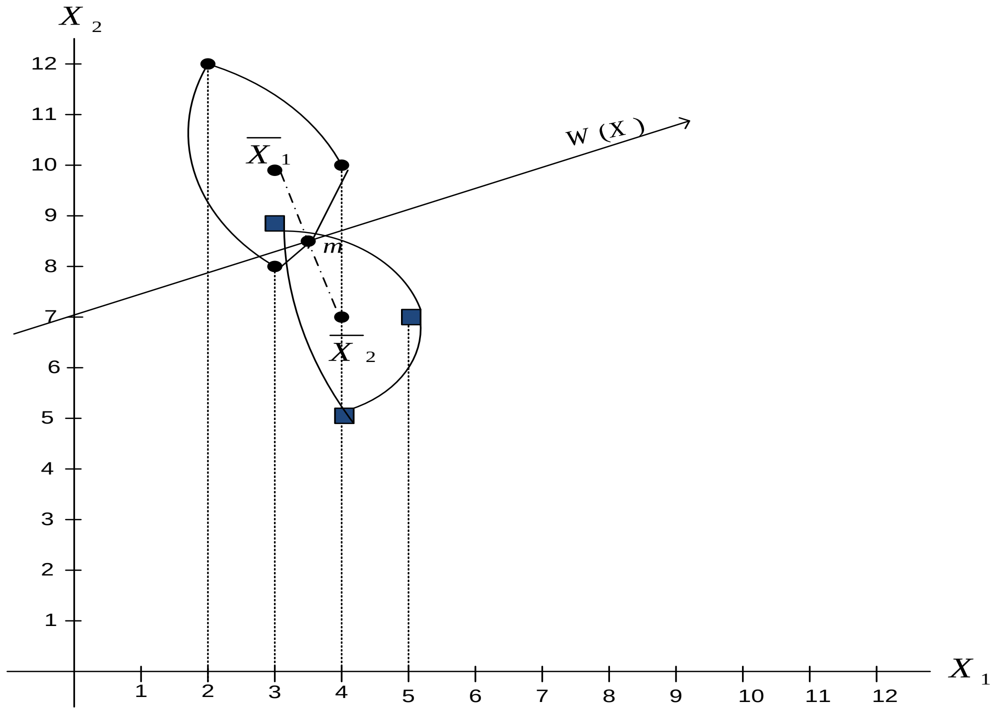
المجموع	التصنيف الجديد			
	G_2	G_1	المجموعة	التوزيع الاصلي
3	$n_{1m} = 1$	$n_{11} = 2$	G_1	
3	$n_{22} = 2$	$n_{2m} = 1$	G_2	

وبناءً على ذلك يمكننا حساب معدل الخطأ الظاهري $APER$ (*Apparent Error Rate*) وهو المعرف بالعلاقة التالية:

$$APER = \frac{n_{1m} + n_{2m}}{n_1 + n_2} = \frac{1+1}{6} = 0.333 \qquad (38-3)$$

وهو احتمال كبير نسبياً، ويمكن أن يكون أكبر من ذلك لو إننا لم نعتبر الاحتمال المسبق والتكاليف متساوية، ويعود ذلك إلى صغر حجمي العينتين n_2و n_1، ويمكن أن تقل قيمة هذا الاحتمال عندما نجعل حجمي العينتين كبيرين .

ويمكننا رسم هذه النقاط على الشكل البياني كما يلي:



الشكل (3-4): توزيع عناصر المجموعتين G_1 وG_2 حيث أن

● **عناصر المجموعة G_1** ■ **عناصر المجموعة G_2**

ومن الشكل نلاحظ أن المستقيم الفاصل بين عناصر هاتين المجموعتين يقسمها إلى مجموعتين جديدتين كما هو مبين في الجدول السابق. ونحصل على معادلة المستقيم الفاصل في المستوى $X_1 0 X_2$ بوضع $W = 0$ فنجد أن :

$$\frac{2}{3}X_2 = \frac{1}{3}X_1 + \frac{27}{6}$$
$$X_2 = \frac{1}{2}X_1 + \frac{27}{4}$$

وهو يمر من النقطة $\left(0\ ,\frac{27}{4}\right)$ ومن القيمة الوسطى $\left(\frac{7}{2}\ ,\ \frac{17}{2}\right)$، ولكنه لا يتعامد مع القطعة المستقيمة الواصلة بين المركزين $\bar{X}_1$ و $\bar{X}_2$ لأن ميله $m' = \frac{1}{2}$ لا يساوي $\left(-\frac{1}{m}\right)$حيث أن $\left(m = \frac{10-7}{3-4} = -3\right)$

- **الحالة الثانية:** وهي التي يكون فيها $V_1 \neq V_2$

وهذه الحالة هي الحالة العامة التي تفترض أن مصفوفتي التباين المشترك للمتحولين $\left(X_1 \text{و } X_2\right)$ في المجموعتين $\left(G_1 \text{و } G_2\right)$ غير متساويتين .

وهنا يمكننا إجراء نفس المعالجات التي أجريناها في الحالة $V_1 = V_2$، ثم نقوم بحساب نسبة التوزيعين $f_1(X)/f_2(X)$ ثم حساب لوغاريتم هذه النسبة من العلاقة :

$$\ell n\frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} = \ell n f_1(\boldsymbol{X}) - \ell n f_2(\boldsymbol{X}) \qquad (38-3)$$

وبعد تعويض معادلتي $f_1(\boldsymbol{X})$ و $f_2(\boldsymbol{X})$ من العلاقتين $(3-20)$ و $(3-21)$ وبإجراء بعض الإصلاحات نحصل على منطقتي التصنيف المعرفين كما يلي:

$$R_1: -\frac{1}{2}X'\left(\boldsymbol{V_1^{-1}} - \boldsymbol{V_2^{-1}}\right) * \boldsymbol{X} + \left(\boldsymbol{\mu_1'V_1^{-1}} - \boldsymbol{\mu_2'V_2^{-1}}\right) * \boldsymbol{X} - K \geq \ell n\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \quad (39-3)$$

$$R_2: -\frac{1}{2}X'\left(\boldsymbol{V_1^{-1}} - \boldsymbol{V_2^{-1}}\right) * \boldsymbol{X} + \left(\boldsymbol{\mu_1'V_1^{-1}} - \boldsymbol{\mu_2'V_2^{-1}}\right) * \boldsymbol{X} - K < ln\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \quad (40-3)$$

حيث أن: K هو عدد يحسب من العلاقة التالية:

$$K = \frac{1}{2}\ell n\left(\frac{|V_1|}{|V_2|}\right) + \frac{1}{2}\left(\boldsymbol{\mu_1'V_1^{-1}\mu_1} - \boldsymbol{\mu_2'V_2^{-1}\mu_2}\right) \qquad (41-3)$$

وهنا نلاحظ أن الحد الأول $\left(-\frac{\boldsymbol{1}}{\boldsymbol{2}}\boldsymbol{X'}\left(\boldsymbol{V_1^{-1}} - \boldsymbol{V_2^{-1}}\right) * \boldsymbol{X}\right)$ في العلاقة $(3-39)$ يعطي بدلالة تابع تربيعي لـ X (من الدرجة الثانية لـ X)، أي أن الفاصل بين المجموعتين –في هذه الحالة– سيكون على شكل منحني من الدرجة الثانية لـ X.

كما نلاحظ أنه عندما: $\boldsymbol{V_1} = \boldsymbol{V_2}$ فإن هذا الحد يختفي وتعود العلاقة $(3-39)$ إلى العلاقة $(3-32)$ ذات التابع الخطي.

وعندما تكون المعالم $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ و $\boldsymbol{V_1}$ و $\boldsymbol{V_2}$ غير معلومة في المجتمع، فإننا نستبدلها بتقديراتها غير المتحيزة المحسوبة من العينتين العشوائيتين $\boldsymbol{\overline{X}_1}$ و $\boldsymbol{\overline{X}_2}$ و $\boldsymbol{S_1}$ و $\boldsymbol{S_2}$ على الترتيب، وعند تعويض ذلك في العلاقات $(3-39)$ و $(3-40)$ و $(3-41)$ نحصل على قاعدة تصنيف أي عنصر من المجتمع مثل X_0 في إحدى المجموعتين كما يلي:

نصنف X_0 في المجموعة G_1 إذا تحققت المتراجحة التالية:

$$-\frac{1}{2}\boldsymbol{X_0'}(\boldsymbol{S_1'} - \boldsymbol{S_2'})\boldsymbol{X_0} + \left(\boldsymbol{\overline{X}_1'S_1^{-1}} - \boldsymbol{\overline{X}_2'S_2^{-1}}\right)\boldsymbol{X_0} - K \geq ln\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \qquad (\boldsymbol{42-3})$$

ونصنفه في المجموعة G_2 إذا كانت المتراجحة $(3-42)$ غير محققة .

علماً بأن K تحسب أو تقدر من العلاقة التالية:

$$K = \frac{1}{2}\ ln\left(\frac{|V_1|}{|V_2|}\right) + \frac{1}{2}\left(\boldsymbol{\overline{X}_1'S_1^{-1}\overline{x}} - \boldsymbol{\overline{X}_2'S_2^{-1}\overline{x}}\right) \qquad (43-3)$$

ملاحظة: كتعقيب على هذه الحالة نلاحظ أن التصنيف وفق التابع التربيعي يكون أكثر تعقيداً في الفضاءات العليا (لأكثر من متحولين) ويمكن أن يقودنا إلى بعض النتائج الغريبة .

وإذا كانت البيانات غير خاضعة للتوزيع الطبيعي المتعدد فهناك رأيان ممكنان لمعالجة ذلك هما:

الأول: يمكن أن نقوم بتحويل البيانات غير الطبيعية إلى بيانات قريبة من الطبيعة، ثم اختبار تساوي مصفوفتي التباين المشترك (قبل وبعد التحويل)، ثم تحديد أي الحالتين أفضل للتصنيف. حالة التابع الخطي أم حالة التابع التربيعي .

وإن أهم المعادلات المستخدمة في التحويل إلى طبيعي أو شبه طبيعي هي:

1- للمتحول الرقمي y الموجب يمكن أن نستخدم أحد التحويلين $\sqrt{y}$ أو $\log y$.

2- للمتحول النسبي $\tilde{P}$ نستخدم التحويل المنطقي: $\text{logit}\tilde{P} = \frac{1}{2}\log\left(\frac{\tilde{P}}{1-\tilde{P}}\right)$

3- لمعاملات الارتباط r نستخدم تحويل فيشر : $\text{Z(r)} = \frac{1}{2}\log\left(\frac{1+\text{r}}{1-\text{r}}\right)$

الثاني: يمكننا استخدام التابع الخطي أو التربيعي بدون قلق أو اهتمام بخصائص المجتمع الأصلي. ولقد قام فيشر بذلك، دون أن يعتمد على خصائص المجتمع الأصلي. وانطلق من فكرة تعظيم المسافة النسبية بين متوسطي مجموعتين طبيعيتين أو غير طبيعيتين. رغم أن نتائج بعض الحالات غير الطبيعية كانت غريبة .

لذلك يصبح من الواجب علينا دائماً أن نقوم باختبار نتائج أي أسلوب يمكن أن يستخدم للتصنيف، وذلك من أجل إنشاء التابع التصنيفي المناسب.

ومن الناحية المثالية يفضل أن يكون حجم البيانات كافياً ليبرهن على صحة العينات (التدريبية) والعينات (التصديقية). حيث أن العينات (التدريبية) تستخدم لإنشاء توابع التصنيف ويجب أن تكون حجومها كبيرة أو كافية لتؤمن ثقة كافية في التوابع المستخدمة. أما العينات (التصديقية) فهي تستخدم لتقييم أداء تلك التوابع. ويمكن أن تكون حجومها صغيرة نسبياً لتؤكد صحة التوابع التصنيفية المستخدمة .

5-3 تقييم توابع التصنيف: [Johnston,Wichernو P. 494 بتصرف وإضافة]

إن أهم وسيلة للحكم على أداء أي أسلوب للتصنيف هو حساب (معدل الخطأ) فيه، أو حساب اجمالي احتمالات التصنيف الخاطئ *TPM* الذي يعطي بالعلاقة (2-12) التالية:

$$TPM = P_1 \int_{R_2} f_1(\boldsymbol{X})dx + P_2 \int_{R_1} f_2(\boldsymbol{X})dx \qquad (44-3)$$

إن أصغر قيمة لـ TPM التي يمكن الحصول عليها يتم بتحديد المنطقتين R_1 و R_2 (أو تحديد الحد الفاصل بينها)، تسمى بمعدل الخطأ المثالي *(Optimum error rate) OER* ونكتب ذلك كما يلي:

$$OER = Min\left[P_1 \int_{R_2} f_1(\boldsymbol{X})dx + P_2 \int_{R_1} f_2(\boldsymbol{X})dx\right] \qquad (45-3)$$

أي أن OER هو معدل الخطأ الأصغر لتابع التصنيف .

ولتحديد المنطقتين R_1 و R_2 نفترض أن تكاليف التصنيف الخاطئ متساوية، ونستخدم العلاقتين (3-22) و(3-23) فنحصل على ما يلي:

$$R_1 : \frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} \geq \frac{P_2}{P_1} \qquad R_2 : \frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} < \frac{P_2}{P_1} \qquad (46-3)$$

ثال (3−2): يطلب منا أن نوضح معنى معدل الخطأ المثالي OER لمجتمع مؤلف من مجموعتين: G_1 وG_2 متساويتين الحجم، أي أن نسبتيهما $\left(p_1 = p_2 = \frac{1}{2}\right)$، وتتأثران بمتحولين X_1 و X_2 خاضعين فيهما للتوزيعين الطبيعيين $f_1(\boldsymbol{X})$ و $f_2(\boldsymbol{X})$ المعرفتين بالعلاقتين (3−20) و(3−21)، وإن تكاليف التصنيف الخاطئ متساوية، وإن $(V_1 = V_2 = V)$.

نلاحظ أن أصغر قيمة لـ *ECM* وأصغر قيمة لـ *TPM* لأي تابع تمييزي تتطابقان عندما يكون $C(1/2) = C(2/1)$ وعندما $p_1 = p_2$ ، ولأنه عندها يكون الطرف الأيمن من العلاقة (3−29) معدوماً لأن:

$$\ell n \left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] = ln(1) = 0$$

ومنها نستنتج أن المنطقتين R_1 و R_2 تتحددان في هذه الحالة من العلاقتين:

$$R_1: (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{X} - \frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \geq 0$$
$$R_2: (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{X} - \frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) < 0$$

وإذا رمزنا للحد الأول فيهما بالرمز y يكون لدينا:

$$y = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{X} = \boldsymbol{\ell}' * \boldsymbol{X} \qquad (47-3)$$

حيث رمزنا بـ $\boldsymbol{\ell}'$ للمقدار $\boldsymbol{\ell}' = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}}$ فيكون لدينا :

$$\boldsymbol{\ell} = \boldsymbol{V^{-1}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (48-3)$$

وبعدها نجد أن المنطقتين R_1 و R_2 تصبحان تابعين لـ y، ويتم تحديدهما من العلاقتين:

$$R_1(y): y \geq \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \qquad (48-3)$$
$$R_2(y): y < \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \qquad (49-3)$$

وبما أن y هو تركيب خطي لمتحولات طبيعية عشوائية، فإن التوزيع الاحتمالي لـ y هو توزيع طبيعي بمتحول واحد y في المجموعتين، ونرمز له بـ $f_1(y)$ في G_1 وبـ $f_2(y)$ في G_2 .

وإن متوسطي y في المجموعتين يساويان:

$$\mu_{1y} = \ell' * \boldsymbol{\mu_1} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{\mu_1} \qquad (50-3)$$
$$\mu_{2y} = \ell' * \boldsymbol{\mu_2} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{\mu_2} \qquad (51-3)$$

وإن تباينه في كل من المجموعتين يساوي: $\sigma_{y_1}^2 = \sigma_{y_2}^2 = \sigma_y^2$ ويحسب من العلاقة:

$$\sigma_y^2 = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{V} * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$
$$\sigma_y^2 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = \Delta^2 \qquad (51-3)$$

وبذلك تأخذ علاقة *TPM* الشكل التالي:

$$TPM(y) = P_1 \int_{R_2} f_1(y)dy + P_2 \int_{R_1} f_2(y)dy \qquad (52-3)$$

وبناءً على العلاقة (3–49) نجد أن التكامل الأول يساوي:

$$P(2/1) = \int_{R_2} f_1(y)dy = P\left[y < \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2})\right]$$

والآن نحول y إلى متحول معياري Z بإجراء التحويل $\left(Z = \frac{y-\mu_{1y}}{\sigma_y}\right)$ فنجد أن (3–50) تعطينا أن:

$$\text{P}(2/1) = P\left[\frac{y - \mu_{1y}}{\sigma_y} < \frac{\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{\mu_1}}{\sigma_y}\right]$$

وبعد الإصلاح واعتماداً على (3–16) نجد أن:

$$\text{P}(2/1) = P\left[Z < \frac{-\frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})}{\Delta}\right]$$

$$P(2/1) = \int_{R_2} f_1(y)dy = P\left(Z < \frac{-\frac{1}{2}\Delta^2}{\Delta}\right) = P\left(Z < \frac{-\Delta}{2}\right) = \phi\left(\frac{-\Delta}{2}\right) \qquad (53-3)$$

حيث $\phi\left(\frac{-\Delta}{2}\right)$ هي قيمة تابع التوزيع الطبيعي المعياري عند النقطة $\left(\frac{-\Delta}{2}\right)$.

وبطريقة مشابهة نجد أن التكامل الثاني يساوي:

$$P(1/2) = \int_{R_1} f_2(y)dy = P\left[y \geq \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2})\right]$$

$$P(1/2) = P\left[Z \geq \left(\frac{\Delta}{2}\right)\right] = 1 - \phi\left(\frac{\Delta}{2}\right) = \phi\left(\frac{-\Delta}{2}\right) \qquad (54-3)$$

ومن (3–53) و(3–54) نجد أن الاحتمال الاجمالي للتصنيف الخاطئ *TPM* في (3–52) يساوي:

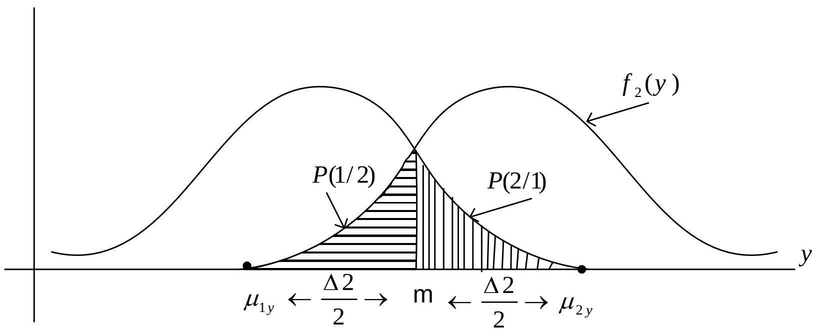
$$TPM = P_1\phi\left(\frac{-\Delta}{2}\right) + P_2\phi\left(\frac{-\Delta}{2}\right) \qquad (55-3)$$

وأخيراً نجد أن معدل الخطأ المثالي *OER* يساوي:

$$OER = Minmum\ TPM = P_1 * \phi\left(\frac{-\Delta}{2}\right) + P_2 * \phi\left(\frac{-\Delta}{2}\right)$$

$$OER = \phi\left(\frac{-\Delta}{2}\right)(P_1 + P_2) = \phi\left(\frac{-\Delta}{2}\right) \qquad (56-3)$$

وإن احتمالات هذا التصنيف الخاطئ مبينة على الشكل التالي:



الشكل (3–5) احتمالات التصنيف الخاطئ حسب *y*

حيث أن:

$$\Delta^2 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$

مثال تطبيقي: لنفترض أن:

$$\Delta^2 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = 2.56$$

فعندها يكون لدينا: $\Delta = \sqrt{2.56} = 1.6$

ومن جداول تابع التوزيع الطبيعي المعياري نجد أن:

$$Minimum\ TPM = OER = \phi\left(\frac{-1.6}{2}\right) = 0.2119$$

وهو معدل غير قليل ويشير إلى توزيع خاطئ لبعض عناصر المجتمع على المجموعة الأولى G_1 أو على المجموعة الثانية G_2 .

ملاحظة: عندما تكون المعالم $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ و $\boldsymbol{V}$ غير معلومة. فإننا نستبدلها بتقديراتها من العينة بـ $\boldsymbol{\bar{X}_1}$ و $\boldsymbol{\bar{X}_2}$ و $\boldsymbol{S_p}$. حيث أن $\boldsymbol{S_p}$ هو التباين المدمج من التباينين S_2 و S_1 وفق العلاقة (3-31) .

3-6 : طريقة فيشر Fisher لمتحولين ومجموعتين: [Johnston,Wichern. P.470 بتصرف وإضافة]

قبل استعراض هذه الطريقة نقدم بعض خواص المتحولات العشوائية ونختصرها بما يلي:

1- إذا كان X_i متحولاً عشوائياً وكان a عدداً ثابتاً فإن التوقع الرياضي لـ (aX_i) يساوي :

$$E(aX_i) = a * E(X_i) = a\mu_{\boldsymbol{i}} \qquad \left(\text{حيث } \mu_i \text{ هو توقع } X_i\right) \qquad (3-57)$$

وإن تباين (aX_i) يساوي :

$$var(aX_i) = E(aX_i - a\mu_i)^2 = a^2E(X_i - \mu_{\boldsymbol{i}})^2 = a^2\sigma_{ii} \qquad (3-58)$$

2- إذا كان a و b عددين ثابتين فإن التباين المشترك للمتحولين aX_1 و bX_2 يساوي:

$$\begin{aligned} cov(aX_1\,, bX_2) &= E(aX_1 - a\mu_1)(bX_2 - a\mu_{\boldsymbol{2}}) \qquad (3-59) \\ &= abE(X_1 - \mu_{\boldsymbol{1}})(X_2 - \mu_{\boldsymbol{2}}) = ab\sigma_{12} \end{aligned}$$

3- إذا كان a و b عددان ثابتين وكان X_1 , X_2 متحولين عشوائيين فإن التوقع الرياضي لأي تركيب خطي لهما مثل $Y = aX_1 + bX_2$ وتباينه يساويان:

$$E(Y) = E\ (aX_1 + bX_2) = aE(X_1) + bE(X_2) = a\mu_{\boldsymbol{1}} + b\mu_{\boldsymbol{2}} \qquad (3-60)$$

$$\begin{aligned} var(Y) &= var\ (aX_1 + bX_2) = E[(aX_1 + bX_2) - (a\mu_1 + b\mu_{\boldsymbol{2}})]^2 \\ &= E[a(X_1 - \mu_1) + b(X_2 - \mu_2)]^2 \\ &= E[a^2(X_1 - \mu_1)^2 + b^2(X_2 - \mu_2)^2 * 2ab(X_1 - \mu_1)(X_2 - \mu_2)] \end{aligned}$$

$$var(y) = a^2var(X_1) + b^2var(X_2) + 2ab\ cov(X_1\,, X_2)$$

أي أن:

$$var(aX_1 + bX_2) = a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \qquad (3-61)$$

وإذا رمزنا لشعاع المتحولين X_2 , X_1 بالرمز $\boldsymbol{X}$ ولشعاع العددين a و b بالرمز C فيكون لدينا:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \boldsymbol{C} = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \boldsymbol{C}' = [a \ , b]$$

وعندها يمكننا كتابة التركيب الخطي السابق كما يلي:

$$Y = aX_1 + bX_2 = [a \ , b] * \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \boldsymbol{C}' * \boldsymbol{X} \qquad (62-3)$$

وبذلك نجد أن توقعه يساوي:

$$E(Y) = E(\boldsymbol{C}' * \boldsymbol{X}) = [a \ , b] * \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \boldsymbol{C}' * \boldsymbol{\mu} \qquad (63-3)$$

وإن تباينه يمكن أن يكتب بدلالة مصفوفة التباين المشترك V للمتحولين X_1 و X_2 كما يلي:

$$var(Y) = var\ (aX_1 + bX_2) = var[\boldsymbol{C}'\boldsymbol{X}] \qquad (64-3)$$

$$= \boldsymbol{C}'\boldsymbol{var}(\boldsymbol{X})\boldsymbol{C} = \boldsymbol{C}' * \boldsymbol{V} * \boldsymbol{C}$$

وهذا يعني أن:

$$var\ [aX_1 + bX_2] = [a \ , b] \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & 22 \end{bmatrix} * \begin{bmatrix} a \\ b \end{bmatrix}$$

وهو يساوي:

$$var\ [aX_1 + bX_2] = a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \qquad (65-3)$$

وانطلاقاً من الأفكار السابقة اقترح (فيشر) أن يتم تحويل التحليل التمييزي لتأثير المتحولين X_1 و X_2 على المجموعتين G_1 و G_2 إلى تحليل بسيط بمتحول واحد Y، ثم العمل على جعل قيم المتحول Y المشتقة من مشاهدات المجموعتين G_1 و G_2 بعيدة عن بعضها بأكبر قدر ممكن .

ولتحقيق هذه الفكرة اقترح (فيشر) أن يتم تشكيل متحول جديد Y بواسطة تركيب خطي للمتحولين X_1 و X_2 كما يلي:

$$Y = \ell_1 X_1 + \ell_2 X_2 = [\ell_1 \ , \ell_2\ ] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \boldsymbol{\ell}' * \boldsymbol{X} \qquad (66-3)$$

وعند تعويض قيم X_1 و X_2 من كل مجموعة على حدة في هذا التركيب الخطي، نحصل على القيم العددية لـ Y المقابلة لمشاهدات كل مجموعة، وبالتالي نحصل على مجموعتين لقيم Y نرمز لهما بما يلي:

$$Y_1 : y_{11} \ \ y_{12} \ \ y_{13} \ldots . y_{1n_1} \qquad (G_1 \text{ للمجموعة})$$

$$Y_2 : y_{21} \ \ y_{22} \ \ y_{23} \ldots . y_{2n_2} \qquad (G_2 \text{ للمجموعة})$$

ومنها يمكننا حساب متوسطي Y في المجموعتين من العلاقتين :

$$\mu_{1y} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i} \qquad (67-3)$$

$$\mu_{2y} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} \qquad (68-3)$$

ومن جهة أخرى نجد أن المتوسطين μ_{1y} و μ_{2y} يمكن أن يحسبان بدلالة توقعي X_1 و X_2 في المجموعتين. فإذا افترضنا أن :

$$\boldsymbol{\mu_1} = E(\boldsymbol{X}/G_1) = \begin{bmatrix}\boldsymbol{\mu_{11}}\\ \boldsymbol{\mu_{12}}\end{bmatrix} \qquad \boldsymbol{\mu_2} = E(\boldsymbol{X}/G_2) = \begin{bmatrix}\boldsymbol{\mu_{21}}\\ \boldsymbol{\mu_{22}}\end{bmatrix} \qquad (69-3)$$

فعندها نجد أن:

$$\mu_{1y} = E(Y/G_1) = E(\boldsymbol{\ell}'\boldsymbol{X}/G_1) = \boldsymbol{\ell}' * \boldsymbol{\mu_1} =$$
$$= (\ell_1\,, \ell_2)\begin{bmatrix}\boldsymbol{\mu_{11}}\\ \boldsymbol{\mu_{12}}\end{bmatrix} = \ell_1\boldsymbol{\mu_{11}} + \ell_2\boldsymbol{\mu_{12}} \qquad (69-3)$$

$$\mu_{2y} = E(Y/G_2) = E(\boldsymbol{\ell}'\boldsymbol{X}/G_2) = \boldsymbol{\ell}' * \boldsymbol{\mu_2} =$$
$$= (\ell_1\,, \ell_2)\begin{bmatrix}\boldsymbol{\mu_{21}}\\ \boldsymbol{\mu_{22}}\end{bmatrix} = \ell_1\boldsymbol{\mu_{21}} + \ell_2\boldsymbol{\mu_{22}} \qquad (70-3)$$

ولحساب تباين Y بدلالة مصفوفة التباينات المشتركة لـ X_1 و X_2 نجد أن:

$$\boldsymbol{V_1} = E[(\boldsymbol{X} - \boldsymbol{\mu_1})(\boldsymbol{X} - \boldsymbol{\mu_1})'] \qquad (71-3)$$

$$\boldsymbol{V_2} = E[(\boldsymbol{X} - \boldsymbol{\mu_2})(\boldsymbol{X} - \boldsymbol{\mu_2})'] \qquad (72-3)$$

والآن لنفترض أن هاتين المصفوفتين متساويتان، أي نفترض أن $\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V}$، فنجد أن تباين Y يساوي حسب (3–64) مايلي:

$$\sigma_y^2 = vor(Y) = vor(\boldsymbol{\ell}'\boldsymbol{X}) = \boldsymbol{\ell}' * \boldsymbol{vor}(\boldsymbol{X}) * \boldsymbol{\ell}$$

ومنها نجد أن :

$$\sigma_y^2 = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} \qquad (73-3)$$

وهكذا نصل إلى مكونات فكرة (فيشر) التي صاغها على شكل علاقة رياضية سميت (معيار فيشر)، وهي تهدف إلى جعل مربع المسافة النسبية بين متوسطي قيم Y في المجموعتين G_1 و G_2 أكبر مايمكن. ولقد صاغ (فيشر) معياره كما يلي:

$$\mathrm{F} = \frac{\left(\mu_{1y} - \mu_{2y}\right)^2}{\sigma_y^2} = \frac{(\boldsymbol{\ell}'\boldsymbol{\mu_1} - \boldsymbol{\ell}'\boldsymbol{\mu_2})^2}{\sigma_y^2} = \frac{[\boldsymbol{\ell}'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})]^2}{\sigma_y^2} = \frac{(\boldsymbol{\ell}' * \boldsymbol{d})^2}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} \qquad (74-3)$$

حيث أن: $\boldsymbol{d}$ هو شعاع الفرق بين المتوسطين $\boldsymbol{d} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ ، والمطلوب الآن تحديد الأعداد في $\boldsymbol{\ell}' = (\ell_1\,, \ell_2)$ أو $\boldsymbol{\ell} = \begin{pmatrix}\ell_1\\ \ell_2\end{pmatrix}$ التي تجعل قيمة المعيار F أكبر ما يمكن .

ولتحديد مركبات $\boldsymbol{\ell}$ نستند إلى مسألة تعظيم المقدار الأخير الواردة في الفصل (2) من الجزء الأول، وحسب متراجحة (كوشي شتوارز) (2–70) فيه وما يليها، نجد أنها تعطينا أن أكبر قيمة للمقدار (3–74) تساوي :

$$Max\left[\frac{(\boldsymbol{\ell}'\boldsymbol{d})^2}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}}\right] = \boldsymbol{d}' * \boldsymbol{V}^{-1} * \boldsymbol{d} \qquad (75-3)$$

وإن هذه القيمة الكبرى تحصل أو تتحقق حسب العلاقة (*70a –2*) فيه) عندما تأخذ ℓ القيمة التالية:

$$\boldsymbol{\ell} = C * \boldsymbol{V}^{-1} * \boldsymbol{d} \qquad (76-3)$$

حيث C عامل تناسب وهو أي عدد حقيقي لا يساوي الصفر $C \neq 0$

وبتعويض قيمة $\boldsymbol{d}$ التي تساوي : $\boldsymbol{d} = \boldsymbol{\mu_1} - \boldsymbol{\mu_2}$ نحصل على أن الحل المناسب $\boldsymbol{\ell}$ لتعظيم (3−75) هو الحل التالي:

$$\boldsymbol{\ell} = C * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (77-3)$$

ومنه نحصل على لا نهاية من الحلول لـ $\boldsymbol{\ell}$ تتناسب مع الجداء $\boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$.

وبوضع $C=1$ وملاحظة أن V^{-1} متناظرة نحصل على الحل الخاص $\boldsymbol{\ell}$ ومنقوله $\boldsymbol{\ell}'$ التالي:

$$\boldsymbol{\ell} = \boldsymbol{V^{-1}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \implies \boldsymbol{\ell}' = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) * \boldsymbol{V^{-1}} \qquad (78-3)$$

وبتعويض ذلك الحل في تابع (فيشر) المعرف في (3−66)، يصبح كما يلي:

$$\tilde{Y} = \boldsymbol{\ell}' * \boldsymbol{X} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{X} \qquad (79-3)$$

وهي علاقة شبيهة بالعلاقة (3−34) السابقة، وهو عبارة عن مستقيم يمر من مبدأ الاحداثيات ويتم اسقاط النقاط عليه وحساب المتوسطين $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ والقيمة الوسطى m .

ولاستخدام التابع (3−79) في عمليات تصنيف المشاهدات على المجموعتين G_1 و G_2، نقوم بحساب القيمة الوسطى للمتوسطين μ_{1y} و μ_{2y} فنحصل على أنها تساوي:

$$m = \frac{1}{2}(\mu_{1y} + \mu_{2y}) = \frac{1}{2}(\boldsymbol{\ell}'\boldsymbol{\mu_1} + \boldsymbol{\ell}'\boldsymbol{\mu_2}) = \frac{1}{2}\boldsymbol{\ell}'(\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \qquad (80-3)$$

وبتعويض $\boldsymbol{\ell}'$ من (3−78) نحصل على أن القيمة الوسطى تساوي:

$$m = \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}}(\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \qquad (81-3)$$

والآن لنفترض أنه لدينا مشاهدة عشوائية مثل $X_0 = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}$ من المجتمع المدروس ونريد تحديد المجموعة التي ستنتمي إليها هذه المشاهدة .

لذلك نقوم بتعويض إحداثيات هذه المشاهدة في العلاقة (3−79) فنحصل منها على قيمة محددة $\tilde{y}_0$

$$\tilde{y}_0 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}}\boldsymbol{X_0} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})'\boldsymbol{V^{-1}} \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} \qquad (82-3)$$

ثم نقوم بمقارنة هذه القيمة $\tilde{y}_0$ مع القيمة الوسطى m ونتخذ قرار التصنيف كما يلي:

إذا كانت $\tilde{y}_0 \geq m$ فإننا نصنف X_0 في المجموعة G_1

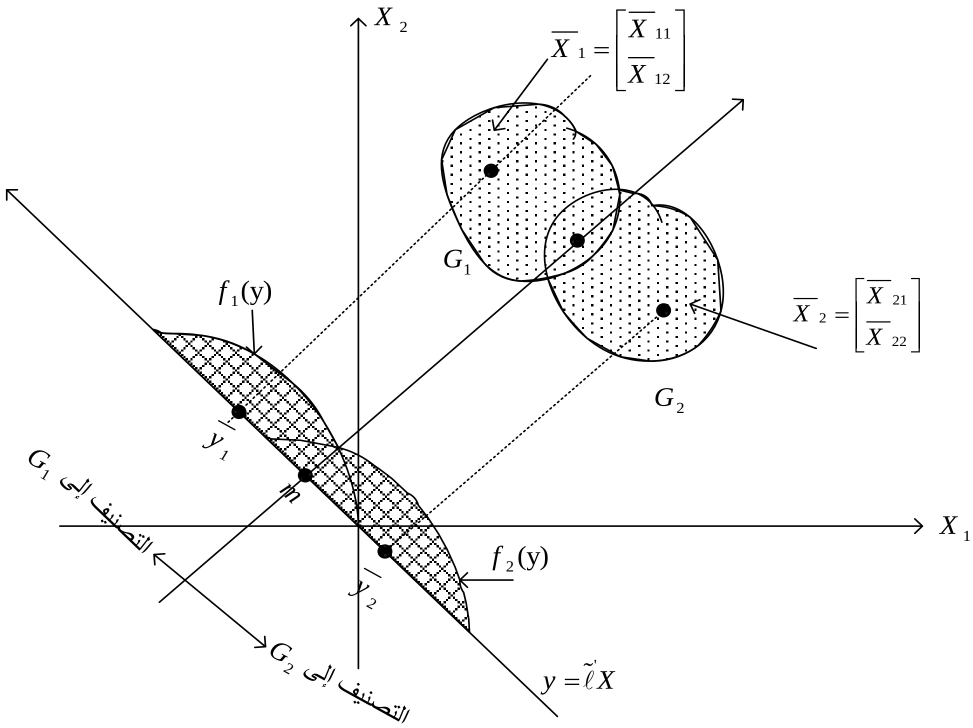
وإذا كانت $\tilde{y}_0 < m$ فإننا نصنف X_0 في المجموعة G_2 $\qquad (83-3)$

وبعبارة أخرى يمكننا أن نصنف العناصر حسب القاعدة التالية :

إذا كانت $(\tilde{y}_0 - m) \geq 0$ فإننا نصنف X_0 في المجموعة G_1

وإذا كانت $(\tilde{y}_0 - m) < 0$ فإننا نصنف X_0 في المجموعة G_2 $\qquad (84-3)$

والشكل البياني التالي يوضح ذلك :



الشكل (3–6): التصنيف حسب معيار (فيشر)

ملاحظة إذا كانت معالم المجتمع $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ و$\boldsymbol{V}$ مجهولة فإننا نقوم بتقدير الكميات $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ و$\boldsymbol{V}$وحساب ℓ وm من خلال معلومات العينات المسحوبة عشوائياً من المجموعتين G_1 و G_2 .

والآن لنفترض إننا سحبنا عينة عشوائية من المجموعة الأولى بحجم n_1 وسحبنا عينة عشوائية أخرى من المجموعة الثانية بحجم n_2 وحصلنا منهما على مصفوفتي القياسات التالية:

المتحولات	قيم عناصر العينة الثانية n_1	المتوسط
X_1	$x_{111}\ \ x_{112}\ \ x_{113}\ \dots\ x_{11n_1}$	$\bar{x}_{11}$
X_2	$x_{211}\ \ x_{212}\ \ 2x_{113}\ \dots\ x_{21n_1}$	$\bar{x}_{21}$

المتحولات	قيم عناصر العينة الثانية n_2	المتوسط
X_1	$x_{121}\ \ x_{122}\ \ x_{123}\ \dots\ x_{12n_2}$	$\bar{x}_{12}$
X_2	$x_{221}\ \ x_{222}\ \ x_{223}\ \dots\ x_{22n_2}$	$\bar{x}_{22}$

ومنهما نحسب شعاعي المتوسطين فنجد من العلاقة $\bar{x} = \frac{1}{n}\sum x_i$ أن :

$$\overline{\boldsymbol{X}}(\boldsymbol{G_1}) = \overline{\boldsymbol{X}}_{\boldsymbol{1}} = \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{21} \end{bmatrix} \qquad \overline{\boldsymbol{X}}(\boldsymbol{G_2}) = \overline{\boldsymbol{X}}_{\boldsymbol{2}} = \begin{bmatrix} \bar{x}_{12} \\ \bar{x}_{22} \end{bmatrix} \qquad (85-3)$$

نعلم أن هذين المتوسطين هما تقديران غير متحيزين للتوقعين $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ في المجموعتين G_1 و G_2، اللتين يتألف منهما المجتمع المدروس، ونكتب ذلك كما يلي:

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{1}} = \overline{\boldsymbol{X}}_{\boldsymbol{1}} \qquad \widetilde{\boldsymbol{\mu}}_{\boldsymbol{2}} = \overline{\boldsymbol{X}}_{\boldsymbol{2}} \qquad (86-3)$$

ولإيجاد تقدير غير متحيز لمصفوفة التباين المشترك $\boldsymbol{V}$، نقوم بحساب مصفوفتي التباين المشترك للمتحولين X_1 و X_2 في كل مجموعة على حدة من العلاقتين التاليتين:

$$S_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)' = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

$$S_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)' = \begin{bmatrix} S'_{11} & S'_{12} \\ S'_{21} & S'_{22} \end{bmatrix} \qquad (87-3)$$

وبما إننا نفترض أن مصفوفتي التباين المشترك في المجموعتين G_1 و G_2 متساويتان، أي أن $(V_1 = V_2 = V)$ ، فإننا نقوم بتركيب مصفوفتي التباين المشترك S_1 و S_2 لنحصل منهما على مصفوفة واحدة، تصلح لأن تكون تقديراً غير متحيز للمصفوفة V المعرفة في العلاقة (3−71) .

وإن أفضل تركيب لهما هو المتوسط المثقل للمصفوفتين S_1 و S_2 بدرجتي الحرية في العينتين $(n_1 - 1)$ و $(n_2 - 1)$، والمسمى بمصفوفة التباين المدمج (*Pooled*) ويرمز لها بـ S_p وتعرف بالعلاقة :

$$S_p = \left[\frac{(n_1 - 1)}{(n_1 - 1) + (n_2 - 1)}\right] S_1 + \left[\frac{(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}\right] S_2 \qquad (88-3)$$

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{\mathrm{n}_1 + n_2 - 2} \qquad (89-3)$$

وإن هذا التقدير هو تقدير غير المتحيز للمصفوفة V، فيما إذا كانت بيانات المصفوفتين X_1 و X_2 مأخوذة من عينتين عشوائيتين مسحوبتين من المجموعتين G_1 و G_2 على الترتيب .

وبذلك نكون قد حصلنا من بيانات هاتين العينتين على تقديرين غير متحيزين لـ μ_1 و μ_2 وللمصفوفة V ، وبناء على ذلك يمكننا حساب القيم التقديرية لـ ℓ من العلاقة (3−78) كما يلي:

$$\ell = S_p^{-1}(\bar{X}_1 - \bar{X}_2) \qquad (90-3)$$

ومنها يمكننا إيجاد تقدير لتابع فيشر التمييزي الخطي من العلاقة:

$$\tilde{Y} = \tilde{\ell}' X = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X \qquad (91-3)$$

ومنها نحصل على تقديري متوسطي Y في العينتين من العلاقتين :

$$\tilde{y}_1 = \tilde{\ell}' * \bar{X}_1 \qquad \tilde{y}_2 = \tilde{\ell}' * \bar{X}_2 \qquad (92-3)$$

ومنهما نحصل على تقدير القيمة الوسطى لها من علاقة شبيهه بالعلاقة (3−81) وهي:

$$\tilde{m} = \frac{1}{2}(\tilde{y}_1 + \tilde{y}_2) = \frac{1}{2}\ell'(\bar{X}_1 + \bar{X}_2)' = \frac{1}{2}(\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * (\bar{X}_1 + \bar{X}_2) \quad (93-3)$$

وهكذا نجد أن قاعدة التصنيف حسب التابع التمييزي (لفيشر) في العينات تصبح كما يلي:

لنأخذ عنصراً $X_0 = \begin{bmatrix} X_{01} \\ X_{02} \end{bmatrix}$ بشكل عشوائي من ذلك المجتمع، ثم نقوم بحساب تقدير قيمة التابع التمييزي y_0 من العلاقة (3−91):

$$\tilde{y}_0 = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X_0$$

ثم نقارن قيمة $\tilde{y}_0$ مع تقدير القيمة الوسطى $\tilde{m}$ ونتخذ قرار التصنيف كما يلي :

إذا كانت $\tilde{y}_0 \geq \tilde{m}$ نصنف X_0 في المجموعة G_1
وإذا كانت $\tilde{y}_0 < \tilde{m}$ نصنف X_0 في المجموعة G_2 $(94-3)$

أو بعبارة أخرى:

إذا كانت $\tilde{y}_0 - \tilde{m} \geq 0$ نصنف X_0 في المجموعة G_1
وإذا كانت $\tilde{y}_0 - \tilde{m} < 0$ نصنف X_0 في المجموعة G_2 $(95-3)$

ملاحظة: يمكننا إعادة صياغة علاقة معيار (فيشر) في (3−75) بعد تعويض $\boldsymbol{d}$ و $\boldsymbol{V^{-1}}$ بتقديرها $\tilde{\boldsymbol{d}} = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)$ و $\tilde{\boldsymbol{V}}^{-1} = \boldsymbol{S}_p^{-1}$، فنحصل على ما يلي:

$$Max \frac{(\boldsymbol{\ell}'\boldsymbol{d})^2}{\boldsymbol{\ell}' * \boldsymbol{S_p} * \boldsymbol{\ell}} = \tilde{\boldsymbol{d}}' * \boldsymbol{S}_p^{-1} * \tilde{\boldsymbol{d}} = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\boldsymbol{S}_p^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) = D^2 \quad (96-3)$$

ونعلم أن هذا يحدث عندما تكون $\tilde{\boldsymbol{d}} = \mathrm{C}\boldsymbol{S}_p^{-1}\boldsymbol{d} = \mathrm{C}\boldsymbol{S}_p^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)$، وهذا يعني أن القيمة العظمى للنسبة (3−75) في حاله مجموعتين تساوي مربع المسافة بين مركزي العينتين المسحوبتين من هاتين المجموعتين والذي رمزنا له بـ D^2 .

وهذا يفيدنا في اختبار فيما إذا كان الفرق $(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ معنوياً أم لا. لأنه يمكن النظر إلى اختبار الفرق بين أشعة المتوسطات كاختبار معنوية الحد الفاصل الذي حصلنا عليه .

لذلك نفترض أن القياسات في المجموعتين G_1 و G_2 خاضعة للتوزيع الطبيعي ولهما تباين مشترك موحد، $\boldsymbol{V}$ ونضع الفرضيتين (العدم H_0 والبديلة H_1) كما يلي:

$$H_0: \boldsymbol{\mu_1} = \boldsymbol{\mu_2} \qquad H_1: \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2} \qquad (97-3)$$

ويتم اختبار H_0 باستخدام مؤشر الاختبار التالي:

$$F = \left[\frac{\mathrm{n}_1 + n_2 - P - 1}{(\mathrm{n}_1 + n_2 - 2) * P}\right]\left(\frac{\mathrm{n}_1 * n_2}{\mathrm{n}_1 + n_2}\right) * D^2 \qquad (98-3)$$

وهو يخضع لتوزيع $F(x)$ بدرجتي حرية $(v_1 = P)$ و $(v_2 = \mathrm{n}_1 + n_2 - P - 1)$، وعندما نرفض H_0 بمستوى دلالة $\propto$ يمكننا أن نستخلص أن الفصل بين المجموعتين G_1 و G_2 يعتبر معنوياً .

ملاحظة: إن الفصل المعنوي لا يؤدي بالضرورة إلى تصنيف جيد. لأن فعالية أسلوب التصنيف يمكن أن تكون مستقلة عن أي من اختبارات الفصل بين المجموعات . ولكن بالمقابل إذا كان الفصل غير معنوي فإن البحث عن تصنيف نافع يعتبر ضياعاً للوقت .

** **المعيرة:** إن شعاع الأمثال الأولى $\boldsymbol{\ell}$ أو تقديره $\tilde{\boldsymbol{\ell}}$ الذي حصلنا عليه من إحدى العلاقتين (3−77) أو (3−87) ليس وحيداً، بل يمكن الحصول منه على لا نهاية من الحلول بضربه بأي عدد ثابت $C \neq 0$ فنحصل على الحلول $C\boldsymbol{\ell}$ أو $C\tilde{\boldsymbol{\ell}}$ التي تعظم قيمة النسبة (3−74).

وحتى نستطيع تفسير هذه الحلول من العلاقة (3−77) أو(3−78)، نقوم بإجراء معيرة للشعاع $\boldsymbol{\ell} = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}$، بحيث نحصل منه على حل خاص يكون طوله مساوياً للواحد، ولإجراء هذه المعيرة يمكن اتباع إحدى الطريقتين التاليتين:

الطريقة الأولى: تتلخص بتقسيم عناصر الشعاع $\boldsymbol{\ell} = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}$ على طوله $||\ell||$ فنحصل على شعاع واحدي $\boldsymbol{e}$، يكون طوله مساوياً للواحد، ونحسب عناصره من العلاقة:

$$\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \dfrac{\ell_1}{||\ell||} \\ \dfrac{\ell_2}{||\ell||} \end{bmatrix} \qquad (99-3)$$

ولذلك نقوم أولاً بحساب طول الشعاع $\boldsymbol{\ell}$ من العلاقة :

$$||\ell|| = \sqrt{\ell_1^2 + \ell_2^2} = \sqrt{\boldsymbol{\ell}' * \boldsymbol{\ell}} \qquad (100-3)$$

ثم نقوم بحساب مركبات الشعاع الواحدي $\boldsymbol{e}$ من العلاقتين :

$$e_1 = \frac{\ell_1}{||\ell||} \qquad e_2 = \frac{\ell_2}{||\ell||} \qquad (101-3)$$

فنحصل على الشعاع الواحدي $\boldsymbol{e}$ المعرف في (3−99) والذي طوله يساوي الواحد لأن:

$$||e|| = \sqrt{e' * e} = \sqrt{e_1^2 + e_2^2} = \sqrt{\left(\frac{\ell_1}{||\ell||}\right)^2 + \left(\frac{\ell_2}{||\ell||}\right)^2} = 1$$

الطريقة الثانية: وهي تعتمد على جعل تباين تابع (فيشر) مساوياً للواحد، أي جعل العلاقة (3−73) تساوي الواحد:

$$vor(y) = \sigma_y^2 = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} = 1 \qquad (102-3)$$

واعتماداً على خواص المصفوفات المتناظرة والمحددة إيجابياً نكتب المصفوفة المتناظرة والمحددة إيجابياً $\boldsymbol{V}$، كما يلي: $\left(\boldsymbol{V} = \boldsymbol{V}^{\frac{1}{2}}\boldsymbol{V}^{\frac{1}{2}}\right)$ ونعالج العلاقة (3−102) كما يلي:

$$\sigma_y^2 = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} = \boldsymbol{\ell}' * \boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell} = \left(\boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell}\right)' * \left(\boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell}\right) = 1 \qquad (103-3)$$

وإذا رمزنا للشعاع الذي ضمن القوس الثاني $\left(\boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell}\right)$ بالرمز $\boldsymbol{e}$ يكون لدينا:

$$\boldsymbol{e} = \boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell} \quad \text{و} \quad \boldsymbol{e}' = \left(\boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell}\right)' \qquad (104-3)$$

ومنه نحصل على أن:

$$\sigma_y^2 = \boldsymbol{e}' * \boldsymbol{e} = ||e||^2 = 1 \qquad (105-3)$$

أي أن الشعاع الجديد $\boldsymbol{e} = \boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell}$ المعرف بالعلاقة $(104-3)$ هو شعاع واحدي (طوله واحد)، وبذلك نحصل على حل خاص ممعير **e** من كل حل عادي $\boldsymbol{\ell}$، ومن المصفوفة $\boldsymbol{V}^{\frac{1}{2}}$ واعتماداً على العلاقة $(58-2)$ في الجزء الأول، نجد ما يلي :

$$\boldsymbol{e} = \boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell} = \boldsymbol{E} * \Lambda^{\frac{1}{2}} * \boldsymbol{E}' * \boldsymbol{\ell} \qquad (106-3)$$

ومنها يمكن استخلاص $\boldsymbol{\ell}$ بدلالة $\boldsymbol{e}$ كما يلي:

$$\boldsymbol{\ell} = \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{e} = \boldsymbol{E} * \Lambda^{-\frac{1}{2}} * \boldsymbol{E}' * \boldsymbol{e} \qquad (107-3)$$

حيث أن E هي مصفوفة الأشعة الذاتية الممعيرة والمتعامدة للمصفوفة $\boldsymbol{V}$. وأن $\Lambda^{\frac{1}{2}}$ هي المصفوفة القطرية لجذور القيم الذاتية للمصفوفة $\boldsymbol{V}$.

أي أنه إذا كان لدينا حل ممعير $\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ لتابع (فيشر) المعياري، فإنه يمكننا منه استخلاص الحل الخاص الخام $\boldsymbol{\ell}$ لذلك التابع من العلاقة :

$$\boldsymbol{\ell} = \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{e} \qquad (107a-3)$$

ملاحظة: ينصح بإجراء المعيرة على $\boldsymbol{\ell}$ إذا كان المتحولان X_1 و X_2 معياريين .

مثال (3−2): في دراسة ميدانية قام بها *[Bouma, B. N.+ others 1975]* لتقدير معدل انتشار مرض الناعور A عند النساء في الولايات المتحدة [المصدر(3) ص477]، تم سحب عينتين من النساء كما يلي:

العينة الأولى: تتألف من $n_1 = 30$ امرأة، وتم سحبها من مجموعة النساء العاديات اللاتي لا يحملن جينات مرض الناعور A ونرمز لها بـ G_1 .

العينة الثانية: تتألف من $n_2 = 22$ امرأة، وتم سحبها من مجموعة النساء اللاتي يحملن مرض الناعور A (بنات الحاملات، أمهات أبناء أو بنات مرضى) وسميت هذه المجموعة بمجموعة الحوامل الإجبارية للمرض ونرمز لها بـ G_2 .

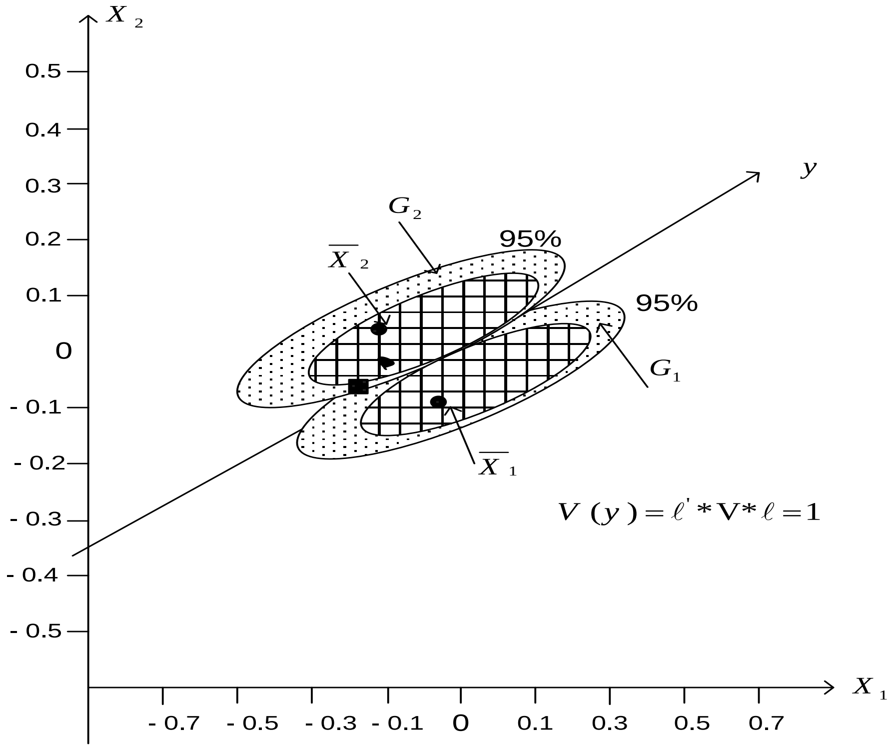
وللتمييز بين هاتين المجموعتين تم الاعتماد على متحولين X_1 و X_2 تم تعريفها وقياسها وتحويلها من نتائج تحليل الدم كما يلي:

$$X_1 = log_{10}(AHF\ Ativity) = \text{فعالية عامل الالتهاب}$$

$$X_2 = log_{10}(AHF - like\ antigen) = \text{المستضد : فعالية المقاومة}$$

ومن أزواج هذين المتحولين $(X_1\ , X_2)$ في كل مجموعة تم الحصول على القياسات المقابلة لذلك، والتي نعرضها ضمن قطعين ناقصين على الشكل $(7-3)$ ولكل مجموعة على حدة، القطع الصغير يضم 50% من المشاهدات، والكبير يضم 95% من المشاهدات . وبعد اختبار خضوع هذه المشاهدات للتوزيع الطبيعي المتعدد في كل مجموعة تم حساب متوسطيهما فكانا كما يلي:

$$\bar{X}_1 = \begin{bmatrix} -0.0065 \\ -0.0390 \end{bmatrix} \qquad \bar{X}_2 = \begin{bmatrix} -0.2483 \\ 0.0262 \end{bmatrix}$$



الشكل (3−7) توزيع قياسات X_1 و X_2

وبعد حساب مصفوفتي التباين المشترك لهما $\boldsymbol{S_1}$ و $\boldsymbol{S_2}$ في المجموعتين G_1 و G_2 تم حساب المصفوفة المدمجة للتباين المشترك $\boldsymbol{S_p}$ من العلاقة (3−86)، ثم تم حساب مقلوبها فكان يساوي كما يلي:

$$\boldsymbol{S_p^{-1}} = \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix}$$

وبعدها نقوم بحساب تابع فيشر التمييزي من العلاقة (3−88) التالية:

$$\tilde{y} = \tilde{\boldsymbol{\ell}}' * \boldsymbol{X} = (\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})' \boldsymbol{S_p^{-1}} * \boldsymbol{X} =$$

لذلك نحسب أولاً الشعاع $[\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}]$ فنجد أنه يساوي:

$$[\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}] = \begin{bmatrix} (\bar{X}_{11} - \bar{X}_{21}) \\ (\bar{X}_{12} - \bar{X}_{22}) \end{bmatrix} = \begin{bmatrix} (-0.0065 + 0.2483) \\ (-0.0390 - 0.0262) \end{bmatrix} = \begin{bmatrix} 0.2418 \\ -0.0652 \end{bmatrix}$$

ثم نعوض ذلك في المعادلة السابقة فنجد أن:

$$\tilde{y} = [0.2148 \quad -0.0652] \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\tilde{y} = [37.61 \quad -28.921] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\bar{y} = 37.61 X_1 - 28.921 X_2 \quad \Rightarrow \quad \tilde{\ell} = \begin{bmatrix} 37.61 \\ -28.921 \end{bmatrix}$$

وهو تابع فيشر التمييزي لهذه المسألة، وهو المستقيم المرسوم على الشكل (3−6) السابق. ثم نقوم بحساب تقديري المتوسطين μ_{1y} و μ_{2y} فنجد أن:

$$\tilde{\mu}_{1y} = \tilde{\boldsymbol{\ell}}' * \overline{\boldsymbol{X}}_{\boldsymbol{1}} = [37.81 \quad -28.921] \begin{bmatrix} -0.0065 \\ -0.0395 \end{bmatrix} = 0.88$$

$$\tilde{\mu}_{2y} = \tilde{\boldsymbol{\ell}}' * \overline{\boldsymbol{X}}_{\boldsymbol{2}} = [37.81 \quad -28.921] \begin{bmatrix} -0.2483 \\ 0.0262 \end{bmatrix} = -10.10$$

ثم نحسب القيمة الوسطى لهما فنجد أن:

$$\tilde{m} = \frac{1}{2}\left(\tilde{\mu}_{1y} + \tilde{\mu}_{2y}\right) = \frac{1}{2}(0.88 - 10.10) = -4.61$$

ولاستخدام هذه النتائج في تصنيف النساء على المجموعتين G_1 (العاديات) و G_2 (الحاملات لمرض الناعور A) نفترض أن نتائج تحليل الدم لإحدى النساء كانت كما يلي: $\boldsymbol{X}_0 = \begin{bmatrix}-0.210\\-0.044\end{bmatrix}$ أي أن:

$$x_{01} = -0.210 \qquad\qquad x_{02} = -0.044$$

نعوض ذلك في تابع فيشر التمييزي فنجد أن قيمته تساوي:

$$\tilde{y}_0 = \tilde{\boldsymbol{\ell}}'\boldsymbol{x_0} = [37.61 \quad -28.921]\begin{bmatrix}-0.210\\-0.044\end{bmatrix} = -6.62$$

وعند مقارنة قيمة $\tilde{y}_0$ مع $\tilde{m}$ نجد أن:

$$\tilde{y}_0 < \tilde{m} \quad \Leftrightarrow \quad -6.62 < -4.61$$

لذلك نصنف هذه المرأة ضمن المجموعة G_2 أي ضمن مجموعة حاملات مرضى الناعور A. ولقد أشرنا إلى هذه الحالة على الشكل (3–7) بإشارة نجمة. ومن الشكل نلاحظ أنها تقع ضمن قطع الـ 50% لمجموعة الحاملات G_2 وضمن قطع الـ95% لمجموعة النساء العاديات G_1، وهذا يعني أن هذا التصنيف ليس واضح المعالم .

والآن نقوم بمعيرة شعاع الثوابت $\boldsymbol{\ell}$ في تابع فيشر التمييزي وذلك باستخدام العلاقة:

$$\tilde{\boldsymbol{e}}^* = \frac{\tilde{\boldsymbol{\ell}}}{\sqrt{\tilde{\boldsymbol{\ell}}'\tilde{\boldsymbol{\ell}}}} = \frac{\tilde{\boldsymbol{\ell}}}{\|\tilde{\boldsymbol{\ell}}\|}$$

لذلك نحسب طول الشعاع $\tilde{\boldsymbol{\ell}}$ فنجد أن:

$$\|\boldsymbol{\ell}\| = \sqrt{\boldsymbol{\ell}' * \boldsymbol{\ell}} = \sqrt{\tilde{\ell}_1^2 + \tilde{\ell}_2^2} = \sqrt{(37.61)^2 + (-28.921)^2} = \sqrt{2250.936} = 47.444$$

ثم نقسم مركبات $\tilde{\boldsymbol{\ell}}$ على $\|\boldsymbol{\ell}\|$ فنحصل على الشعاع المميز التالي:

$$\tilde{\mathrm{e}} = \begin{bmatrix}\frac{\tilde{\ell}_1}{\|\tilde{\boldsymbol{\ell}}\|}\\[4pt]\frac{\tilde{\ell}_2}{\|\tilde{\boldsymbol{\ell}}\|}\end{bmatrix} = \begin{bmatrix}\frac{37.61}{47.444}\\[4pt]\frac{-28.921}{47.444}\end{bmatrix} = \begin{bmatrix}0.7927\\-0.6096\end{bmatrix} \approx \begin{bmatrix}0.79\\-0.61\end{bmatrix}$$

وبذلك تصبح المعادلة المميزة لتابع فيشر التمييزي كما يلي:

$$\tilde{y}^* = \tilde{\boldsymbol{e}}' * \boldsymbol{X} = 0.791X_1 - 0.61X_2$$

ومنه نستنتج أن الأهمية النسبية لـ X_1 أكبر من X_2 في التأثير على $\tilde{y}^*$، وبغض النظر عن الإشارة فإن الفرق بينهما صغيراً .

وحتى نطبق عملية التصنيف بعد هذه المعيرة، نقوم بحساب نقطة القيمة الوسطى الممعيرة من العلاقة:

$$\tilde{m}^* = \frac{1}{2}\left(\mu_{1y}^* + \mu_{2y}^*\right) = \frac{1}{2}(\tilde{\boldsymbol{e}}'\overline{\boldsymbol{X}}_{\boldsymbol{1}} + \tilde{\boldsymbol{e}}'\overline{\boldsymbol{X}}_{\boldsymbol{2}}) =$$

$$\tilde{m}^* = \frac{1}{2}\left\{\left[(0.79, -0.61)\begin{bmatrix}-0.0065\\-0.0390\end{bmatrix} + (0.79, -0.61)\begin{bmatrix}-0.2483\\0.0262\end{bmatrix}\right]\right\} = -0.0973$$

وإذا كنا نريد أن نصنف المشاهدة السابقة $X_0' = [-0.210\ ,\ \ -0.044]$ نعوضها في تابع فيشر التمييزي الممعير فنحصل على أن قيمته تساوي:

$$\tilde{y}^* = \boldsymbol{e}^{*\prime} * \boldsymbol{X} = (0.791\,, -0.61)\begin{bmatrix}-0.210\\-0.044\end{bmatrix} = -0.1397$$

ثم نقارن هذه القيمة مع القيمة الوسطى الممعيرة $\tilde{m}^*$ فنجد أن: $\tilde{y}^* < \tilde{m}^*$ (لأن $-0.1397 < -0.097$)

لذلك نصنف هذه المشاهدة ضمن المجموعة G_2 (مجموعة الحاملات للمرض) .

ملاحظة: يمكن حساب قيمتي $\tilde{y}_0^*$ و $\tilde{m}_0^*$ الممعيرتين مباشرة من القيمتين الأصليتين لهما $\tilde{y}_0$ و m_0، وذلك بضربهما بثابت المعيرة، وهو مقلوب طول الشعاع $\boldsymbol{\ell}$ والذي يساوي:

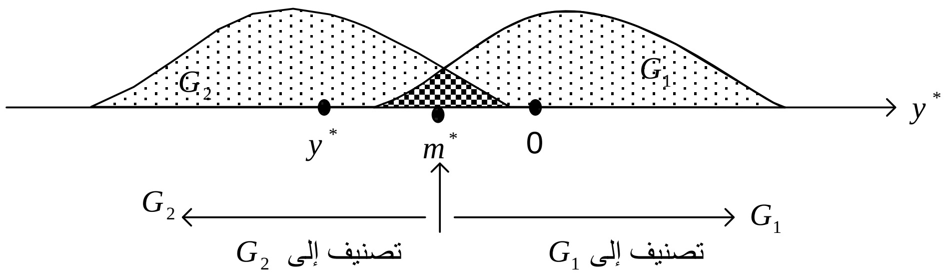
فنجد أن:

$$C = \frac{1}{\|\boldsymbol{\ell}\|} = \frac{1}{47.44} = 00211$$

$$\tilde{y}^* = C * \tilde{y} = 0.0211(-6.62) = -0.1397$$

$$\tilde{m}^* = C * \tilde{m} = 0.0211(-4.61) = -0.0973$$

ويمكن توضيح عملية التصنيف بيانياً كما يلي:



الشكل (3−8) محور التصنيف y

- **إضافات أخرى للبرهان على تابع (فيشر)** :هناك أساليب وطرائق أخرى لتحديد قيمة الشعاع $\boldsymbol{\ell}$ في العلاقة (3−74) التي تجعل النسبة F أكبر ما يمكن :
- **الطريقة الأولى:** لتطبيق هذه الطريقة نقوم بكتابة المعيار F كما يلي:

$$F = \frac{(\boldsymbol{\ell}'\boldsymbol{d})^2}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} = (\boldsymbol{\ell}'\boldsymbol{d})^2 * \frac{\mathbf{1}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} \qquad (108-3)$$

ثم نقوم باشتقاق F بالنسبة لـ $\boldsymbol{\ell}$ (حسب قواعد اشتقاق المصفوفات)، ونضع ذلك المشتق مساوياً للصفر فنجد أن:

$$\frac{dF}{d\ell} = 2(\boldsymbol{\ell}'\boldsymbol{d}) * \boldsymbol{d} * \frac{\mathbf{1}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} - \frac{\mathbf{2}\boldsymbol{V} * \boldsymbol{\ell}}{(\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell})^2} * (\boldsymbol{\ell}'\boldsymbol{d})^{\mathbf{2}} = 0$$

نختصر على (2) ثم نخرج المقدار العددي $\left(\frac{\ell' d}{\ell' * V * \ell}\right)$ خارج قوس فنحصل على أن:

$$\frac{\boldsymbol{\ell}'\boldsymbol{d}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}}\left[\boldsymbol{d} - \left(\frac{\boldsymbol{\ell}'\boldsymbol{d}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}}\right) * \boldsymbol{V} * \boldsymbol{\ell}\right] = 0 \qquad (109-3)$$

وبما أن المقدار $\left(\frac{\ell' d}{\ell' * V * \ell}\right)$ هو عبارة عن عدد حقيقي سلمي لذلك نرمز له بـ K فنحصل على أن:

$$K[\boldsymbol{d} - K * \boldsymbol{V} * \boldsymbol{\ell}] = 0 \qquad (110-3)$$

وبعد الاختصار على K نحصل على المعادلة التالية:

$$\boldsymbol{d} = K * \boldsymbol{V} * \boldsymbol{\ell} \qquad (111-3)$$

ولتحديد قيمة $\boldsymbol{\ell}$ نضرب الطرفين من اليسار بـ $\boldsymbol{V^{-1}}$ ثم نقسم على k فنحصل على أن $\boldsymbol{\ell}$ يساوي:

$$\boldsymbol{\ell} = \frac{\boldsymbol{1}}{K}\boldsymbol{V^{-1}} * \boldsymbol{d} = C * \boldsymbol{V^{-1}} * \boldsymbol{d} \qquad : (c = \frac{1}{k}) \qquad (112-3)$$

وبما أن $\boldsymbol{d} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ فإننا نستنتج أن:

$$\boldsymbol{\ell} = C * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (113-3)$$

وهي نفس العلاقة التي حصلنا عليها في (3–77)، ومنها نستنتج أن الشعاع $\boldsymbol{\ell}$ يتناسب طرداً مع الشعاع $(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) * \boldsymbol{V^{-1}}$، وهنا يعطينا لا نهاية من الحلول المتناسبة لـ $\boldsymbol{\ell}$ وذلك حسب قيم $C \neq 0$.

وإذا قمنا بوضع $C = 1$ نحصل على الحل الخاص التالي:

$$\boldsymbol{\ell} = \boldsymbol{V^{-1}} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \quad \Longrightarrow \quad \boldsymbol{\ell}' = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} \qquad (114-3)$$

وبتعويض ذلك في تابع (فيشر) نحصل على العلاقة التالية:

$$y = \boldsymbol{\ell}' * \boldsymbol{X} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * \boldsymbol{V^{-1}} * \boldsymbol{X} \qquad (115-3)$$

وفي حالة العينات نقدرها بالعلاقة التالية:

$$\tilde{y} = \tilde{\boldsymbol{\ell}}' * \boldsymbol{X} = (\bar{\boldsymbol{X}}_{\boldsymbol{1}} - \bar{\boldsymbol{X}}_{\boldsymbol{2}}) * \boldsymbol{S_p^{-1}} * \boldsymbol{X} \qquad (116-3)$$

ولاستخدام هذه العلاقة في عمليات تصنيف المشاهدات على المجموعتين نقوم بحساب القيمة الوسطى للمتوسطين μ_{1y} و μ_{2y} أو لتقديرهما $\bar{X}_{1y}$ و $\bar{X}_{2y}$ فنحصل على أن :

$$\widetilde{m} = \frac{1}{2}\left(\tilde{\mu}_{1y} + \tilde{\mu}_{2y}\right) = \frac{1}{2}\left(\bar{X}_{1y} + \bar{X}_{2y}\right) \qquad (117-3)$$

وبعدها نطبق القاعدة (3–94) على أية مشاهدة مثل X_0 .

- **الطريقة الثانية:** ومن جهة أخرى نلاحظ أنه إذا كنا نشترط أن يكون $\sigma^2_{(y)} = 1$ فإن ذلك يعني أن نضيف إلى المعيار F الشرط التالي:

$$\sigma^2_{(y)} = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} = 1 \qquad (118-3)$$

وعندها فإن معيار (فيشر) في (3–74) يأخذ شكلاً مبسطاً كما يلي:

$$F = \frac{(\boldsymbol{\ell}' * \boldsymbol{d})^2}{1} \Longrightarrow Max \qquad (119-3)$$

ويصبح المطلوب منا أن نحسب عناصر الشعاع $\boldsymbol{\ell} = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}$ التي تجعل المعيار F في (3–119) أكبر ما يمكن ضمن الشرط الجديد (3–118) .

ولإيجاد الشعاع $\boldsymbol{\ell}$ نشكل تابع (لاغرانج) من التابع (3– 119) والشرط (3–118) كما يلي:

$$L = (\boldsymbol{\ell}' * \boldsymbol{d})^{\boldsymbol{2}} - \boldsymbol{\lambda}(\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} - \boldsymbol{1}) \qquad (120-3)$$

ثم نشتق التابع L بالنسبة لـ $\boldsymbol{\ell}$ (حسب قواعد اشتقاق المصفوفات) ونضع ذلك المشتق مساوياً للصفر فنحصل على أن:

$$\frac{\partial L}{\partial \ell} = 2(\boldsymbol{\ell}' * \boldsymbol{d}) * \boldsymbol{d} - \lambda * 2\boldsymbol{V} * \boldsymbol{\ell} = 0 \qquad (121-3)$$

وبعد الاختصار على (2) نحصل منها على المعادلة التالية:

$$(\boldsymbol{\ell}' * \boldsymbol{d}) * \boldsymbol{d} - \lambda * \boldsymbol{V} * \boldsymbol{\ell} = 0 \qquad (122-3)$$

وبما أن $\boldsymbol{\ell}' * \boldsymbol{d}$ هو عدد حقيقي سلمي لأن :

$$\boldsymbol{\ell}' * \boldsymbol{d} = (\ell_1 , \ell_2) \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \ell_1 d_1 + \ell_2 d_2 = C_1 = \text{عدد سلمي}$$

ولذلك نرمز له بـ C_1 فنجد أن العلاقة(121−3) تأخذ الشكل التالي:

$$C_1 * \boldsymbol{d} = \lambda * \boldsymbol{V} * \boldsymbol{\ell} \qquad (123-3)$$

ثم نضرب الطرفين من اليسار بـ $\boldsymbol{V^{-1}}$ ونقسمهما على λ فنحصل على أن الشعاع $\boldsymbol{\ell}$ يساوي:

$$\boldsymbol{\ell} = \frac{C_1}{\lambda} \boldsymbol{V^{-1}} * \boldsymbol{d} = C\boldsymbol{V^{-1}}\boldsymbol{d} \quad : \quad \left(C = \frac{C_1}{\lambda} \right) \qquad (124-3)$$

وبما أن $\boldsymbol{d} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ فإن $\boldsymbol{\ell}$ يساوي:

$$\boldsymbol{\ell} = C\boldsymbol{V^{-1}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (125-3)$$

وهي نفس العلاقة التي حصلنا عليها في (3−77)، أي أن $\boldsymbol{\ell}$ يتناسب طرداً مع الشعاع، $\boldsymbol{V^{-1}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ وأن العلاقة (3−125) تعطينا لا نهاية من الحلول. ولكن بوضع $C = 1$ نحصل على حل خاص له هو:

$$\boldsymbol{\ell} = \boldsymbol{V^{-1}}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (126-3)$$

وفي حالة العينات نحسب تقدير $\boldsymbol{\ell}$ من العلاقة:

$$\tilde{\boldsymbol{\ell}} = \boldsymbol{S_p^{-1}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \qquad (127-3)$$

وبتعويض ذلك في تابع (فيشر) نحصل على تابع (فيشر) التمييزي التالي:

$$y = \tilde{\boldsymbol{\ell}}'\boldsymbol{X} = (\boldsymbol{X_1} - \boldsymbol{X_2})' * \boldsymbol{S_p^{-1}} * \boldsymbol{X} \qquad (128-3)$$

ثم نتعامل معه كالعادة ونحسب القيمة الوسطى لـ y، ثم نطبق القاعدة من أجل تصنيف أي عنصر جديد من المجتمع مثل $\boldsymbol{X_0}$.

الفصل الرابع
التحليل التمييزي الخطي المتعدد (لعدة مجموعات ولعدة متحولات)

إن التحليل التمييزي المتعدد هو تعميم للتحليل التمييزي (بمجموعتين ومتحولين) . لذلك سنستخدم وسنعمم نفس الرموز والطرائق الرياضية السابقة، لكن نظراً لصعوبة التعامل مع بعض الطرائق في الفضاءات العليا، فإننا سنقتصر على أهم طريقتين هما: طريقة ECM وطريقة (فيشر) مع التعريج على بعض الطرائق الأخرى، وسنعرضها كما يلي:

4-1 : طريقة تخفيض التكلفة المتوقعة للتصنيف الخاطئ ECM:
[Johnson,Wichern. P. 502 بتصرف وإضافة]

لنفترض أنه لدينا مجتمعاً مؤلفاً من g مجموعة منفصلة ومتكاملة، نرمز لها بالرموز التالية: $G_1\ G_2\ G_3 \dots G_g$، وإن كل منها يتأثر بجملة من المتحولات المستقلة هي $X_1\ X_2\ X_3 \dots X_P$، وإن التوزيعات المشتركة للمتحولات X ضمن كل مجموعة G_j هي $f_j(x)$ [ومن المفترض أن تكون هذه التوزيعات توزيعات طبيعية متعددة. ولكن هذا الشرط لا يعتبر ضرورياً في مجرى سير النظرية العامة] .
كما نفترض أن الاحتمالات السابقة لانتماء العناصر إلى هذه المجموعات (نسبة كل مجموعة من أصل المجتمع) معلومة، ونرمز لها بالرموز التالية:

$$P_1\ , P_2, P_3, \dots\ P_g \qquad : \quad \left(\sum_{j=1}^{g} P_j = 1\right) \qquad (1-4)$$

كما نفترض أن تكلفة التصنيف الخاطئ لعنصر من المجموعة G_j ضمن المجموعة G_k معلومة، ونرمز لها بالرموز :

$$C(k/j) \qquad k \neq j: 1\ 2\ 3 \dots . g \qquad (2-4)$$

كما نفترض أن تكلفة التصنيف الصحيح في نفس المجموعة معدومة(لأنها بدون ضرر)، ونرمز لها بالرموز :

$$C(j/j) = 0 \qquad j: 1\ 2\ 3 \dots . g \qquad (3-4)$$

وأخيراً لنفترض أن R_jهي المنطقة المؤلفة من جملة قيم X في الفضاءR^p، التي تضم بعض العناصر المصنفة في المجموعة k وهي أصلاً من المجموعة j، وبذلك نجد أن احتمال التصنيف الخاطئ لعنصر x_0 من G_j في المجموعة G_k يساوي :

$$P(k/j) = P[G_k\ \text{تصنيف العنصر في} / G_j\ \text{أصله من}] = \int_{R_k} f_j(x)dx \qquad (4-4)$$

حيث أن: $k \neq\ j: 1\ 2\ 3 \dots . g$

أما احتمال التصنيف الصحيح له في G_j نفسها فيساوي:

$$P(j/j) = 1 - \sum_{\substack{k=1 \\ k\neq j}}^{g} P(k/j) \qquad K \neq j \qquad (5-4)$$

فإذا أخذنا عنصراً من المجموعة الأولى G_1 (كمجموعة شرطية أولى)، فإننا نجد أن التكلفة الشرطية للتصنيف الخاطئ لعنصر x_0 من G_1 في إحدى المجموعات الأخرى $G_2\ G_3 \dots G_g$ ، يساوي ما يلي:

$$ECM(1) = P(2/1)C(2/1) + P(3/1)C(3/1) + \cdots P(g/1)C(g/1)$$

$$ECM(1) = \sum_{k=2}^{g} P(k/1) * C(k/1) \qquad (6-4)$$

وباتباع الأسلوب نفسه يمكننا حساب التكاليف الشرطية للتصنيف الخاطئ لأي عنصر من المجموعة G_j في المجموعات الأخرى $G_k(k \neq j)$ ، وذلك من العلاقة التالية :

$$ECM(j) = \sum_{\substack{k=1 \\ k\neq j}}^{g} P(k/j) * C(k/j) \qquad (7-4)$$

حيث أن: $j: 1\ 2\ 3 \dots g$

وبذلك نحصل على التكاليف المتوقعة للتصنيف الخاطئ في تلك المجموعات وهي التالية: $ECM(1)\ ECM(2) \dots ECM(g)$. وبما أن كل من هذه التكاليف يحدث في المجموعة G_j بما يتناسب مع الاحتمال السابق المقابل لها P_j، لذلك فإننا نضرب كل من هذه التكاليف باحتمالاتها المسبقة P_j، ثم نأخذ مجموع تلك الجداءات فنحصل على التكلفة الاجمالية المتوقعة للتصنيف الخاطئ ECM والتي نكتبها كما يلي:

$$ECM = P_1ECM(1) + P_2ECM(2) + P_3ECM(3) + \cdots + P_gECM(g) \qquad (8-4)$$

$$ECM = P_1\left[\sum_{k=2}^{g} P(k/1) * C(k/1)\right] + P_2\left[\sum_{\substack{k=1 \\ k\neq 2}}^{g} P(k/2) * C(k/2)\right] +$$

$$+ P_3\left[\sum_{\substack{k=1 \\ k\neq 3}}^{g} P(k/3) * C(k/3)\right] + \cdots + P_g\left[\sum_{\substack{k=1 \\ k\neq g}}^{g-1} P(k/g) * C(k/g)\right] \quad (9-4)$$

وباختصار نكتبها كما يلي :

$$ECM = \sum_{j=1}^{g} P_j\left[\sum_{\substack{k=1 \\ k\neq j}}^{g} P(k/j) * C(k/j)\right] \Rightarrow Min \qquad (10-4)$$

وباستخدام أي أسلوب للتصنيف المثالي، يمكننا التوصل إلى تحديد مناطق التصنيف العامة والخاصة المقابلة لها $R_1\ R_2\ R_3 \dots R_g$ ، والتي تجعل التكلفة الاجمالية في العلاقة (4−10) أصغر ما يمكن .

لذلك نقوم بمعالجة العلاقة (4−10)، قياساً على العلاقة (2−14)، ونستبدل الاحتمال $P(k/j)$ بالتكامل $\int_{R_k} f_j(x)dx$ المعرف على المنطقة R_k فنحصل على أن:

$$ECM = \sum_{j=1}^{g} P_j \left[\sum_{\substack{k=1 \\ k \neq j}}^{g} \int_{R_k} f_j(x) * C(k/j) dx \right] \qquad (11-4)$$

ثم نقوم بتبديل مواقع المجاميع والرموز والتكامل فنحصل على أن:

$$ECM = \sum_{k=1}^{g} \int_{R_k} \left[\sum_{\substack{j=1 \\ j \neq k}}^{g} P_j * f_j(x) * C(k/j) \right] dx \qquad (12-4)$$

والآن علينا أن نبحث عن المناطق R_k التي تجعل هذه القيمة أصغر ما يمكن، وحتى نستطيع تحقيق ذلك يجب أن نجعل قيمة التكامل على R_k أصغر ما يمكن، أي أن نحدد المنطقة R_k التي تجعل التكامل أصغر ما يمكن .

وحتى يحدث ذلك يجب أن يكون المقدار المكامل (الذي تحت التكامل) أصغر ما يمكن. أي يجب أن يكون المقدار التالي أصغر ما يمكن .

$$H_k = \sum_{\substack{j=1 \\ j \neq k}}^{g} P_j * f_j(x) * C(k/j) \rightarrow Min \qquad (13-4)$$

حيث أن: H_k تحسب لجميع المجموعات $k: 1\ 2\ 3 \dots g$، ولكن المجموع في كل H_k يحسب على جميع المجموعات باستثناء المجموعة k . وهكذا نجد أنه يمكننا صياغة القاعدة العامة لتصنيف العناصر وتوزيعها على المجموعات كما يلي:

نصنف أي عنصر x_0 في المجموعة k إذا كان المقدار:

$$H_k = \sum_{\substack{j=1 \\ j \neq k}}^{g} P_j * f_j(x_0) * C(k/j) \rightarrow Min \qquad (14-4)$$

أصغر ما يمكن من بين المقادير H_j في المجموعات الأخرى .

الحالات الخاصة: إذا كانت تكاليف التصنيف الخاطئ $C(k/j)$ متساوية أو كانت تساوي الواحد فإن قاعدة تصنيف العنصر x_0 في المجموعة k تصبح كما يلي:

نصنف أي عنصر x_0 في المجموعة k إذا كان المقدار:

$$h_k = \sum_{\substack{j=1 \\ j \neq k}}^{g} P_j * f\!\!\!/_j(x_0) \rightarrow Min \qquad (15-4)$$

أصغر ما يمكن من بين المقادير h_j في المجموعات الأخرى .

ملاحظة /1/: نلاحظ في العلاقة (4–15) أن حداً محذوفاً من المجموع العام وهو الحد: $P_k * f\!\!\!/_k(x_0)$، لذلك يمكننا أن نكتب العلاقة (4–15) كما يلي:

$$h_k = \sum_{j=1}^{g} P_j * f\!\!\!/_j(x_0) - P_k * f\!\!\!/_k(x_0) \qquad (16-4)$$

وحتى يأخذ المقدار h_k أصغر قيمة له، يجب أن يأخذ الحد الأخير أكبر قيمة له. وهكذا نجد أنه يمكننا تعديل قاعدة التصنيف الخاصة (4–15) وصياغة قاعدة خاصة أخرى كما يلي:

إذا كان المقدار $P_k f\!\!\!/_k(x_0)$ أكبر من جميع الجداءات المقابلة للمجموعات الأخرى $P_j f\!\!\!/_j(x_0)$ ، فإننا نصنف العنصر x_0 في المجموعة G_k ونكتب ذلك كما يلي:

نصنف العنصر x_0 في المجموعة G_k إذا كان:

$$P_k * f\!\!\!/_k(x_0) = \max\left[P_j * f\!\!\!/_j(x_0)\right] \quad : For \;\; j \neq k \quad j: 1\,2\,3 \dots g \qquad (17-4)$$

ويمكن استخدام القاعدة المكافئة لها في التوزيعات الأسية وهي:

نصنف العنصر x_0 في المجموعة G_k، إذا كان :

$$\ell n\,[P_k * f\!\!\!/_k(x_0)] = \max \ell n \left[P_j * f\!\!\!/_j(x_0)\right] \quad : For \;\; j \neq k \qquad (18-4)$$

وذلك بفرض أن تكاليف التصنيف الخاطئ $C(k/j)$ متساوية .

ملاحظة /2/ : يمكننا الاستفادة من العلاقة (2–34) لاستخراج الاحتمالات اللاحقة وتحديد أكبرها ثم تطبيق طريقة (بايز) لتصنيف عناصر المجتمع. فنجد أن الاحتمالات اللاحقة هي:

$$P(G_k/X) = P\left(G_k \text{ من } X \text{ حيث } X \text{ مشاهدة معطيه/احتمال أن تكون}\right)$$

ومن علاقة (بايز) نجد أن:

$$P(G_k/X) = \frac{P_k * f\!\!\!/_k(x)}{\sum_{j=1}^{g} P_j * f\!\!\!/_j(x)} = \frac{P_k \text{ التوزيع * الاحتمال السابق}}{P_j * f\!\!\!/_j(x) \text{ مجموع الجداءات}} \qquad (19-4)$$

حيث أن: $k: 1\,2\,3 \dots g$، ومنها نحصل على القاعدة التالية:

نصنف x_0 في المجموعة G_k إذا كان الاحتمال اللاحق:

$$P(G_k/x_0) = \text{Max}\frac{P_k * f\!\!\!/_k(x_0)}{\sum P_j f\!\!\!/_j(x_0)} \qquad (20-4)$$

يأخذ أكبر قيمة ممكنة من بين الاحتمالات اللاحقة الأخرى .

وذلك بفرض أن تكاليف التصنيف الخاطئ $C(k/j)$ متساوية .

وأخيراً نشير إلى أن قاعدة التكاليف المتوقعة للتصنيف الخاطئ ECM تعتمد على ثلاثة مركبات هي: الاحتمالات السابقة P_k وتكاليف التصنيف الخاطئ $C(k/j)$ وقانون التوزيع الاحتمالي $f_k(x)$ ، وإن هذه المركبات يجب أن تكون معلومة (أو مقدرة) قبل تطبيق هذه القاعدة . وعندما تكون مجهولة يتم تقديرها من العينات العشوائية، التي يتم سحبها من كل مجموعة من المجموعات $G_1\ G_2\ G_3 \dots G_g$، التي يتألف منها المجتمع المدروس .

مثال (4−1): لنفترض إننا نريد تصنيف عنصر محدد x_0 إلى إحدى ثلاث مجموعات هي: $G_1\ G_2\ G_3$، ولنفترض أن الاحتمالات السابقة الافتراضية P_j وتكاليف التصنيف الخاطئ $C(k/j)$ وقيم قوانين التوزيع معلومة كما في الجدول التالي:

جدول (4−1) **بيانات المثال [من** Johnson,Wishern, P.505]

المجموعات	G_1	G_2	G_3
G_1	$C(1/1) = 0$	$C(1/2) = 500$	$C(1/3) = 100$
G_2	$C(2/1) = 10$	$C(2/2) = 0$	$C(2/3) = 50$
G_3	$C(3/1) = 50$	$C(3/2) = 200$	$C(3/3) = 0$
الاحتمالات السابقة	$P_1 = 0.05$	$P_2 = 0.60$	$P_3 = 0.35$
قيم التوزيعات الاحتمالية عند القيمة المحددة x_0	$f_1(x_0) = 0.01$	$f_2(x_0) = 0.85$	$f_3(x_0) = 0.2$

ولإجراء عملية التصنيف نستخدم أولاً الطريقة العامة لـ ECM ، لذلك نقوم بحساب قيم الجداءات الواردة في العلاقة (4−14)، فنجد أنه عندما نعطي k القيم المتتالية 3، 2، 1، نحصل على أن:

$$k = 1 \Longrightarrow H_1 = P_2 * f_2(x_0) * C(1/2) + P_3 * f_3(x_0) * C(1/3) = \\ = (0.60)(0.85)(500) + (0.35)(2)(100) = 325$$

$$k = 2 \Longrightarrow H_2 = P_1 * f_1(x_0) * C(2/1) + P_3 * f_3(x_0) * C(2/3) = \\ = (0.05)(0.01)(10) + (0.35)(2)(50) = 35.055$$

$$k = 3 \Longrightarrow H_3 = P_1 * f_1(x_0) * C(3/1) + P_2 * f_2(x_0) * C(3/2) = \\ = (0.05)(0.01)(50) + (0.60)(0.85)(200) = 102.025$$

وعندما مقارنة قيم $H_1\ H_2\ H_3$، نجد أن أصغرها هي القيمة $H_2 = 35.055$، لذلك نصنف العنصر المفروض x_0 ضمن المجموعة G_2 .

ملاحظة: إذا كانت تكاليف التصنيف الخاطئ متساوية، فإنه يمكننا استخدام الطريقة الخاصة، وتصنيف x_0 حسب العلاقة (4−17). وهذا الأمر لا يتطلب –في هذه الحالة– سوى حساب الجداءات التالية:

$$P_1 * f_1(x_0) = (0.05)(0.1) = 0.0005$$
$$P_2 * f_2(x_0) = (0.60)(0.85) = 0.510$$
$$P_3 * f_3(x_0) = (0.35)(2) = 0.700$$

وبما أن هذه الطريقة تعتمد على أكبر قيمة للجداءات $P_j * f_j(x_0)$ فإننا نبحث عن أكبرها فنجد بعد المقارنة أن:

$$P_3 * f_3(x_0) = 0.700 > P_j * f_j(x_0) \qquad j: 1\ ,2$$

لذلك نقوم بتصنيف x_0 في المجموعة G_3 (خلافاً للتصنيف في ECM)، وذلك لأننا استبعدنا التكاليف.

كما يمكننا في هذه الحالة استخدام طريقة (بايز) لتصنيف x_0 . لذلك نقوم بحساب الاحتمالات اللاحقة ثم نختار أكبرها فنجد أن:

$$P(G_1/x_0) = \frac{P_1 * f_1(x_0)}{\sum_{j=1}^{3} P_j * f_j(x_0)}$$

$$P(G_1/x_0) = \frac{(0.05)(0.01)}{(0.05)(0.01) + (0.60)(0.85) + (0.35)(2)} = \frac{0.0005}{1.2105} = 0.0004$$

$$P(G_2/x_0) = \frac{P_2 * f_2(x_0)}{\sum P_j f_j(x_0)} = \frac{(0.60)(0.85)}{1.2105} = \frac{0.510}{1.2105} = 0.421$$

$$P(G_3/x_0) = \frac{P_3 * f_3(x_0)}{\sum P_j f_j(x_0)} = \frac{(0.35)(2)}{1.2105} = \frac{0.700}{1.2105} = 0.578$$

وعند مقارنة هذه الاحتمالات نجد أن أكبرها هو الاحتمال $P(G_3/x_0)$، لذلك نصنف العنصر x_0 في المجموعة G_3 المقابلة للاحتمال اللاحق الأكبر. وهذا يتفق مع الطريقة الخاصة السابقة. ولكنه يختلف عن القاعدة العامة بسبب اختلاف الشروط المتعلقة باستبعاد التكاليف .

4-2 : توابع التصنيف في المجتمعات الطبيعية المتعددة حسب ECM:

إن هذه الحالة الخاصة تحدث عندما تكون قوانين التوزيع الاحتمالية المتعددة في المجموعات $G_1\ G_2\ G_3 \ldots G_g$ للمتحولات X طبيعية ومعرفة بالعلاقة العامة التالية:

$$f_j(X) = \frac{1}{(2\pi)^{\frac{P}{2}} |V_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu_j})' * \boldsymbol{V_j^{-1}} * (\boldsymbol{X}-\boldsymbol{\mu_j})} \qquad (21-4)$$

حيث أن $j: 1\ 2\ 3 \ldots g$، وأن μ_j هو شعاع المتوسطات في المجموعة G_j، و V_j هي مصفوفة التباين المشترك لـ X في المجموعة G_j .

ولنفترض الآن أن تكاليف التصنيف الخاطئ متساوية أو تساوي الواحد $[C(k/j) = 1]$ وأن تكاليف التصنيف الصحيح معدومة $[C(j/j) = 0]$، وعندها نجد أن قاعدة التصنيف للتوزيعات الأسية (4-18) لأصغر تكلفة متوقعة *ECM* تأخذ الشكل التالي:

$$\ell n\big(P_k * f_k(X)\big) = Max\, \ell n\big[P_j * f_j(X)\big] = Max\big[\ell n\, P_j + \ell n\, f_j(X)\big] = \qquad (22-4)$$

$$= Max\left[\ell n\, P_j - \frac{P}{2}\ell n(2\pi) - \frac{1}{2}\ell n|V_j| - \frac{1}{2}\big(\boldsymbol{X}-\boldsymbol{\mu_j}\big)' * \boldsymbol{V_j^{-1}} * \big(\boldsymbol{X}-\boldsymbol{\mu_j}\big)\right]$$

وهنا نلاحظ أن الحد $\left[-\frac{P}{2}\ln(2\pi)\right]$ هو عدد ثابت ومشترك بين جميع المجموعات، لذلك نقوم بتجاهله وحذفه من الطرف الأيمن. ثم نعرف من الحدود المتبقية ما يسمى (بالتابع التمييزي التربيعي– QDS) [Quadratic Discrimination Score] ونرمز له بالنسبة للمجموعة G_j بالرمز :

$$QD_j(X) = -\frac{1}{2}\ln|V_j| - \frac{1}{2}\left(\boldsymbol{X}-\boldsymbol{\mu_j}\right)' * \boldsymbol{V_j^{-1}} * \left(\boldsymbol{X}-\boldsymbol{\mu_j}\right) + \ln P_j \qquad (23-4)$$

حيث أن $j: 1\ 2\ 3 \ldots g$.

وهنا نلاحظ أن التابع التربيعي $QD_j(x)$ مؤلف من عدة مركبات هي: لوغاريتم الاحتمالات السابقة P_j– نصف قيمة لوغاريتم محدد مصفوفة التباين المشترك $|V_j|$– نصف مربع المسافة من النقطةX إلى μ_j شعاع متوسط المجموعةG_j.

وبالاعتماد على هذا التابع التمييزي فإن قاعدة التصنيف (4–18) السابقة تصبح كما يلي:

قاعدة الاحتمال الإجمالي الأصغر للتصنيف الخاطئ TPM في حالة المجتمعات الطبيعية: نصنف النقطة X_0 في المجموعة G_k . إذا كان : $QD_k(X_0) = \text{Max}\left[QD_j(X)\right] \qquad j: 1\ 2\ 3 \ldots g \qquad (24-4)$ حيث أن $QD_j(X)$ تحسب من العلاقة (4–23) وذلك بفرض أن تكاليف التصنيف الخاطئ متساوية .

ولكن تطبيق هذه العلاقات تقتضى أن تكون معالم المجتمع الأصلي μ_j وV_j في كل مجموعة G_j معلومة .

أما عندما تكون هذه المعالم مجهولة – وهي الحالة الغالبة في التطبيقات العملية– نقوم بتقديرها من العينات العشوائية المسحوبة من تلك المجموعات، ونقدر المتوسط μ_j بمتوسط العينة $\boldsymbol{\bar{X}_j}$ ومصفوفة التباين المشترك $\boldsymbol{V_j}$ بمصفوفة التباين المشترك $\boldsymbol{S_j}$ المحسوبة من العينة المسحوبة من المجموعة G_j .

ثم نعوض ذلك في العلاقة (4–23) فنحصل على تقدير للتابع التربيعي $QD_j(X)$ ونرمز له بـ $qd_j(X)$ (أو $\widetilde{QD}_j(X)$) ونكتبه كما يلي:

$$qd_j(X) = -\frac{1}{2}\ln|S_j| - \boldsymbol{\frac{1}{2}}\left(\boldsymbol{X}-\boldsymbol{\bar{X}_j}\right)' * \boldsymbol{S_j^{-1}} * \left(\boldsymbol{X}-\boldsymbol{\bar{X}_j}\right) + \ln P_j \qquad (25-4)$$

حيث أن $j: 1\ 2\ 3 \ldots g$.

وبذلك تأخذ قاعدة TPM للتصنيف في حالة العينات وللتوزيع الطبيعي المتعدد الشكل المقدر التالي :

قاعدة التصنيف لأصغر TPM في المجتمعات الطبيعية: نصنف التقطة X_0 ضمن المجموعة G_k إذا كان : $qd_k(X_0) = \text{Max}\left[qd_j(X_0)\right] \qquad j: 1\ 2\ 3 \ldots g \qquad (26-4)$ حيث أن $qd_j(X)$ تحسب من العلاقة (4–25) السابقة .

حالة خاصة: إذا كانت مصفوفات التباين المشترك $\boldsymbol{V_j}$ متساوية، يمكن تبسيط القواعد السابقة، أي أنه إذا كانت $\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V_3} = \boldsymbol{V}$.

فإننا نضع $\boldsymbol{V_j} = \boldsymbol{V}$ مهما تكن $j: 1\ 2\ 3 \dots g$، ونعوضها في العلاقة (4−23) فنحصل على أن :

$$QD_j(X) = -\frac{1}{2}\ln|V| - \frac{\mathbf{1}}{\mathbf{2}}\left(\boldsymbol{X} - \boldsymbol{\mu_j}\right)' * \boldsymbol{V^{-1}} * \left(\boldsymbol{X} - \boldsymbol{\mu_j}\right) + \ln P_j \qquad (27-4)$$

وبعد حساب مفكوك الحد الأوسط وإصلاحه نحصل على الصيغة التالية:

$$QD_i(X) = -\frac{1}{2}\ln|V| - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{X}' * \boldsymbol{V^{-1}} * \boldsymbol{X} + \boldsymbol{\mu_j'} * \boldsymbol{V^{-1}} * \boldsymbol{X} - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\mu_j'} * \boldsymbol{V^{-1}} * \boldsymbol{\mu_j} + \ln P_j \quad (\mathbf{28-4})$$

وهنا نلاحظ أن الحدين الأول والثاني ثابتين في كل المجموعات G_j (اللوغاريتم ومربع المسافة) .لذلك يمكن تجاهلهما في عمليات التصنيف . أما الحدود المتبقية فتتألف من جزأين: جزء عددي وهو $\left[-\frac{1}{2}\mu_j' V^{-1}\mu_j + \ln P_j\right]$، وجزء هو عبارة عن تركيب خطي من مركبات X هو$\left(\mu_j' * V^{-1} * X\right)$.

وبناء عليه نعرف التابع التمييزي الخطي المؤلف من الحدود المتبقية بالعلاقة التالية:

$$D_j(X) = \boldsymbol{\mu_j'} * \boldsymbol{V^{-1}} * \boldsymbol{X} - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\mu_j'} * \boldsymbol{V^{-1}} * \boldsymbol{\mu_j} + \ln P_j \qquad (29-4)$$

حيث أن $j: 1\ 2\ 3 \dots g$. ويسمى هذا التابع بالتابع التصنيفي للمجموعة G_j

ومنها نحصل على قاعدة لتصنيف العناصر حسب هذا التابع نصيغها كما يلي:

قاعدة التصنيف لأصغر TPM في حالة تساوي مصفوفات التباينات المشتركة في المجتمعات الطبيعية: نصنف النقطة X_0 في المجموعة G_k، إذا كانت قيمة التابع الخطي التصنيفي $D_k(X_0)$ أكبر ما يمكن. أي إذا كان:

$$D_k(X_0) = \mathrm{Max}\left[D_j(X_0)\right] \qquad j: 1\ 2\ 3 \dots g \qquad (30-4)$$

حيث ان $D_j(x_0)$ تحسب من العلاقة (4−29) .

وإذا كانت معالم المجتمع مجهولة فإننا نقدرها من بيانات العينات المسحوبة من المجموعات $G_1\ G_2 \dots G_g$ ونستبدل كل $\boldsymbol{\mu_j}$ بـ$\boldsymbol{\bar{X}_j}$، ولكننا نستبدل مصفوفة التباين المشترك الموحدة **V** (بفرض $\boldsymbol{V_j} = \boldsymbol{V}$) بالتباين المدمج والمركب من مصفوفات التباين المشترك $\boldsymbol{S_j}$ باستخدام العلاقة :

$$\boldsymbol{S_P} = \frac{(n_1 - 1)\boldsymbol{S_1} + (n_2 - 1)\boldsymbol{S_2} + \dots + \left(n_g - 1\right)\boldsymbol{S_g}}{n_1 + n_2 + n_3 + \dots + n_g - g} \qquad (31-4)$$

وعندها تأخذ العلاقة (4−29) للتابع التصنيفي الخطي الشكل المقدر التالي :

$$d_j(x) = \boldsymbol{\bar{X}_j'} * \boldsymbol{S_P^{-1}} * \boldsymbol{X} - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\bar{x}_j'} * \boldsymbol{S_P^{-1}} * \boldsymbol{\bar{X}_j} + \ln P_j \qquad (32-4)$$

وإن قاعدة التصنيف للعينات تأخذ الشكل التالي:

قاعدة أصغر TPMفي العينات للتباينات المشتركة المتساوية وللتوزيعات الطبيعية: نصنف العنصر X_0 في المجموعة G_k، إذا كان التابع التصنيفي الخطي$[d_k(X_0)]$ أكبر ما يمكن. أي إذا كان :

$$d_k(X_0) = \mathrm{Max}\left[d_j(X_0)\right] \qquad j: 1\ 2\ 3 \dots g \qquad (33-4)$$

حيث ان $d_j(X)$ تحسب من العلاقة (4−32) السابقة .

ملاحظة1: نلاحظ أن العلاقتين (4−29) و(4−32) تضمان تابع خطي لـ **X** يقوم بدور المصنف الفاصل لعناصر المجتمع حسب أكبر قيمة له عند **X** . لذلك تسمى هذه التوابع بتوابع التصنيف .

ملاحظة2: في هذه الحالة الخاصة (حالة تساوي مصفوفات التباين المشترك $\left(\boldsymbol{V_j}=\boldsymbol{V}\right)$ يمكننا أن نحصل من العلاقة (4−27) على مصنف آخر كما يلي:

نقوم بحذف الحد الثابت $\frac{1}{2}|\boldsymbol{V}|$ أو نتجاهله ثم نستبدل**V** بـ $\boldsymbol{S_P}$ في العلاقة (4−27) فنحصل على مايلي:

$$\widetilde{D}_j(X)=-\frac{1}{2}\left(\boldsymbol{X}-\overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)'*\boldsymbol{S_P^{-1}}*\left(\boldsymbol{X}-\overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)+\ln P_j \qquad (34-4)$$

وحتى يأخذ التابع $D_j(x)$ أكبر قيمة له حسب المجموعات يجب أن تكون قيمة الحد الأول أصغر ما يمكن . ولذلك نرمز لهذا الحد بالرمز $D^2(X)$ (لأنه مربع المسافة بين **X** والمركز $\overline{\boldsymbol{X}}_{\boldsymbol{j}}$)، فيكون لدينا :

$$D_j^2(X)=\left(\boldsymbol{X}-\boldsymbol{X_j}\right)'*\boldsymbol{S_P^{-1}}*\left(\boldsymbol{X}-\overline{\boldsymbol{X}}_{\boldsymbol{j}}\right) \qquad (35-4)$$

ومنها نحصل على أن:

$$\widetilde{\mathrm{D}}_j(X)=-\frac{1}{2}D^2(X)+\ln P_j \qquad (36-4)$$

بما أن المقدار $D_j^2(x)$ هو عبارة عن مربع المسافة الاحصائية من النقطة X حتى المتوسط $\bar{X}_j$، فإن هذا يعني أن النقطة X_0 تصنف ضمن المجموعةG_k الأقرب إليها، أي إننا نصنف النقطة X_0 إلى المجموعة G_k ، إذا كانت X_0 أقرب ما يمكن إلى متوسط نلك المجموعة $\bar{X}_k$.

وهكذا نجد أنه يمكننا من العلاقتين (4−35) و(4−36) صياغة قاعدة جديدة لتصنيف العناصر في حالة تساوي مصفوفات التباين المشترك $\boldsymbol{V_1}=\boldsymbol{V_2}\ldots.=\boldsymbol{V}$ كما يلي:

قاعدة أصغر المسافات عن متوسطات المجموعات عندما $\boldsymbol{V_1}=\boldsymbol{V_2}=\boldsymbol{V_3}\ldots=\boldsymbol{V}$ كما يلي:

نصنف العنصر X_0 ضمن المجموعة G_k، إذا كان مربع المسافة $D_j^2(X_0)$ أصغر ما يمكن :

$$D_k^2(X_0)=Min\left[\left(\boldsymbol{X_0}-\overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)'*\boldsymbol{S_P^{-1}}*\left(\boldsymbol{X_0}-\overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)\right]=Min\left[D_j^2(X_0)\right] \qquad (37-4)$$

أو إذا كانت قيمة التابع التصنيفي $\widetilde{\mathrm{D}}_k(X_0)$ التالي أكبر ما يمكن :

$$\widetilde{\mathrm{D}}_k(X_0)=Max\left[-\frac{1}{2}D_j^2(X_0)+\ln P_j\right]=Max\left[\mathrm{D}_j(X_0)\right] \qquad (37a-4)$$

حيث أن $j:1\ 2\ 3\ldots g$ وأن مربع المسافة $D_j^2(x)$ يحسب من العلاقة (4−35) وإن التابع التصنيفي $\widetilde{\mathrm{D}}_j(X_0)$ يحسب من العلاقة (4−36) .

وأخيراً نشير إلى أنه إذا كانت الاحتمالات السابقةP_j مجهولة أو إذا كانت متساوية $P_1=P_2=..P_g=\frac{1}{g}$، فعندها يمكن اهمال الحد الأخير $\left(\ell n\,P_j\right)$ من العلاقتين (4−32) و$(37a-4)$ لأنه يصبح ثابتاً في كل المجموعات، وتصبح عملية التصنيف معتمدة على مربع المسافة فقط $D_j^2(x)$، وعندها يتم تصنيف العناصر **X** إلى المجموعات الأقرب إلى مركزها $\bar{x}_j$ حسب (4−37) . (راجع فقرة الصيغة التربيعية والمسافات في الفصل الثاني من الجزء الأول) .

مثال (4−2): لنفترض أننا نريد إيجاد توابع التصنيف الخطية لمجتمع مؤلف من /3/ مجموعات، وتتأثر بمتحولين $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ خاضعين للتوزيع الطبيعي في R^2، ولهما مصفوفة تباين مشتركة موحدة **V**، لذلك سحبنا عينات عشوائية من هذه المجموعات $G_1\ G_2\ G_3$ بحجوم $n_1 = 3\ ,\ n_2 = 3\ ,\ n_3 = 3$ على الترتيب، فكانت أشعة قيم المتحولات ومتوسطاتها ومصفوفات التباين المشترك كما يلي:

$$\boldsymbol{X}G_1 = \begin{array}{c c} \text{المشاهدت} & \begin{array}{ccc} Q_1 & Q_2 & Q_3 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \end{array} & \begin{bmatrix} -2 & 0 & 1 \\ 5 & 3 & 1 \end{bmatrix} \end{array} \Rightarrow \overline{\boldsymbol{X}}G_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \qquad \boldsymbol{S_1} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\boldsymbol{X}G_2 = \begin{array}{c} X_1 \\ X_2 \end{array} \begin{bmatrix} 0 & 2 & 1 \\ 6 & 4 & 2 \end{bmatrix} \Rightarrow \overline{\boldsymbol{X}}G_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \qquad \boldsymbol{S_2} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\boldsymbol{X}G_3 = \begin{array}{c} X_1 \\ X_2 \end{array} \begin{bmatrix} 1 & 0 & -1 \\ -2 & 0 & -4 \end{bmatrix} \Rightarrow \overline{\boldsymbol{X}}G_3 = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \qquad \boldsymbol{S_3} = \begin{bmatrix} 1 & +1 \\ +1 & 4 \end{bmatrix}$$

ولنفترض أن الاحتمالات السابقة P_j معطية كما يلي:

$$P_1 = 0.25\ , \qquad P_2 = 0.25\ , \qquad P_3 = 0.50$$

والمطلوب تصنيف العنصر $\boldsymbol{X}_0 = \begin{bmatrix} X_{01} \\ X_{02} \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$ في المجموعة المناسبة، وذلك حسب القاعدة (4−33) لأصغر الاحتمالات TPM .

الحل: لتطبيق تلك القاعدة علينا أن نحسب التباين المشترك المدمج S_P من العلاقة (4−31) فنجد أن:

$$\boldsymbol{S_P} = \frac{(3-1)\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} + (3-1)\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} + (3-1)\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}}{9-3}$$

$$\boldsymbol{S_P} = \frac{2}{6}\begin{bmatrix} 1+1+1 & -1-1+1 \\ -1-1+1 & 4+4+4 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{3} \\ -\frac{1}{3} & 4 \end{bmatrix}$$

ومنها نحسب المقلوب $\boldsymbol{S_P^{-1}}$ فنجد أن:

$$\boldsymbol{S_P^{-1}} = \frac{9}{35}\begin{bmatrix} 4 & +\frac{1}{3} \\ +\frac{1}{3} & 1 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix}$$

ثم نحسب الجداء $\boldsymbol{\overline{X}_1'S_P^{-1}}$ الخاص بالمجموعة الأولى G_1 فنجد أن :

$$\boldsymbol{\overline{X}_1'S_P^{-1}} = (-1\ ,3) * \frac{1}{35}\begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} = \frac{1}{35}(-27\ ,24)$$

ثم نحسب الجداء $\boldsymbol{\overline{X}_1'} * \boldsymbol{S^{-1}} * \boldsymbol{\overline{X}_1}$ الخاص بالمجموعة G_1 فنجد أن :

$$\boldsymbol{\overline{X}_1'} * \boldsymbol{S_P^{-1}} * \boldsymbol{\overline{X}_1} = \frac{1}{35}(-27\ ,24)\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \frac{99}{35}$$

وبذلك نجد أن التابع التصنيفي الأول يساوي حسب (4−32) ما يلي:

$$d_1(X_0) = \overline{\boldsymbol{X}}_{\boldsymbol{1}}' * \boldsymbol{S}^{-\boldsymbol{1}} * \boldsymbol{X}_{\boldsymbol{0}} - \frac{\boldsymbol{1}}{\boldsymbol{2}}\overline{\boldsymbol{X}}_{\boldsymbol{1}}' * \boldsymbol{S}^{-\boldsymbol{1}} * \overline{\boldsymbol{X}}_{\boldsymbol{1}} + \ln P_1$$

$$d_1(X_0) = \frac{1}{35}(-27\,,24)\begin{bmatrix}X_{01}\\X_{02}\end{bmatrix} - \frac{1}{2} * \frac{99}{35} + \ln(0.25)$$

$$d_1(X_0) = \left(\frac{-27}{35}\right)X_{01} + \left(\frac{24}{35}\right)X_{02} - \frac{99}{70} - 1.386 \qquad (a)$$

وبطريقة مشابهة نحسب التابع التصنيفي الثاني لـ G_2 فنجد أن:

$$\overline{\boldsymbol{X}}_{\boldsymbol{2}}' * \boldsymbol{S}_{\boldsymbol{P}}^{-\boldsymbol{1}} = (1\,,4)\frac{1}{35}\begin{bmatrix}36 & 3\\3 & 9\end{bmatrix} = \frac{1}{35}(48, 39)$$

$$\overline{\boldsymbol{X}}_{\boldsymbol{2}}' * \boldsymbol{S}_{\boldsymbol{P}}^{-\boldsymbol{1}} * \overline{\boldsymbol{X}}_{\boldsymbol{2}} = \frac{1}{35}(48, 39) * \begin{bmatrix}1\\4\end{bmatrix} = \frac{204}{35}$$

وهكذا نجد من العلاقة (4−32) أن:

$$d_2(X_0) = \frac{1}{35}(48, 39)\begin{bmatrix}X_{01}\\X_{02}\end{bmatrix} - \frac{1}{2}\frac{204}{35} + \ln(0.25)$$

$$d_2(X_0) = \left(\frac{48}{35}\right)X_{01} + \left(\frac{39}{35}\right)X_{02} - \frac{204}{70} - 1.386 \qquad (b)$$

ولحساب التابع التصنيفي الثالث لـ G_3 فنجد أن:

$$\overline{\boldsymbol{X}}_{\boldsymbol{3}}' * \boldsymbol{S}_{\boldsymbol{P}}^{-\boldsymbol{1}} = (0\,,-2)\frac{1}{35}\begin{bmatrix}36 & 3\\3 & 9\end{bmatrix} = \frac{1}{35}(-6, -18)$$

$$\overline{\boldsymbol{X}}_{\boldsymbol{3}}' * \boldsymbol{S}_{\boldsymbol{P}}^{-\boldsymbol{1}} * \overline{\boldsymbol{X}}_{\boldsymbol{3}} = \frac{1}{35}(-6, -18) * \begin{bmatrix}0\\-2\end{bmatrix} = \frac{36}{35}$$

وهكذا نجد من العلاقة (4−32) أن:

$$d_3(X_0) = \frac{1}{35}(-6\,,-18)\begin{bmatrix}X_{01}\\X_{02}\end{bmatrix} - \frac{1}{2} * \frac{36}{35} + \ln(0.5)$$

$$d_3(X_0) = \left(\frac{-6}{35}\right)X_{01} + \left(\frac{-18}{35}\right)X_{02} - \frac{36}{70} - 0.693 \qquad (C)$$

ولتحديد انتماء العنصر $X_0 = \begin{bmatrix}X_{01}\\X_{02}\end{bmatrix} = \begin{bmatrix}-2\\-1\end{bmatrix}$، نعوض قيم ذلك العنصر في كل من التوابع التصنيفية (a) و(b) و(C) السابقة ونحسب قيمته العددية فنجد أن:

$$d_1(X_0) = \left(\frac{-27}{35}\right)(-2) + \left(\frac{24}{35}\right)(-1) - \frac{99}{70} - 1.386 = -1.943$$

$$d_2(X_0) = \left(\frac{48}{35}\right)(-2) + \left(\frac{39}{35}\right)(-1) - \frac{204}{70} - 1.386 = -8.158$$

$$d_3(X_0) = \left(\frac{-6}{35}\right)(-2) + \left(\frac{-18}{35}\right)(-1) - \frac{36}{70} - 0.693 = -0.350$$

وبمقارنة هذه القيم نجد أن أكبرها هي $d_3(X_0) = -0.350$ ولذلك فإننا نصنف العنصر $X_0 = \begin{bmatrix}-2\\-1\end{bmatrix}$ ضمن المجموعة G_3 .

- **طريقة ثانية سنقوم بحل هذا المثال** بطريقة أصغر المسافات بين العنصر X_0 ومراكز المجموعات.

لتطبيق هذه الطريقة نستخدم العلاقة (4-35) لحساب مربعات المسافات للعنصر $X_0 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$ عن مراكز المجموعات، مع افتراض أن الاحتمالات السابقة متساوية $P_1 = P_2 = P_3 = \frac{1}{3}$، فنجد أن مربع المسافة لـx_0 عن مركز المجموعة الأولى $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ ، يساوي حسب العلاقة (4-35) ما يلي:

$$D_1^2(X_0) = (X_0 - \bar{X}_1)' * S_P^{-1} * (X_0 - \bar{X}_1) = [(-2+1), (-1-3)] \frac{1}{35} \begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} (-2+1) \\ (-1-3) \end{bmatrix}$$

$$D_1^2(X_0) = \frac{1}{35} (-1, -4) \begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} -1 \\ -4 \end{bmatrix} = \frac{1}{35} (-1, -4) \begin{bmatrix} -48 \\ -39 \end{bmatrix} =$$

$$D_1^2(X_0) = \frac{1}{35} [48 + 156] = \frac{204}{35} = 5.8286$$

وكذلك نجد أن مربع المسافة لـx_0 عن مركز المجموعة الثانية $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ ، يساوي:

$$D_2^2(X_0) = (X_0 - \bar{X}_2)' * S^{-1} * (X_0 - \bar{X}_2) = [(-2-1), (-1-4)] \frac{1}{35} \begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} (-2-1) \\ (-1-4) \end{bmatrix}$$

$$D_2^2(X_0) = \frac{1}{35} (-3, -5) \begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} -3 \\ -5 \end{bmatrix} = \frac{1}{35} (-3, -5) \begin{bmatrix} -123 \\ -54 \end{bmatrix} =$$

$$D_2^2(X_0) = \frac{1}{35} [369 + 270] = \frac{639}{35} = 18.257$$

ولإيجاد مربع المسافة لـx_0 عن مركز المجموعة الثالثة $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$ نجد أن:

$$D_3^2(X_0) = (X_0 - \bar{X}_3)' * S_P^{-1} * (X_0 - \bar{X}_3) = [(-2-0), (-1+2)] \frac{1}{35} \begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} (-2-0) \\ (-1+2) \end{bmatrix}$$

$$D_3^2(X_0) = \frac{1}{35} (-2, 1) \begin{bmatrix} 36 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{1}{35} (-2, 1) \begin{bmatrix} -69 \\ 3 \end{bmatrix} =$$

$$D_3^2(X_0) = \frac{1}{35} [138 + 3] = \frac{141}{35} = 4.02857$$

وعند مقارنة مربعات هذه المسافات نجد أن أصغرها هو $D_3^2(x_0) = 4.02857$، لذلك نصنف العنصر $x_0 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$ في المجموعة الثالثة G_3 (وهو نفس الجواب السابق) .

ملاحظة: هناك صيغة أخرى لقاعدة التصنيف (4-33) . ونحصل عليها من مقارنة التوابع $d_j(X)$ مثنى مثنى وتحديد أكبرها وليكن $d_k(X)$.

وعندها فإن شرط أن يكون $d_k(X)$ أكبر التوابع $d_1(X), d_2(X), \dots, d_g(X)$ وهو أن يكون:

$$d_k(X) - d_j(X) \geq 0 \qquad For\ :\ j: 1\ 2\ 3 \dots g \qquad (38-4)$$

وبتعويض $d_k(X)$ و $d_j(X)$ **من العلاقة** (4-29) **نجد أن (بفرض** $V_1 = V_2$) **الشرط** (4-38) **يأخذ الشكل التالي:**

$$\mu_k' * V^{-1} * X - \frac{1}{2}\mu_k' * V^{-1} * \mu_k + \ln P_k - \mu_j' * V^{-1} * X + \frac{1}{2}\mu_j' * V^{-1} * \mu_j - \ln P_j \geq 0 \quad (39-4)$$

وبعد الاصلاح ووضع $\left(\ln P_k - \ln P_j = \ln\left(\frac{P_k}{P_j}\right) = -\ln\left(\frac{P_j}{P_k}\right)\right)$ ونقلها إلى الطرف الثاني، نحصل على الصيغة البديلة لقاعدة التصنيف التي تصغر الاحتمال الاجمالي للتصنيف الخاطئ *(TPM)* وهي التي تأخذ الشكل التالي:

نصنف العنصر X_0 في المجموعة G_k إذا كان:

$$d_{kj}(X_0) = (\mu_k - \mu_j)' * V^{-1} * X_0 - \frac{1}{2}(\mu_k - \mu_j)' * V^{-1} * (\mu_k - \mu_j) \geq \ln\left(\frac{P_j}{P_k}\right) \quad (\mathbf{40-4})$$

وذلك من أجل جميع قيم $j: 1\ 2\ 3 \dots g$

وإذا رمزنا للطرف الأيسر من العلاقة $(4-40)$ بالرمز $d_{kj}(X_0)$ فإن الشرط السابق $(4-40)$ يصبح كما يلي:

نصنف العنصر x_0 في المجموعة G_k إذا كان :

$$d_{kj}(X_0) \geq \ln\left(\frac{P_j}{P_k}\right) \qquad j: 1\ 2\ 3 \dots g \qquad (41-4)$$

حيث $d_{kj}(x)$ تحسب من $(4-40)$.

إن الشروط الأخيرة $(4-41)$، هي التي ستحدد لنا مناطق التصنيف $R_1, R_2, R_3, \dots, R_g$، والتي تفصلها عن بعضها بواسطة مستويات فوقية في الفضاء R^P، وذلك لأن $d_{kj}(x)$ هو عبارة عن تركيب خطي من مركبات X . فعلى سبيل المثال نجد أنه عندما يكون لدينا ثلاث مجموعات $(g = 3)$، فإن منطقة التصنيف الأولى R_1 تتألف من جميع قيم X التي تحقق العلاقات التالية:

$$R_1: d_{1j} \geq \ln\left(\frac{P_j}{P_1}\right) \qquad (j = 2, 3) \qquad (42-4)$$

أي أن المنطقة R_1 في الفضاء R^3 تتألف من قيم X التي تحقق معاً المتراجحتين التاليتين :

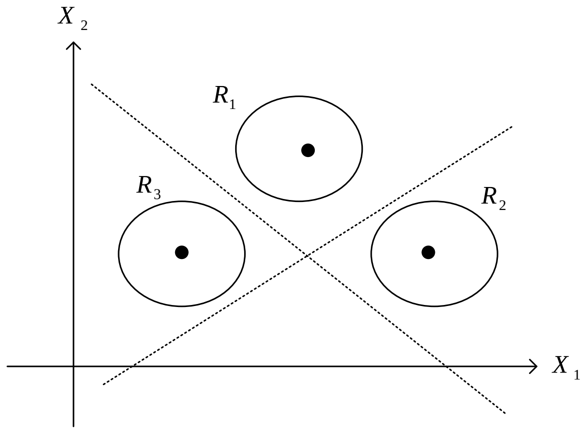
$$d_{12}(x) = (\mu_1 - \mu_2)' * V^{-1} * X - \frac{1}{2}(\mu_1 - \mu_2)' * V^{-1} * (\mu_1 - \mu_2) \geq \ln\left(\frac{P_2}{P_1}\right) \quad (43-4)$$

$$d_{13}(x) = (\mu_1 - \mu_3)' * V^{-1} * X - \frac{1}{2}(\mu_1 - \mu_3)' * V^{-1} * (\mu_1 - \mu_3) \geq \ln\left(\frac{P_3}{P_1}\right) \quad (44-4)$$

وبفرض أن المتوسطات: μ_1, μ_2, μ_3 لاتقع على خط مستقيم . فإن المعادلة $d_{12}(x) = \ln\left(\frac{P_2}{P_1}\right)$ والمعادلة $d_{13} = \ln\left(\frac{P_3}{P_1}\right)$ تحددان مستويين فوقيين ومتقاطعين ويشكلان فيما بينهما منطقة التصنيف الأولى R_1 في الفضاء R^3 .

وهنا نلاحظ أن الحد $\ln\left(\frac{P_2}{P_1}\right)$ يجعل المستوى أقرب إلى μ_1 من μ_2 إذا كان P_2 أكبر من P_1 .

والشكل التالي يوضح مثال لمناطق التصنيف الثلاثة R_1, R_2, R_3 لحالة ثلاثة مجموعات ومتحولين X_1 و X_2 .



الشكل (4−1) مناطق التصنيف لمتحولين وثلاثة مجموعات

وعندما تكون معالم المجموعات μ_1 , μ_2 , μ_3 وV مجهولة فإننا نقدرها من عينات عشوائية مسحوبة منها بواسطة المؤشرات $\bar{x}_1 , \bar{x}_2 , \bar{x}_3$ و S_P (المدمج)، وتصبح قاعدة التصنيف (4−40) في حالة العينات كمايلي:

نصنف العنصر X_0 في المجموعة G_k إذا كان :
$\tilde{\mathrm{d}}_{kj}(X_0) = (\bar{X}_k - \bar{X}_j)' * S_P^{-1} * X_0 - \frac{1}{2}(\bar{X}_k - \bar{X}_j)' * S_P^{-1} * \left((\bar{X}_k - \bar{X}_j)\right) \geq \ln\left(\frac{P_j}{P_k}\right) \quad (45-4)$
وذلك من أجل $(j \neq k)$, $j: 1\ 2\ 3 \ldots g$ $\qquad d_{ki}(x_0) \geq \ell n\left(\frac{P_j}{P_k}\right)$

وبتعويض بيانات العينة $\bar{x}_j$ و S_P في (4−45) نحصل على توابع التميز الخطية بدلالة مركبات X ويكون عددها (g−1) تابعاً لأن $(j \neq k)$، وإن هذه التوابع تحدد لنا مناطق التصنيف R_1 , R_2 , R_3 حسب العلاقات السابقة وكما هو مبين على الشكل (4−1) السابق .

ملاحظة: إذا كانت الاحتمالات P_j مجهولة أو يصعب الحصول عليها. فإننا نعتبرها جميعاً متساوية $P_1 = P_2 = \cdots\ P_g = \frac{1}{g}$، وعندها فإن الطرف الأيمن من العلاقة (4−45) يصبح معدوماً لجميع الأزواج، أي يصبح $\ln\left(\frac{P_i}{P_1}\right) = 0$ لجميع قيم (j , k)، وعندها تأخذ العلاقة (4−45) الشكل التالي:
نصنف X_0 في G_k إذا كان: $d_{kj}(X_0) \geq 0$.

- **طريقة العزل أو الاستبعاد (*Holdout*) أو أسلوب (*Lachenbroch*)** : إن استخدامنا لتقديرات معالم المجموعات يجعل قاعدة التصنيف في حالة العينات غير مثالية، وحتى نحسن من التمثيل في هذه الحالة نقوم بتطبيق أسلوب (لاشينبروش *Lachenbroch*) أو أسلوب العزل (*holdout*) والذي يلخص بما يلي:

1− نبدأ بعناصر المجموعة الأولى G_1 ونحذف أحد المشاهدات (العناصر) منها، ثم نقوم بإجراء التحليل التمييزي على العناصر المتبقية منها $(n_1 - 1)$ مع عناصر المجموعات الأخرى فنحصل على توابع تمييز معينة .

2- نعود ونعمل على تصنيف العنصر الذي حذفناه من G_1 في المجموعة المناسبة له باستخدام التوابع التي حصلنا عليها في الخطوة (1) .

3- نكرر الخطوتين (1) و(2) حتى تشمل جميع عناصر المجموعة G_1 ونحصل على تصنيفاتها الجديدة ولنفترض أن عدد التصنيفات الخاطئة في G_1 كان: n_{1m}^H عنصراً .

4- نكرر الخطوات (1) و(2) و(3) على عناصر جميع المجموعات الأخرى G_2 وG_3 … G_g، فنحصل على تصنيفات جديدة لها ولنفترض أن عدد العناصر المصنفة خطأً في G_2 هو n_{2m}^H .

وعندها نقدر احتمالات التصنيف الخاطئ في كل مجموعة كما يلي:

$$\tilde{P}(2/1) = \frac{n_{1m}^H}{n_1} \qquad (46-4)$$

$$\tilde{P}(1/2) = \frac{n_{2m}^H}{n_2} \qquad (47-4)$$

وهكذا نتابع العمل مع باقي المجموعات .

وأخيراً يمكننا أن نحصل على تقدير غير متحيز لمعدل الخطأ الفعلي المتوقع ونحسبه من العلاقة :

$$\tilde{E}(AER) = \frac{n_{1m}^H + n_{2m}^H + n_{3m}^H + \cdots}{n_1 + n_2 + n_3 + \cdots} = \frac{n_m}{n} \qquad (48-4)$$

مثال (4-3): سنقوم بتطبيق أسلوب العزل (لاشينبروش) على مجموعتين ومتحولين لتقدير معدل التصنيف الخاطئ، وذلك بفرض أن الاحتمالات السابقة متساوية، وأن تكاليف التصنيف الخاطئ متساوية. ولنفترض إننا سحبنا من المجموعتين G_1 وG_2 عينتين بحجمين $n_1 = 3$ و $n_2 = 3$، وأخذنا قياسات المتحولين X_1 وX_2 فيهما. فكانت بيانات المشاهدات ومتوسطاتها ومصفوفة التباين المشترك لهما كمايلي: [*Johnson R. A. P. 499*] .

$$\boldsymbol{X}G_1 = \begin{array}{c} \begin{matrix} Q_1 & Q_2 & Q_3 \end{matrix} \\ \begin{bmatrix} 2 & 4 & 3 \\ 12 & 10 & 8 \end{bmatrix} \end{array} \qquad \bar{\boldsymbol{X}}_1 = \begin{bmatrix} 3 \\ 10 \end{bmatrix} \qquad \boldsymbol{S_1} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\boldsymbol{X}G_2 = \begin{bmatrix} 5 & 3 & 4 \\ 7 & 9 & 5 \end{bmatrix} \qquad \bar{\boldsymbol{X}}_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix} \qquad \boldsymbol{S_2} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

ومن المصفوفتين S_1 وS_2 نشكل مصفوفة التباين المشترك المدمج كما يلي:

$$\boldsymbol{S_P} = \frac{(3-1)\boldsymbol{S_1} + (3-1)\boldsymbol{S_2}}{4} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

وعند تطبيق قاعدة التصنيف (4-26)على هذه البيانات في المثال(4-2) وجدنا أنها تعطينا التصنيف التالي:

جدول (4-2): نتائج التصنيف المقارن

	المجموعات	التصنيف الجديد G_1	التصنيف الجديد G_2	المجموع
التصنيف الأصلي	G_1	2	1	3
	G_2	1	2	3

ومنه نجد أن تقدير معدل الخطأ الظاهري يساوي:

$$APER = \frac{1+1}{6} = 0.33$$

وإن هذا التقدير لمعدل الخطأ يسمى ظاهرياً، لأن القيمة الحقيقية لمعدل الخطأ، قد تختلف عنه بسبب حجوم العينات وأسلوب سحبها. وللكشف عن حقيقة هذا المعدل نقوم بإجراء التحليل التمييزي بأسلوب العزل *Holdout* (لاشينبروش) كما يلي:

1- لنبدأ من المجموعة الأولى G_1 ونختار العنصر الأول $\begin{bmatrix} 2 \\ 12 \end{bmatrix}$ من G_1 ونرمز له بـ $\boldsymbol{X}_{11} = \begin{bmatrix} 2 \\ 12 \end{bmatrix}$

ونعزله عنها ثم نطبق إجراءات التحليل على العناصر المتبقية في المجموعتين فنجد أن:

$$\boldsymbol{XG}_1^{1H} = \begin{bmatrix} 4 & 3 \\ 10 & 8 \end{bmatrix} \qquad \overline{\boldsymbol{X}}_{1H} = \begin{bmatrix} 3,5 \\ 9 \end{bmatrix} \qquad \boldsymbol{S_{1H}} = \begin{bmatrix} 0.5 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\boldsymbol{XG}_2 = \begin{bmatrix} 5 & 3 & 4 \\ 7 & 9 & 5 \end{bmatrix} \qquad \overline{\boldsymbol{X}}_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix} \qquad \boldsymbol{S_2} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

ومنها نجد أن مصفوفة التباين المدمج الجديدة S_P تساوي :

$$\boldsymbol{S_P} = \frac{\boldsymbol{S}_{1H} + 2\boldsymbol{S_2}}{3} = \frac{1}{3}\begin{bmatrix} 0.5 & 1 \\ 1 & 2 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2.5 & -1 \\ -1 & 10 \end{bmatrix}$$

وأن المقلوب $\boldsymbol{S_P^{-1}}$ يساوي :

$$\boldsymbol{S_P^{-1}} = \frac{1}{8}\begin{bmatrix} 10 & +1 \\ +1 & 2.5 \end{bmatrix}$$

وعلينا الآن أن نقوم بتصنيف العنصر المعزول $\boldsymbol{X_{11}} = \begin{bmatrix} 2 \\ 10 \end{bmatrix}$ في المجموعة المناسبة .

إن أسرع طريقة لتصنيف العنصر المعزول هي طريقة أصغر مربع للمسافة بين ذلك العنصر والمركز الجديد للمجموعة الأولى $\overline{\boldsymbol{X}}_{1H}$ ومركز المجموعة الثانية $\overline{\boldsymbol{X}}_{\boldsymbol{2}}$.

والآن نحسب مربع تلك المسافة الاحصائية لـ X_1 عن المتوسط الجديد للمجموعة الأولى من العلاقة (4-35) التالية :

$$D_1^2(\boldsymbol{X}_{11}) = (\boldsymbol{X_{11}} - \overline{\boldsymbol{X}}_{\boldsymbol{1H}})' * \boldsymbol{S_P^{-1}} * (\boldsymbol{X_{11}} - \overline{\boldsymbol{X}}_{\boldsymbol{1H}})$$

فنجد أن:

$$D_1^2(\boldsymbol{X}_{11}) = [2-3.5\,, 12-9]\tfrac{1}{8}\begin{bmatrix} 10 & 1 \\ 1 & 2.5 \end{bmatrix}\begin{bmatrix} 2-3.5 \\ 12-9 \end{bmatrix} = 4.5$$

ثم نحسب مربع تلك المسافة لـ X_{11} عن المتوسط الأصلي للمجموعة الثانية من العلاقة :

$$D_2^2(\boldsymbol{X}_{11}) = (\boldsymbol{X_{11}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})' * \boldsymbol{S_P^{-1}}(\boldsymbol{X_{11}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})$$

$$D_2^2(\boldsymbol{X}_{11}) = [2-4\,, 12-7]\frac{1}{8}\begin{bmatrix} 10 & 1 \\ 1 & 2.5 \end{bmatrix}\begin{bmatrix} 2-4 \\ 12-7 \end{bmatrix} = 10.3$$

وعند مقارنة مربعي هاتين المسافتين نجد أن $D_1^2(\boldsymbol{X}_{11}) < D_2^2(\boldsymbol{X}_{11})$، أي أن العنصر المعزول $X_{11} = \begin{bmatrix} 2 \\ 10 \end{bmatrix}$ هو أقرب إلى المتوسط الجديد للمجموعة الأولى $\overline{\boldsymbol{X}}_{1H} = \begin{bmatrix} 3.5 \\ 9 \end{bmatrix}$ ، منه إلى متوسط المجموعة الثانية $\overline{\boldsymbol{X}}_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$، لذلك نصنف العنصر المعزول $\boldsymbol{X}_{11} = \begin{bmatrix} 2 \\ 10 \end{bmatrix}$ في المجموعة G_1 . علماً بأن هذا التصنيف هو تصنيف صحيح لأن $\boldsymbol{X}_{11}$ هو أصلاً من G_1 .

ثم ننتقل إلى عزل العنصر الثاني من G_1 وهو $\boldsymbol{X}_{12}=\begin{bmatrix}4\\10\end{bmatrix}$ ونقوم بحساب مربع المسافة منه حتى متوسطي المجموعتين . فنجد أن بيانات المجموعة الأولى الجديدة ومقلوب المصفوفةS_P تصبح كما يلي:

$$\boldsymbol{X}G_1^{2H}=\begin{bmatrix}2&3\\12&8\end{bmatrix}\ ,\ \ \overline{\boldsymbol{X}}_{2H}=\begin{bmatrix}2.5\\10\end{bmatrix}\ \ ,\ \ \boldsymbol{S_P^{-1}}=\frac{1}{8}\begin{bmatrix}16&4\\4&2.5\end{bmatrix}$$

ومنها نجد أن مربع المسافة لـ X_{12} عن المتوسط الجديد للمجموعة الأولى يساوي:

$$D_1^2(\boldsymbol{X}_{12})=(\boldsymbol{X_{12}}-\overline{\boldsymbol{X}}_{\boldsymbol{2H}})'*\ \boldsymbol{S_P^{-1}}*(\boldsymbol{X_{12}}-\overline{\boldsymbol{X}}_{\boldsymbol{2H}})=$$

$$D_1^2(\boldsymbol{X}_{12})=[4-2.5\ ,10-10]\frac{1}{8}\begin{bmatrix}16&4\\4&2.5\end{bmatrix}\begin{bmatrix}4-2.5\\10-10\end{bmatrix}=4.5$$

وكذلك نجد أن مربع المسافة لـ X_{12} عن المتوسط الأصلي للمجموعة الثانية يساوي :

$$D_2^2(\boldsymbol{X}_{12})=(\boldsymbol{X_{12}}-\overline{\boldsymbol{X}}_{\boldsymbol{2}})'*\ \boldsymbol{S_P^{-1}}(\boldsymbol{X_{12}}-\overline{\boldsymbol{X}}_{\boldsymbol{2}})=$$

$$D_2^2(\boldsymbol{X}_{12})=[4-4\ ,10-7]\frac{1}{8}\begin{bmatrix}16&4\\4&2.5\end{bmatrix}\begin{bmatrix}4-4\\10-7\end{bmatrix}=2.8$$

وبمقارنة مربعي هاتين المسافتين نجد أن $D_1^2(\boldsymbol{X}_{12})>D_2^2(\boldsymbol{X}_{12})$، وهذا يعني أن مربع المسافة من العنصر المعزول X_{12} إلى متوسط المجموعة الثانية $\bar{X}_{12}$، أصغر منه إلى المتوسط الجديد للمجموعة الأولى $\bar{X}_{2H}$. لذلك نصنف العنصر المعزول $\boldsymbol{X}_{12}=\begin{bmatrix}4\\10\end{bmatrix}$ ضمن المجموعة الثانية G_2، وهو تصنيف خاطئ لأنه أصلاً من G_1 .

ثم ننتقل إلى العنصر الثالث من G_1 وهو $\boldsymbol{X}_{13}=\begin{bmatrix}3\\8\end{bmatrix}$ ونعزله منها ثم نقوم بإجراء الحسابات اللازمة لتصنيفه في إحدى المجموعتين. فنجد أن العنصر المعزول سيصنف في المجموعة الثانية G_2 وهو تصنيف خاطئ أيضاً، لأنه أصلا من G_1 . وهكذا نجد أن عدد العناصر المصنفة خطأ من G_1 في المجموعة الثانية G_2 أصبح يساوي: $n_{1M}^H=2$.

والآن ننتقل إلى المجموعة الثانية G_2 ونقوم بعزل العنصر الأول منها وهو $\boldsymbol{X}_{21}=\begin{bmatrix}5\\7\end{bmatrix}$ فيبقى لدينا فيها عنصران هما:

$$\boldsymbol{X}G_2^{2H}=\begin{bmatrix}3&4\\9&5\end{bmatrix},\ \ \overline{\boldsymbol{X}}_{2H}=\begin{bmatrix}3.5\\7\end{bmatrix},\ \ \boldsymbol{S}_{2H}=\frac{1}{8}\begin{bmatrix}0.5&-2\\-2&8\end{bmatrix}$$

ثم نحسب مصفوفة التباين المدمج فنجد أن:

$$\boldsymbol{S_P}=\frac{1}{3}[2\boldsymbol{S_1}+\boldsymbol{S_2}]=\frac{2}{3}\begin{bmatrix}1&-1\\-1&4\end{bmatrix}+\begin{bmatrix}0.5&-2\\-2&8\end{bmatrix}=\frac{1}{3}\begin{bmatrix}2.5&-4\\-4&10\end{bmatrix}$$

وإن مقلوبها يساوي:

$$\boldsymbol{S_P^{-1}}=\tfrac{3}{24}\begin{bmatrix}16&4\\4&2.5\end{bmatrix}$$

وإن مربع المسافة للعنصر المعزول $\boldsymbol{X}_{21}=\begin{bmatrix}5\\7\end{bmatrix}$ عن المتوسط الأصلي للمجموعة الأولى $\overline{\boldsymbol{X}}_1$ يساوي:

$$D_1^2(\boldsymbol{X}_{21})=(5-3\ ,7-10)\frac{3}{24}\begin{bmatrix}16&4\\4&2.5\end{bmatrix}*\begin{bmatrix}5-3\\7-10\end{bmatrix}=4.8$$

وأن مربع المسافة للعنصر المعزول $X_{21}=\begin{bmatrix}5\\7\end{bmatrix}$ عن المتوسط الجديد للمجموعة الثانية X_{2H} يساوي:

$$D_2^2(\boldsymbol{X}_{21})=(5-3.5\ \ ,7-7)\frac{3}{24}\begin{bmatrix}16&4\\4&2.5\end{bmatrix}*\begin{bmatrix}5-3.5\\7-7\end{bmatrix}=4.5$$

وبمقارنة مربعي هاتين المسافتين نجد أن $D_2^2(\boldsymbol{X}_{21}) < D_1^2(\boldsymbol{X}_{21})$، وهذا يعني أن العنصر المعزول $\boldsymbol{X}_{21} = \begin{bmatrix}5\\7\end{bmatrix}$ سيصنف في المجموعة G_2، وهو تصنيف صحيح لأنه أصلاً من G_2 .

ثم نقوم بعزل العنصر الثاني من G_2 وهو العنصر $\boldsymbol{X}_{22} = \begin{bmatrix}3\\9\end{bmatrix}$ فنجد أن:

$$\boldsymbol{X}G_2^{2H} = \begin{bmatrix}5 & 4\\7 & 5\end{bmatrix} \quad , \quad \overline{\boldsymbol{X}}_{2H} = \begin{bmatrix}4.5\\6\end{bmatrix} \quad \boldsymbol{S_P^{-1}} = \frac{3}{24}\begin{bmatrix}10 & 1\\1 & 2.5\end{bmatrix}$$

وبذلك نجد أن مربع المسافة للعنصر $\boldsymbol{X}_{22} = \begin{bmatrix}3\\9\end{bmatrix}$ عن المتوسط الأصلي للمجموعة الأولى $\overline{\boldsymbol{X}}_1$ يساوي :

$$D_1^2(\boldsymbol{X}_{22}) = (3-3\ ,9-10)\frac{3}{24}\begin{bmatrix}10 & 1\\1 & 2.5\end{bmatrix} * \begin{bmatrix}3-3\\9-10\end{bmatrix} = 0.33$$

وإن مربع المسافة للعنصر $X_{22} = \begin{bmatrix}3\\9\end{bmatrix}$ عن المتوسط الجديد للمجموعة الثانية يساوي :

$$D_2^2(\boldsymbol{X}_{22}) = (3-4.5\ ,9-6)\frac{3}{24}\begin{bmatrix}10 & 1\\1 & 2.5\end{bmatrix} * \begin{bmatrix}3-4.5\\9-6\end{bmatrix} = 4.5$$

وبمقارنة مربعي هاتين المسافتين نجد أن: $D_1^2(\boldsymbol{X}_{22}) < D_2^2(\boldsymbol{X}_{22})$، وهذا يعني أن العنصر $\boldsymbol{X}_{22} = \begin{bmatrix}3\\9\end{bmatrix}$ سيصنف في المجموعة G_1 وهذا تصنيف خاطئ لأنه أصلاً من G_2 .

ثم ننتقل إلى العنصر الثالث من G_2 وهو $\boldsymbol{X}_{23} = \begin{bmatrix}4\\5\end{bmatrix}$ ونجري الحسابات اللازمة المشابهة لما سبق .

فنجد أنها ستقودنا إلى تصنيف العنصر $\boldsymbol{X}_{23} = \begin{bmatrix}4\\5\end{bmatrix}$ ضمن المجموعة G_2، وهو تصنيف صحيح لأنه أصلاً من G_2 . وبذلك نجد أن عدد العناصر المصنفة خطأ من G_2 في G_1 يساوي $N_{2M}^H = 1$.

ومنه يمكننا أن نقدر المعدل الفعلي المتوقع لخطأ التصنيف من خلال العلاقة :

$$E(AER) = \frac{N_{1M}^H + N_{2M}^H}{n_1 + n_2} = \frac{2+1}{3+3} = 0.50$$

وهكذا نجد أن تقديرنا لمعدل الخطأ الظاهري المساوي لـ $ADER = 0.33$ كان تقديراً متفائل لقياس الأداء وإن الفرق بينهما ملحوظ، ولكن هذا الفرق يعود إلى صغر حجوم العينات، وإن الفرق بين هذين المقياسين سيصبح أصغر كلما أصبحت حجوم العينات كبيرة .

وأخيراً نشير إلى أن تطبيق أسلوب العزل ثم التصنيف يحتاج إلى حسابات معقدة وطويلة. وخاصة إذا كان عدد المجموعات أكثر من $g > 2$، أو كان عدد المتحولات X أكثر من $(P > 2)$، ولذلك تستخدم برامج حاسوبية خاصة بذلك .

3-4 : طريقة (فيشر) للتحليل التمييزي المتعدد (لعدة مجموعات ولعدة متحولات) :

[المصدر : Johnston,Wichern. P.514 بتصرف وإضافة] .

تعتبر هذه الحالة تعميماً لطريقة (فيشر) للتحليل التمييزي في حالة مجموعتين، وإن الهدف الأساسي لهذا التعميم هو الحاجة الماسة إلى الفصل بين المجموعات في المجتمع المدروس بواسطة أقل عدد ممكن من التراكيب الخطية للمتحولات المستقلة X مثل:

$$Y_1 = \boldsymbol{\ell}_1'\boldsymbol{X} \qquad Y_2 = \boldsymbol{\ell}_2'\boldsymbol{X} \qquad Y_3 = \boldsymbol{\ell}_3'\boldsymbol{X} \dots\dots \qquad (49-4)$$

كما يهدف إلى إيجاد أداة رياضية لتصنيف العناصر المشاهدة وتوزيعها على المجموعات بأقل خطأ ممكن .
وهنا نشير إلى أن هذه الطريقة لا تشترط أن تكون بيانات المتحولات المستقلة X في المجموعات خاضعة للتوزيع الطبيعي المتعدد . ولكنها تشترط أن تكون مصفوفات التباين المشترك لها في المجموعات متساوية وأن تكون رتبها كاملة (من الرتبة P). أي تشترط أن يكون :

$$\boldsymbol{V_1} = \boldsymbol{V_2} = \boldsymbol{V_3} = \cdots = \boldsymbol{V} \qquad (50-4)$$

وأن تكون رتبة (rank) كل منها تساوي p (لأن V من المرتبة ($p \times p$)، وهذا يعني أن تكون V نظامية (غير شاذة) .

والآن لنفترض أن المجتمع المدروس مؤلف من g مجموعة منفصلة هي:

$$G_1, G_2, G_3, \ldots, G_g \qquad (51-4)$$

وأن المتحولات المستقلة X التي تؤثر على تمييزها هي:

$$X_1, X_2, X_3, \ldots, X_p \qquad (52-4)$$

ولنفترض أن متوسط هذه المتحولات في المجموعة G_j هو $\boldsymbol{\mu_j}$ حيث $\boldsymbol{\mu}_j = E\left(\boldsymbol{X}/G_j\right)$.

وأن المتوسط العام لهذه المتوسطات هو $\overline{\boldsymbol{\mu}}$ وهو يحسب من العلاقة:

$$\overline{\boldsymbol{\mu}} = \frac{1}{g}\sum_{j=1}^{g}\boldsymbol{\mu_j} \qquad \mathbf{(53-4)}$$

ولنرمز الآن بـ B لمصفوفة مجاميع مربعات الانحرافات SSB المتقاطعة لمتوسطات المجموعات $\boldsymbol{\mu_j}$ عن المتوسط العام $\overline{\boldsymbol{\mu}}$، وهي التي تحسب من العلاقة التالية:

$$B = \sum_{j=1}^{g}\left(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}\right)' \qquad \mathbf{(54-4)}$$

$$\boldsymbol{B_{p*p}} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{g}}\begin{bmatrix}\boldsymbol{\mu_{j1}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{1}}\\ \boldsymbol{\mu_{j2}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{2}}\\ \cdots\cdots\cdots\cdots\\ \boldsymbol{\mu_{jp}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{p}}\end{bmatrix} * \left[(\boldsymbol{\mu_{j1}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{1}}), (\boldsymbol{\mu_{j2}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{2}}), \ldots (\boldsymbol{\mu_{jp}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{p}})\right] =$$

$$\boldsymbol{B} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{g}}\begin{bmatrix}\left(\boldsymbol{\mu_{j1}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{1}}\right)^{\boldsymbol{2}}, \left(\boldsymbol{\mu_{j1}} - \boldsymbol{\mu_1}\right)\left(\boldsymbol{\mu_{j2}} - \boldsymbol{\mu_2}\right)\cdots\cdots\cdots\left(\boldsymbol{\mu_{j1}} - \boldsymbol{\mu_1}\right)\left(\boldsymbol{\mu_{jP}} - \boldsymbol{\mu_P}\right)\\ \left(\boldsymbol{\mu_{j2}} - \boldsymbol{\mu_2}\right)\left(\boldsymbol{\mu_{j1}} - \boldsymbol{\mu_1}\right), \left(\boldsymbol{\mu_{j2}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{2}}\right)^{\boldsymbol{2}}\cdots\cdots\cdots\left(\boldsymbol{\mu_{j2}} - \boldsymbol{\mu_2}\right)\left(\boldsymbol{\mu_{jP}} - \boldsymbol{\mu_P}\right)\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \left(\boldsymbol{\mu_{jp}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{p}}\right)\left(\boldsymbol{\mu_{j1}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{1}}\right), \left(\boldsymbol{\mu_{jp}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{p}}\right)\left(\boldsymbol{\mu_{j2}} - \overline{\boldsymbol{\mu}}_{\boldsymbol{2}}\right)\cdots\cdots\cdots\left(\boldsymbol{\mu_{jP}} - \boldsymbol{\mu_P}\right)^{\boldsymbol{2}}\end{bmatrix} \quad \mathbf{(55-4)}$$

أي أن B هي مجموع مصفوفات مربعات الانحرافات الخاصة بكل مجموعةG_j، وهي من المرتبة$p * p$، والآن نفترض أننا نريد البحث عن التركيب (أو التركيبات) الخطية Y المؤلفة من **X** والفاصلة بين المجموعات والمعرفة بالعلاقة:

$$Y = \boldsymbol{\ell}' * \boldsymbol{X} = \left(\ell_1, \ell_2 \ldots . \ell_p\right)\begin{bmatrix}X_1\\ X_2\\ \vdots\\ X_p\end{bmatrix} \qquad (56-4)$$

ومنها نجد أن القيمة المتوقعة لـ Y في كل مجموعة G_j تساوي :

$$E_j(Y) = E_j(\boldsymbol{\ell}' * \boldsymbol{X}) = \boldsymbol{\ell}' * E\left(\boldsymbol{X}/G_j\right) = \boldsymbol{\ell}'\boldsymbol{\mu_j} = \boldsymbol{\mu_{jy}} \qquad (57-4)$$

وهكذا نجد أن القيمة المتوقعة $\boldsymbol{\mu_{jy}} = \boldsymbol{\ell}'\boldsymbol{\mu_j}$ تتغير من مجموعة لأخرى حسب المجموعة التي سحبت منها X، وأن تباين Y في كل مجموعة G_j يساوي:

$$\sigma_j^2(Y) = var_j(Y) = var_j(\boldsymbol{\ell}'\boldsymbol{X}) = \boldsymbol{\ell}' var\left(\boldsymbol{X}/G_j\right) * \boldsymbol{\ell} = \boldsymbol{\ell}' * \boldsymbol{V_j} * \boldsymbol{\ell} \qquad (58-4)$$

وبما أننا افترضنا أو اشترطنا أو يكون $var\left(X/G_j\right) = \boldsymbol{V_j} = \boldsymbol{V}$ فإننا نجد أن:

$$\sigma_j^2(Y) = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} \qquad (59-4)$$

وأخيراً نجد أن المتوسط العام لقيم Y ونرمز له بـ $\bar{\mu}_y$ يساوي بناءً على (4−53) ما يلي:

$$\bar{\mu}_y = \frac{1}{g}\sum_{j=1}^{g} \boldsymbol{\mu_{jy}} = \frac{1}{g}\sum_{j=1}^{g} \ell'\boldsymbol{\mu_j} = \ell'\left(\frac{1}{g}\sum \boldsymbol{\mu_j}\right) = \boldsymbol{\ell}' * \boldsymbol{\bar{\mu}} \qquad (60-4)$$

والآن نعرف معيار (فيشر) بالعلاقة التالية:

$$\boldsymbol{F} = \frac{\sum_{\boldsymbol{j=1}}^{\boldsymbol{g}}\left(\mu_{jy} - \bar{\mu}_y\right)^{\boldsymbol{2}}}{\sigma^2(Y)} = \frac{\begin{bmatrix}\text{مجموع مربعات المسافات من متوسطات} \\ \text{المجموعات إلى المتوسط العام } \bar{\mu}_Y\end{bmatrix}}{\text{تباين المتحول}(\boldsymbol{Y})} \qquad (\boldsymbol{61-4})$$

واعتماداً على العلاقات السابقة (4−54) و(4−56) و(4−57) و(4−59) و(4−60) نجد أن:

$$F = \frac{\sum_{j=1}^{g}\left(\boldsymbol{\ell}'\boldsymbol{\mu_j} - \boldsymbol{\ell}'\overline{\boldsymbol{\mu}}\right)^2}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} = \frac{\boldsymbol{\ell}'\left[\sum_{j=1}^{g}\left(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}\right)'\right] * \boldsymbol{\ell}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}}$$

وبذلك نجد أن معيار (فيشر) اعتماداً على (4−54) يأخذ الشكل التالي:

$$F = \frac{\sum_{j=1}^{g}\left(\mu_{jy} - \bar{\mu}_y\right)^2}{\sigma^2(y)} = \frac{\boldsymbol{\ell}' * \boldsymbol{B} * \boldsymbol{\ell}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} \Longrightarrow \max \qquad (62-4)$$

وهنا نلاحظ أن النسبة التي في الطرف الأيمن لـ (4−62) هي تعميم على المجموعات المتعددة للنسبة الواردة في العلاقة (3−74)، وهي تقيس التغيرات بين المجموعات (Between Group) الناجمة عن قيمY منسوبة إلى التغيرات العامة داخل المجموعات (Within Groups)، وكما فعلنا في حالة المجموعتين . يجب علينا الآن تحديد شعاع الأعداد ℓ، الذي يجعل النسبة في العلاقة (4−62) أكبر ما يمكن .

ولتحقيق هذا الهدف نضيف إلى التابع (4−62) شرط المعيرة المفروض على الشعاع $\boldsymbol{\ell}$ المعرف بالعلاقة (3−105)، والذي يجعل تباينات التوابع Y_i مساوية للواحد . أي نضيف الشرط التالي:

$$\sigma^2(Y) = \boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} = 1 \qquad (63-4)$$

وبذلك يصبح هدفنا تعظيم المقدار :

$$F = \frac{\boldsymbol{\ell}' * \boldsymbol{B} * \boldsymbol{\ell}}{\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell}} \Longrightarrow \max \qquad (64-4)$$

مع وجود الشرط (4–63) المفروض على ℓ .

ولإيجاد حل لهذه المسألة نشكل تابع (لاغرانج) من التابع (4–64)، مع الأخذ بعين الاعتبار الشرط (4–63) كما يلي:

$$L = \frac{\boldsymbol{\ell}' * \boldsymbol{B} * \boldsymbol{\ell}}{1} - \lambda(\boldsymbol{\ell}' * \boldsymbol{V} * \boldsymbol{\ell} - 1) \qquad (65-4)$$

ثم نقوم باشتقاق هذا التابع بالنسبة لـ $\boldsymbol{\ell}$ (حسب قواعد اشتقاق المصفوفات)، ونضع ذلك المشتق مساوياً للصفر فنحصل على أن:

$$\frac{\partial L}{\partial \boldsymbol{\ell}} = \boldsymbol{2B}\,\boldsymbol{\ell} - \lambda(\boldsymbol{2V} * \boldsymbol{\ell}) = 0 \qquad (66-4)$$

ومنها نحصل على المعادلة التالية:

$$\boldsymbol{B}\,\boldsymbol{\ell} - \lambda \boldsymbol{V} * \boldsymbol{\ell} = 0 \qquad (67-4)$$

ثم نضرب طرفي العلاقة (4–67) من اليسار بـ $\boldsymbol{V^{-1}}$ (بفرض أنها نظامية) فنحصل على أن:

$$\boldsymbol{V^{-1}} * \boldsymbol{B} * \boldsymbol{\ell} - \lambda \boldsymbol{\ell} = 0 \qquad (68-4)$$

ثم نكتب هذه العلاقة على الشكل التالي:

$$[\boldsymbol{V^{-1}} * \boldsymbol{B} - \lambda \boldsymbol{I}] * \ell = 0 \qquad (69-4)$$

وهذا هو الشكل النموذجي للمعادلة المميزة لمصفوفة الجداء $(\boldsymbol{V^{-1}} * \boldsymbol{B})$، وهذا يعني أن λ هي مجموعة القيم الذاتية المميزة للمصفوفة $(\boldsymbol{V^{-1}} * \boldsymbol{B})$، ونحصل على تلك القيم بوضع محدد المصفوفة السابقة$[\boldsymbol{V^{-1}} * \boldsymbol{B} - \lambda I]$ مساوياً للصفر، أي بوضع المحدد :

$$\left|\boldsymbol{V^{-1}} * \boldsymbol{B} - \lambda \boldsymbol{I}\right| = 0 \qquad (70-4)$$

نقوم بفك هذا المحدد فنحصل على معادلة جبرية (كثير حدود) من المرتبة p بالنسبة لـ λ . ثم نقوم بإيجاد جذور المعادلة (4–70)، فنحصل على s قيمة ذاتية موجبة لـ λ ، نرتبها تنازلياً كما يلي:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_s \qquad (71-4)$$

حيث أن $s = \min[p\,, g-1]$ وهو عدد الجذور غير المعدومة لـ (4–70) .

ملاحظة حول s: نلاحظ أن المصفوفة $\boldsymbol{V^{-1}B}$ هي من المرتبة $p * p$، وهكذا نتوقع أن يكون عدد الجذور غير المعدومة للمعادلة (4–70) مساوياً لـ p جذراً، أي أن عدد الجذور غير المعدومة يحقق المتراجحة التالية:

$$s \leq p \qquad (72-4)$$

ومن (5–54) نجد أن المصفوفة B تتألف من أشعة الفروقات التالية:

$$(\boldsymbol{\mu_1} - \overline{\boldsymbol{\mu}})\,, (\boldsymbol{\mu_2} - \overline{\boldsymbol{\mu}})\,, (\boldsymbol{\mu_3} - \overline{\boldsymbol{\mu}})\,, \dots\dots\dots, (\boldsymbol{\mu_g} - \overline{\boldsymbol{\mu}}) \qquad (73-4)$$

وهي مرتبطة مع بعضها بواسطة العلاقة التالية:

$$(\boldsymbol{\mu_1} - \overline{\boldsymbol{\mu}}) + (\boldsymbol{\mu_2} - \overline{\boldsymbol{\mu}}) + (\boldsymbol{\mu_3} - \overline{\boldsymbol{\mu}}) \dots + (\boldsymbol{\mu_g} - \overline{\boldsymbol{\mu}}) = \sum_{j=1}^{g} \boldsymbol{\mu_j} - g\overline{\boldsymbol{\mu}} = 0 \qquad (74-4)$$

وهذا يعني أن شعاع الفرق الأول $(\boldsymbol{\mu_1} - \boldsymbol{\bar{\mu}})$ أو غيره يمكن أن يكتب بدلالة الـ $(g-1)$ شعاعاً الأخرى، أي أن رتبة (rank) المصفوفة B تساوي q وأن $q \leq (g-1)$، وهذا يعني أن عدد القيم الذاتية λ غير المعدومة للمصفوفة B يساوي q (أو أقل)، وإن عدد القيم الذاتية λ المعدومة يساوي $(P-q)$ (أو أكثر)، ومنها نستنتج أن المعادلة المميزة (4 − 70) يمكن أن تعطينا q، قيمة غير معدومة لـ λ (أو أقل)، و$(p-q)$ قيمة معدومة لـ λ (أو أكثر) .

ولتوضيح ذلك نأخذ شعاعاً ما $\boldsymbol{e}$، متعامداً مع كل من أشعة الفروق $(\boldsymbol{\mu_j} - \boldsymbol{\bar{\mu}})$، فإنه يكون لدينا:

$$(\boldsymbol{\mu_j} - \boldsymbol{\bar{\mu}})' * \boldsymbol{e} = 0 \qquad j: 1\ 2\ 3\ ..\ g \qquad (75-4)$$

وبناءً على (4−54) فإن الجداء $B * \boldsymbol{e}$ يعطينا النتيجة التالية:

$$B * \boldsymbol{e} = \sum_{j=1}^{g} (\boldsymbol{\mu_j} - \boldsymbol{\bar{\mu}})(\boldsymbol{\mu_j} - \boldsymbol{\bar{\mu}})' * \boldsymbol{e} = \sum_{j=1}^{g} (\boldsymbol{\mu_j} - \boldsymbol{\bar{\mu}}) * 0 = 0$$

نضرب العلاقة السابقة من اليسار بـ V^{-1} فنحصل على أن:

$$\boldsymbol{V^{-1}} * \boldsymbol{B} * \boldsymbol{e} = 0 * \boldsymbol{e} \qquad (76-4)$$

$$[\boldsymbol{V^{-1}}\,\boldsymbol{B} - 0\boldsymbol{I}]\boldsymbol{e} = 0 \implies \lambda = 0 \qquad (77-4)$$

أي أنه يوجد للمصفوفة $V^{-1}\,B$ قيم ذاتية معدومة $\lambda = 0$، وإن عددها يساوي أو يزيد عن $(p-q)$ قيمة كما شرحنا سابقاً .

وهذا يعني أن هناك عدداً من الأشعة المتعامدة والمقابلة لتلك القيم الذاتية المعدومة يساوي $(p-q)$ ومن جهة اخرى نجد أن عدد الأشعة المتعامدة والمقابلة للقيم الذاتية غير المعدومة s يساوي q (أو أقل). وبذلك يكون عدد التوابع التمييزية المقابلة لها يساوي q أيضاً .

وبذلك يكون لدينا دائماً أن: $s \leq q \leq (g-1)$، ومن (4−72) نستنتج أن عدد القيم الذاتية λ غير المعدومة s يجب أن يحقق العلاقة:

$$s \leq min[p\,,(g-1)] \qquad (78-4)$$

وهكذا نجد أنه يمكننا أن نحدد عدد القيم الذاتية غير المعدومة λ_s وعدد الأشعة المتعامدة المقابلة لها ℓ_s وعدد التوابع التمييزية الموافقة لها (على الأكثر) كما يلي:

جدول (4−3): عدد التوابع التمييزية s:

عدد المتحولات المستقلة	عدد المجموعات $g \geq 2$	عدد التوابع التمييزية S (على الأكثر)
أي عدد p	$g = 2$	1
أي عدد $p > 2$	$g = 3$	2
أي عدد $p > 2$	أي عدد $g > 2$	$s = \min\ (p\,, g-1)$

وبعد إيجاد الجذور غير المعدومة $\lambda_1, \lambda_2, \lambda_3, \dots \lambda_s$، نقوم بإيجاد الأشعة الذاتية المقابلة لهذه القيم الذاتية $\ell_1, \ell_2\,, \ell_3 \dots. \ell_s$ واحداً بعد الآخر .

فلإيجاد الشعاع الذاتي الأول أو الحل التمييزي الأول $\boldsymbol{\ell_1}$، نقوم بتعويض القيمة الكبرى λ_1 في المعادلة (4−69)، فنحصل على جملة من المعادلات الخطية المتجانسة تتضمن عناصر الشعاع $\boldsymbol{\ell}$، ثم نقوم بحلها فنحصل من بين لا نهاية من الحلول على الحل التمييزي الأول ℓ_1 المقابل للقيمة الذاتية الأولى λ_1.
إن هذا الحل ℓ_1 يشكل مع مضاعفاته $(C\boldsymbol{\ell_1})$ لا نهاية أخرى من الحلول المقبولة (حيث أن C أي عدد حقيقي لا يساوي الصفر)، وكل منها يعطينا أكبر قيمة لمعيار (فيشر) في (4−62) ويحقق شرط التباين المميز المعرف في (4−63)، أي أنه يحقق العلاقة :

$$\sigma^2(Y_1) = \boldsymbol{\ell}'_{\boldsymbol{1}} * \boldsymbol{V} * \boldsymbol{\ell} = 1 \qquad (79-4)$$

ولإيجاد الحل التمييزي الثاني ℓ_2 المقابل للقيمة الذاتية λ_2، نشترط أن يكون مستقلاً عن الحل التمييزي الأول ℓ_1. أي نشترط أن يكون $\text{cov}(\boldsymbol{\ell}'_{\boldsymbol{1}}\overline{\boldsymbol{X}}, \boldsymbol{\ell}'_{\boldsymbol{2}}\boldsymbol{X}) = 0$، ثم نعوض λ_2 في المعادلة (4−69)، ومنها نقوم بحساب عناصر الشعاع $\boldsymbol{\ell_2}$، فنحصل من بين لا نهاية من الحلول على الحل التمييزي الثاني $\boldsymbol{\ell_2}$ المقابل للقيمة الذاتية λ_2 . وإن هذا الحل يشكل مع مضاعفاته $(C\boldsymbol{\ell_2})$ لا نهاية من الحلول المقبولة (حيث $C \neq 0$)، وهو يعطينا قيمة أخرى لمعيار فيشر في (4−62) ويحقق شرط التباين المميز المعرف في (4−63) أي أنه يحقق العلاقة :

$$\sigma^2(Y_2) = \boldsymbol{\ell}'_{\boldsymbol{2}} * \boldsymbol{V} * \boldsymbol{\ell_2} = 1 \qquad (80-4)$$

كما إنه يجب أن يحقق شرط الاستقلال عن $\boldsymbol{\ell_1}$ وهو أن يكون :

$$\text{cov}(\boldsymbol{\ell}'_{\boldsymbol{1}}\boldsymbol{X}, \boldsymbol{\ell}'_{\boldsymbol{2}}\boldsymbol{X}) = \boldsymbol{\ell}'_{\boldsymbol{1}} * \boldsymbol{V} * \boldsymbol{\ell_2} = 0 \qquad (81-4)$$

ولإيجاد الحل التمييزي الثالث $\boldsymbol{\ell_3}$ المقابل للقيمة الذاتية λ_3، نفترض أنه مستقل عن الحلين السابقين $\boldsymbol{\ell_1}$ و$\boldsymbol{\ell}_2$، وإنه يحقق الشرط (4−63)، ولذلك نعوض λ_3 في المعادلة (4−69) ونحلها بالنسبة لعناصر $\boldsymbol{\ell}_3$ فنحصل من بين لا نهاية من الحلول على الحل التمييزي الثالث $\boldsymbol{\ell}_3$ المقابل للقيمة الذاتية λ_3. وهو يعطينا قيمة أخرى لمعيار (فيشر) في (4−62) ويحقق الشرط (4−63) التالي :

$$\sigma^2(y_3) = \boldsymbol{\ell}'_{\boldsymbol{3}} * \boldsymbol{V} * \boldsymbol{\ell_3} = 1 \qquad (82-4)$$

كما أنه يجب أن يحقق شروط الاستقلال عن $\boldsymbol{\ell_1}$ و $\boldsymbol{\ell_2}$ التالية :

$$\begin{aligned} &cov(\boldsymbol{\ell}'_{\boldsymbol{3}}\boldsymbol{X}\,, \boldsymbol{\ell}'_{\boldsymbol{2}}\boldsymbol{X}\,) = 0 \\ &cov(\boldsymbol{\ell}'_{\boldsymbol{3}}\boldsymbol{X}\,, \boldsymbol{\ell}'_{\boldsymbol{1}}\boldsymbol{X}\,) = 0 \qquad (83-4) \\ &cov(\boldsymbol{\ell}'_{\boldsymbol{1}}\boldsymbol{X}\,, \boldsymbol{\ell}'_{\boldsymbol{2}}\boldsymbol{X}\,) = 0 \end{aligned}$$

وهذه الشروط تعني أن الحلول التمييزية الثلاثة مستقلة عن بعضها البعض (أي أنها متعامدة)
وهكذا نتابع العمل لنحصل على الحل التمييزي الكائي $\boldsymbol{\ell_k}$ المستقل عن الحلول السابقة .
فنقوم بتعويض λ_k في المعادلة (4−69). ومنها نحصل على عناصر الشعاع $\boldsymbol{\ell_k}$ المقابل للقيمة الذاتية λ_k، وهو يعطينا قيمة معينة لمعيار (فيشر) من (4−62) ويحقق شرط المعيرة (4−63) التالي:

$$\sigma^2(Y_k) = \boldsymbol{\ell}'_{\boldsymbol{k}} * \boldsymbol{V} * \boldsymbol{\ell_k} = 1 \qquad (84-4)$$

كما إنه يجب أن يحقق شروط الاستقلال التالية:

$$cov(\boldsymbol{\ell}'_{\boldsymbol{k}}\boldsymbol{X}\,, \boldsymbol{\ell}'_{\boldsymbol{i}}\boldsymbol{X}\,) = 0 \qquad i \neq k \;\; i{:}\, 1\; 2\; 3 \ldots s \qquad (85-4)$$

وإذا وضعنا العلاقتين (4−84) و(4−85) ضمن مصفوفة واحدة تحت الرمز $cov(Y\ )$ نجد أن:

$$cov(Y\ ) = cov\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_s \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = I \qquad (86-4)$$

أي أن مصفوفة التباين المشترك للتوابع التمييزية المختلفة Y_i هي مصفوفة واحدية **I** لأن هذه التوابع مستقلة، ويستفاد من هذه الخاصة في التحقيق من نتائج المسائل العملية .

وفي النهاية نحصل على جميع الحلول التمييزية المقابلة للقيم الذاتية λ غير المعدومة، ونضعها في جدول خاص كما يلي:

جدول (4−4) معاملات التوابع التمييزية :

التوابع / المتحولات	Y_1	Y_2	Y_3	……	Y_s
	$\boldsymbol{\ell_1}$	$\boldsymbol{\ell_2}$	$\boldsymbol{\ell_3}$	……	$\boldsymbol{\ell_s}$
X_1	ℓ_{11}	ℓ_{21}	ℓ_{31}	……	ℓ_{s1}
X_2	ℓ_{12}	ℓ_{22}	ℓ_{32}	……	ℓ_{s2}
X_3	ℓ_{13}	ℓ_{23}	ℓ_{33}	……	ℓ_{s3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	……	$\vdots$
X_p	ℓ_{1p}	ℓ_{2p}	ℓ_{3p}	……	ℓ_{sp}

ومن الجدول السابق نحصل على التوابع المميزة المختلفة، فمثلاً نحصل على التابع التمييزي الأول من عناصر العمود الأول بدلالة المتحولات **X** من خلال العلاقة التالية :

$$Y_1 = \boldsymbol{\ell'_1 X} = \ell_{11}X_1 + \ell_{12}X_2 + \ell_{13}X_3 + \cdots\ + \ell_{1p}X_p \qquad (87-4)$$

وهو يشكل مستوى فوقي في الفضاء R^p ويفصل بين بعض المجموعات G_j عن المجموعات الأخرى .

وكذلك نحصل على التابع التمييزي الثاني من عناصر العمود الثاني ونكتبه كما يلي:

$$Y_2 = \boldsymbol{\ell'_2 X} = \ell_{21}X_1 + \ell_{22}X_2 + \ell_{23}X_3 + \cdots\ + \ell_{2p}X_p \qquad (88-4)$$

وهكذا دواليك ……

وأخيراً نستخلص أن هذه التوابع التمييزية تتمتع بالخواص التالية:

1− إن أي حل خاص لـ (4−69) مثل $\boldsymbol{\ell_k}$ يرتبط مع الأشعة المعيارية الواحدية $\boldsymbol{e_k}$ من خلال العلاقة:

$$\boldsymbol{\ell_k} = \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{e_k} \qquad \Leftrightarrow \qquad \boldsymbol{e_k} = \boldsymbol{V}^{\frac{1}{2}} * \boldsymbol{\ell_k} \qquad (89-4)$$

وبما أن الحلول ℓ_k متعامدة فإن الحلول المميزة $\boldsymbol{e_k}$ تكون متعامدة وتحقق العلاقتين :

$$\boldsymbol{e'_k} * \boldsymbol{e_i} = 0 \qquad i \neq k \qquad \boldsymbol{e'_k} * \boldsymbol{e_i} = 1 \qquad (90-4)$$

2- إن تباين كل تابع تمييزي $Y_k = \boldsymbol{\ell'_k X}$ يساوي الواحد أي يحقق العلاقة:

$$\sigma^2(Y_k) = \boldsymbol{\ell'_k} * \boldsymbol{V} * \boldsymbol{\ell_k} = 1 \qquad (91-4)$$

3- إن التباين المشترك لأي زوج من التوابع التمييزية $Y_i \ , \ Y_k$ يساوي الصفر أي أن :

$$cov(Y_i\,, Y_k) = cov(\boldsymbol{\ell'_i X}\,, \boldsymbol{\ell'_k X}\,) = 0 \quad for \ \ i \neq k \qquad (92-4$$

وهذا يعني أن التوابع التمييزية مستقلة عن بعضها البعض .

4- إن أي حل $\boldsymbol{\ell_k}$ يحقق المعادلة المميزة الأساسية (4-69) أو المعادلة (4-68) كما يلي:

$$[\boldsymbol{V^{-1}} * \boldsymbol{B} - \lambda \boldsymbol{I}]\ell_k *= 0 \quad \Leftrightarrow \quad \boldsymbol{V^{-1} B \ell_k} = \lambda\, \boldsymbol{\ell_k} \qquad (93-4)$$

وبتعويض $\boldsymbol{\ell_k}$ بـ $\boldsymbol{\ell_k} = \boldsymbol{V^{-\frac{1}{2}} e_k}$ نجد أن :

$$\boldsymbol{V^{-1}} * \boldsymbol{B} * \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_k} = \lambda \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_k}$$

وبضرب الطرفين من اليسار بـ $\boldsymbol{V^{+\frac{1}{2}}}$ نجد أن :

$$\left(\boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{-\frac{1}{2}}}\right) * \boldsymbol{e_k} = \lambda \boldsymbol{e_k} \qquad (94-4)$$

أي أن λ هي مجموعة القيم الذاتية للمصفوفة $\left(\boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{\frac{1}{2}}}\right)$ التي أشعتها الذاتية $\boldsymbol{e_k}$، ولكن كما رأينا سابقاً فإن λ هي نفس الأعداد والقيم الذاتية للمصفوفة $(\boldsymbol{V^{-1}} * \boldsymbol{B})$، والتي أشعتها الذاتية $\boldsymbol{\ell_k}$.

ولكن الأشعة الذاتية ℓ_k للمصفوفة الأساسية $(\boldsymbol{V^{-1}} * \boldsymbol{B})$ تكون متناسبة مع الأشعة $\boldsymbol{e_k}$ للمصفوفة $\left(\boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{+\frac{1}{2}}}\right)$ وترتبط معها العلاقة : $\left(\boldsymbol{\ell_k} = \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_k}\right)$ وتشكل مجموعة متكاملة من الحلول التمييزية التي تحقق الشروط التالية:

$$var(\boldsymbol{\ell' X}) = 1$$
$$cov\left(\boldsymbol{\ell'_k X} \ \ , \boldsymbol{\ell'_j X}\,\right) = 0$$

وذلك من اجل $i \neq k$ و $i{:}\,1\ 2\ 3 \ldots .s$

5- إن مصفوفة التباين المشترك لجملة التوابع التمييزية $Y_1, Y_2, Y_3, \ldots Y_s$، هي مصفوفة واحدية، أي أن:

$$cov(Y) = cov\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_s \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = I \qquad (95-4)$$

وأخيراً نتوصل إلى النتيجة التالية:

- **النتيجة:** من العلاقة (4-94) نستنتج أن القيم الذاتية λ_k للمصفوفة $\left(\boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{-\frac{1}{2}}}\right)$، هي نفس القيم الذاتية للمصفوفة $(\boldsymbol{V^{-1}} * \boldsymbol{B})$، والتي رمزنا لها بعد ترتيبها تنازلياً كما يلي:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_s \qquad (96-4)$$

حيث أن s هو أصغر العددين $\Big[(g-1)$ أو $p\Big]$.

ولكن الأشعة الذاتية $\boldsymbol{a_k}$ للمصفوفة $\left[\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}}\right]$ المقابلة للقيم الذاتية λ_k، ترتبط مع الأشعة الذاتية $\boldsymbol{\ell_k}$ للمصفوفة $(\boldsymbol{V}^{-1} * \boldsymbol{B})$، بواسطة العلاقة : $\left(\boldsymbol{\ell_k} = \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a_k}\right)$ وبصورة عامة يكون لدينا: $\boldsymbol{\ell} = \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a}$.

وبتعويض ذلك في العلاقة (4−62) نجد أن معيار (فيشر) يأخذ الشكل التالي :

$$F = \frac{\left(\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a}\right)' * \boldsymbol{B} * \left(\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a}\right)}{\left(\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a}\right)' * \boldsymbol{V} * \left(\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a}\right)} = \frac{\boldsymbol{a}' * \left[\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}}\right] * \boldsymbol{a}}{\boldsymbol{a}' * \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{V} * \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{a}} =$$

$$F = \frac{\boldsymbol{a}' * \left[\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}}\right] * \boldsymbol{a}}{\boldsymbol{a}' * \boldsymbol{a}} \Longrightarrow Max \qquad (97-4)$$

وبذلك يأخذ معيار (فيشر) شكل الصيغة $\frac{\boldsymbol{X}' * \boldsymbol{B} * \boldsymbol{X}}{\boldsymbol{X}' * \boldsymbol{X}}$ المعرفة في الملحق (2) بالعلاقة (2−71)، والتي تؤكد على أن أكبر قيمة لها تساوي القيمة الذاتية الأولى λ_1، وهي التي تقابل الشعاع الذاتي المميز الأول $\boldsymbol{e_1}$ (أي عندما $\boldsymbol{a} = \boldsymbol{e_1}$) وبذلك يكون لدينا أن:

$$Max\ F = \frac{\boldsymbol{e_1}' * \left[\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}}\right] * \boldsymbol{e_1}}{\boldsymbol{e_1}' * \boldsymbol{e_1}} = \lambda_1 \qquad (98-4)$$

أي أن الشعاع الذاتي $\boldsymbol{e_1}$ للمصفوفة $\left[\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}}\right]$، هو الذي يعظم معيار (فيشر). وهو الذي يقابل الشعاع الذاتي $\left(\boldsymbol{\ell_1} = \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{e_1}\right)$ للمصفوفة$(\boldsymbol{V}^{-1} * \boldsymbol{B})$.

ومنه نحصل على التركيب الخطي الأول: $Y_1 = \boldsymbol{\ell_1}'\boldsymbol{X}$، الذي يسمى بالتابع التمييزي الأول، ونكتبه كمايلي:

$$Y_1 = \boldsymbol{\ell_1}'\boldsymbol{X} = \ell_{11}X_1 + \ell_{12}X_2 + \ell_{13}X_3 + \cdots \ldots + \ell_{1p}X_p \qquad (99-4)$$

وهو يتصف بأن تباينه يساوي الواحد، أي أن:

$$\sigma^2(Y_1) = \boldsymbol{\ell_1}' * \boldsymbol{V} * \boldsymbol{\ell} = 1 \qquad (100-4)$$

أما الشعاع الذاتي ℓ_2 فهو الشعاع الذي يساوي $\boldsymbol{\ell_2} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{e_2}$، ومنه نحصل على التركيب الخطي $Y_2 = \boldsymbol{\ell_2}'\boldsymbol{X}$، والذي يسمى بالتابع التمييزي الثاني ونكتبه كما يلي:

$$Y_2 = \boldsymbol{\ell_2}'\boldsymbol{X} = \ell_{21}X_1 + \ell_{22}X_2 + \ell_{23}X_3 + \cdots\ \ell_{2p}X_p \qquad (101-4)$$

وهو يتصف بأنه مستقل عن Y_1 ويحقق الشرطين التاليين :

$$\sigma^2(Y_2) = \boldsymbol{\ell_2}' * \boldsymbol{V} * \boldsymbol{\ell_2} = 1 \qquad (102-4)$$

$$cov(Y_1\,, Y_2) = cov(\boldsymbol{\ell_2}'\boldsymbol{X}\,, \boldsymbol{\ell_1}'\boldsymbol{X}) = 0 \qquad (103-4)$$

وهكذا نجد أن الشعاع الذاتي المطلوب $\boldsymbol{\ell_k}$ هو الشعاع الذي يساوي $\boldsymbol{\ell_k} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{e_k}$، ومنه نحصل على التركيب الخطي $Y_k = \boldsymbol{\ell_k}'\boldsymbol{X}$، والذي يسمى بالتابع التمييزي الكائي ونكتبه كما يلي:

$$Y_k = \boldsymbol{\ell_k}'\boldsymbol{X} = \ell_{k1}X_1 + \ell_{k2}X_2 + \cdots\ \ell_{kp}X_p \qquad (104-4)$$

وهو يتصف بأنه مستقل عن التوابع السابقة له $Y_1, Y_2, \dots Y_{k-1}$ ويحقق الشروط التالية:

$$\sigma^2(Y_k) = \boldsymbol{\ell}'_{\boldsymbol{k}} * \boldsymbol{V} * \boldsymbol{\ell}_{\boldsymbol{k}} = 1 \qquad (105-4)$$

$$cov(Y_k\,, Y_i) = cov(\boldsymbol{\ell}'_{\boldsymbol{k}}\boldsymbol{X}\,, \boldsymbol{\ell}'_{\boldsymbol{i}}\boldsymbol{X}) = 0 \qquad (106-4)$$

وذلك من أجل جميع $S \dots .3\ 2\ 1 \quad : \quad i < k$

ملاحظة: إن حساب الحلول الخاصة $\boldsymbol{\ell}_{\boldsymbol{i}}$ أو المميزة$\boldsymbol{e}_{\boldsymbol{i}}$ المقابلة للقيم الذاتية الموجبة للمصفوفة $(\boldsymbol{V}^{-1}\boldsymbol{B})$ يتطلب معرفة عناصر المصفوفة$\boldsymbol{V}$ وقيم المتوسطات $\boldsymbol{\mu}_{\boldsymbol{i}}$.

وإذا كان هذا الأمر غير متوفر، فإننا نقوم باستبدال تلك العناصر والمتوسطات بتقديراتها المحسوبة من العينات المسحوبة من المجموعات المفروضة$G_1, G_2, \dots G_g$ بحجوم $n_1, n_2, \dots n_g$ على الترتيب، ومنها نحسب التقديرات التالية:

$$\widetilde{\boldsymbol{\mu}}_i = \overline{\boldsymbol{X}}_{\boldsymbol{i}} = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{\boldsymbol{ij}} \qquad (107-4)$$

$$\widetilde{\boldsymbol{\mu}} = \overline{\boldsymbol{X}} = \frac{\sum_{i=1}^{g} n_i * \overline{\boldsymbol{X}}_{\boldsymbol{i}}}{\sum n_i} = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i} \boldsymbol{x}_{\boldsymbol{ij}}}{n} \qquad (108-4)$$

حيث أن $n = \sum_{i=n}^{g} n_i$ هو حجم العينة الكلية

وعندها نجد أن مصفوفة المربعات بين المجموعات تقدر بما يلي:

$$\boldsymbol{B}_{\boldsymbol{i}} = \sum_{j=1}^{g} (\overline{\boldsymbol{X}}_{\boldsymbol{j}} - \overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}_{\boldsymbol{j}} - \overline{\boldsymbol{X}})' \qquad (109-4)$$

وإن مصفوفة التباين الداخلي للمجموعات $\boldsymbol{V}$ تقدر بتركيب مصفوفة مربعات الانحرافات الداخلية $\boldsymbol{S}_{\boldsymbol{j}}$ لهذه المجموعات كما يلي:

$$\widetilde{\boldsymbol{V}} = \boldsymbol{S}_{\boldsymbol{p}} = \frac{\sum_{j=1}^{g}(n_j - 1)\boldsymbol{S}_{\boldsymbol{j}}}{\sum^{g} n_j - g} = \frac{\boldsymbol{W}}{n-g} \qquad (110-4)$$

وعندها نأخذ العلاقة (4–62) الشكل التالي:

$$\widetilde{F} = \frac{\boldsymbol{\ell}' * \boldsymbol{B} * \boldsymbol{\ell}}{\boldsymbol{\ell}' * \boldsymbol{S}_{\boldsymbol{p}} * \boldsymbol{\ell}} \qquad (111-4)$$

ومن العلاقة (4–110) نلاحظ أن المصفوفة$\boldsymbol{W}$ هي مصفوفة مربعات الانحرافات الداخلة وهي مرتبطة مع المصفوفة $\boldsymbol{S}_{\boldsymbol{p}}$ بالعلاقة:

$$\boldsymbol{W} = (n-g)\boldsymbol{S}_{\boldsymbol{p}} = \sum_{j=1}^{g}(n_j - 1)\boldsymbol{S}_{\boldsymbol{j}} \qquad (110\,a-4)$$

وعندها فإن العلاقة (4–111) تأخذ الشكل التالي:

$$\widetilde{F} = \frac{\boldsymbol{\ell}' * \boldsymbol{B} * \boldsymbol{\ell}}{\boldsymbol{\ell}' * \boldsymbol{S}_{\boldsymbol{p}} * \boldsymbol{\ell}} = (n-g)\frac{\boldsymbol{\ell}' * \boldsymbol{B} * \boldsymbol{\ell}}{\boldsymbol{\ell}' * \boldsymbol{W} * \boldsymbol{\ell}} \qquad (111a-4)$$

وهنا نلاحظ أن عملية تعظيم النسبة الأخيرة يمكن أن تتم بعد حذف العدد الثابت $(n-g)$ ، وبذلك يأخذ معيار (فيشر) الشكل المعدل التالي:

$$\tilde{f}=\frac{\tilde{\boldsymbol{\ell}}' * \boldsymbol{B} * \tilde{\boldsymbol{\ell}}}{\tilde{\boldsymbol{\ell}}' * \boldsymbol{W} * \tilde{\boldsymbol{\ell}}} \qquad (112-4)$$

وهكذا نجد أنه يمكننا تقديم الحل المثالي $\tilde{\boldsymbol{\ell}}$ بصيغة شعاع ذاتي ممعير $\tilde{\boldsymbol{e}}$ للمصفوفة $(\boldsymbol{W}^{-1}\boldsymbol{B})$، لأنه إذا كان $\tilde{\boldsymbol{e}}$ يحقق العلاقة :

$$(\boldsymbol{W}^{-1}\boldsymbol{B})\tilde{\boldsymbol{e}}=\tilde{\lambda}\tilde{\boldsymbol{e}} \qquad (113-4)$$

فإنه يحقق العلاقة التالية: [لأن $\boldsymbol{W}^{-1}=\frac{1}{n-g}\boldsymbol{S}^{-1}$]

$$\left(\boldsymbol{S}_{\boldsymbol{p}}^{-1}\boldsymbol{B}\right)\tilde{\boldsymbol{e}}=\tilde{\lambda}(n-g)\tilde{\boldsymbol{e}} \qquad (114-4)$$

وأخيراً نصيغ قاعدة معيار فيشر للتوابع التمييزية في حالة العينات كما يلي :

- **نتيجة:** إذا كانت الجذور الذاتية الموجبة للمصفوفة $(\boldsymbol{W}^{-1}\boldsymbol{B})$: $\lambda_1\ge\lambda_2\ge\lambda_3\ge\cdots\lambda_s>0$، حيث أن عددها s هو أصغر العددين $\left(g-1 \text{ أو } P\right)$، وكانت الأشعة الذاتية $\tilde{\boldsymbol{e}}_{\boldsymbol{1}},\tilde{\boldsymbol{e}}_{\boldsymbol{2}},\tilde{\boldsymbol{e}}_{\boldsymbol{3}},\dots\tilde{\boldsymbol{e}}_{\boldsymbol{s}}$ المقابلة لها (والممعيرة لتحقيق العلاقة $\tilde{\boldsymbol{e}}' * \boldsymbol{S}_{\boldsymbol{p}} * \tilde{\boldsymbol{e}}=1$)، فإن شعاع الأمثال$\boldsymbol{\ell}$ الذي يعظم النسبة التالية :

$$\tilde{f}=\frac{\tilde{\boldsymbol{\ell}}' * \boldsymbol{B}_{\boldsymbol{0}} * \tilde{\boldsymbol{\ell}}}{\tilde{\boldsymbol{\ell}}' * \boldsymbol{W} * \tilde{\boldsymbol{\ell}}}=\frac{* \tilde{\boldsymbol{\ell}}'\left(\sum_{J=1}^{g}\widetilde{(\overline{\boldsymbol{x}}_{\boldsymbol{J}}-\overline{\boldsymbol{x}})}(\overline{\boldsymbol{x}}_{\boldsymbol{J}}-\overline{\boldsymbol{x}})'\right)\tilde{\boldsymbol{\ell}}}{\tilde{\boldsymbol{\ell}}'\left[\sum_j\sum_i(\boldsymbol{x}_{\boldsymbol{ij}}-\overline{\boldsymbol{x}}_{\boldsymbol{i}})(\overline{\boldsymbol{x}}_{\boldsymbol{ij}}-\overline{\boldsymbol{x}}_{\boldsymbol{j}})'\right] * \tilde{\boldsymbol{\ell}}} \qquad (114-4)$$

هو الشعاع $\tilde{\boldsymbol{\ell}}_{\boldsymbol{1}}$ الذي يساوي $\tilde{\boldsymbol{e}}_{\boldsymbol{1}}$ (أي يكون $\tilde{\boldsymbol{\ell}}_{\boldsymbol{1}}=\tilde{\boldsymbol{e}}_{\boldsymbol{1}}$)، وإن التركيب الخطي $\left(\tilde{\boldsymbol{\ell}}_{\boldsymbol{1}}'\boldsymbol{X}\right)$ يسمى بالتابع التمييزي الممعير الأول للعينة المسحوبة . كما إن الشعاع $\tilde{\boldsymbol{\ell}}_{\boldsymbol{2}}=\tilde{\boldsymbol{e}}_{\boldsymbol{2}}$ يعطينا التابع التمييزي الممعير الثاني$\left(\tilde{\boldsymbol{\ell}}'\boldsymbol{X}\right)$ للعينة. وكذلك الشعاع $\tilde{\boldsymbol{\ell}}_{\boldsymbol{k}}=\tilde{\boldsymbol{e}}_{\boldsymbol{k}}$ يعطينا التابع التمييزي $\left(\tilde{\boldsymbol{\ell}}_{\boldsymbol{k}}'\boldsymbol{X}\right)$ حيث أن $k\le s$.

حيث أن الأشعة $\tilde{\boldsymbol{\ell}}_{\boldsymbol{1}},\dots\tilde{\boldsymbol{\ell}}_{\boldsymbol{s}}$ هي أشعة مستقلة ومتعامدة . وحسب القواعد السابقة فمن غير المحتمل أن تكون التباينات المشتركة للتوابع التمييزية $\left(\tilde{\boldsymbol{\ell}}_{\boldsymbol{k}}'\boldsymbol{X}\right)$ معدومة تماماً أو شبه معدومة لكل عينة عشوائية، ويكون من الأفضل ان تحقق أمثالها $\tilde{\boldsymbol{\ell}}_{\boldsymbol{i}}$ الشروط التالية:

$$\tilde{\boldsymbol{\ell}}_{\boldsymbol{i}}' * \boldsymbol{S}_{\boldsymbol{p}} * \tilde{\boldsymbol{\ell}}_{\boldsymbol{k}}=\begin{cases}1 & i=k\le s\\ 0 & \text{غير ذلك}\end{cases} \qquad (115-4)$$

علماً بأن استخدامنا لـ$\boldsymbol{S}_{\boldsymbol{p}}$ كان مشروطاً بأننا افترضنا مؤقتاً أن مصفوفات التباين المشترك للمجموعاتG_i كانت متساوية .

مثال (4-4): سنتابع دراسة المجتمع المؤلف من ثلاث مجموعات $(g=3)$ ونريد أن نميز عناصره بواسطة متحولين فقط$(P=2)$، ولنفترض إننا سحبنا عينات من المجموعات بحجوم متساوية كما يلي: $n_1=n_2=n_3=3$ ، واعطتنا البيانات التالية :

$$\boldsymbol{X}G_1=\begin{array}{c} \text{المشاهدات} \\ X_1 \\ X_2 \end{array}\begin{array}{c} \begin{array}{ccc} Q_1 & Q_2 & Q_3 \end{array} \\ \begin{bmatrix} -2 & 0 & 1\\ 5 & 3 & 1\end{bmatrix}\end{array} \qquad \boldsymbol{X}G_2=\begin{bmatrix}0 & 2 & 1\\ 6 & 4 & 2\end{bmatrix} \qquad \boldsymbol{X}G_3=\begin{bmatrix}1 & 0 & -1\\ -2 & 0 & -4\end{bmatrix}$$

ومنها وجدنا أن أشعة المتوسطات هي:

$$\bar{\boldsymbol{X}}G_1 = \begin{bmatrix}-1\\3\end{bmatrix} \qquad \bar{\boldsymbol{X}}G_2 = \begin{bmatrix}1\\4\end{bmatrix} \qquad \bar{\boldsymbol{X}}G_3 = \begin{bmatrix}0\\-2\end{bmatrix}$$

وكذلك نجد أن شعاع المتوسط العام $\bar{\boldsymbol{X}}$ يساوي $\bar{\boldsymbol{X}} = \begin{bmatrix}0\\\frac{5}{3}\end{bmatrix}$ ، ومنها نقوم بحساب المصفوفة B فنجد أن:

$$\widetilde{\boldsymbol{B}} = \sum_{j=1}^{g}(\boldsymbol{X_i}-\bar{\boldsymbol{X}})(\boldsymbol{X_i}-\bar{\boldsymbol{X}})' = \begin{bmatrix}2 & 1\\1 & \frac{62}{3}\end{bmatrix}$$

فمثلاً نشرح كيفية حساب المصفوفة B_1 عن المجموعة الأولى فنجد من (4-54) أن:

$$\boldsymbol{B_1} = \begin{bmatrix}(-1-0)\\\left(3-\frac{5}{3}\right)\end{bmatrix}\left[(-1-0)\,,\left(3-\frac{5}{3}\right)\right] = \begin{bmatrix}-1\\\frac{4}{3}\end{bmatrix}\left(-1\,,\frac{4}{3}\right) = \begin{bmatrix}1 & -\frac{4}{3}\\-\frac{4}{3} & \frac{16}{9}\end{bmatrix}$$

وكذلك نجد أن المصفوفتين B_2 و B_3 تساويان :

$$\boldsymbol{B_2} = \begin{bmatrix}(+1-0)\\\left(4-\frac{5}{3}\right)\end{bmatrix}\left[(1-0)\,,\left(4-\frac{5}{3}\right)\right] = \begin{bmatrix}1\\\frac{7}{3}\end{bmatrix}\left[1\,,\frac{7}{3}\right] = \begin{bmatrix}1 & \frac{7}{3}\\\frac{7}{3} & \frac{49}{9}\end{bmatrix}$$

$$\boldsymbol{B_3} = \begin{bmatrix}(0-0)\\\left(-2-\frac{5}{3}\right)\end{bmatrix}\left[(0-0)\,,\left(-2-\frac{5}{3}\right)\right] = \begin{bmatrix}0\\-\frac{11}{3}\end{bmatrix}\left[0,\frac{-11}{3}\right] = \begin{bmatrix}0 & 0\\0 & \frac{121}{9}\end{bmatrix}$$

ومنه نجد أن المصفوفة B تساوي :

$$\boldsymbol{B} = \boldsymbol{B_1}+\boldsymbol{B_2}+\boldsymbol{B_3} = \begin{bmatrix}2 & 1\\1 & \frac{62}{3}\end{bmatrix}$$

ولقد وجدنا في المثال (4-3) أن التباين المدمج S_P يساوي :

$$\boldsymbol{S_P} = \frac{1}{9-3}[2\boldsymbol{S_1}+2\boldsymbol{S_2}+2\boldsymbol{S_3}] = \begin{bmatrix}1 & -\frac{1}{3}\\-\frac{1}{3} & 4\end{bmatrix}$$

ومنها نجد أن مصفوفة مربعات الانحرافات الداخلية $\boldsymbol{W}$ تساوي:

$$\boldsymbol{W} = (n-g)\boldsymbol{S_P} = (9-3)\begin{bmatrix}1 & -\frac{1}{3}\\-\frac{1}{3} & 4\end{bmatrix} = \begin{bmatrix}6 & -2\\-2 & 24\end{bmatrix}$$

ومنها نجد أن مقلوبها $\boldsymbol{W^{-1}}$ يساوي : $\boldsymbol{W^{-1}} = \frac{1}{140}\begin{bmatrix}24 & +2\\+2 & 6\end{bmatrix}$

وإن مصفوفة الجداء $W^{-1} * B$ تساوي :

$$\boldsymbol{W^{-1}} * \boldsymbol{B} = \frac{1}{140}\begin{bmatrix}24 & +2\\+2 & 6\end{bmatrix}\begin{bmatrix}2 & 1\\1 & \frac{62}{3}\end{bmatrix} = \begin{bmatrix}0.3571 & 0.4667\\0.0714 & 0.9000\end{bmatrix}$$

وحتى نحصل على القيم الذاتية لهذه المصفوفة ، نلاحظ أن عدد تلك القيم الموجبة يساوي أو لا يتجاوز $s = \min(g-1\,, p) = 2$، وبذلك نجد أن المعادلة المميزة (4–69) تأخذ الشكل النموذجي التالي:

$$[\boldsymbol{W^{-1}} * \boldsymbol{B} - \lambda \boldsymbol{I}]\, \ell = 0 \qquad (a)$$

لإيجاد قيم λ من هذه المعادلة نضع محددها مساوياً للصفر فنحصل على المعادلة التالية:

$$\left|\boldsymbol{W^{-1}} * \boldsymbol{B} - \lambda \boldsymbol{I}\right| = \begin{vmatrix} 0.3571-\lambda & 0.4667 \\ 0.0714 & 0.9000-\lambda \end{vmatrix} = 0$$

ومنها نحصل على كثير حدود من الدرجة الثانية لـ λ يساوي:

$$(0.3571-\lambda)(0.9000-\lambda) - (0.4667)(0.0714) = 0$$
$$\lambda^2 - 1.2571\,\lambda + 0.2881 = 0$$

وباستخدام قواعد حل معادلات الدرجة الثانية نحصل على الجذرين التاليين لها وهما:

$$\lambda_1 = 0.9556 \qquad\qquad \lambda_2 = 0.3015$$

وهنا نلاحظ أن λ_1 هي القيمة الذاتية الكبرى وهي التي تعطينا أكبر قيمة ممكنة للنسبة (4–62) المستخلصة من معيار (فيشر)، ومنها نحسب عناصر الشعاع ℓ كما يلي :

نعوض القيمة الذاتية الأولى $\lambda_1 = 0.9556$ في المعادلة النموذجية (a) فنحصل على أن:

$$[\boldsymbol{W^{-1}} * \boldsymbol{B} - \lambda \boldsymbol{I}]\, \tilde{\boldsymbol{\ell}}_{\boldsymbol{1}} = \begin{bmatrix} 0.3571-0.9556 & 0.4667 \\ 0.0714 & 0.9000-0.9556 \end{bmatrix} * \begin{bmatrix} \tilde{\ell}_{11} \\ \tilde{\ell}_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

ومنها نحصل على المعادلتين التاليتين اللتين تتضمنان عناصر الحل $\boldsymbol{\ell_1}$:

$$\begin{aligned} -0.5985\ell_{11} + 0.4667\ell_{12} &= 0 \\ 0.0714\ell_{11} - 0.0556\ell_{12} &= 0 \end{aligned} \qquad (b)$$

وهما معادلتان غير مستقلتين ومن الأولى نجد أن:

$$\ell_{11} = 0.77978\ell_{12} \qquad (c)$$ وهي تعطينا لا نهاية من الحلول

ولكن بما أن الشعاع ℓ_1 يجب أن يحقق شرط المعيرة التالي: $\boldsymbol{\ell'_1 S_p \ell_1} = 1$ فإننا نجد أن:

$$\boldsymbol{\ell'_1 S_p \ell_1} = (\ell_{11}\,, \ell_{12}) \begin{bmatrix} 1 & -\frac{1}{3} \\ -\frac{1}{3} & 4 \end{bmatrix} \begin{bmatrix} \ell_{11} \\ \ell_{12} \end{bmatrix} = 1$$

$$= (\ell_{11}\,, \ell_{12}) \begin{bmatrix} \left(1\ell_{11} - \frac{1}{3}\ell_{12}\right) \\ \left(-\frac{1}{3}\ell_{11} 4\ell_{12}\right) \end{bmatrix} = 1$$

$$\ell_{11}^2 - \frac{1}{3}\ell_{11}\,\ell_{12} - \frac{1}{3}\ell_{11}\,\ell_{12} + 4\ell_{12}^2 = 1$$

$$\ell_{11}^2 + 4\ell_{12}^2 - \frac{2}{3}\ell_{11}\,\ell_{12} = 1 \qquad (116-4)$$

ومن المعادلة الأولى من (b) وجدنا أن $(\ell_{11} = 0.77978\ell_{12})$ نعوض ذلك في (4–116) فنحصل على أن:

$$(0.77978)^2\ \ell_{12}^2 + 4\ \ell_{12}^2 - \frac{2}{3}(0.77978)\ \ell_{12}^2 = 1$$

$$0.60805\ \ell_{12}^2 + 4\ell_{12}^2 - 051985\ \ell_{12}^2 = 1$$

$$4.08820\ \ell_{12}^2 = 1$$

$$\ell_{12} = \pm\sqrt{\frac{1}{4.08820}} = \pm\sqrt{0.244606} = \pm 0.49458$$ (ونعتبره موجباً)

ومنها ومن (b) نجد أنه عندما تكون $\ell_{12} > 0$ فإن :

$$\ell_{11} = 0.77978(0.49458) = 0.38566$$

وهكذا نحصل على الحل الخاص الأول:
$$\tilde{\boldsymbol{\ell}}_{\mathbf{1}} = \begin{bmatrix} 0.38566 \\ 0.49458 \end{bmatrix}$$

ومنه نحصل على التابع التمييزي الأول:

$$\tilde{Y}_1 = \tilde{\boldsymbol{\ell}}'_{\mathbf{1}}\boldsymbol{X} = (0.38566\ \ ,0.49458)\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\tilde{Y}_1 = 0.38566X_1 + 0.49458\ X_2 \qquad (117-4)$$

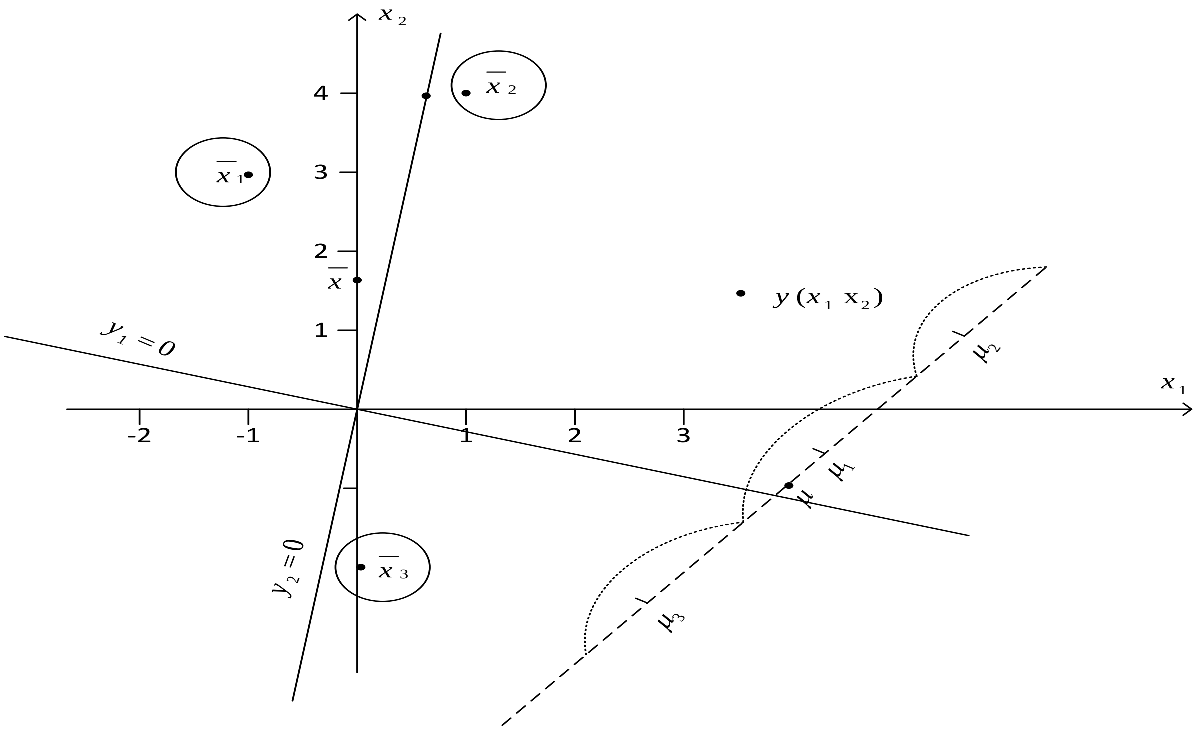
وبطريقة مشابهة نحصل على الحل الخاص الثاني $(\tilde{\boldsymbol{\ell}}_{\mathbf{2}})$ من جراء تعويض قيمة $\lambda_2 = 0.3015$ في المعادلة التمييزية (a)، ثم حساب مركبتي الحل $\boldsymbol{\ell}_{\mathbf{2}}$ مع ضرورة تحقق الشرط $\tilde{\boldsymbol{\ell}}'_{\mathbf{2}}\ \boldsymbol{S}_{\boldsymbol{p}}\ \tilde{\boldsymbol{\ell}}_{\mathbf{2}} = 1$، فنحصل على أن:
$$\tilde{\boldsymbol{\ell}}_{\mathbf{2}} = \begin{bmatrix} 0.988 \\ -0.112 \end{bmatrix}$$

ومنه نحصل على التابع التمييزي الثاني:

$$\tilde{Y}_2 = \tilde{\boldsymbol{\ell}}'_{\mathbf{2}}\boldsymbol{X} = (0.938\ \ , -0.112)\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\tilde{Y}_2 = 0.938\ X_1 - 0.112\ X_2 \qquad (118-4)$$

وهما يمثلان الحدود الفاصلة بين المجموعات الثلاثة ويمكن تمثيلهما بيانياً على الشكل التالي:



الشكل (4−2): تمثيل التوابع التمييزية $\tilde{Y}_1$ و $\tilde{Y}_2$ الفاصلة بين المجموعات

4−3−1: استخدام توابع (فيشر) في التصنيف: [Johnson, Wichern. P.522 بتصرف وإضافة **].**

لقد أشرنا سابقاً إلى أن الهدف الأساسي من توابع (فيشر) التمييزية هو تخفيض عدد أبعاد تمثيل البيانات التي تفصل بين المجموعات، وهي رغم إنها مستخلصة من الاعتبارات الانفصالية، إلا أنها تضع الأسس الرياضية لصياغة قواعد التصنيف .

وفي البداية نعيد توضيح وترميز المصطلحات المستخدمة في التوابع التمييزية$\boldsymbol{\ell}' * \boldsymbol{X}$ ونصيغها كما يلي:

نرمز للتابع التمييزي الكائي بالرمز:

$$Y_k = \boldsymbol{\ell}'_{\boldsymbol{k}} * \boldsymbol{X} \qquad k \leq s \qquad (119-4)$$

ومنه نستخلص شعاع التوابع التمييزية Y ونرمز له بما يلي:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{\ell}'_1\boldsymbol{X} \\ \boldsymbol{\ell}'_2\boldsymbol{X} \\ \vdots \\ \boldsymbol{\ell}'_s\boldsymbol{X} \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}'_1\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X} \\ \boldsymbol{e}'_2\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X} \\ \vdots \\ \boldsymbol{e}'_s\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X} \end{bmatrix} \qquad (120-4)$$

وذلك على اعتبار أن $\boldsymbol{\ell_i} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{e_i}$ وأن $\boldsymbol{\ell}'_{\boldsymbol{i}} = \boldsymbol{e}'_{\boldsymbol{i}}\boldsymbol{V}^{-\frac{1}{2}}$ لأن $\boldsymbol{V}^{-\frac{1}{2}}$ متناظرة .

ونرمز لشعاع متوسطاته (توقعاته) في المجموعة G_j بالرمز μ_{jy}، وهو ما يطلق عليه مراكز التوابع التمييزية في المجموعة G_j فيكون لدينا ما يلي:

$$E(\boldsymbol{Y}/G_j) = \boldsymbol{\mu_{jy}} = \begin{bmatrix} \mu_{jy1} \\ \mu_{jy2} \\ \vdots \\ \mu_{jys} \end{bmatrix} = \begin{bmatrix} \ell'_1\mu_j \\ \ell'_2\mu_j \\ \vdots \\ \ell'_s\mu_j \end{bmatrix} = \begin{bmatrix} e'_1V^{-\frac{1}{2}}\mu_j \\ e'_2V^{-\frac{1}{2}}\mu_j \\ \vdots \\ e'_sV^{-\frac{1}{2}}\mu_j \end{bmatrix} \qquad (121-4)$$

حيث أن: $\boldsymbol{\mu_j} = E(\boldsymbol{X}/G_j)$

ولقد وجدنا من (4−86) أن مصفوفة التباين المشترك للتوابع التميزية Y في جميع المجموعات G_j تساوي I، أي أن:

$$V_y = [cov(Y_k \ ,Y_i)] = I \qquad (122-4)$$

كما وجدنا أن عدد التوابع التمييزية يساوي s وأن:$s = \min[P\ ,(g-1)]$

وضمن هذه الاعتبارات إذا أخذنا مربع المسافة من نقطة $Y = y$ إلى المركز μ_{jy} كمقياس لتصنيف المشاهدات، فإن ذلك المقياس يساوي (حسب (4−35)) ما يلي :

$$D_j^2(y) = (\boldsymbol{y} - \boldsymbol{\mu_{jy}})' * I^{-1} * (\boldsymbol{y} - \boldsymbol{\mu_{jy}}) = \sum_{i=1}^{s}(\boldsymbol{y_i} - \boldsymbol{\mu_{jyi}})^2 \qquad (123-4)$$

حيث أن: $j: 1\ 2\ 3 \dots g$

وبذلك نتوصل إلى قاعدة منطقية لتصنيف العنصر y في المجموعة G_k، نصيغها كما يلي:

قاعدة: إذا كان مربع المسافة من y إلى المركز μ_{ky} أصغر من أي مربع آخر للمسافة بين y وأي مركز آخر μ_{jy}، وذلك من أجل $k \neq j: 1\ 2\ 3 \dots g$، فإننا نصنف العنصر y في المجموعة G_k . كما هو موضح بالشكل (4−2) .

وبعبارة أخرى نكتب :

نصنف y في المجموعة G_k إذا كان :
$D_k^2(y) = \min[D_j^2(y)] \qquad k \neq j: 1\ 2\ 3 \dots g \qquad (124-4)$

وهنا نشير إلى أن المجموع في (4−123) مأخوذ على جميع التوابع التمييزية في كل مجموعة G_j، ولكن الأسس النظرية والتطبيقات العملية تؤكد لنا إنه يمكن اختصار حدود هذا المجموع إلى عدد محدود منها r حيث $r <$s . وذلك لأن التوابع التمييزية الأولى تكون أكثر أهمية من التوابع الأخيرة، ويتم التعبير عن أهميتها النسبية أو التراكمية من خلال ترتيبها حسب مقياس رقمي معين على امتداد جميع المجموعات، وبما أن هذه التوابع التمييزية $(Y_i = \boldsymbol{\ell'_i X})$ تقابل الأشعة الذاتية $\boldsymbol{\ell_i}$ التي تقابل القيم الذاتية للمصفوفة $(\boldsymbol{V^{-1}B})$ والمرتبة تنازلياً كما يلي:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_s \geq \lambda_{s+1} \dots . \lambda_p \geq 0$$

فإن عدد القيم الذاتية غير المعدومة يساوي (على الأكثر) s قيمة، وإن عدد القيم الذاتية المعدومة يساوي (على الأقل) (p−s) قيمة، وإن التوابع التمييزية Y_i تقابل القيم الذاتية غير المعدومة . أما القيم الذاتية المعدومة فتقابلها توابع لها قيم ثابتة .

ولتوضيح ذلك نأخذ مقياس المسافة الاحصائية المميزة بين متوسطات المجموعات $\boldsymbol{\mu_j}$ والمتوسط العام $\bar{\mu}$ والمعرف بالعلاقة التلية:

$$\Delta^2 = \sum_{j=1}^{g} (\boldsymbol{\mu_j} - \bar{\boldsymbol{\mu}})' * \boldsymbol{V^{-1}} * (\boldsymbol{\mu_j} - \bar{\boldsymbol{\mu}}) \qquad (125-4)$$

حيث أن : $\bar{\boldsymbol{\mu}} = \frac{1}{g} \sum_{j=1}^{g} \boldsymbol{\mu_j}$

ولنفترض أن القيم المرتبة التالية: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_s \dots \geq \lambda_p$، هي القيم الذاتية للمصفوفة $(\boldsymbol{V^{-1}B})$ أو لـ $\left(\boldsymbol{V^{-\frac{1}{2}}BV^{-\frac{1}{2}}}\right)$، وأن القيم$s$ الأولى منها موجبة: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_s > 0$، وأن القيم الأخيرة معدومة: $\lambda_{s+1} = \lambda_{s+2} = \lambda_{s+3} = \cdots \lambda_p = 0$، ولنبرهن الآن على أن Δ^2 يساوي مجموع القيم الذاتية، أي أن:

$$\Delta^2 = \lambda_1 + \lambda_2 + \lambda_3 + \cdots \lambda_s + \lambda_{s+1} + \cdots . \lambda_p \qquad (126-4)$$

البرهان: لنفترض أن E هي المصفوفة المتعامدة المؤلفة من الأشعة الذاتية الواحدية $\boldsymbol{e_i}$ للمصفوفة $\left(\boldsymbol{V^{-\frac{1}{2}} * B * V^{-\frac{1}{2}}}\right)$ والمقابلة للقيم الذاتية λ_i . حيث أن: $i: 1\ 2\ 3 \dots P$

ولنرمز الآن لشعاع التوابع التميزية بـY ونعالجه كما يلي:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\ell'_1 X} \\ \boldsymbol{\ell'_2 X} \\ \vdots \\ \boldsymbol{\ell'_s X} \\ \vdots \\ \boldsymbol{\ell'_p X} \end{bmatrix} = \begin{bmatrix} \boldsymbol{e'_1 V^{-\frac{1}{2}} * X} \\ \boldsymbol{e'_2 V^{-\frac{1}{2}} * X} \\ \vdots \\ \boldsymbol{e'_s V^{-\frac{1}{2}} * X} \\ \vdots \\ \boldsymbol{e'_p V^{-\frac{1}{2}} * X} \end{bmatrix} = \boldsymbol{E' * V_y^{-\frac{1}{2}} * X} \qquad (127-4)$$

وذلك على اعتبار أن : $\boldsymbol{\ell_i = V^{-\frac{1}{2}} e_i}$ و $\boldsymbol{\ell'_i = e'_i V^{-\frac{1}{2}}}$

وبما أنه لدينا μ_{jy} يساوي التوقع $E(\boldsymbol{Y}/G_j)$ وأن: $E(\boldsymbol{X}/G_j) = \boldsymbol{\mu}_j$ فيكون لدينا :

$$\boldsymbol{\mu_{jy}} = E(\boldsymbol{Y}/G_j) = E\left(\boldsymbol{E' * V^{-\frac{1}{2}} * X}/G_j\right) = \boldsymbol{E' * V^{-\frac{1}{2}}} * E(\boldsymbol{X}/G_j) = \boldsymbol{E' * V^{-\frac{1}{2}} * \mu_j} \quad (\mathbf{128-4})$$

كما نجد أن المتوسط أو المركز العام $\bar{\boldsymbol{\mu}}_{\boldsymbol{y}}$ يساوي: $\bar{\boldsymbol{\mu}} = \boldsymbol{E'} * V^{-\frac{1}{2}} * \bar{\boldsymbol{\mu}}$ وأن الفرق بينهما يساوي:

$$(\boldsymbol{\mu_{jy}} - \bar{\boldsymbol{\mu}}_{\boldsymbol{y}}) = \boldsymbol{E' * V_y^{-\frac{1}{2}}} * \bar{\boldsymbol{\mu}}_{\boldsymbol{j}} - \boldsymbol{E' * V^{-\frac{1}{2}}} * \bar{\boldsymbol{\mu}}$$
$$= \boldsymbol{E' * V^{-\frac{1}{2}}}(\boldsymbol{\mu_j} - \boldsymbol{\mu}) \qquad (129-4)$$

وهكذا نجد أن المسافات الاحصائية من المراكز $\boldsymbol{\mu_{jy}}$ إلى المركز العام$\bar{\boldsymbol{\mu}}_{\boldsymbol{y}}$ تساوي حسب(4−123) ما يلي:

$$D_j^2(y) = (\boldsymbol{\mu_{jy}} - \bar{\boldsymbol{\mu}}_{\boldsymbol{y}})' * I^{-1} * (\boldsymbol{\mu_{jy}} - \bar{\boldsymbol{\mu}}_{\boldsymbol{y}}) = \left[\boldsymbol{E' * V_y^{-\frac{1}{2}}}(\boldsymbol{\mu_j} - \bar{\boldsymbol{\mu}})\right]' * \left[\boldsymbol{E' * V_y^{-\frac{1}{2}}}(\boldsymbol{\mu_j} - \bar{\boldsymbol{\mu}})\right]$$

$$= (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})' * \boldsymbol{V_y}^{-\frac{1}{2}} * \boldsymbol{E} * \boldsymbol{E}' * \boldsymbol{V_y}^{-\frac{1}{2}} * (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})$$

ولكن بما أن: $\boldsymbol{E} * \boldsymbol{E}' = \boldsymbol{I}$ لأن $\boldsymbol{E}$ متعامدة، وأن $\boldsymbol{V_y}^{-\frac{1}{2}} * \boldsymbol{V_y}^{-\frac{1}{2}} = \boldsymbol{V_y^{-1}}$ لأن $\boldsymbol{V_y^{-1}}$ محددة إيجابياً، فإننا نجد أن :

$$D_j^2(y) = (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})' * \boldsymbol{V_y^{-1}} * (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}) \qquad (130-4)$$

وبمقارنة (4–130) مع العلاقة (4–125)، مع ملاحظة أن: $\boldsymbol{V_y} = \boldsymbol{I}$، نحصل على أن:

$$\Delta^2 = \sum_{j=1}^{g} D_j^2(y) = \sum_{j=1}^{g} (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})'(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}) = \sum_{j=1}^{g}\sum_{i=1}^{p} (\mu_{jyi} - \bar{\mu}_{yi})^2$$

نبدل موقعي المجموع فنجد أن:

$$\Delta^2 = \sum_{i=1}^{p}\sum_{j=1}^{g} (\mu_{jyi} - \bar{\mu}_{yi})^2 = \sum_{i=1}^{p} \Delta_i^2 \qquad (131-4)$$

والآن لنأخذ التابع التمييزي الأول $y_1 = \boldsymbol{\ell'_1 X}$ ونضع $i = 1$ في (4–131)، فنجد أن الحد الأول من المجموع الاخير يساوي:

$$\Delta_1 = \sum_{j=1}^{g} (\mu_{jy1} - \bar{\mu}_{y1})^2$$

ومن العلاقة (4–121) نجد أن :

$$\Delta_1^2 = \sum_{j=1}^{g} \left[\boldsymbol{e'_1 V^{-\frac{1}{2}} \mu_j} - \boldsymbol{e'_1 V^{-\frac{1}{2}}} \overline{\boldsymbol{\mu}}\right]^2 = \sum_{j=1}^{g} \boldsymbol{e'_1} * \boldsymbol{V^{-\frac{1}{2}}} * (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})^2 * \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_1}$$

$$\Delta_1^2 = \sum_{j=1}^{g} \boldsymbol{e'_1} * \boldsymbol{V^{-\frac{1}{2}}} * (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})'(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}}) * \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_1}$$

$$\Delta_1^2 = \boldsymbol{e'_1 V^{-\frac{1}{2}}} \left[\sum_{j=1}^{g} (\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})'(\boldsymbol{\mu_j} - \overline{\boldsymbol{\mu}})\right] \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_1} \qquad \mathbf{(132-4)}$$

وبناء على العلاقة (4–54) والعلاقة (4–98) وعلى مسألة التعظيم في الفصل (2) من الجزء الأول، نحصل على أن: (علماً بأن $\boldsymbol{e'_1} * \boldsymbol{e} = 1$):

$$\Delta_1^2 = \boldsymbol{e'_1} * \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_1} = \lambda_1 \qquad (133-4)$$

وذلك لأن الشعاع $\boldsymbol{e_1}$ يقابل القيمة الذاتية الكبرى λ_1 للمصفوفة $\left(\boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{-\frac{1}{2}}}\right)$ أو للمصفوفة المكافئة لها $(\boldsymbol{V^{-1}} * \boldsymbol{B})$، ولمتابعة البرهان نأخذ التابع التمييزي الثاني $Y_2 = \boldsymbol{\ell'_2 X}$ فنحصل بطريقة مشابهة على:

$$\Delta_2^2 = \sum_{j=1}^{g} (\mu_{jy2} - \bar{\mu}_{y2})^2 = \boldsymbol{e'_2 V^{-\frac{1}{2}}} * \boldsymbol{B} * \boldsymbol{V^{-\frac{1}{2}}} * \boldsymbol{e_2} = \lambda_2 \qquad \mathbf{(134-4)}$$

حيث أن: λ_2 هي القيمة الذاتية الثانية للمصفوفة $\left(\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}}\right)$ أو للمصفوفة المكافئة لها $(\boldsymbol{V}^{-1} * \boldsymbol{B})$.

ولمتابعة البرهان نأخذ التابع الأخير $Y_p = \boldsymbol{\ell}'_p\boldsymbol{X}$، ومنه نحصل بطريقة مشابهة على أن:

$$\Delta_p^2 = \sum_{j=1}^{g}\left(\mu_{jyp} - \bar{\mu}_{yp}\right)^2 = \boldsymbol{e}'_{\boldsymbol{p}}\boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{B} * \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{e}_{\boldsymbol{p}} = \lambda_p \qquad (135-4)$$

وذلك لأن الشعاع $\boldsymbol{e}_{\boldsymbol{p}}$ يقابل القيمة الذاتية الأخيرة λ_p، وبالعودة إلى العلاقة (4−131) نجد أن :

$$\Delta^2 = \sum_{i=1}^{g}\Delta_i^2 = \lambda_1 + \lambda_2 + \lambda_3 + \cdots \lambda_s + \cdots + \lambda_p \qquad (136-4)$$

وهو المطلوب، ونكتب ذلك كما يلي:

$$\Delta^2 = \sum_{i=1}^{g}\left(\boldsymbol{\mu}_{\boldsymbol{j}} - \overline{\boldsymbol{\mu}}\right)' * \boldsymbol{V}^{-\mathbf{1}} * \left(\boldsymbol{\mu}_{\boldsymbol{j}} - \overline{\boldsymbol{\mu}}\right) = \sum_{i=1}^{p}\lambda_i \qquad (137-4)$$

علماً بأنه لدينا أيضاً أن: $\sum_{i=1}^{p}\lambda_i = trace(\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B})$ (انظر الفصل (2) من الجزء الأول).

نظرية: إذا كانت بعض القيم الذاتية الأولى للمصفوفة $(\boldsymbol{V}^{-\mathbf{1}} * \boldsymbol{B})$ موجبة والمتبقية معدومة كما يلي:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_s > 0 \qquad \text{(الموجبة)}$$

$$\lambda_{s+1} = \lambda_{s+2} = \lambda_{s+3} = \cdots \lambda_p = 0 \qquad \text{(المعدومة)}$$

فإن المجموع المقابل للقيم الذاتية المعدومة والمعرف بالعلاقة التالية:

$$C = \sum_{i=s+1}^{g}\left(y_i - \mu_{jyi}\right)^2 = constant \qquad (138-4)$$

يأخذ قيمة ثابتة في جميع المجموعات G_j حيث $j: 1\ 2\ 3 \ldots g$

وإن القيم الموجبة s الأولى لـ λ_i فقط ، هي التي تساهم في عملية التصنيف، أي أن المجموع المعرف بالعلاقة :

$$S = \sum_{i=1}^{s}\left(y_i - \mu_{jyi}\right)^2 \qquad j: 1\ 2\ 3 \ldots g \qquad (139-4)$$

هو فقط الذي يساهم في عملية تصنيف العناصر على المجموعات، وإضافة إلى ذلك، إذا كانت الاحتمالات السابقة للمجموعات متساوية وتساوي: $P_1 = P_2 = P_3 = \cdots P_g = \frac{1}{g}$ فإن القاعدة (4−124) تكافي قاعدة أصغر الاحتمالات TPM المعرفة في (4−37) .

البرهان: لنأخذ المسافة المربعة من **X** حتى $\boldsymbol{\mu}_{\boldsymbol{j}}$ ونكتبها كما يلي: [Johnson Wichern. P.524]:

$$\begin{aligned} D_j^2(X) &= \left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{j}}\right)' * \boldsymbol{V}^{-\mathbf{1}} * \left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{j}}\right) = \left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{j}}\right)' * \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{V}^{-\frac{1}{2}} * \left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{j}}\right) \\ &= \left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{j}}\right)' * \boldsymbol{V}^{-\frac{1}{2}} * \boldsymbol{E} * \boldsymbol{E}' * \boldsymbol{V}^{-\frac{1}{2}} * \left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{j}}\right) \qquad (140-4) \end{aligned}$$

حيث أن : $E = [e_1 , e_2 \dots e_p]$ وهي المصفوفة المتعامدة والتي أعمدتها مؤلفة من الأشعة الذاتية e_i للمصفوفة $\left(V^{-\frac{1}{2}}BV^{-\frac{1}{2}}\right)$، وبما أنه لدينا مما سبق أن:

$$\ell_i = V^{-\frac{1}{2}}e_i \qquad \ell_i' = e_i' * V^{-\frac{1}{2}} \;\Rightarrow\; \ell' = E'V^{-\frac{1}{2}}$$

فإنه يمكننا كتابة النصف الأيمن من الطرف الايمن من (4−140) كما يلي:

$$E' * V^{-\frac{1}{2}}(X-\mu_j) = \ell'(X-\mu_j) = \begin{bmatrix} \ell_1'(X-\mu_j) \\ \ell_2'(X-\mu_j) \\ \vdots \\ \ell_p'(X-\mu_j) \end{bmatrix} \qquad (141-4)$$

ومنها نجد أنه يمكننا كتابة $D_j^2(X)$ كما يلي:

$$D_j^2(X) = \left[(X-\mu_j)'V^{-\frac{1}{2}} * E\right] * \left[E' * V^{-\frac{1}{2}} * (X-\mu_j)\right] = \left[E' * V^{-\frac{1}{2}} * (X-\mu_j)\right]' * \left[E' * V^{-\frac{1}{2}} * (X-\mu_j)\right]$$

$$= [\ell'(X-\mu_j)]'[\ell'(X-\mu_j)] = \sum_{i=1}^{p}[\ell_i(X-\mu_j)]^2 \qquad (142-4)$$

بما أن الأشعة $\ell_i = V^{-\frac{1}{2}}e_i$ من أجل جميع قيم $i > s$ هي أشعة ذاتية غير ممعيرة (عادية) للمصفوفة $(V^{-1}B)$، ومقابلة للقيم الذاتية المعدومة . وبما أن الشعاع ℓ_i (حسب العلاقة (4−74)) يتعامد مع كل شعاع $(\mu_j - \mu)$، ويكون لدينا: $\ell_i'(\mu_j - \bar{\mu}) = 0$.

وبما أن الفرق بين أي شعاعين $(\mu_k - \bar{\mu})$ و $(\mu_j - \bar{\mu})$ يساوي ما يلي:

$$(\mu_k - \bar{\mu}) - (\mu_j - \bar{\mu}) = (\mu_k - \mu_j) \qquad j,k = 1\ 2\ 3 \dots g$$

فإننا نستنتج أن:

$$\ell_i'(\mu_k - \bar{\mu}) + \ell_i'(\mu_j - \bar{\mu}) = \ell_i'(\mu_k - \mu_j) = 0 \qquad (143-4)$$

ولكن من جهة أخرى نجد أن نتيجة الجداء السابق تساوي:

$$\ell_i'(\mu_k - \mu_j) = \mu_{kyi} - \mu_{jyi} = 0 \qquad For \qquad i > s$$

ومنها نستنتج أن:

$$\mu_{kyi} = \mu_{jyi} = constant \qquad For \qquad i > s \qquad (144-4)$$

ثم نطرح من الطرفين القيمة العددية y_i فنجد أن :

$$(y_i - \mu_{kyi}) = (y_i - \mu_{jyi}) = constant \quad : \quad i > s \qquad (145-4)$$

ومما سبق نستنتج أن المجموع التالي يساوي مقداراُ ثابتاً:

$$\sum_{i=S+1}^{p}(y_i - \mu_{kyi})^2 = constant \qquad (146-4)$$

وذلك في جميع المجموعات G_j ولا يؤثر في عمليات التمييز .

وهكذا نستنتج أن التوابع التميزية الأولى فقط هي التي نحتاجها للاستخدام في عملية التصنيف . وذلك لأن المجموع (4−146) يدل على أن التوابع الأخيرة $Y_i(i > s)$ المقابلة للقيم الذاتية المعدومة لا تلعب دوراً في عملية التصنيف لأن تأثيرها ثابت في جميع المجموعات .

وهكذا نجد أنه إذا أخذنا الـ r تابعاً الأولى من التوابع التمييزية (حيث $r \leq s$) فإنه يمكننا أن نصيغ قاعدة التصنيف كما يلي:

نقوم بتصنيف العنصر $\boldsymbol{x}$ في المجموعة G_k إذا كان:

$$D_k^2(X) = \sum_{i=1}^{r} \left(y_i - \mu_{kyi}\right)^2 = \sum_{i=1}^{r} [\boldsymbol{\ell}_i'(\boldsymbol{X} - \boldsymbol{\mu_k})]^2 \leq \sum_{i=1}^{r} \left[\boldsymbol{\ell}_i'\left(\boldsymbol{X} - \boldsymbol{\mu_j}\right)\right]^2 \qquad (147-4)$$

وذلك من أجل جميع قيم : $g \ldots 3\ 2\ 1\ :\ k \neq j$

وبعبارة أخرى نقوم بتصنيف العنصر X في المجموعة G_k إذا كان :

$$D_k^2(X) = \sum_{i=1}^{r} [\boldsymbol{\ell}_i'(\boldsymbol{X} - \boldsymbol{\mu_k})]^2 = \min_j \left[\sum_{i=1}^{r} \left[\boldsymbol{\ell}_i'\left(\boldsymbol{X} - \boldsymbol{\mu_j}\right)\right]^2\right] \qquad (148-4)$$

وذلك لأجل $g \ldots 3\ 2\ 1\ :\ k \neq j$ و $r < s$

وعندما تكون معالم المجتمع مجهولة فإننا نقدرها من مؤشرات العينات المسحوبة من المجموعات، فتصبح قاعدة (فيشر) لتصنيف العناصر في حالة العينات كما يلي:

نصنف العنصر x في المجموعة G_k إذا كان:

$$d_k^2(X) = \sum_{i=1}^{r} \tilde{\boldsymbol{\ell}}_i'(\boldsymbol{X} - \overline{\boldsymbol{X}}_{\boldsymbol{k}}) = \min \left[\sum_{i=1}^{r} \left[\boldsymbol{\ell}_i'\left(\boldsymbol{X} - \overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)\right]^2\right] \qquad (149-4)$$

وذلك من أجل $g \ldots 3\ 2\ 1\ :\ k \neq j$ و $r < s$

وأخيراً نشير إلى أنه إذا كان r عدد التوابع التمييزية المستخدمة في التصنيف أصغر من s $(r \leq s)$, فإننا نفقد جزءاً صغيراً من القيمة الكلية للمسافة يساوي $\sum_{i=r+1}^{p}\left[\boldsymbol{\ell}_i'\left(\boldsymbol{X} - \overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)\right]^2$، وذلك من أجل كل مجموعة G_j أو من القيمة المختصرة $\sum_{i=1}^{r}\left[\boldsymbol{\ell}_i'\left(\boldsymbol{X} - \overline{\boldsymbol{X}}_{\boldsymbol{j}}\right)\right]^2$، والتي تسمى بالجزء النافع المستخدم في عملية التصنيف .

وهكذا نجد أن التوابع التمييزية القليلة الأولى تكون أكثر أهمية من التوابع الأخيرة . وإن التابع التمييزي الأول $\left(y_1 = \boldsymbol{e}_1' * \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\right)$ يشكل المساهمة الكبرى في تفسير قيمة مربع المسافة Δ^2 وتبلغ مساهمته فيها مقداراً يساوي λ_1 .

وكذلك نجد أن التابع التمييزي الكائي $y_k = \boldsymbol{e}_k' * \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}$ يساهم في تفسير قيمة مربع المسافة Δ^2 بمقدار λ_k، وإن التوابع التمييزية الأولى من 1 حتى r تشكل مساهمة شبه كاملة في تفسير قيمة Δ^2، وإن جميع التوابع التمييزية الأخرى $(s - r)$ تشكل مساهمة قدرها $(\lambda_{r+1} + \lambda_{r+2} + \cdots \lambda_s)$، وإن قيمة هذه المساهمة تكون صغيرة مقارنة مع مساهمات التوابع التمييزية الأولى والبالغة $(\lambda_1 + \lambda_2 + \lambda_3 +$

$(\lambda_r \cdots$ لذلك فإنه يمكننا اهمال التوابع التمييزية الأخيرة $(y_{r+1} + y_{r+2} + \cdots y_s)$ بدون خسارة ملحوظة في عمليات التصنيف .

مثال (4−5): لنعود إلى المثال (4−4) السابق ولنأخذ التابعين المميزين (لفيشر) اللذين حصلنا عليهما منه وهما:

$$\tilde{Y}_1 = \tilde{\boldsymbol{\ell}}'_1 \boldsymbol{X} = 0.385x_1 + 0.495x_2$$
$$\tilde{Y}_2 = \boldsymbol{\ell}'_2 \boldsymbol{X} = 0.938x_1 - 0.112x_2$$

وهما يشكلان مستوى $Y_1 0 Y_2$ في الفضاء R^3 ونسميه المستوى التمييزي . ولتصنيف أي عنصر جديد مثل $x_0 = \begin{bmatrix}1\\3\end{bmatrix}$ حسب القاعدة الأخيرة (4−148) نحسب إحداثياته في المستوى التمييزي $Y_1 0 Y_2$ فنجد بعد التعويض أن:

$$\tilde{Y}_1 = 0.385x_{01} + 0.495x_{02} = 0.385(1) + 0.495(3) = 1.87$$
$$\tilde{Y}_2 = 0.938x_{01} - 0.112x_{02} = 0.938(1) - 0.112(3) = 0.60$$

وهما إحداثيات العنصر x_0 في المستوى التمييزي $Y_1 0 Y_2$.

ومن جهة أخرى نقوم بحساب إحداثيات مراكز المجموعات $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ في المستوى التمييزي $Y_1 0 Y_2$، فنجد أن إحداثيات مركز المجموعة الأولى $\bar{x} = \begin{bmatrix}-1\\3\end{bmatrix}$ في المستوى $Y_1 0 Y_2$ تساوي: $\bar{y}_1$ ونحسبها كمايلي :

$$\left.\begin{aligned}\bar{y}_{11} &= \tilde{\ell}'_1 \bar{x}_1 = [0.385\,, 0.495]\begin{bmatrix}-1\\3\end{bmatrix} = 1.10\\ \bar{y}_{12} &= \tilde{\ell}'_2 \bar{x}_1 = [0.938\,, -0.112]\begin{bmatrix}-1\\3\end{bmatrix} = -1.27\end{aligned}\right. \Longrightarrow \bar{y}_1 = \begin{bmatrix}1.10\\-1.27\end{bmatrix}$$

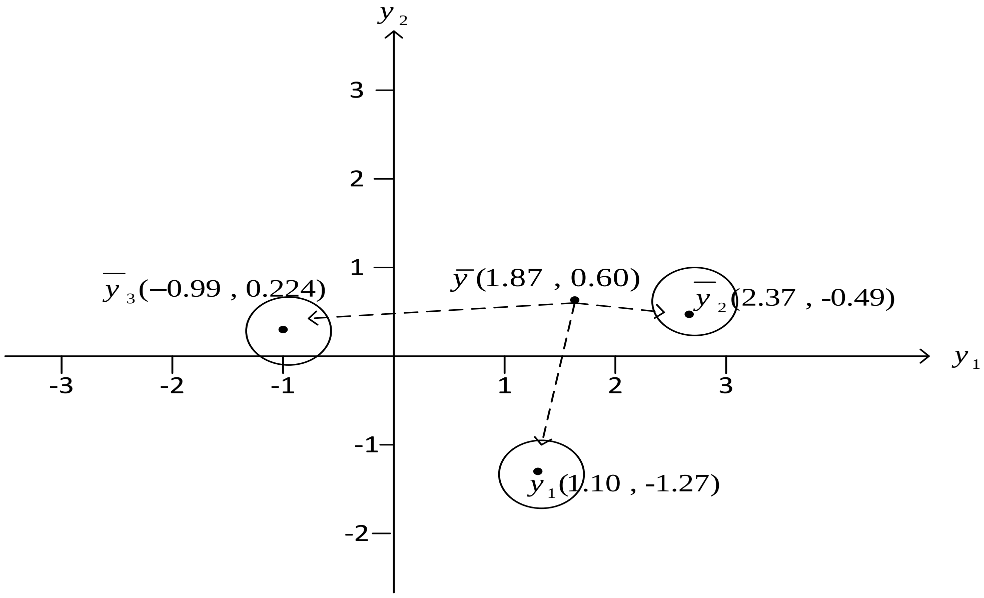
وبطريقة مشابهة نجد أن إحداثيات مركز المجموعة الثانية $\bar{x}_2 = \begin{bmatrix}1\\4\end{bmatrix}$ في المستوى $Y_1 0 Y_2$، تساوي: $\bar{y}_2$ ونحسبها كما يلي:

$$\left.\begin{aligned}\bar{Y}_{21} &= \tilde{\ell}'_1 \bar{x}_2 = [0.385\,, 0.495]\begin{bmatrix}1\\4\end{bmatrix} = 2.37\\ \bar{Y}_{22} &= \tilde{\ell}'_2 \bar{x}_2 = [0.938\,, -0.112]\begin{bmatrix}1\\4\end{bmatrix} = 0.49\end{aligned}\right. \Longrightarrow \quad \bar{y}_2 = \begin{bmatrix}2.37\\0.49\end{bmatrix}$$

وكذلك نجد أن إحداثيات مركز المجموعة الثالثة $\bar{x}_3 = \begin{pmatrix}0\\-2\end{pmatrix}$ في المستوى $Y_1 0 Y_2$، تساوي: $\bar{y}_3$ ونحسبها كما يلي:

$$\left.\begin{aligned}\bar{y}_{31} &= \tilde{\ell}'_1 \bar{x}_3 = [0.385\,, 0.495]\begin{bmatrix}0\\-2\end{bmatrix} = -0.99\\ \bar{y}_{32} &= \tilde{\ell}'_2 \bar{x}_3 = [0.938\,, -0.112]\begin{bmatrix}0\\-2\end{bmatrix} = 0.224\end{aligned}\right. \Longrightarrow \quad \bar{y}_3 = \begin{bmatrix}-0.99\\0.224\end{bmatrix}$$

وإن هذه الاحداثيات تمثل على المستوى التمييزي $Y_1 0 Y_2$ كما يلي:



الشكل $(4-3)$: التمثيل البياني لإحداثيات المراكز على المستوى التمييزي $Y_1 0 Y_2$

ومن الشكل نلاحظ أن الإحداثيات المقابلة للعنصر $X = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ في $Y_1 0 Y_2$ هي إحداثيات النقطة $\bar{y}(1.87\ , 0.60)$ ، وإن أصغر مسافة (أو مربع المسافة) بينها وبين إحداثيات مراكز المجموعات $\bar{y}_1$ ، $\bar{y}_2$ ، $\bar{y}_3$ هي المسافة منها حتى المركز $\bar{y}_2$، لذلك نصنف العنصر x في المجموعة الثانية G_2 لأنها أقرب إلى مركزها $\bar{y}_2$.

وبما أن الأسلوب البياني غير متوفر في الفضاءات العليا. فإننا نقوم بحساب مربعات المسافات من النقطة $\bar{y}(1.87\ , 0.60)$ إلى مراكز المجموعات المختلفة باستخدام العلاقة :

$$D_j^2(y) = \sum_{j=1}^{2} \left(\tilde{y}_i - \bar{y}_{kj}\right)^2 = \sum_{j=1}^{2} \left[\tilde{\boldsymbol{\ell}}_j'(\boldsymbol{X} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})\right]^2$$

وذلك من أجل جميع قيم k الخاصة بالمجموعات الثلاثة .

فمن أجل المجموعة الأولى نضع $k = 1$ فنجد أن:

$$D_1^2(y) = \sum_{i=1}^{2} (\tilde{y}_i - \bar{y}_{1i})^2 = (1.87 - 1.10)^2 + (0.60 + 1.27)^2 = 4.09$$

ومن أجل المجموعة الثانية نضع $k = 2$ فنجد أن:

$$D_2^2(y) = \sum_{i=1}^{2} (\tilde{y}_i - \bar{y}_{2i})^2 = (1.87 - 2.37)^2 + (0.60 - 0.49)^2 = 2.26$$

ومن أجل المجموعة الثالثة نضع $k = 3$ فنجد أن:

$$D_3^2(y) = \sum_{i=1}^{2} (\tilde{y}_i - \bar{y}_{3i})^2 = (1.87 + 0.99)^2 + (0.60 - 0.224)^2 = 8.22$$

وعند مقارنة مربعات المسافات السابقة نجد أن أصغرها هو المربع $D_2^2(y)$ لذلك نصنف العنصر $x_0 = \begin{bmatrix}1\\3\end{bmatrix}$ المقابل لـ y في المجموعة G_2 .

4-4 : خطوات بناء النموذج التمييزي:

تتألف عملية بناء وتحليل النموذج التمييزي من الخطوات التالية:

1- تحديد المتحولات المستقلة *X* وتحديد المجموعات في المتحول التابع Y .

2- جمع البيانات اللازمة عن المتحولات والمجموعات وإنشاء مصفوفة البيانات .

3- إجراء الاختبارات اللازمة على البيانات النهائية. وذلك للتأكد من تحقيق الشروط المفروضة على النموذج فيها، وأهم هذه الاختبارات هي:

- اختبار التوزيع الطبيعي، وهو خضوع المتحولات الكمية لتوزيعات طبيعية، وذلك باستخدام اختبار *ks* (كولموغوروف- سميرنوف *k-s*)، أو غيره .
- اختبار معنوية الفرق بين متوسطات المجموعات، ويستخدم في ذلك اختبار(*Wilk's lambda*) مع اختبار (*Chi- square*) أو اختبار *t* أو T^2 .
- اختبار تجانس أو تساوي مصفوفات التباين المشترك، ويستخدم اختبار (*Box's – M*)، أو غيره.
- اختبار معنوية أمثال التابع التمييزي، ويستخدم الاختبار (*Wilk's lambda*) والتابع *F* .

علماً بأن هذه الاختبارات معروفة في كثير من المراجع الاحصائية، وموجودة في معظم البرامج الحاسوبية المختصة .

4- إجراء الحسابات اللازمة لتقدير أمثال التوابع التمييزية ثم حساب القيم النظرية لها وحساب مراكزها.

5- حساب توابع التصنيف المقابلة لكل مجموعة .

6- إعداد جدول التصنيف المقارن وحساب معامل جودة التصنيف .

7- إجراء التحليلات اللازمة في كل مرحلة من هذه المراحل .

مثال (4-6): نشر الباحثان (العلي+ علوش. 2018) مقالاً حول تصنيف المشروعات الصغيرة التي كانت قائمة في سورية خلال عام 2016، فكان عددها الاجمالي (784) مشروعاً، وكانت الجهات المسؤولة عنها قد قامت بتقييمها وتصنيفها إلى (3) مجموعات كما يلي:

جدول (4-5)

تقييم المشروع	خاسر G_1	متعثر G_2	ناجح G_3	المجموع
الرمز	1	2	3	-
العدد	113	236	435	784

ومن أجل نمذجة تصنيف هذه المشروعات الصغيرة باستخدام التحليل التمييزي المتعدد، قام الباحثان بتحديد عدة متحولات مؤثرة على تصنيف تلك المشروعات، ثم تم اختصارها إلى سبعة متحولات وتحويلها إلى متحولات رقمية مرتبة كما يلي:

X_1= مكان المشروع (المحافظة): دمشق (1)- ريف دمشق (2)- طرطوس (3)- اللاذقية (4) .
X_2= عدد العمال في المشروع: (1 ، 2 ، 3 ، 4 ، 5) .
X_3= حجم الميزانية الافتتاحية(ل.س): أقل من (1000000)=1، من(1000001-2000000) =2،
من (2000001- 3000000) =3، من (3000001- 4000000) =4
من (4000001- 5000000) =5، أكثر من (5000000) =6
X_4= عمر المشروع بالسنوات: (أقل من سنة) =1، من (سنة إلى سنتين) =2،
من (سنتين إلى ثلاث سنوات) =3، أكثر من (ثلاث سنوات) =4 .
X_5= المؤهل العلمي لصاحب المشروع: أمي (1)، ابتدائي (2)، إعدادي (3)، ثانوي (4)، معهد (5)، جامعي (6) .
X_6= عمر القائم بالإدارة: (أقل من 30 سنة) =1، من (30-40) =2، من (41-50) =3، أكثر من (50 سنة) = 4 .
X_7= درجة المنافسة في السوق: ضعيفة (1)، متوسطة (2)، قوية (3) .

وبعد جمع البيانات عن هذه المتحولات وتنظيمها في جداول خاصة تم حساب مواصفاتها ودراسة خواصها والتحقق من إنها تحقق شروط تطبيق التحليل التمييزي كما يلي:

1- **وصف المتحولات المستقلة:** إن الاحصاءات الوصفية لهذه المتحولات في العينة ككل (ويمكن حسابها في كل مجموعة من المجموعات الثلاثة) كانت كما يلي:

جدول (4-6): الاحصاءات الوصفية للمتحولات

المتحولات	المتوسط *Mean*	انحراف المعياري *St. D*	عدد العناصر *n*	معامل الاختلاف
المحافظة X_1	2,6059	1,0377	784	39,82
عدد العمال X_2	2,6633	1.2701	784	47,69
الموازنة X_3	3,5306	1,0297	784	29,11
عمر المشروع X_4	2,8176	0,9257	784	32,85
المؤهل العلمي للمدير X_5	4,0203	1,3316	784	33,05
عمر القائم بالإدارة X_6	2,7156	08852	784	32,60
المنافسة X_7	2,1492	0,9425	784	43,85

2- **اختبار استقلال المتحولات:** ويمكن أن يتم من خلال المصفوفة الارتباطية، أو من خلال تضخم التباين (*VIF*)، أو من خلال معامل التحمل *Tolerance*، اللذين يساويان :

$$VIF(x_i , x_j) = \frac{1}{1 - r_{ij}^2} = \frac{1}{Tolerance}$$

وعند حساب هذين المعاملين لعلاقة المتحولات المستقلة X مع المتحول التابع Y، كانت جميع قيم (*VIF*) أقل من (5)، وهذا يدل على عدم وجود ارتباط قوي بين المتحولات المستقلة .

3- **اختبار التوزيع الطبيعي:** وعند استخدام اختبار (كولموغوروف- سميرنوف)، لم تكن النتائج إيجابية لبعض المتحولات، ولكن بما أن حجم العينة المدروسة كبير جداً (784 مشروعاً). يمكننا التساهل في ذلك واعتبار أن هذه المتحولات خاضعة تقاربياً للتوزيع الطبيعي. وذلك استناداً إلى نظرية النهاية المركزية وإلى قانون الأعداد الكبيرة في نظرية الاحتمالات .

4- **اختبار تجانس مصفوفات التباين المشترك $(V_1 = V_2 = V_3)$:**

بينت مخرجات هذا الاختبار أن لوغاريتمات المحددات تساوي :

جدول (4-7): قيم لوغاريتمات محددات مصفوفات التباين المشترك

نوع المشروع	الرتبة *rank*	*Log Determinant*
خاسر G_1	4	−1,763
متعثر G_2	4	−1,990
متعثر G_3	4	−1,374
المدمجة ضمن المجموعات	4	−1,474

ومن هذا الجدول نلاحظ أن قيم لوغاريتمات المحددات متقاربة، لذلك يمكن اعتبار أن مصفوفات التباين المشترك متجانسة ، ومنه نلاحظ أيضاً أن رتبة المحدد فيه تساوي (4)، وهذا يشير إلى أن عدد المتحولات المعنوية في النموذج سيكون مساوياً لـ (4) متحولات فقط .

ولاختبار فرضية تجانس مصفوفات التباين المشترك تم استخدام اختبار *Box's M* فحصلنا على ما يلي:

جدول (4-8): اختبار Box's M لتجانس المصفوفات

قيمة اختبار *Box's M*		110,032
F	$Approx_0$	5,445
	df_1	20
	df_2	452960,861
	Sig	0,063

ونلاحظ من هذا الجدول أن قيمة احتمال الدلالة Sig أكبر من $\propto = 0{,}05$، لذلك نقبل فرضية العدم التي تقول: لا يوجد اختلاف بين مصفوفات التباين المشترك: $H_0: V_1 = V_2 = V_3$

5- تحديد المتحولات التي ستدخل في النموذج التمييزي: تمت مراجعة البيانات وإدخالها في برنامج *SPSS* حسب أسلوب (*F to Remove*)، ومن خلال (4) خطوات فقط ثم تحديد المتحولات المؤثرة على النموذج التمييزي . والاقتصار على أربعة متحولات هي كما في الجدول التالي:

جدول (4–9): نتائج خطوات ادخال المتحولات حسب *F to Remove* – من مخرجات *SPSS*

الخطوة	المتحولات الداخلة	*Tolerance*	*F to Remove*	*Min F*	*Between Groups*
1	عدد العمال X_2	1,000	294,575	–	–
2	عدد العمال X_2	1,000	224,412	70,044	متعثر *and* خاسر
	المنافسة X_7	1,000	563,170	89,932	ناجح *and* خاسر
3	عدد العمال X_2	0,990	220,610	91,126	متعثر *and* خاسر
	المنافسة X_7	1,000	560,125	106,904	ناجح *and* خاسر
	الميزانية X_3	0,990	60,329	209,868	ناجح *and* خاسر
4	عدد العمال X_2	0,980	224,887	61,298	متعثر *and* خاسر
	المنافسة X_7	0,997	561,388	72,605	ناجح *and* خاسر
	الميزانية X_3	0,987	57,944	140,882	ناجح *and* خاسر
	عمر المشروع X_4	0,983	3,857	183,358	ناجح *and* خاسر

ونلاحظ من الجدول (4–9) السابق أن برنامج ***SPSS*** وباستخدام أسلوب ***F to Remove*** قد أبقى على (4) متحولات فقط هي: X_2 عدد العمال، X_7 المنافسة، X_3 الميزانية، X_4 عمر المشروع .
ولدراسة مدى تأثير هذه المتحولات على نموذج التصنيف تم حساب قيم الاختبار (Wilk's Lambda) في كل خطوة من خطوات الادخال فحصلنا على الجدول التالي:

جدول (4–10): نتائج اختبار (*Wilk's Lambda*) حسب خطوات الإدخال – من *SPSS*

رقم الخطوة	*Number of variables*	*Wilk's Lambda*	df_1	df_2	df_3	*Exact F*	*Sig*
1	1	0,570	1	2	781	294,578	0,000
2	2	0,233	2	2	781	417,564	0,000
3	3	0,202	3	2	781	318,161	0,000
4	4	0,200	4	2	781	240,455	0,000

ونلاحظ من هذا الجدول أن قيم *Lambda* تناقصت حتى بلغت قيمتها الدنيا عند القيمة 0,200 . وبما أن قيم احتمالات الدلالة *Sig* المقابلة لتلك الخطوات أصغر من 0,05، فإن تأثير هذه المتحولات معنوي.
ولاستخراج معاملات التابعين التمييزيين بين هذه المجموعات الثلاثة، نبحث عن القيم الذاتية للمصفوفة $(W^{-1} * B)$ للمتحولات المدخلة X_4, X_3, X_2, X_1، فنجد أن لها قيمتين ذاتيتين λ_2, λ_1 وتقابلان تابعين تمييزيين F_2, F_1 . وهما معروضتان في الجدول التالي :

جدول (4–11): القيم الذاتية للمصفوفة $(\mathbf{W^{-1} * B})$ – من SPSS

Function	Eigenvalue	% of variance	Cumulative %	Canonical correlation
F_1	2,102	77,5	77,5	0,823
F_2	0,612	22,5	100,0	0,616

وهنا نلاحظ وجود تابعين تمييزيين فقط لأن $(S = g - 1 = 3 - 1 = 2)$، ومن الجدول (4–11) نجد أن التابع الأول F_1 يفسر 77,5% من التباين الكلي، بينما يفسر التابع الثاني F_2 فقط 22,5% ، ولدراسة القوة التمييزية لهذين التابعين تم تطبيق اختبار (*Wilk's lambda*) مع اختبار χ^2 فحصلنا على الجدول التالي:

(4– 12): قيم (Wilk's lambda) مع اختبار χ^2 – من SPSS

اختبار التوابع	Wilk's *lambda*	Chi– square	*df*	Sig
1 through 2	0,200	1254,698	8	0,000
2	0,620	372,177	3	0,000

ومن الجدول نلاحظ أن القوة التميزية لهذين التابعين جيدة ومعنوية ويفسران نسبة كبيرة من التباين بين المجموعات الثلاث .

وأخيراً ننتقل إلى حساب المعاملات العددية (الأمثال) لهذين التابعين التمييزيين، فنحصل على معاملات شبيهة بمعاملات الانحدار المتعدد المعياري وهي المعاملات المعروضة في الجدول التالي:

جدول (4–13): معاملات التابعين التمييزيين القانونيين المعيارين $\mathbf{Z_1}$, $\mathbf{Z_2}$

Standardized canonical Discriminant Function Coefficients		
المتحولات المعيارية *x*	أمثال Z_1	أمثال Z_2
عدد العمال x_2	0,486	−0,750
الميزانية الافتتاحية x_3	−0,078	0,579
عمر المشروع x_4	−0,094	0,102
المنافسة x_7	0,875	0,441

ومن الجدول السابق نحصل على التابعين التمييزيين المعياريين التاليين :

$$Z_1 = 0{,}486x_2 - 0{,}078x_3 - 0{,}094x_4 + 0{,}875x_7$$
$$Z_2 = -0{,}750x_2 + 0{,}579x_3 + 0{,}102x_4 + 0{,}114x_7$$

ومن هذين التابعين نلاحظ أن المتحول الأكثر تأثيراً على التابع Z_1 هو متحول المنافسة X_7 . ثم يأتي متحول عدد العمال X_2 . أما بالنسبة للتابع التمييزي Z_2 فنجد أن عدد العمال الم و X_2 هو المتحول الأكثر تأثيراً عليه، ثم متحول الميزانية الافتتاحية X_3 ثم متحول المنافسة X_7 .

ثم نقوم بحساب المعاملات (الأمثال) القانونية غير المعيارية لهذين التابعين المتميزين فنحصل على تابعين جديدين بدلالة القيم الأساسية للمتحولات X_2 , X_3 , X_4 , X_7 كما هو موضح في الجدول التالي:

جدول (4-14): معاملات التابعين التمييزين F_1 , F_2 القانونيين الخام (غير المعياريين)- من SPSS

canonical Discriminant Function Coefficients		
المتحولات العادية	التابع التمييزي F_1	التابع التمييزي F_2
عدد العمال X_2	0,507	−0,781
الميزانية X_3	−0,081	0,606
عمر المشروع X_4	−0,102	0,111
المنافسة X_7	1,529	0,770
الحد الثابت *constant*	−4,062	−2,029

ومن الجدول السابق نحصل على التابعين التميزيين القانونيين F_1 , F_2 التاليين :

$$F_1 = -4{,}062 + 0{,}507X_2 - 0{,}081X_3 - 0{,}102X_4 + 1{,}529X_7$$
$$F_2 = -2{,}029 - 0{,}781X_2 + 0{,}606X_3 + 0{,}111X_4 + 0{,}770X_7$$

ونلاحظ منهما أن أكثر المتحولات تأثيراً في F_1 هو X_7 ، ثم X_2، وأن أكثر المتحولات تأثيراً في F_2 هي: X_2 و X_7 و X_3 .

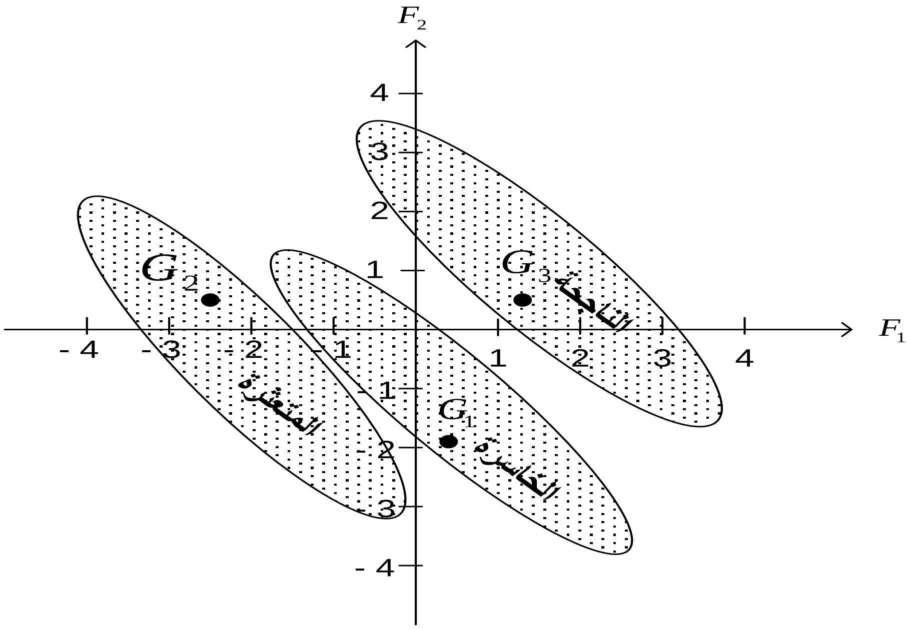
ثم نقوم بحساب القيم النظرية لهذين التابعين F_1 , F_2 من المعادلتين السابقتين، وذلك بتعويض قيم المتحولات الداخلة في معادلة كل منهما، فنحصل على عمودين جديدين يضمان *n* زوجاً متقابلاً (F_1 , F_2) .

وإن قيم هذين التابعين تقابل مجموعات التصنيف الثلاثة (خاسر- متعثر- ناجح)، لذلك نقوم بحساب مركز كل تابع (متوسط قيمه) في كل مجموعة على حدة فنحصل على الجدول التالي:

جدول (4-15): مراكز التابعين في المجموعات الثلاثة- من SPSS

Function at Group canonical		
مجموعة المشاريع	مراكز التابع F_1	مراكز التابع F_2
الخاسر G_1	0.056	−1,902
المتعثر G_2	−2,133	0,302
الناجحة G_3	1,143	0,330

وعندما قمنا بتمثيل القيم النظرية للتابعين F_1 , F_2 وحسبنا مراكز هذه المجموعات على المستوى فحصلنا على الشكل التالي:



الشكل (4–4): شكل انتشار قيم F_1 , F_2 على المستوى

ولحساب توابع التصنيف المقابلة لهذه المجموعات اعتمدنا على العلاقة (4–33) التي تنص على التالي:

يتم تصنيف X_0 في المجموعة k إذا كان: $d_k(X_0) = \max[d_j(X_0)]$

حيث أن $d_j(X)$ تحسب من العلاقة:

$$d_j(X) = \bar{x}'_j * \bar{S} * P * X - \frac{1}{2}\bar{x}'_j * S_P^{-1} * \bar{x}_j + \ln P_j$$

وبذلك وجدنا أن معاملات (أمثال) توابع التصنيف تساوي ما يلي:

جدول (4–16): معاملات (أمثال) توابع التصنيف – من SPSS

Classification Function Coefficients			
المتحولات	تقييم نوع المشروع		
	الخاسر G_1	المتعثر G_2	الناجحة G_3
عدد العمال X_2	3,764	0,933	2,571
الميزانية X_3	2,383	3,898	3,649
عمر المشروع X_4	2,388	2,855	2,525
المنافسة X_7	5,065	3,415	8,446
الثابت	−19,121	−15,209	−26,959

وبذلك نكون قد حصلنا على توابع التصنيف المقابلة للمجموعات الثلاثة G_1 , G_2 , G_3 التالية:

$$D_1 = -19{,}121 + 3{,}764X_2 + 2{,}383X_3 + 2{,}388X_4 + 5{,}065X_7$$

$$D_2 = -15{,}209 + 0{,}933X_2 + 3{,}898X_3 + 2{,}855X_4 + 3{,}415X_7$$

$$D_3 = -26{,}956 + 2{,}571X_2 + 3{,}649X_3 + 2{,}525X_4 + 8{,}446X_7$$

واعتماداً على هذه التوابع الثلاثة: تم حساب القيمة التصنيفية لكل مشروع X من مشروعات العينة (784 مشروعاً). ثم المقارنة بينها واختيار القيمة الأكبر منها (من التوابع الثلاثة)، ومنها تم تحديد انتماء ذلك المشروع X إلى المجموعة المقابلة لتلك القيمة الكبرى من قيم $D_j(X)$.

ونتيجة لهذه العمليات تم الحصول على تصنيف جديد لمشاريع العينة يسمى بالتصنيف المتنبأ به (*Predicted*)، وبعدها تمت مقارنة التصنيف الجديد مع التصنيف الأصلي (*Original*) وتبويب نتائج المقارنة في جدول التقاطع التالي:

جدول (4– 17): نتائج تقاطع التصنيف المتنبأ به مع التصنيف الأصلي

Classification Results						
المجموع الأفقي n_i	الناجحة G_3	المتعثر G_2	الخاسر G_1	تقييم نوع المشروع		
113	14	3	96	الخاسر G_1	التكرارات counts (الأعداد)	التصنيف الأصلي *Original*
236	286	200	18	المتعثر G_2		
435	386	21	28	الناجحة G_3		
784	**418**	**224**	**142**	**المجموع العمودي C_i ←**		
100,0	12,389	2,655	84,956	الخاسر G_1	النسب المئوية %	
100,0	7,627	84,627	7,627	المتعثر G_2		
100,0	88,736	4,828	6,437	الناجحة G_3		
A 86,98% of original grouped cases correctly classified						

نلاحظ من هذا الجدول أن: 84,956% من المشروعات الخاسرة أصلاً قد تم تصنيفها خاسرة (أي أن نسبة التصنيف الصحيح فيها بلغت (84,956%) . كما أن نسبة التصنيف الصحيح للمشروعات المتعثرة أصلاً قد بلغت (84,627%) وكذلك نجد أن نسبة التصنيف الصحيح للمشروعات الناجحة أصلاً قد بلغت (88,736%) . أما النسبة الكلية للتصنيف الصحيح في المجموعات الثلاثة فنحسبها من تقسيم مجموع العناصر القطرية على حجم العينة n فنجد أنها تساوي:

$$R = \frac{96 + 200 + 386}{784} * 100 = 86,98\ \%$$

وهي نسبة جيدة جداً، وتدل على أنه يمكننا الاعتماد على توابع التصنيف السابقة في تصنيف أي مشروع جديد قبل إعطاء الموافقة على إنشائه ،وذلك باستخدام المتحولات الأربعة المذكورة.

اما النسبة الكلية للتصنيف غير الصحيح (معدل التصنيف الخاطئ) فنحسبها من العلاقة:

$$Q = 100 - R = 100 - 86,98 = 13,02\ \%$$

وهو معدل معقول ومقبول في عمليات التصنيف الاقتصادية .

وهناك مقاييس أخرى لجودة التصنيف مثل المؤشر *Kappa* الذي يعرف بالعلاقة:

$$Kappa = \frac{n\sum n_{ii} - \sum r_i * C_i}{n^2 - \sum r_i * C_i}$$

حيث أن: n_{ii} عدد العناصر القطرية . و n حجم العينة

وأن: n_i المجموع الأفقي و C_i المجموع العمودي

ومن الجدول السابق نجد أن:

$$\sum n_{ii} = 96 + 200 + 386 = 682$$

كما نجد أن:

$$\sum r_i * C_i = 113 * 142 + 236 * 224 + 435 * 418 = 250740$$

وبالتعويض في العلاقة السابقة نجد أن:

$$Kappa = \frac{784(682) - 250740}{(782)^2 - 250740} = 0{,}78$$

وهي قيمة جيدة وتدل على درجة توافق قوية بين التصنيفين (الأصلي والمتنبأ به) .

الفصل الخامس

التحليل التمييزي النوعي (Qualitive)

(شجرة التصنيف والانحدار CART)

5-1 : تمهيد:

يتناول التحليل التمييزي النوعي دراسة تأثير المتحولات النوعية أو المختلطة (الكمية والنوعية) على تصنيف عناصر المجتمع المدروس إلى مجموعات منفصلة (متجانسة أو نقية)، ويعتمد هذا التحليل على تصميم شجرة تشبه شجرة القرارات تسمى شجرة التصنيف والانحدار (CART)، المأخوذة من الكلمات (Classification And Regression Tree)، وينطلق هذا التحليل من اعتبار أن جميع عناصر المجتمع المدروس تشكل مجموعة واحدة، ونريد تصنيفها إلى مجموعتين عمليتين أو أكثر، ثم تفريع هاتين المجموعتين إلى مجموعتين جديدتين أو أكثر، وتتم عملية التفريع واتخاذ القرارات من خلال فواصل توضع على مفاصل الشجرة، وتعطينا عند كل مفصل (عقدة) فرعين أو أكثر، وتستمر عملية التفريع حتى يتحقق أمر التوقف الذي يحدده الباحث .

وتبدأ شجرة التصنيف من جذر واحد (يسمى بعقدة الأصل أو عقدة الأب Parent)، ويتفرع عنه أغصان على شكل أسهم لتتصل بعقد فرعية أخرى تسمى بالعقد الداخلية Internal node (أو بعقد الأبناء والأحفاد)، ثم تنتهي بعقد ختامية تسمى بالعقد الخارجية Terminal Node، وهي تضم نتائج التصنيف النهائية ويطلق عليها أيضاً اسم الورقة (laef) وهي ثمرة التصنيف .

وترسم شجرة التصنيف بصورة مقلوبة، وتمثل العقدة الأصلية والعقد الداخلية بدوائر (أو بقطوع ناقصة) ويكتب بداخلها شروط التصنيف أو التفريع. أما العقد الخارجية فتمثل على شكل مربعات وتتضمن نتائج التصنيف وتضم عناصر (واحد على الأقل) مفروزة من المجتمع المدروس (انظر الشكل 5-1)، وتتم عملية الفرز في كل عقدة داخلية حسب الجواب على السؤال الذي فيها. فإذا كان السؤال ثنائي الجواب (نعم أو لا) وكان الجواب بـ (نعم) فإننا نضع (yes) على السهم اليساري، أما إذا كان الجواب بـ (لا) فإننا نضع (No) على السهم اليميني .

ولكن إذا كان جواب السؤال متعدد الحالات فإننا نعمل على وضعها في نظام معين في كل العقد الداخلية التي تتضمن مثل ذلك السؤال .

مثال (5−1): لنفترض إننا نريد تصنيف عينة من موظفي الجامعة (حجمها $n = 100$) حسب ثلاثة متحولات هي: الجنس− الحالة الزوجية− حالة المسكن. إلى مجموعات متجانسة من حيث المسكن، فوجهنا إليهم ثلاثة أسئلة مع أجوبتها الممكنة، وطلبنا من كل منهم تحديد الجواب الذي ينطبق عليه، وكانت الأسئلة كما يلي:

س1: الجنس : ذكر | أنثى

س2: الحالة الزواجية: متزوج | غير متزوج (بدون تفاصيل)

س3: حالة السكن: ملك | إيجار | عند الأهل أو الأقارب

ولنفترض أن تكرارات الإجابات كانت كما في الجدول التالي:

جدول (5−1) تصنيف أفراد العينة انطلاقاً من حالة الجنس ثم الحالة الزواجية ثم حالة المسكن:

الجنس	العدد	الحالة الزوجية	العدد	حالة السكن	العدد
ذكر	60	متزوج	45	ملك	20
				إيجار	15
				عند الأهل	10
		غير متزوج	15	ملك	5
				إيجار	3
				عند الأهل	7
أنثى	40	متزوجة	12	ملك	7
				إيجار	3
				عند الأهل	2
		غير متزوجة	28	ملك	2
				إيجار	1
				عند الأهل	25
المجموع	100	المجموع	100	المجموع	100

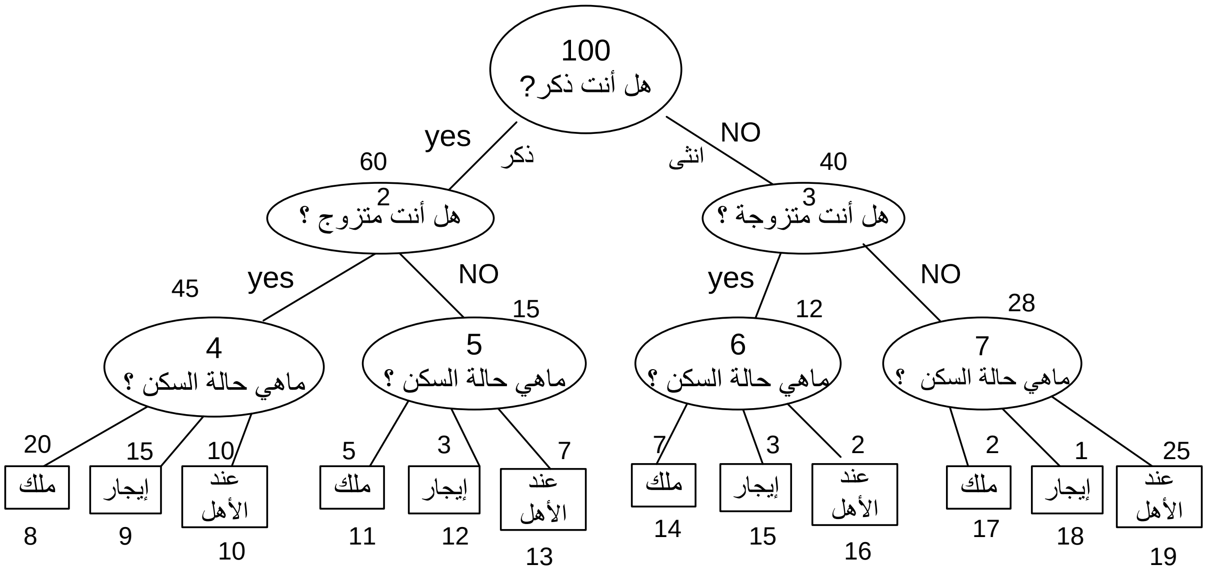
ومن الجدول السابق نلاحظ أنه يمكننا فرز أو تصنيف أفراد هذه العينة إلى مجموعات متجانسة وبأشكال مختلفة. وهنا بدأت عملية فرز هذه المجموعات إلى مجموعتي الجنس (ذكور اناث)، ثم إلى مجموعتي المتزوجين وغير المتزوجين (من كل جنس ومن الجنسين)، ثم إلى مجموعات حالات السكن (ملك، إيجار، عند الأهل)، وذلك حسب كل حالة من الحالات السابقة . كما نلاحظ أنه يمكننا استخلاص وتشكيل المجموعات المتجانسة من حيث المسكن كما يلي:

G_1 **مجموعة المالكين من جميع الحالات وعدد عناصرها يساوي:** $N_1 = 20 + 5 + 7 + 2 = 34$

G_2 **مجموعة المستأجرين من جميع الحالات وعدد عناصرها يساوي:** $N_2 = 15 + 3 + 3 + 1 = 22$

G_3 **مجموعة الساكنين عند الأهل لجميع الحالات وعدد عناصرها أنه يساوي:** $N_3 = 10 + 7 + 2 + 25 = 44$

وأخيراً يمكننا رسم شجرة التصنيف لهذه العينة كما يلي:



الشكل (5-1): شجرة التصنيف لأفراد العينة المدروسة

كما يمكننا تشطير الشجرة إلى شجيرات فرعية بشرط أن يكون لكل شجيرة فرعية جذر واحد، أي أن يكون على رأس كل شجيرة فرعية أصل واحد، فمثلاً يمكننا أن نشطر هذه الشجرة إلى شجرتين فرعيتين هما: شجيرة الذكور وأصلها في العقدة (2)، وشجيرة الإناث وأصلهما في العقدة (3). وكذلك يمكننا اعتبار المتزوجين الذكور شجيرة فرعية أصلها العقدة (4)، كما يمكننا اعتبار الإناث المتزوجات شجيرة فرعية أصلها في العقدة (6).

وأخيراً نشير إلى أن هذا التصنيف يقسم فضاء العينة إلى (12) مجموعة متجانسة هي عبارة عن مجموعات العقد الخارجية كما هو مبين على الشكل التالي:

الاناث		**الذكور**	
غير المتزوجات	**المتزوجات**	**غير متزوج**	**متزوج**
المجموعة 17	**المجموعة** 14	**المجموعة** 11	**المجموعة** 8
المجموعة 18	**المجموعة** 15	**المجموعة** 12	**المجموعة** 9
المجموعة 19	**المجموعة** 16	**المجموعة** 13	**المجموعة** 10

الشكل (5-2): فضاء العينة المقسم

5-2 : مراحل تصميم شجرة التصنيف :

لقد لاحظنا من المثال السابق أن شجرة التصنيف تأخذ شكلاً هرمياً، يبدأ من عقدة واحدة تسمى عقدة الأصل (الجذر)، ويتفرع عنها عبر الأغصان عقد داخلية متعددة وذات مستويات مختلفة، وينتهي التفرع بما يسمى بالعقد الخارجية، وإن حجم أو عمق الشجرة يتوقف على أهداف البحث أو على الشروط التي يضعها الباحث على عملية التصنيف .

إن عملية تصميم شجرة التصنيف تمر بعدة مراحل هي كما يلي:

1- مرحلة الإنشاء أو البناء وتشمل هذه المرحلة عمليات النمو وأهمها عمليات التفرع أو الانشطار Splitting وتتألف من الخطوات التالية:

أ- تحديد المتحول التابع Y والمؤلف من المجموعات التصنيفية $Y = G_1, G_2, \dots G_g$.

ب- اختيار المتحولات التصنيفية أو التفسيرية وتحديد تسلسل تطبيقها عند كل عقدة، بما في ذلك عقدة الأصل. وإذا كان عدد المتحولات كبيراً فإنه يجب اختصارها باختيار المتحولات الهامة منها وحذف المتحولات غير الهامة منها، وذلك وفق معايير محددة. ولنرمز لهذه المتحولات بـ $X_1, X_2, X_3 \dots X_p$.

ج- اختيار عقدة الأصل بحيث تكون مناسبة لأهداف البحث، ومتوافقة مع معايير الانشطار المتسلسلة والمميزة لاستنباط الأبناء والأحفاد من ذلك الأصل، وهنا نشير إلى أنه يمكننا الحصول على أشجار متعددة إذا غيرنا الجذور المعتمدة في عقدة الأصل .

د- تحديد قواعد التفرع (الانشطار) عند كل عقدة t ابتداء من عقدة الأصل. وتحديد الحدود العددية أو الفئات النوعية لكل متحول X_i عند كل عقدة t .

فإذا كان المتحول X_i كمياً فإنه يمكننا أن نضع قاعدة التفرع على شكل متراجحات كما يلي:

$$X_i \geq C_1 \quad \text{أو} \quad X_k \geq C_k \quad \text{أو} \quad X_i + X_k \geq C_3 \qquad (1-5)$$

حيث C_3, C_k, C_1 أعداد حقيقية من مجال تحول $X_i\ , X_k$.

أما إذا كان المتحول X_i نوعياً فإنه يمكننا أن نضع قاعدة التفرع على شكل إشارة انتماء كما يلي:

$$X_i \in G_j \quad \text{أو} \quad X \notin G_j \qquad (2-5)$$

حيث أن: G_j هي احدى مجموعات تابع المخرجات النوعي Y، والمؤلف من المجموعات المحددة التالية :

$$Y = \{G_1, G_2, \dots G_g\}$$

وهنا يجب أن نشير إلى أنه يجب صياغة واختيار قاعدة التفرع بحيث تكون البيانات الناتجة عن عملية التفرع في العقد التالية متجانسة وأكثر نقاوة من البيانات التي كانت في العقدة الأولى . وهناك معايير خاصة لاختيار قاعدة التفرع المناسبة سنعرضها لاحقاً في فقرة خاصة .

2- عملية التقسيم أو التجزئة (Partition): وفيها تتم عملية تقسيم البيانات X في كل عقدة إلى مجموعتين أساسيتين منفصلتين (أو أكثر) حسب قاعدة التفرع ، وبذلك يتم تقسيم الفضاء R^P إلى قسمين (أو أكثر)، لكل منها سمات خاصة. ثم نقوم بتكرار ذلك التقسيم مرة أخرى فنحصل على مجموعتين منفصلتين جديدتين(أو أكثر) مقابل كل مجموعة سابقة .

وهكذا نكرر عمليات التقسيم حتى تتحقق القاعدة المخصصة للتوقف عن التقسيم .

3- عملية التوقف (Stoping) وتتم حسب قاعدة معينة يضعها الباحث حسب طبيعة وهدف البحث، ويمكن أن تكون من أحد الأشكال التالية:

- إذا أصبح التغير في قيمة تابع الخطأ (الشوائبية) صغير جداً أو أقل من الحد المفروض عليه .
- إذا أصبح عدد العناصر عند التقسيم الأخير في إحدى المجموعات الناتجة عنه صغيراً (أقل من 5) أو أصبح يساوي الواحد . أو أصبحت نسبتهم صغيرة بالنسبة للمجموعات الناتجة الأخرى .
- إذا أصبح حجم الشجرة أو عمقها كبيراً (يحدده الباحث) والمقصود بعمقها عدد مستويات التفرع فيها. أما حجم الشجرة فيقاس بعد العقد الخارجية فيها .
- إذا أصبح مستوى الدقة محققاً أو أصبحت المجموعات الناتجة نقية تماماً .

4- عملية التقليم أو التشذيب (Pruning): وهي عبارة عن إسقاط بعض الفروع السابقة، الصادرة عن بعض العقد الداخلية (مع العقد الخارجية التابعة لها). والإبقاء على الفروع الضرورية اللازمة لأغراض البحث. ولكن ذلك يجب أن يتم وفق قواعد محددة سنتعرض لها لاحقاً .

5- التجميع (Grouping) وهو عبارة عن تجميع العقد الخارجية النهائية في مجموعات مناسبة لكل منها. بحيث يكون معدل التصنيف الخاطئ فيها أصغر ما يمكن .

6- رسم الشجرة وترقيم العقد الداخلية والخارجية ترقيماً تصاعدياً حسب مستويات التفرع، وتبدأ عملية الترقيم بوضع الرقم $t = 1$ للعقدة الأصلية (عقدة الجذر)، ثم إعطاء العقد الناتجة عنها أرقاماً متتالية متصاعدة مثل (2) و(3) و(4) و(5)، وتوضع أرقام العقد ضمن دائرتها. وأحياناً توضع أرقام العقد الخارجية تحتها. ويفضل أن تتم عملية الترقيم حسب مستويات العمق ومن اليسار إلى اليمين كما هو موضح على الشكل (5-1). فإذا كان رقم العقدة الداخلية t، فإن رقم العقدة الناتجة عنها يمكن أن يكون أي رقم t' أكبر من الرقم t، أي يجب أن يكون $t' > t$.

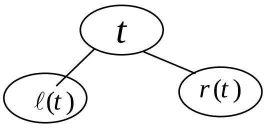
5-3 : كيفية تصميم شجرة التصنيف (حسب التفرع الثنائي Binary)

لنفترض أنه لدينا شجرة T مؤلفة من جملة من العقد الداخلية والخارجية، والمرقمة تصاعدياً بأرقام صحيحة موجبة وحسب مستويات التفرع، وتحقق العلاقة $t' > t$، حيث t' هو رقم العقدة المتفرعة عن العقدة t، فإذا كان الرقم $t = 1$ للعقدة الأصلية (عقدة الجذر) فإن الأرقام (2) و(3) أو(4) أو(5) ستكون للعقد الناتجة عنها .

ولنعرف الآن على كل عقدة t، تابعين $\ell(\mathrm{t})$ و$\mathrm{r(t)}$ كما يلي:

$\ell(\mathrm{t})$ – تابع يعبر عن رقم العقدة اليسارية الناتجة عن التفرع عند النقطة t (من كلمة left) .

$\mathrm{r(t)}$– تابع يعبر عن رقم العقدة اليمينية الناتجة عن التفرع عند النقطة t (من كلمة right) .



بحيث يحقق هذان التابعان في حالة التفرع الثنائي الخواص التالية:

1- من أجل أية عقدة $t \in T$ فإن $\ell(\mathrm{t}) > 0$ و$\mathrm{r(t)} > 0$.

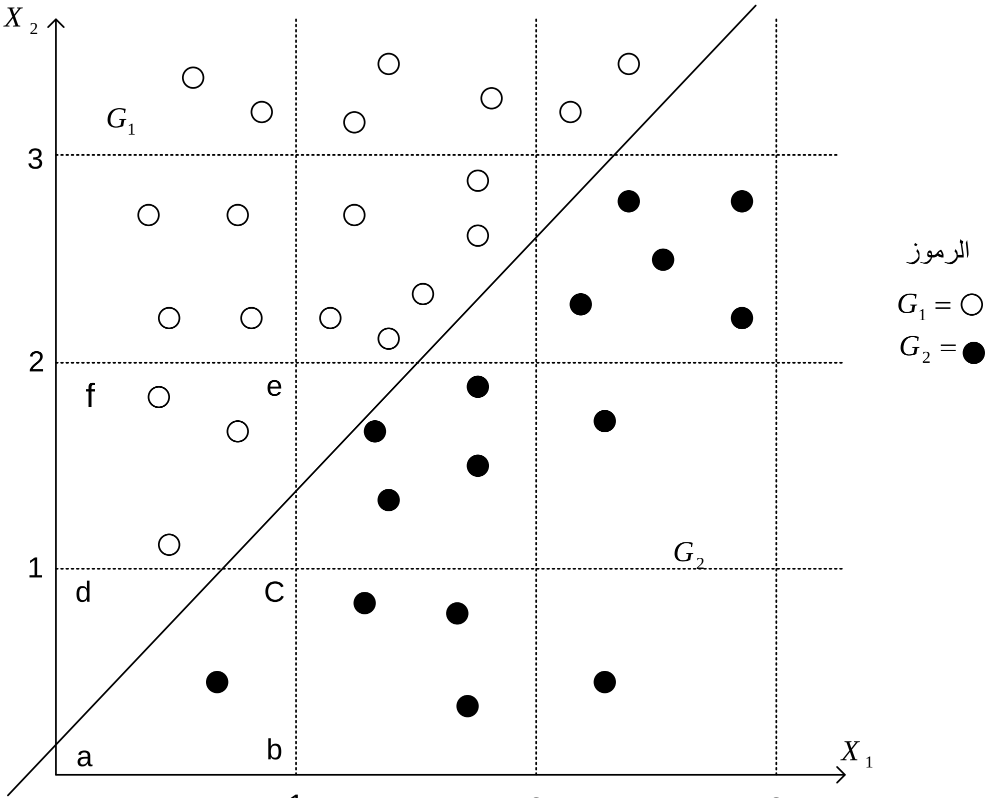
وإذا كانت العقدة الناتجة عن t عقدة داخلية هيt' فإنه يكون لدينا $\ell(t) >$t و $\mathrm{r(t)} > t$.

أما إذا كانت العقدة t' خارجية (ورقة) فإننا نجعل هذين التابعين يأخذان قيمة الصفر، أي نجعلهما يساويان $\ell(t) = 0$ و $\mathrm{r}(t) = 0$. وذلك للدلالة على عدم وجود عقد بعد الأوراق .

2- لكل عقدة t من الشجرة T (ماعدا عقدة الأصل حيث $t = 1$) يوجد أصل وحيد S من T، ولا توجد عقدة بدون أصل، أي أنه يوجد أصل وحيد s لكل عقدة t $(t \neq 1)$ بحيث يكون: $t = \ell(s)$ أو $t = \mathrm{r}(s)$.

مثال (5-2): لنفترض إننا نريد تصنيف (35) عنصراً إلى مجموعتين متجانستين، حسب تغيرات متحولين كميين X_1 و X_2 معرفين في المستوى على أن: $X_1 \geq 0$ و $X_2 \geq 0$، علماً بأن هذه العناصر تؤلف قبل التصنيف مجموعتين G_1 و G_2 فيهما (20) و(15) عنصراً على الترتيب .

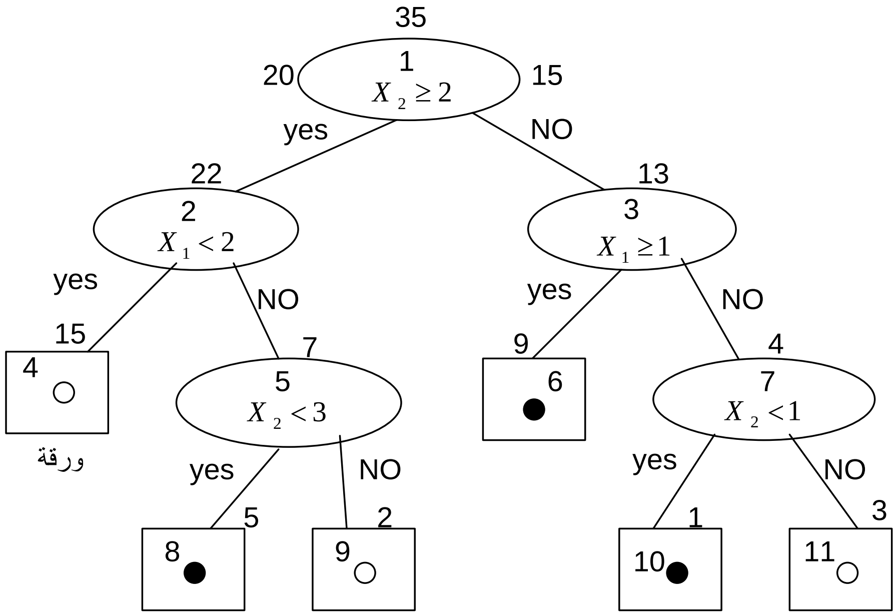
ولنفترض أنه بعد الدراسة والرسم تبين لنا أن تلك العناصر تتوزع على المستوى $X_1 0 X_2$ كما يلي: (المصدر: *Webb. A. R p.227* بتصرف)



الشكل (5-3) التوزيع البياني لعناصر العينة: $G_1 =$ ○ و $G_2 =$ ●

ولنفترض أن الباحث قد قام بتصميم شجرة التصنيف انطلاقاً من الشرط $(X_2 \geq 2)$ كمايلي:

t	$\ell(t)$	$r(t)$
1	2	3
2	4	5
3	6	7
4	0	0
5	8	9
6	0	0
7	10	11
8	0	0
9	0	0
10	0	0
11	0	0



الشكل (5−4): شجرة التصنيف للمثال (5−2) : $G_1 = \bigcirc$ و $G_2 = \bullet$

والآن نعرف التابعين $\ell(t)$ و $r(t)$ على الشجرة المرسومة على الشكل (5−4)، فنجد أن قيمتهما عند كل نقطة t تساويان كما في الجدول المرافق لذلك الشكل . ولقد أشرنا سابقاً إلى أن هذه التفرعات تقسم الربع الموجب للفضاء R^2 حسب رقم المتحولين X_1 و X_2 إلى مناطق منفصلة، ونحصل عليهما بشكل متتالٍ كما يلي:

بما أن قاعدة التفرع في العقدة الأولى الأصلية هي: $(X_2 \geq 2)$ فهي تقسم الربع الموجب إلى قسمين: قسم تحت الخط $(X_2 = 2)$ وقسم فوقه، وضمن هذا التقسيم نأخذ الفرع الأيمن من الشجرة حيث الجواب (No) والموافق لـ $(X_2 < 2)$، فنجد أن قاعدة التفرع في العقدة (3) هي: $(X_1 \geq 1)$ وهي تعطينا بتقاطعها مع $(X_2 < 2)$ (حيث الفرع (No)) منطقتين تتحددان كما يلي:

إذا كان الجواب في العقدة (3) على $(X_1 \geq 1)$ بـ (yes)، فإننا نحصل على المنطقة التي نقع تحت المستقيم $(X_2 = 2)$ وعلى يمين المستقيم $(X_1 = 1)$، وهي تقابل العقدة الداخلية (6) لذلك رمزنا لها على الشكل (5−5) بـ $R(6)$ وهي تضم (9) عناصر من G_2 فقط .

أما إذا كان الجواب على $(X_1 \geq 1)$ بـ (No)، فإننا نحصل على المنطقة الواقعة تحت المستقيم $(X_2 = 2)$ وعلى يسار المستقيم $(X_1 = 1)$، وهي تشمل المستطيل $a\ b\ e\ f$ المقابل للعقدة الداخلية (7) .

وضمن هذا المستطيل نختبر نتيجة قاعدة التفرع في العقدة (7) والتي هي: $(X_2 < 1)$ فنجد أنه: إذا كان الجواب بـ (yes)، فإننا نحصل على المستطيل $a\ b\ c\ d$، والذي يضم عنصراً واحداً من G_1. وهو يقابل العقدة الخارجية (10)، ولذلك رمزنا لها على الشكل (5−5) بـ$R(10)$، أما إذا كان الجواب بـ (No) فإننا نحصل على المستطيل $d\ c\ e\ f$ ورمزنا له على الشكل (5−5) بـ $R(11)$، لأنه يقابل العقدة الخارجية (11)، وهي تضم (3) عناصر من G_1 .

والآن لنذهب إلى الفرع الأيسر حيث الجواب بـ (yes) على القاعدة الأصلية $(X_2 \geq 2)$، فنجد أن قاعدة التفرع في العقدة (2) هي: $(X_1 < 2)$، فإذا كان الجواب بـ (yes) فإننا نحصل على المنطقة المفتوحة فوق $(X_2 = 2)$ وعلى يسار $(X_1 = 2)$، وهي تقابل العقدة الخارجية (4) لذلك نرمز لها بـ R(4) على الشكل (5−5)، وهي تضم (17) عنصراً من G_1 .

أما إذا كان الجواب بـ (No) فإننا نحصل على منطقة مفتوحة تقع فوق $(X_2 = 2)$، وعلى يمين المستقيم $(X_1 = 2)$، وهي تقابل العقدة الداخلية (5) .

وأخيراً نقوم باختيار نتيجة قاعدة التفرع في العقدة (5) والتي هي $(X_2 < 3)$، وندرس تقاطعها مع نتيجتي القاعدتين السابقتين لها وهما: (No) للقاعدة $(X_1 < 2)$ و(yes) للقاعدة $(X_2 \geq 2)$ فنجد أنه: إذا كان الجواب على $(X_2 < 3)$ بـ (yes) فإننا نحصل على المنطقة المفتوحة الواقعة فوق $(X_2 = 2)$ وتحت المستقيم $(X_2 = 3)$ وعلى يمين $(X_1 = 2)$، وهي تقابل العقدة الخارجية (8) لذلك نرمز لها بـ R(8) على الشكل (5−5)، وهي تضم (5) عنصراً من G_2 .

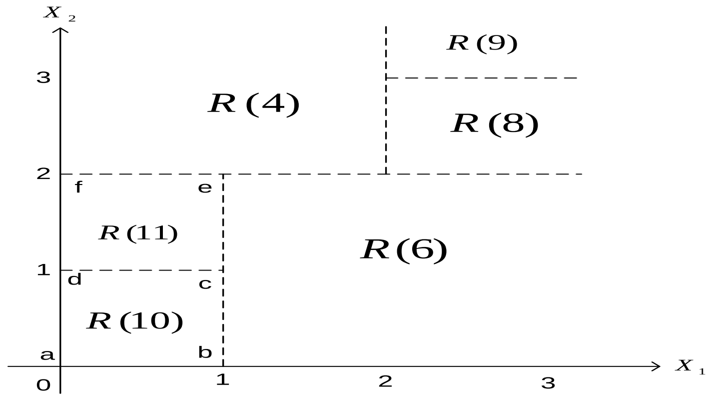
أما إذا كان الجواب على $(X_2 < 3)$ بـ (No) فإننا نحصل على المنطقة المفتوحة الواقعة فوق $(X_2 = 3)$، وعلى يمين المستقيم $(X_1 = 2)$، وهي تقابل العقدة الخارجية (9) لذلك نرمز لها بـ R(9) على الشكل (5−5)، وهي تضم عنصرين من G_1 .

وهكذا نجد أن عملية التصنيف على الشجرة أنجزت عمليتين بآن واحدهما:

1− تقسيم منطقة تعريف المتحولات X إلى مناطق منفصلة تقابل العقد الخارجية وهي: $R(4), R(6), R(8), R(9), R(10), R(11)$.

2− توزيع عناصر العينة على مجموعات مناسبة حسب إحداثياتها ومواقعها في المناطق السابقة وسندرسها لاحقاً في فقرة خاصة .

وأخيراً نرسم هذه المناطق كما يلي :



الشكل (5−5): المناطق المنفصلة لفضاء العينة

وأخيراً يمكننا تجميع هذه المناطق ضمن المجموعتين المفروضتين كما يلي:

المجموعة الأولى G_1 : وتضم المناطق $R(9), R(4), R(11)$ ونكتبها كما يلي:

$$G_1 = R(11) \cup R(4) \cup R(9) \qquad (3-5)$$

المجموعة الثانية G_2 : وتضم المناطق $R(8), R(6), R(10)$ ونكتب ذلك كما يلي :

$$G_2 = R(10) \cup R(6) \cup R(8) \qquad (4-5)$$

ملاحظة 1: عند تصميم الشجرة يجب الانتباه إلى عدم تعارض الشروط الواردة في قواعد التفرع (ضمن العقد الداخلية) مع بعضها البعض وخاصة على المسار الواحد .

فمثلاً لا يجوز أن تكون قاعدة الفصل في العقدة (5) متعارضة مع قاعدة الفصل في (2) أو في (1) وكأن نضع فيها الشرط $(X_2 < 2)$.

ملاحظة 2: إن المناطق $R(t)$ التي حصلنا عليها غير متقاطعة ومتكاملة أي أنها تحقق العلاقتين :

$$R(t) \cap R(s) = \phi = \quad \left(\text{مجموعة خالية}\right) \qquad (5-5)$$

$$\bigcup_{t=1}^{|T|} R(t) = \Omega \qquad : R^P \text{ فضاء العينة في} \qquad (5-6)$$

حيث $|T|$: هو عدد العقد الخارجية ويسمى حجم الشجرة .

ملاحظة 3: يمكن رسم الشجرة بطرائق مختلفة، وذلك بوضع نفس قواعد التفرع بطرائق مختلفة، فنحصل على شجرات جديدة تقسم فضاء العينة بطرائق مختلفة، ولهذا كان لابد من اختيار طريقة مثالية للتصنيف ولتوزيع قواعد التفرع على العقد الداخلية، بحيث يعطينا التقسيم الناتج مناطق نقية أو صافية، أي بحيث تضم كل منطقة عناصر من إحدى المجموعات فقط ، أو على مناطق تتضمن أقل عدد من العناصر الغريبة فيها (ارسم الشجرة السابقة بطريقة أخرى) .

4-5 : التصنيف حسب احتمالات الانتماء والتوزيع :

بما أن تصميم شجرة التصنيف يتم باستخدام خواص مجموعة البيانات المأخوذة من عناصر عينة حجمها n، والتي تتضمن قيم المتحولات X فيها، وقيم مستويات المجموعات المقابلة لها في التابع Y، فإنه يكون لدينا n مجموعة من القياسات المتقابلة $X_{1i}\, X_{2i} \dots X_{pi}\ , G_i$، حيث أن *i* هي دليل عناصر العينة ويأخذ $(i{:}\, 1\ 2\ 3 \dots n)$، وحيث أن *n* هو حجم العينة المدروسة، ويتم تنظيم ذلك في جدول البيانات الأولية المتقابلة .

والآن لنرمز بـ $N(t)$ لعدد عناصر العينة التي تنتمي إلى المنطقة $R(t)$ المقابلة للعقدة t، وهو يمكن أن يكون موزعاً على عدد من المجموعات G_j التي يتألف منها التابع Y .

ولنرمز بـ $N_j(t)$ لعدد عناصر العينة التي تنتمي إلى المنطقة $R(t)$ وإلى المجموعة G_j معاً، والتي تحقق العلاقة: $\sum_{j=1}^{g} N_j(t) = N(t)$، وعندها نجد أنه يمكننا حساب احتمالات الانتماء والتوزيع كمايلي:

- إن احتمال انتماء أي عنصر *i* من العينة *n* عند العقدة t إلى المنطقة $R(t)$ يساوي :

$$P(t) = \frac{N(t)}{n} \qquad (7-5)$$

– إن احتمال انتماء أي عنصر *i* من المنطقة $R(t)$ المقابلة للعقدة *t*، إلى المجموعة G_j يساوي :

$$P\big(G_j/i \in R(t)\big) = \frac{N_j(t)}{N(t)} \qquad (8-5)$$

– إن احتمال أن يتم توزيع عناصر العينة عند العقدة *t*، إلى عقدة اليسار $\ell(t)$ وإلى عقدة اليمين $r(t)$ يساويان :

$$P[\ell(t)] = \frac{N[\ell(t)]}{n} \qquad (9-5)$$

$$P[r(t)] = \frac{N[r(t)]}{n} \qquad (10-5)$$

– إن احتمال أن يذهب أي عنصر *i* متواجد في العقدة *t*، إلى إحدى عقدتي اليسار أو اليمين يساوي :

$$P_\ell = \frac{P[\ell(t)]}{P(t)} = \frac{N[\ell(t)]}{N(t)} \qquad \text{إلى العقدة اليسارية :} \quad (11-5)$$

$$P_r = \frac{P[r(t)]}{P(t)} = \frac{N[r(t)]}{N(t)} \qquad \text{إلى العقدة اليمينية :} \quad (12-5)$$

وهكذا يمكننا أن نحدد ملامح كل مجموعة G_j في كل عقدة *t*، وذلك حسب تناسبها مع عدد عناصر العينة التي تنتمي إليها من المنطقة $R(t)$ ، أي حسب احتمالاتها فيها، ثم نقارن الاحتمالات المقابلة لهذه المجموعات ونختار المجموعة G_k التي تقابل أكبر الاحتمالات .

القاعدة: نصنف العنصر *i* من العقدة *t* إلى المجموعة G_k إذا كان الاحتمال الشرطي المقابل لها أكبر من الاحتمالات الشرطية المقابلة للمجموعات الأخرى، ونكتب ذلك كما يلي. إذا كان:

$$P(G_k/t) = \max_{j=1}^{g} P\big(G_j/t\big) \qquad (13-5)$$

فإننا نصنف العنصر *i* من العقدة t في المجموعة G_k .

مثال (5–3): لنأخذ المثال (5–2) السابق ولنفترض أن عناصر العينة $(n = 35)$ مؤلفة من مجموعتين $(n_1 = 20)$, $(n_2 = 15)$ وتتوزع على العقدتين اللاحقتين $r(t)$,$\ell(t)$. ثم نقوم بحساب الاحتمالات السابقة حسب العلاقات (5–7) ، (5–8) ، (5–9+10) ، (5–11+12) ونضعها في الجدول (5–4) التالي، فنجد أن :

$$P(1) = \frac{N(1)}{n} = \frac{35}{35} = 1 \quad , \quad P(2) = \frac{N(2)}{n} = \frac{22}{35} \ldots\ldots\ldots\ P(11) = \frac{N(11)}{n} = \frac{1}{35}$$

$$P(G_1/1 \in R(1)) = \frac{N_1(1)}{N(1)} = \frac{20}{35} \quad , P(G_2/x \in R(2)) = \frac{N_2(2)}{N(1)} = \frac{15}{35}$$

$$P[\ell(1)] = \frac{N[\ell(1)]}{n} = \frac{22}{35} \ , \ P[r(1)] = \frac{N[r(1)]}{n} = \frac{13}{35} \qquad \left(\text{غير موجودة في الجدول}\right)$$

$$P_{\ell 1} = \frac{P[\ell(1)]}{P(1)} = \frac{N[\ell(1)]}{N(1)} = \frac{22}{35} \quad , \qquad P_{r1} = \frac{N[r(1)]}{N(1)} = \frac{13}{35} \ldots\ldots\ldots\ldots\ldots$$

$$P_{\ell 2} = \frac{P[\ell(2)]}{P(2)} = \frac{N[\ell(2)]}{N(2)} = \frac{15}{22} \quad , \qquad P_{r2} = \frac{N[r(2)]}{N(2)} = \frac{7}{22} \ldots\ldots\ldots\ldots\ldots$$

جدول (5-4): احتمالات الانتماء والتوزيع :

t	قاعدة الفصل	$\ell(t)$	$r(t)$	$N(t)$	التوزيع على G_j		$P(t)$	$P(y_1/t)$	$P(y_2/t)$	P_ℓ	P_r
					$N_1(t)$	$N_2(t)$					
1	$x_2 \geq 2$	2	3	35	20	15	1	$\frac{20}{35}$	$\frac{15}{35}$	$\frac{22}{35}$	$\frac{13}{35}$
2	$x_1 < 2$	4	5	22	17	5	$\frac{22}{35}$	$\frac{17}{22}$	$\frac{5}{22}$	$\frac{15}{22}$	$\frac{7}{22}$
3	$X_1 \geq 1$	6	7	13	4	9	$\frac{13}{35}$	$\frac{4}{13}$	$\frac{9}{13}$	$\frac{9}{13}$	$\frac{4}{13}$
4	○	0	0	15	10	5	$\frac{15}{35}$	$\frac{10}{15}$	$\frac{5}{15}$	–	–
5	$X_2 \geq 3$	8	9	7	2	5	$\frac{7}{35}$	$\frac{2}{7}$	$\frac{5}{7}$	$\frac{5}{7}$	$\frac{2}{7}$
6	●	0	0	9	0	9	$\frac{9}{35}$	0	1	–	–
7	$x_2 < 1$	10	11	4	3	1	$\frac{4}{35}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{4}$
8	●	0	0	5	1	4	$\frac{5}{35}$	$\frac{1}{5}$	$\frac{4}{5}$	–	–
9	○	0	0	2	2	0	$\frac{2}{35}$	1	0	–	–
10	●	0	0	3	3	0	$\frac{3}{35}$	1	0	–	–
11	○	0	0	1	0	1	$\frac{1}{35}$	0	1	–	–

المصدر: Webb. A. R. (2002) Statistical Pattern Recigrition, P.231

ومن الجدول السابق نلاحظ أن العقد الخارجية هي: (4) و(6) و(8) و(9) و(10) و(11)، ولقد تم تحديد انتماء كل منها إلى إحدى المجموعتين حسب الاحتمال الأكبر $P(y_j/t)$، في تلك العقدة . فمثلاً نجد أن العقدة (4) تقابل الاحتمالين $P(y_1/4) = \frac{10}{15}$ $P(y_2/4) = \frac{5}{15}$، فلذلك يتم تصنيف العقدة (4) إلى المجموعة الأولى G_1 المرموز لها بالرمز ○، لأن الاحتمال المقابل لها $\left(\frac{10}{15}\right)$ هو أكبر الاحتمالين، وهكذا نقوم بتصنيف العقد الخارجية الأخرى .

5-5 معايير التفرع أو الانشطار :

لقد لاحظنا أن عملية التفرع تحتاج إلى قواعد ومعايير دقيقة تفرض على المتحولات X وعلى حدودها C في كل عقدة داخلية (t). فمثلاً يمكن أن نضع شرط التفرع كما يلي:

$$S_P = \left[X \in R^P \quad و \quad X_4 \leq 8,2\right] \qquad (14-5)$$

فنحصل على الفضاء الجزئي S_P من R^P الذي يحقق الشرط $X_4 \leq 8{,}2$.

وعندما يكون لدينا $X \in S_P$ فإن الجواب على الشرط $(X_4 \leq 8{,}2)$ يكون بـ (yes)، لذلك نصنف ذلك العنصر X على الفرع الأيسر باتجاه العقدة $\ell(t)$. أما إذا كان العكس فإننا نصنفه باتجاه العقدة اليمنى $R(t)$.

والسؤال الآن هو: بكم طريقة يمكننا أن نفرع البيانات X المتواجدة في المنطقة $R(t)$ المقابلة للعقدة t ؟

الجواب هو: بطرائق كثيرة ونحتاج إلى حساب معقدة وطويلة .

ويكفي أن نتذكر أن عدد طرائق التجزئة لأي مجتمع A يحتوي على k فئة خاصة، إلى مجموعتين غير خاليتين يساوي $(2^{k-1} - 1)$ طريقة. عدا عن أنه يمكننا أن نضع في كل عقدة أحد المتحولات $X_1\ X_2 \ldots\ X_p$، وأن نضع عليه أي شرط ممكن وأن نختار له أي حد من مجاله C_i ...الخ، وهكذا يكون لدينا عدد كبير من الاختيارات في كل عقدة (t)، وذلك حسب قيم المتحول المختار X_i وحسب حدوده الممكنة C. وللخروج من هذا النفق تم وضع معايير للتفرع والانشطار عند كل عقدة t، ونقدم لها بمايلي:

لنفترض أنه لدينا في كل عقدة t عدة مخارج ممكنة هي: $G_1\ G_2 \ldots\ G_g$ وتشكل تابعاً نوعياً Y نرمز له بـ

$$Y: G_1\ G_2\ G_3 \ldots\ G_g \qquad (18-5)$$

وأنه لدينا P متحولاً تفسيرياً مؤثراً على Y نرمز لها بـ

$$X: X_1\ X_2\ X_3 \ldots\ X_p \qquad (19-5)$$

وإذا أخذنا أحد هذه المتحولات X وحددنا شرط التفرع عليه في العقدة t، ثم قمنا بحساب الاحتمالات المقابلة لتلك المجموعات من (5–8) و(5–9)، ورمزنا للتوزيع الاحتمالي المقابل لتلك المجموعات المتفرعة كما يلي:

$$\begin{aligned} &Y: G_1\ G_2\ G_3 \ldots\ G_g \\ &P(G_j/t): P_1\ P_2\ P_3 \ldots\ P_g \end{aligned} \qquad (20-5)$$

وهو يحقق الخاصتين التاليتين : $\sum_{j=1}^{g} P_j = 1 \quad , \quad P_j \geq 0$

علماً بأن كل الاحتمالات P_j تحسب من العلاقة (5–8) التالية: $P_j = \frac{N_j(t)}{N(t)}$، وهنا نعرف على هذا التوزيع الاحتمالي المعايير التالية :

1- **معيار تابع الخطأ للتصنيف الخاطئ:** وهو مشتق من (5–13) ويسمى بتابع الشوائبية (عكس النقاوة) ويعرف في العقدة (t) بالعلاقة التالية :

$$Q_1(t) = 1 - \max_{j=1}[P_j] \qquad (21-5)$$

2- **مؤشر (جيني** *Gini*) والذي يعرف (في حالة التفرع الثنائي) وفي العقدة (t) بالعلاقة :

$$Q_2(t) = \sum_{j=1}^{g} P_j(1 - P_j) = 1 - \sum_{j=1}^{g} P_j^2 \qquad (22-5)$$

3- **تابع القصور** (*entropy*) **أو التشتت** (*deviance*) ويعرف في العقدة (t) بالعلاقة:

$$Q_3(t) = -\sum_{j=1}^{g} P_j lg_2\left(P_j\right) \qquad (23-5)$$

حيث أن: lg_2 هو اللورغاريتم للأساس (2)، وهو يرتبط مع اللوغاريتم الطبيعي بالعلاقة :

$$lg_2 x = \frac{\ell n\, x}{\ell n\, 2}$$

حالة خاصة: إذا كان عدد المجموعات الممكنة عند العقدة t يساوي $g = 2$، فعندها يكون لدينا مخرجان فقط G_1 و G_2 ويقابلهما الاحتمالان P_1 و $(1-P_1)$، وعندها فإن المقاييس الثلاثة السابقة تأخذ الشكل التالي :

$$Q_1(t) = 1 - \max[P_1\ , 1-P_1] \qquad \text{تابع الخطأ} \qquad (24-5)$$

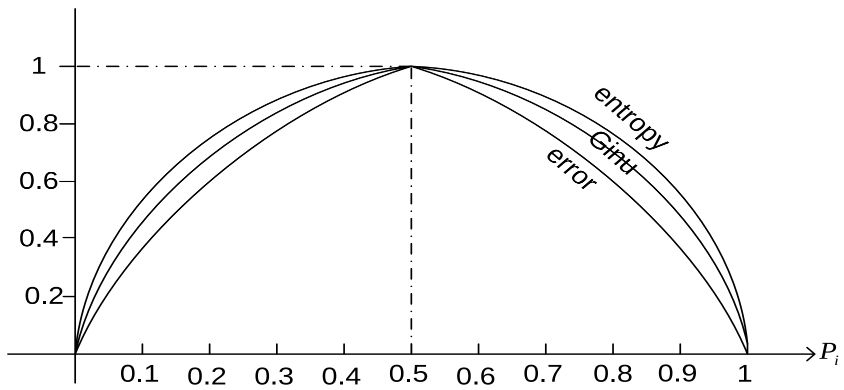
$$Q_2(t) = P_1(1-P_1) + (1-P_1) * P_1 = 2P_1(1-P_1) \qquad (25-5)$$

$$Q_3(t) = -P_1 lg_2(P_1) - (1-P_1) lg_2(1-P_1) \qquad (26-5)$$

أما حالة الفشل فيعبر عنها بالعلاقة :

$$0 * lg_2 0 = 0 \qquad (27-5)$$

وأن هذه التوابع (للمجموعتين في التفرع الثنائي) ترسم المنحنيات والمستقيمات التالية :



الشكل (5-5): منحنيات المعايير السابقة في حالة مجموعتين G_1 و G_2

4- **التابع الاجمالي للشوائبية :** لتعريف التابع الاجمالي للشوائبية على الشجرة ككل، نفترض إننا نستخدم مقياس معين للشوائبية من المقاييس المذكورة سابقاً. وليكن $Q_i(t)$ على شجرة T، وهنا نعرف بعض المصطلحات ونرمز لها كما يلي:

- حجم الشجرة: وهو عدد العقد الخارجية فيها ونرمز له بـ $|T|$.
- المناطق المقابلة للعقد الخارجية في فضاء العينة R^P ونرمز لها بالرموز $R(1), R(2), \dots R(t) \dots R(|T|)$، وعددها يساوي $|T|$ منطقة منفصلة ومتكاملة .
- حجم عناصر العينة في كل منطقة $R(t)$ ونرمز له بالرمز N_t .
- قيمة تابع الشوائبية في العقدة t وليكن $Q(Rt)$.

وبناء على ذلك نعرف التابع الاجمالي للشوائبية في الشجرة ككل بواسطة العلاقة التالية:

$$Q = \sum_{t=1}^{|T|} N_t \, Q(R_t) \qquad (28-5)$$

وتحسب قيمته للشجرة ككل، للاستفادة منها في الحكم على جودة التفرع وعلى كفاية حجم أو عمق الشجرة . كما يمكننا استخدامه في عملية التوقف عن عمليات التفرع في كل مراحله وبدائله .

فكلما كانت قيمة Q صغيرة كانت جودة التفرع جيدة (علماً بأن أصغر قيمة له هي الصفر)، فعندما تبلغ قيمته أصغر من حد معين مثل $(0,01)$ أو $(0,05)$ نقرر التوقف عن متابعة التفرع ، ولكن هناك نقطة ضعف لاستخدام هذا التابع في عملية التوقف، وهي إنه عندما تكون الحدود صغيرة جداً أو معدومة القيمة تقودنا إلى تفريع متشعب جداً، وعندها تنشأ لدينا شجرة كبيرة جداً، ولكن بما أن الشجرة الكبيرة يمكن أن تشوه البيانات، وأن الشجرة الصغيرة يمكن أن تجسد فقط الهيكل العام للبيانات، فإنه يمكننا أن نضع ضابطاً آخر للتوقف، وهو أن نأمر بالتوقف عندما يبلغ عدد عناصر العينة في المنطقة R_t المقابلة للعقدة الخارجية t أقل من عدد محدد (5 مثلاً) أو عندما يساوي (1) حتماً .

5- تابع المكسب المعلوماتي *(Gain)*: وهناك معيار آخر (ولكنه معقد) لإجراء عمليات التفرع، يعتمد على حساب مقدار المكسب المعلوماتي الذي نحققه في كمية المعلومات اللازمة عند التفرع في كل عقدة t، وينطلق ذلك المعيار من حساب تابع القصور (*entropy*) للتوزيع P_i بواسطة الفاصل X والحد C من العلاقة:

$$H(X\ ,C) = -\sum_{j=1}^{g} P_j * lg_2 P_j \qquad (29-5)$$

وهو يعبر عن كمية المعلومات اللازمة (الناقصة) لتوصيف حالة التفرع في العقدة t، ثم نقوم بحساب تابع القصور الشرطي في العقد المتفرعة عن t من العلاقة :

$$H_i(X/C) = -\sum_{j=1}^{g} P_j(X_i/C) * lg_2 P_j(X_i/C) \qquad (30-5)$$

حيث أن الاحتمالات الشرطية تحسب من العلاقة:

$$P_j(X/C) = \frac{N_j(t)}{N(t)}$$

وبعدها نحسب التوقع الرياضي لـ $H_i(X/C)$ من العلاقة:

$$E[H_i(X/C)] = P_1(t_1')H_1 + P_2(t_2')H_2 + P_3(t_3')H_3 + \cdots \qquad (31-5)$$

وأخيراً يتم حساب المكسب المعلوماتي في قيمة تابع القصور، التي سيتم تحقيقها جراء التفريع عند العقدة t من العلاقة:

$$Gain\ (X) = \mathrm{H}(X, C) - E[\mathrm{H}_i(X/C) \qquad (32-5)$$

ثم نختار النتيجة التي تقابل أكبر قيمة للمكسب $(Gain)$، ونقرر اعتماد الفاصل $(X < C)$ للتفريع عند تلك العقدة .

مثال (5−4): لحساب المكسب المعلوماتي للتفرع عند عقدة الجذر (1) في المثال (5−3) المشروط بـ $(X_2 \geq 2)$، نجد أن معيار القصور يعطينا من الجدول (5−4) ما يلي:

$$H = -P_1 lg_2 P_1 - P_2 lg_2 P_2 = -\left[\frac{20}{35}\ln\frac{20}{35} + \frac{15}{35}\ln\frac{15}{35}\right] * \frac{1}{\ln 2}$$

$$H = \frac{0.319780 + 0.3631277}{0.69314718} = 0.985228 \ \ (bit)$$

ثم نقوم بحساب $H_1(X/C)$ و$H_2(X/C)$ في العقدتين (2) و(3) فنجد أن:

$$H_2 = -\frac{17}{22}\lg\frac{17}{22} - \frac{5}{22}\lg\frac{5}{22} = \frac{0.1992316 + 0.336728}{0.69314718} = 0.7732267$$

$$H_3 = -\frac{4}{13}\lg\frac{4}{13} - \frac{9}{13}\lg\frac{9}{13} = \frac{0.3626631 + 0.2545787}{0.69314718} = 0.890492$$

ثم نحسب التوقع الرياضي لهما فنجد أن:

$$E[\mathrm{H}_i(X/C)] = \frac{22}{35}\ H_1 + \frac{13}{35}\ H_2 = 0.4860282 + 0.3307542$$
$$= 0.816782 \ \ (bit)$$

وبذلك نجد أن مقدار المكسب المعلوماتي يساوي :

$$Gain\ (X/C) = H - E[\mathrm{H}(X/C)] = 0.985228 - 0.816782$$
$$Gain\ (X/C) = 0.168446 \quad (bit)$$

فإذا كانت قيمة هذا المكسب أكبر من جميع المكاسب الممكنة عن الشروط الممكنة للمتحولات X، نقرر اعتماد الشرط المستخدم (مثل $X_2 \geq 2$) للمتحول X_2 لتفرع البيانات والمشاهدات عند العقدة الأصلية (1). وهكذا نجد أن عملية تفريع العقدة (1) حسب$(X_2 \geq 2)$ إلى (2) و(3) جعلتنا نكسب (0.168446) بايتا (Bit) من كمية المعلوماتية، وهذا يزيد من قدرتنا على التصنيف الصحيح وتقلل من معدل التصنيف الخاطئ، وتجعلنا نحصل على عقدتين أكثر نقاوة من العقدة (1) .

ولكن حتى نحصل على نتيجة عامة تشمل جميع العقد الداخلية في الشجرة، يجب علينا أن نكرر مثل تلك الحسابات من أجل جميع شروط الفرز $(X \leq C)$، التي يمكن فرضها على المتحولات X وعند كل عقدة داخلية، وبما أن تلك الشروط قد تكون غير محدودة، لأنها تتعلق بعدد المتحولات X وبقيمتها العددية الممكنة وبالحدود المفروضة عليها C، فإن حجم الحسابات سيكون كبيراً جداً، ولا يمكن تنفيذه إلا بواسطة الحواسيب وباستخدام برنامج خاصة لذلك. ولكن يمكن تخفيض حجم هذه الحسابات باستخدام بعض الأساليب المفيدة. فمثلاً إذا كان X متحولاً مستمراً ضمن مجال معين فإننا نقوم بتبويب قيمه ضمن مجالات جزئية محدودة ونعتبره متحولاً مرتباً، ذا فئات محددة، وبذلك ينخفض حجم الحسابات السابقة كثيراً

ملاحظة 1: إن المعايير الثلاثة الأولى تتصف بالخواص التالية:

1- إن التوابع $Q(t)$ تأخذ أكبر قيمة لها وهي الواحد، وذلك عندما تكون الاحتمالات P_i متساوية. أي عندما: $P_1 = P_2 = P_3 = \cdots = \frac{1}{g}$، وهذه الحالة هي أسوأ الخيارات لأنها تجعل معدل التصنيف الخاطئ أكبر ما يمكن، وهذا يقابل أعلى درجة للشوائبية وتساوي الواحد. فمثلاً عندما يكون لدينا Y مؤلفاً من مجموعتين فقط فإن $g = 2$، وعندما يكون الاحتمالان متساويان $P_1 = P_1 = 0.5$ فإن هذه التوابع تعطينا أكبر قيمة لتابع الشوائبية وتساوي الواحد، كما هو مبين على الشكل (5-5) .

2- إن التوابع $Q(t)$ تأخذ أصغر قيمة لها عندما يكون أحد الاحتمالات مساوياً للواحد، وتكون الاحتمالات الأخرى معدومة، أي عندما يأخذ التوزيع الاحتمالي أحد الأشكال التالية: (1,0,0,0,0) أو(0,1,0,0,0) أو(0,0,0,0,1)، وهذا يعني أن عناصر العينة تجمعت في مجموعة واحدة من مجموعات التابع Y، لأنها من صنف واحد، وبذلك تصبح درجة النقاوة 100% ويصبح تابع الشوائبية مساوياً للصفر $Q_n(t) = 0$.

3- إن التوابع$Q(t)$ تأخذ قيماً متناظرة مقابل الاحتمالات$(P_1, P_2, P_3, .., P_g)$ كما في الشكل(5-5).

ملاحظة 2: عند استخدام أحد هذه المعايير في عمليات التفرع عند أية عقدة t، يتم اختيار المتحول X_k والشروط C_k (أو الفئات C_k)، التي تجعل قيمة المعيار المستخدم أصغر ما يمكن، وعادة يتم استخدام $Q_2(t)$ و$Q_3(t)$ في عمليات الإنشاء والتفرع عند العقد الداخلية . أما $Q_1(t)$ فيستخدم لإنشاء قاعدة للتوقف، لأنه يعبر عن جودة التصنيف عند العقدة (t)، كما يستخدم في عملية تقليم الشجرة حسبما سنرى لاحقاً .

5-6 التفريع بواسطة الاحتمالات السابقة والتكاليف : [انظر webb. A.R. P.238]

لقد تعرفنا في العلاقتين (5-7) (5-8) على الاحتمالات السابقة $P(t)$ و$P(j/t)$ ، والآن لنفترض أن الاحتمالات السابقة لكل مجموعة G_j يساوي $\pi(j)$ ويعرف بالعلاقة :

$$\pi(j) = \frac{N_j}{n} \qquad , \qquad P(j/t) = \frac{N_j(t)}{N_j} \qquad (33-5)$$

حيث أن: N_j هو عدد عناصر المجموعة G_j من أصل العينة ذات الحجم n. وهذه الاحتمالات هي عبارة عن نسبة المجموعة G_j في العينة المدروسة .

وإذا كان التوزيع الاحتمالي لمجموعة القرار لا يتناسب مع حدوث تلك المجموعة، فإن التقديرات الحصينة الكلية لاحتمالات أن تقع عناصر تلك العينة في داخل العقدة t يساوي $P(t)$، الذي يحسب من العلاقة (حسب نظرية الاحتمال الكلي) :

$$P(t) = \sum \pi(j) * \frac{N_j(t)}{N_j} \qquad (34-5)$$

حيث أن: $N_j(t)$ هو عدد عناصر المجموعة G_j الواقعة ضمن العقدة t، وعندما تكون G_j معطية مسبقاً فإن الاحتمالات اللاحقة تحسب من العلاقة:

$$P(t/j) = \frac{\pi(j) * \frac{N_j(t)}{N_j}}{\sum_{j=1}^{g} \pi(j) * \frac{N_j(t)}{N_j}} \qquad (35-5)$$

وإذا رمزنا لنسبة عناصر العينة التي من المجموعة G_j وصنفت خطأ في المجموعة G_i بالرمز $q(i/j)$، وتجاهلنا تكاليف تلك الأخطاء (أو اعتبرناها متساوية) فإن معدل التصنيف الخاطئ لكامل الشجرة يحسب من العلاقة :

$$R(T) = \sum_{i,j}^{n} q(i/j) * \pi(j) \qquad (36-5)$$

وبعد حساب قيم $R(T)$ نختار أصغر القيم الممكنة لها لاختيار التفريع الأفضل .

وإذا افترضنا أن تكاليف التصنيف الخاطئ لأي عنصر من المجموعة G_j في المجموعة G_i تساوي C_{ij}، فعندها نجد أن تابع تكاليف التصنيف الخاطئ على كامل الشجرة يساوي :

$$C(T) = \sum_{i\,j}^{n} C_{ij} * q(i/j) * \pi(J) \qquad (37-5)$$

ثم نقوم بحساب قيم هذا التابع، ونختار الحالة التي يأخذ فيها أصغر قيمة ممكنة لاختيار التفريع الأقل تكلفة .

5-7 التفريع بواسطة الانحدار التجميعي (MARS): (انظر Friadman P.1991 أو webb. P.242)

وهي مأخوذة من الكلمات الإنكليزية (Multivoriate Adaptive Regression Splines) لنفترض أنه لدينا بيانات مؤلفة من n قياساً لـ p متحولاً هي $(X_1, X_2, X_3, \dots, X_p)$ ويقابلها n قياساً لتابع الاستجابة Y وتشكل مصفوفة من المرتبة $(p, n+1)$، كما نفترض أنه يمكننا توليد أو تمثيل هذه البيانات بواسطة علاقة انحدار من الشكل :

$$Y_i = f(X_{1i}, X_{2i}, \dots, X_{pi}) + \varepsilon_i \qquad (38-5)$$

حيث أن ε_i هو حد الخطأ المرتكب (البواقي)، والمطلوب منا أن نجد تقديراً معيناً لـ f مثل $\tilde{f}$ بحيث يكون الخطأ ε_i أقل ما يمكن .

إن نموذج التجزئة المتكررة المتتالية الثنائية يعرف بواسطة العلاقة التالية :

$$\tilde{f}(X) = \sum_{t=1}^{|T|} a_t B_t(X) \qquad (39-5)$$

حيث أن التابع الأساسي $B_t(X)$ يساوي :

$$B_t(X) = I[x \in R(t)] \qquad (40-5)$$

حيث أن I هو تابع ثنائي يأخذ القيمة (1) إذا كان الشرط $x \in R(t)$ صحيحاً، ويأخذ القيمة (0) إذا كان الشرط غير صحيح .

وحيث أن $R(t)$ هي المناطق المفصولة، المقابلة للعقد الخارجية وعددها $|T| \ldots 3\ 2\ 1 :i :t$ وهي تحقق العلاقتين: $R_t \cup R_j = \phi$ من أجل $t \neq j$. وأن $\cup R_j = \Omega$.

أما مجموعة الرموز $\mathrm{a_t}$ (حيث لـ $t = 1,2$) فهي الأمثال العددية للتابع (5−39) والتي تحسب قيمها بطريقة المربعات الصغرى أو الإمكانية العظمى وذلك لتمثيل تلك البيانات .

ولكن في مسائل التصنيف فإنه يمكن إعادة صياغة تابع الانحدار في كل مجموعة (باستخدام المتحولات الثنائية التي تأخذ إحدى القيمتين (1) إذا كانت $x_i \in G_j$ وتأخذ الصفر إذا حدث غير ذلك) إلى عدة توابع $\tilde{f}_j$ حسب القواعد المتبقية في التمييز . وهكذا يمكننا من إنتاج التابع الأساسي بواسطة خوارزمية (*Friedman 1991*) وتقديمه كمنتج مؤلف من جداء عدة توابع خطية .

ولتوضيح ذلك نأخذ التجزئة الناتجة عن الشجرة المعطية في المثال (5−2)، فنجد أن التجزئة الأولى قامت على المتحول X_2 وقسمت المستوى حسب الشرط $(X_2 \geq b_2)$ إلى منطقتين وأعطتنا التابعين التاليين :

$$H[(X_2 - b_2)] \quad , \quad H[-(X_2 - b_2)] \qquad (41-5)$$

حيث أن التابع $H(X)$ يساوي :

$$H(X) = \begin{cases} 1 & X \geq 0 \\ 0 & \text{غير ذلك} \end{cases} \qquad (42-5)$$

حيث أن: $X = X_2 - b_2$

ثم تمت تجزئة المنطقة المقابلة $(X_2 < b_2)$ مرة تانية استناداً إلى المتحول X_1 والشرط $(X_1 < a_1)$، وهذا يعطينا تابعين آخرين أساسيين يشكلان مع التابع الثاني السابق تابعين جديدين (بواسطة تقاطعها) ويمكن كتابتهما على الشكل التالي :

$$H[-(X_2 - b_2)] * H[+(X_1 - a_1)] \text{ و } H[-(X_2 - b_2)] * H[-(X_1 - a_1)] \ (43-5)$$

وهكذا يمكننا أن نحصل على عدة توابع أساسية نهائية لكامل الشجرة تتألف من الجداءات التالية:

$$\begin{array}{l} H[-(X_2 - b_2)] * H[-(X_1 - a_1)] * H[+(X_2 - b_1)] \\ H[-(X_2 - b_2)] * H[-(X_1 - a_1)] * H[-(X_2 - b_1)] \\ H[-(X_2 - b_2)] * H[+(X_1 - a_1)] \\ H[+(X_2 - b_2)] * H[-(X_1 - a_2)] \\ H[+(X_2 - b_2)] * H[+(X_1 - a_2)] * H[+(X_2 - b_3)] \\ H[+(X_2 - b_2)] * H[+(X_1 - a_2)] * H[-(X_2 - b_3)] \end{array} \qquad (44-5)$$

وهنا نلاحظ أن كل تابع من هذه التوابع هو عبارة جداء عدة توابع H، وبصورة عامة فإن التوابع الأساسية الناتجة عن خوارزمية التجزئة المتتالية يكون لها الشكل التالي :

$$B_t(X) = \prod_{k=1}^{kt} H\left[S_{kt}\left(X_{v(kt)} - C_{kt}\right)\right. \qquad (45-5)$$

حيث أن S_{kt} هو تابع الإشارة ويأخذ إحدى القيمتين (± 1)

أما kt فهو عدد التفريعات التي تؤدي بنا إلى $B_m(X)$

وأن $X_{v(kt)}$ فهو متحول التفريع، والرمز C_{kt} هو قيمة حد الفصل المفروض على المتحول X .

إن أسلوب MARS هو توليد لأسلوب التجزئة المتتالية حسب الطريقة التالية .

- مسألة الاستمرارية في MARS :

إن نموذج التجزئة المتتالية MARS ينقطع في منطقة الحدود . وهذا يكون بسبب تابع الخطوة H، ولأن أسلوب MARS يستبدل توابع الخطوة بواسطة توابع التفريع . وإن الجانبين يشكلان توابع أسية أساسية للفواصل من المرتبة q كما يلي:

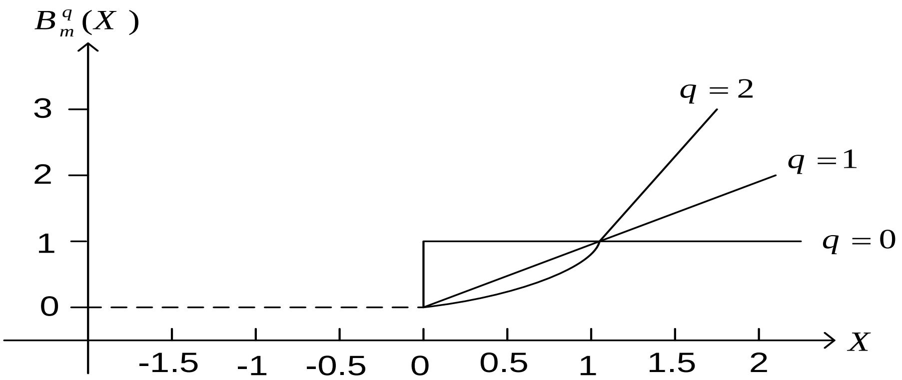
$$b_q^{\pm}(X - t) = [\pm(X - t)]_+^q \qquad (46-5)$$

حيث أن الرمز $[\ ]_+$ يرمز إلى الجزء الموجب للقطعة المعتبرة . أما التابع الأساسي $b_q^+(X)$ فهو موضح في الشكل (5–6) . وبذلك يكون تابع الخطوة H هو حالة خاصة للقيمة $q = 0$. كما يمكن استخدام خوارزمية MARS عندما $q = 1$، وهذا يؤدي إلى استمرارية تابع التقريب ولكنه يتقاطع مع مشتقاته الأولى .

وعندها فإن التابع الأساسي يأخذ الصيغة التالية :

$$B_t^q(X) = \prod\left[S_{kt}\left(X_{v(kt)} - C_{kt}\right)\right]^q \qquad (47-5)$$

حيث أن C_{kt} تشير إلى زمرة الحدود الفاصلة .



الشكل (5–6) توابع التفريع من أجل $q = 0$ $q = 1$ $q = 2$

8−5 التقليم (Pruning): [webb. A.R, P.233 بتصرف] .

إن التقليم يطبق على أشجار التصنيف الجاهزة، للتخلص من الفروع الزائدة، ولمعالجة هذه المسألة نفترض أنه يوجد عند كل عقدة t من الشجرة المعطية T، عدد حقيقي $q(t)$، يمثل احتمال التصنيف الخاطئ في العقدة t، فإذا كانت t عقدة خارجية (أي أن $t \in |T|$) وكان$M(t)$ عدد عناصر العينة، الذين تم تصنيفهم خطأ في تلك العقدة الخارجية t، فإن $q(t)$ يعبر عن نسبة عدد عناصر العينة من المنطقة $R(t)$، الذين لا ينتمون إلى الفئة المقابلة لتلك العقدة الخارجية t. وإن $q(t)$ يحسب من العلاقة التالية:

$$q(t) = \frac{M(t)}{n} \ : \quad t \in |T| \qquad \text{للعقد الخارجية :} \qquad (48-5)$$

حيث أن: $|T|$ هو عدد العقد الخارجية في الشجرة المدروسة، وبذلك نجد أن معدل التصنيف الخاطئ لكامل الشجرة T يساوي مجموع معدلات الخطأ في العقد الخارجية فيها، ولنرمز له بـ $Q(T)$ ونحسبه من العلاقة :

$$Q(T) = \sum_{t \in |T|} q(t) \qquad (47-5)$$

وهو يعبر عن تابع الشوائبية (عدم النقاء) في الشجرة الكلية T.

ولنفترض الآن أن هناك تكاليف أخرى قد تفرض على العقد الخارجية للشجرة T. (مثل الضرائب أو الغرامات أو تخفيض الأسعار أو التلف)، وإن قيمة كل منها تساوي ∝ (قيمة موحدة لجميع العقد الخارجية)، فعندها نجد أن معدل خطأ التصنيف المركب لكل عقدة خارجية t قد أصبح يساوي:

$$q_{\propto}(t) = q(t) + \propto \qquad (48-5)$$

وبذلك يصبح معدل خطأ التصنيف المركب لكامل الشجرة T مساوياً لـ $Q_{\propto}(t)$ ويحسب من العلاقة :

$$Q_{\propto}(t) = \sum_{t \in |T|} q_{\propto}(t) = \sum_{t \in |T|} [q(t) + \propto]$$

وبناء على (5−47) نجد أن:

$$Q_{\propto}(t) = Q(T) + \propto |T| \qquad (49-5)$$

وبما أن عملية التقليم تهدف إلى حذف بعض فروع الشجرة T والعقد الخارجية المتعلقة بها، وإن هذه الفروع تشكل شجيرات فرعية، وتكون أصولها في أحد العقد الداخلية t، لذلك يجب إعادة حساب $q(t)$ و $Q(T)$ و $q_{\propto}(t)$ و $Q_{\propto}(t)$ بعد كل عملية تقليم، وذلك من أجل التمهيد لعملية التقليم التالية .

والآن لنرمز للشجرة الفرعية التي أصلها في العقدة الداخلية t من الشجرة T بالرمز T_t. ولعدد العقد الخارجية المتعلقة بها بـ $|T_t|$، ولمعدل خطأ التصنيف المركب فيها بالرمز $Q_{\propto}(T_t)$. وبطريقة مشابهة للمعالجة السابقة نجد أنه قياساً على (5−49) أن:

$$Q_{\propto}(T_t) = Q(T_t) + \propto |T_t| \qquad (50-5)$$

ومن جهة أخرى نجد أن العقدة t بعد التقليم ستصبح عقدة خارجية في الشجرة الجديدة، وبالتالي فإن معدل خطأ التصنيف المركب فيها يصبح مساوياً لـ

$$q_{\propto}(t) = q(t) + \propto \qquad (51-5)$$

وهو عبارة عن معدل خطأ التصنيف المركب للشجرة المتفرعة من العقدة t، ولذلك فإنه بعد عملية التقليم يجب أن يتساوى المعدلان $Q_{\propto}(T_t)$ و $q_{\propto}(t)$، لأنهما يعبران عن معدل خطأ التصنيف المركب لنفس الشجرة الفرعية T_t. لذلك نضع بينهما المعادلة التالية:

$$Q_{\propto}(T_t) = q_{\propto}(t)$$
$$Q(T_t) + \propto |T_t| = q(t) + \propto$$

ومنها نجد أن $\propto$ تساوي:

$$\propto = \frac{q(t) - Q(T_t)}{|T_t| - 1} \qquad (52-5)$$

فإذا كانت قيمة $\propto$ صغيرة بالنسبة لشجيرة محددة T_t، فهذا يعني أن الفرق بين $[q(t) - Q(T_t)]$ يكون صغيراً، وبالعكس. وعندها يكون لدينا شجرة فرعية كبيرة ولها عدد كبير $|T_t|$ من العقد الخارجية التي ستجعل المقدار $Q(T_t)$، قريباً من $q(t)$، وهذا ما يجعل عملية التقليم مجدية .

أما عندما تكون قيمة $\propto$ كبيرة نسبياً فإنها تقابل شجيرة فرعية صغيرة T_t من T، ويكون لها عدد قليل من العقد الخارجية $|T_t|$، وبالتالي يكون الفرق $[q(t) - Q(T_t)]$ كبيراً نسبياً، وعندها تكون عملية التقليم ليست ذات أهمية .

لذلك اقترح (بريمان Breiman 1984) تعريف تابع جديد $g(t)$ مشابهاً للعلاقة (5−52)، لاستخدامه في تحديد العقدة t، التي تصلح لأن تكون أصلاً لشجرة فرعية T_t، وهي العقدة التي تقابل أصغر قيمة لذلك التابع $g(t)$. وإن $g(t)$ يساوي:

$$g(t) = \frac{q(t) - Q(T_t)}{|T_t| - 1} \qquad (53-5)$$

ولكنه قدم تعريفاً آخر لحساب المعدل $q(t)$ وهو:

$$q(t) = P(t)\left[1 - \max P(G_j/t)\right] \qquad (54-5)$$

ونلخص خوارزمية (Breiman 1984) للتقليم بما يلي:

أولاً نتأكد من أن شجرة التصنيف جاهزة للتقليم ثم نقوم بما يلي:

1− نقوم بحساب معدلات خطأ التصنيف العادي $q(t)$ لجميع العقد الداخلية والخارجية في الشجرة من العلاقة (5−54)، ونسجلها على يسار العقد الداخلية ونضعها بأسفل العقد الخارجية (كما في الشكل (5−7)) .

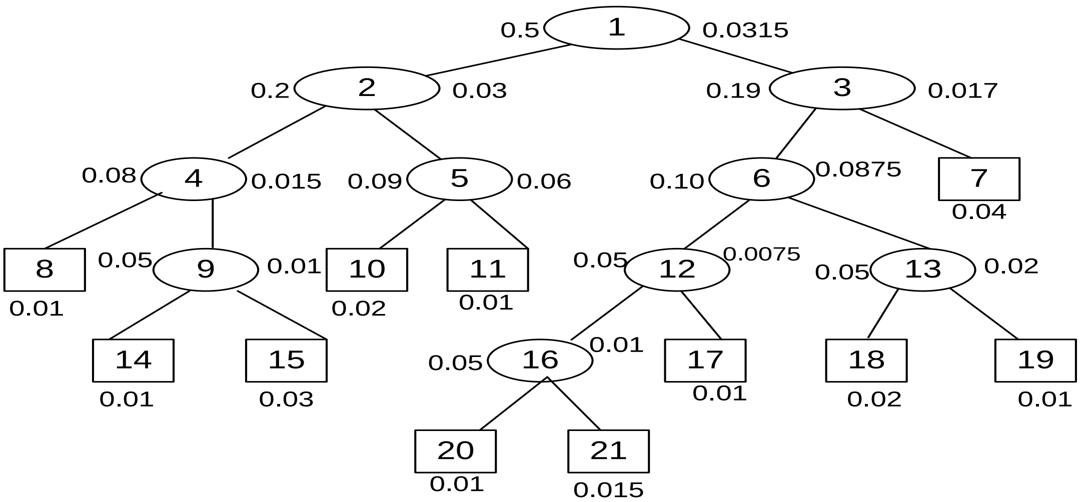
2− نقوم بحساب قيم التابع $g(t)$ لجميع العقد الداخلية في الشجرة من العلاقة (5−53) ونسجلها على يمين العقد الداخلية (كما في الشكل (5−7)) .

3- نبحث عن أصغر قيمة للتابع $g(t)$ ومنها نحدد العقدة المقابلة t التي ستكون أصلاً للشجيرة الفرعية T_t، التي يجب تقليمها، ثم نحدد العقد المتفرعة عنها استعداداً لحذفها (يمكن أن نصادف أكثر من عقدة t وأكثر شجيرة فرعية T_t) .

4- نحذف العقد المتفرعة عن العقدة t المحددة في (3) ونعيد رسم الشجرة المتبقية مع الحفاظ على أرقام العقد السابقة وعلى قيم المعدلات $q(t)$ المسجلة على يسار كل عقدة داخلية لأنها لا تتغير .

5- نعود إلى الخطوة (2) ونكرر إعادة حساب قيم التابع $g(t)$ لجميع العقد الداخلية في الشجرة المتبقية، ونكرر هذه الحلقات حتى نتوصل إلى الشجيرة الفرعية المؤلفة من العقدة الأصلية (الجذر) فقط .

مثال (5-5) : الآن لنأخذ شجرة التصنيف التالية :



الشكل (5-7) شجرة تصنيف افتراضية

ومن الشكل (5-7) نلاحظ أن كل عقدة خارجية قد ميزت تحتها برقم واحد هو قيمة الاحتمال $q(t)$ المحسوبة من (5-54)، وهو عبارة عن احتمال التصنيف الخاطئ، الذي يعبر عن نسبة مساهمة تلك العقدة t في معدل الخطأ الاجمالي، ولكن كل عقدة داخلية قد ميزت برقمين كتبا على يسارها ويمينها وهما:

إن الرقم المكتوب على يسار العقدة t هو قيمة المقدار $q(t)$، وهو يعبر عن نسبة مساهمة تلك العقدة في معدل الخطأ الاجمالي، فيما إذا أصبحت تلك العقدة عقدة خارجية (ورقة). وهنا نلاحظ أن قيم $q(t)$ تتناقص كلما تفرعت الشجرة إلى فروع أو أوراق متجانسة أو صافية وتصبح قيمه صغيرة جداً في العقد الخارجية .

أما الرقم المكتوب على يمين العقدة t، فهو قيمة التابع $g(t)$ المحسوب من العلاقة (5-53)، فمثلاً نجد أن قيمة $g(t)$ عند العقدة $t = 2$ تحسب على العقد الخارجية: (8) و(10) و(11) و(14) و(15) من العلاقة (5-53) فنجد أن :

$$g(2) = \frac{q(2) - Q(T_2)}{|T_2| - 1} = \frac{0.2 - (0.01 + 0.01 + 0.03 + 0.02 + 0.01)}{5 - 1} = \frac{0.12}{4} = 0.03$$

أما في العقدة $t=3$ فإن $g(3)$ يحسب على العقد الخارجية: (7) و(17) و(18) و(19) و(20) و(21) فنجد من العلاقة (5–53) أن :

$$g(3)=\frac{q(3)-Q(T_3)}{|T_3|-1}=\frac{0.19-(0.04+0.01+0.02+0.01+0.01+0.015)}{6-1}=\frac{0.085}{5}=0.017$$

وكذلك نجد أن قيمة $g(t)$ في العقدة $(t=1)$ تحسب على جميع العقد الخارجية في الشجرة T وهي تساوي:

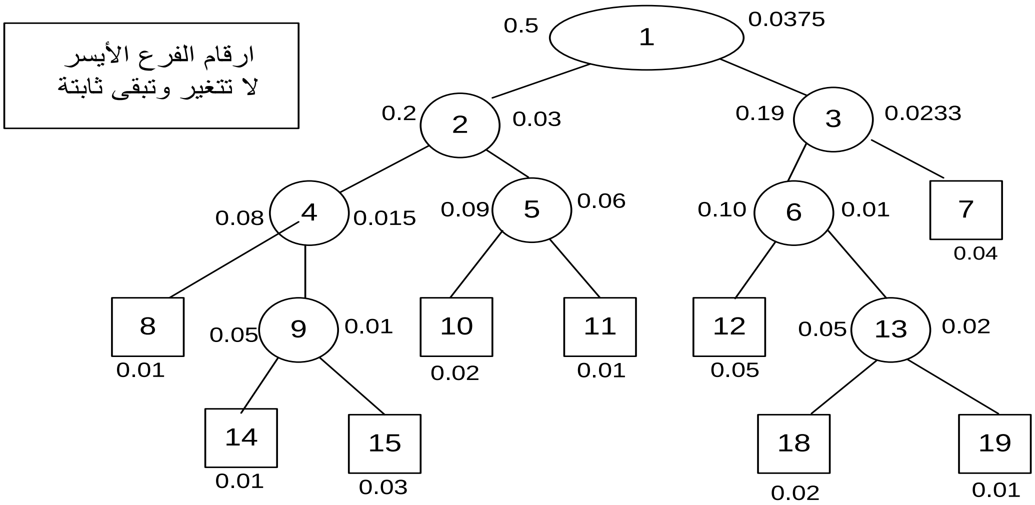
$$g(1)=\frac{0.5-(0.01+0.01+0.03+0.02+0.01+0.04+0.01+0.015+0.01+0.02+0.01)}{11-1}=$$

$$=\frac{0.5-0.185}{10}=0.0315$$

ولإجراء عملية التقليم نتبع الخطوات التالية :

1– ندرس جميع قيم التابع $g(t)$ المكتوبة على يمين العقد الداخلية ثم نحدد أصغرها، فتكون العقدة الداخلية t المقابلة لأصغر قيمة لـ $g(t)$ هي العقدة المرشحة لتكون عقدة خارجية (ورقة)، لذلك نعتبر تلك العقدة عقدة خارجية ونحذف جميع الفروع والعقد الصادرة عنها، ونرمز للشجرة الناتجة عن هذه الخطوة الأولى بـ T^1 . وفي مثالنا هذا نجد أن أصغر قيمة لـ $g(t)$ هي 0.0075 المقابلة للعقدة (12)، لذلك نعتبر هذه العقدة عقدة خارجية ونحذف جميع الفروع والعقد الصادرة عنها، ونضع قيمة $q(t)$ تحتها . علماً بأن $q(12)=0.05$

2– نعود ونقوم بحساب قيم التابع $g(t)$ لجميع العقد الداخلية السابقة لتلك العقدة (12) (أسلافها)، بما في ذلك عقدة الجذر الأصلي (1)، فنحصل على الشجرة المقابلة لهذه الخطوة وهي الشجرة T^1 التالية:



الشكل (5–8) نتيجة التقليم الأول T^1

وهنا نشير إلى أن قيم $g(6)$ و$g(3)$ و$g(1)$ قد تم حسابهما كما يلي :

$$g(6)=\frac{0.10-(0.05+0.02+0.01)}{3-1}=0.01$$

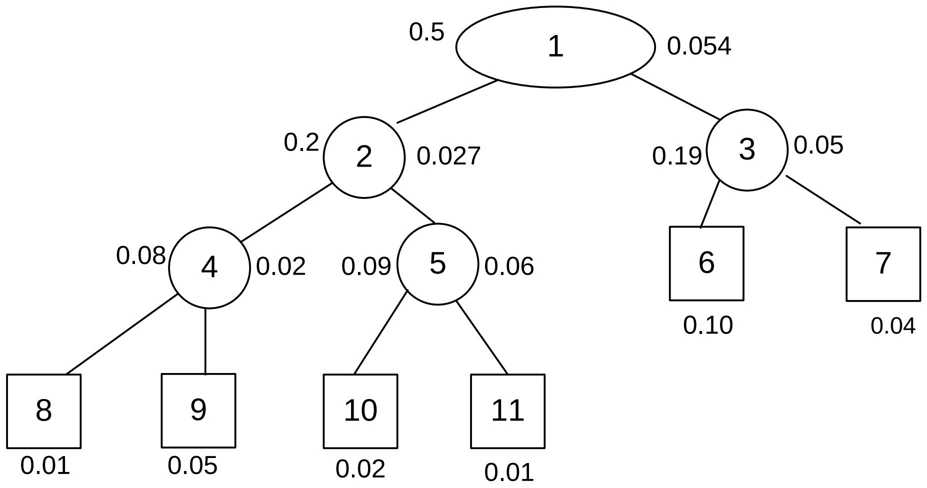
$$g(3)=\frac{0.19-(0.05+0.02+0.01+0.04)}{4-1}=0.02333$$

$$g(1)=\frac{0.5-(0.05+0.02+0.01+0.04+0.01+0.01+0.03+0.02+0.01)}{9-1}=0.0375$$

أما قيم $g(t)$ على العقد الداخلية التي في الفرع الأيسر فتبقى ثابتة كما كانت عليه في الشكل (5−7) السابق .

3− نعود الآن إلى الخطوة (1) وندرس من جديد قيم التابع $g(t)$ عند جميع العقد الداخلية المتبقية، ونحدد أصغر تلك القيم، ونعتبر العقدة الداخلية (أو العقد) المقابلة لها مرشحة لأن تكون عقدة خارجية، ثم يتم حذف العقد المتفرعة عنها. وفي هذه الخطوة نجد أن أصغر قيم $g(t)$ هي $g(6) = 0.01$ وكذلك $g(9) = 0.01$ (من الفرع الأيسر) .

لذلك نعتبر هاتين العقدتين عقدتين خارجيتين (ورقتين) ونحذف العقد التي تتفرع عنهما، فنحصل على الشجرة المقابلة لهذه الخطوة، والتي سنرمز لها بـ T^2 التالية :



الشكل (5−9) نتيجة التقليم الثاني T^2

ثم نقوم بحساب قيم $g(t)$ المقابلة للعقد السابقة للعقدتين (6) و(7) فنجد مثلاً أن:

$$g(3) = \frac{0.19 - (0.10 + 0.04)}{2 - 1} = 0.05$$

$$g(1) = \frac{0.5 - (0.10 + 0.04 + 0.01 + 0.05 + 0.02 + 0.01)}{6 - 1} = 0.054$$

وكذلك حسبنا قيم $g(t)$ المقابلة للعقد الأخرى على الفرع الأيسر ووضعناها على الشكل (5−9) فحصلنا على الشجرة T^2 .

4− نعود مرة أخرى إلى الخطوة (1) ونكرر تعليماتها، فنجد أن أصغر قيمة لـ $g(\mathrm{t})$ هي (0,02) التي تقابل العقدة الداخلية (4)، لذلك نجعل العقدة (4) عقدة خارجية، ونحذف العقدتين المتفرعتين عنها (8) و(9) من الشجرة T، فنحصل على الشجرة المقلمة T^3. وهكذا نتابع ونعود ونكرر تعليمات الخطوة (1)، ونكرر الحسابات، ثم نبحث عن العقدة الداخلية (العقد) المقابلة لأصغر قيمة لـ$g(\mathrm{t})$ فنجد أنها هي المقابلة للعقدة (2)، لذلك نحذف العقد (4) و(5) و(10) و(11) فنحصل على الشجرة المقلمة T^4 التي لا تحتوي على العقد (4) و(5) و(10) و(11) . ثم نكرر ذلك فنحصل على

الشجرة المقلمة T^5، التي لا تحتوي على العقدتين (6) و(7) . وأخيراً نحصل على الشجرة المقلمة T^6 التي تتألف من العقدتين (2) و(3) فقط . وإذا حذفنا العقدتين (2) و(3) نحصل على الشجرة المقلمة الأخيرة، التي تتألف من عقدة واحدة فقط هي عقدة الأصل .

وأخيراً نلخص نتيجة عمليات التقليم المتتالية للشجرة السابقة (5−7) في الجدول التالي :

جدول (5−5) نتائج التقليم المتتالي من T^1 حتى $T^{\bar{x}}$

رقم الشجرة المقلمة	العقد التي يجب حذفها من T^{k-1}	احتمال الخطأ الاجمالي $Q(T^{k-1})$	عدد العقد الخارجية $\|T\|$	رقم العقدة المقابلة للأصغر ∝	أصغر قيمة لـ ∝ أو للتابع $g(t)$	رقم الخطوة k
T^0	قيد الدراسة	0.185	11	البداية	0	0
T^1	10+17+20+2	0.20	9	12	0.0075	1
T^2	12+13+14+ 15+18+19	0.22	6	6+9	0.01	2
T^3	8+9	0.25	5	4	0.02	3
T^4	4+5+10+11	0.34	3	2	0.045	4
T^5	6+7	0.39	2	3	0.05	5
T^6	2+3	0.50	1	1	0.11	6

ملاحظة: نلاحظ أن احتمالات الخطأ الاجمالي المبين في العمود الخامس من الجدول (5−5) تتزايد كلما ازدادت خطوات التعليم. وهذا يعني أن التقليم الشديد يزيد من احتمال الخطأ الاجمالي لذلك فإن عملية التقليم تطبق على الأشجار الكبيرة، وتتوقف عن حد معين لاحتمال الخطأ الاجمالي .

5−9 طرائق اختبار جودة تصميم شجرة التصنيف [Webb. 2002 P.236]:

بعد إنجاز عمليتي التفريع والتقليم لشجرة التصنيف وتقدير معدل الخطأ فيهما . سنستعرض بعض طرائق اختبار جودة التفريع والتقليم وأهمها:

1− **طريقة استقلال التجارب:**

لنفترض أنه لدينا مجموعة بيانات ميدانية L_t ومجموعة من الاختبارات L_s لمصداقية بيانات العينة، وهذه المجموعة تتجمع بواسطة مجموعة من القرارات ضمن مجموعتين تقريباً . وذلك وفق الخطوات التالية :

أ− نستخدم المجموعة L_t لتشكيل الشجرة T بواسطة تفريع جميع العقد الممكنة، حتى تصبح جميع العقد الخارجية نقية وصافية، أي حتى تكون جميع عناصر العينة في كل عقدة خارجية تنتمي إلى مجموعة واحدة (فئة واحدة)، ولكن هذا يمكن أن لا يتحقق بسبب تداخل التوزيعات الاحتمالية. لذلك يمكن اتباع بديل آخر لإيقاف التفريع وذلك عندما يصبح عدد العناصر في كل عقدة خارجية أقل من عدد محدد مسبقاً، أو عندما يصبح عدد العناصر في العقدة اليسرى أو اليمنى مساوياً للصفر أو الواحد .

ب- نستخدم خوارزمية التقليم لتوليد أعشاش (خصائل) من الأشجار الفرعية نرمز لها بـ T^k، وذلك اعتماداً على مجموعة الاختبارات L_s .

ج- نختار الشجرة الفرعية T^k التي تكون قيمة تابع الشوائبية $Q(T^k)$ فيها أصغر ما يمكن .

2- طريقة المصداقية الاجمالية (Cors – Validation) :

إن طريقة المصداقية الاجمالية تقتضي أن نقوم بتقسيم المجموعة L_t إلى v مجموعة أو منطقة منفصلة هي:

$$L_1 \ , L_2 \ , L_3 \ , \dots , L_v \qquad (55-5)$$

وبحيث تكون أحجامها متساوية تقريباً (حجم كل منها يساوي $\frac{n}{v}$) .

ثم نفترض أن الفرق بين المجموعة الكلية L وأية مجموعة جزئية L_v يساوي :

$$L^v = L - L_v \qquad v: 1\ 2\ 3\ \dots . . v \qquad (56-5)$$

وسنرمز بـ $T(\ltimes)$ للشجرة الفرعية المقلمة، التي تكون كل العقد فيها لديها قيمة للتابع $g(t)$ تحقق الشرط التالي $\ltimes > g(t)$، حيث $\ltimes$ هو أصغر قيمة التابع $g(t)$ في الشجرة . وعندها تكون الشجرة $T(\ltimes)$ مساوية للشجرة T^k (الشجرة الفرعية المقلمة في المرحلة k) . وحيث أن k يتم اختياره أو تحديده بحيث يحقق العلاقة التالية:

$$\ltimes_k \le \ltimes \le \ltimes_{k+1} \qquad : (\ltimes_{k+1} \to \infty) \qquad (56-5)$$

وعندها نجد أن خطوات اختبار المصداقية الاجمالية تكون كما يلي:

1- نستخدم المجموعة L_t لتوليد الشجرة T بواسطة تفريع كل العقد كما ورد في طريقة استقلال التجارب السابق .

2- نستخدم خوارزمية التقليم لتوليد أعشاش (خصائل) من الأشجار المقلمة نرمز لها على الترتيب كما يلي:

$$T = T^0 \ge T^1 \ge T^2 \ge T^3 \ge \cdots \ge T^k = root(k) \qquad (57-5)$$

3- نستخدم المجموعة الجزئية L_v لتوليد الشجرة الفرعية T_v، ونحدد المجموعة التي تصنف في العقدة الخارجية المناسبة، وذلك من أجل : $v\ :\ 1\ 2\ 3\ \dots . \mathrm{v}$.

4- نستخدم خوارزمية التقليم لتوليد أعشاش من الأشجار الفرعية المقلمة T_v .

5- نحسب قيمة مؤشر المصداقية الاجمالية $Q^{cv}(T^k)$ (تقدير المصداقية الاجمالية لمعدل التصنيف الخاطئ) من العلاقة :

$$Q^{cv}(T^k) = \frac{1}{V}\sum_{v=1}^{V} Q_v\left[T_v\left(\sqrt{\ltimes_k * \ltimes_{k+1}}\right)\right] \qquad (58-5)$$

حيث أن: Q_v هو تقدير معدل التصنيف الخاطئ في المنطقة L_v، وذلك للشجرة الفرعية $T_v\left(\sqrt{\ltimes_k * \ltimes_{k+1}}\right)$.

6- نختار الشجرة الفرعية المقلمة T^* الصغرى التي يكون فيها :

$$Q^{cv}(T^*) = \min[Q^{cv}(T^k) \qquad (59-5)$$

7- نقدر معدل التصنيف الخاطئ من العلاقة :

$$\tilde{Q}(T^*) = Q^{cv}(T^*) \qquad (60-5)$$

إن الإجراءات السابقة تستخدم في عدة تطبيقات لشجرة التصنيف . وذلك خلال عمليات الإنشاء والتقليم، ويمكن دمجهما في تقنية واحدة للحصول على الحجم الصحيح للشجرة الناتجة عن المتحولات المستمرة أو المنقطعة أو المرتبة أو الأسمية أو المتضمنة مجموعات بيانات مختلفة من أنواع هذه المتحولات .

الفصل السادس

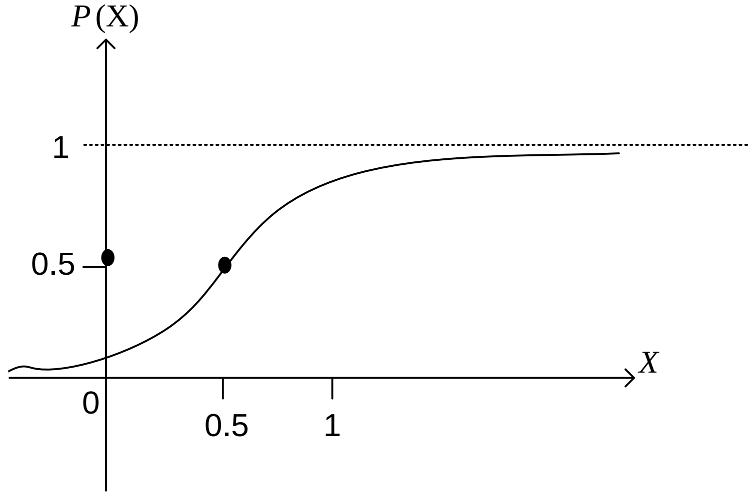
التحليل اللوجستي

6-1- تمهيد:

يهدف التحليل اللوجستي إلى تصنيف عناصر المجتمع المدروس إلى مجموعتين أو أكثر، وذلك باستخدام التوزيع الاحتمالي اللوجستي المعرف بالعلاقة التالية:

$$P(X) = \frac{1}{1 + \bar{e}^{(\beta_0 + \beta_X)}} \qquad -\infty < X < \infty \qquad (1-6)$$

حيث X هو شعاع المتحولات المؤثرة في عمليات التصنيف و $P(x)$ هو الاحتمال المقابل له ويأخذ قيمه في المجال $[0,1]$، وهو يرسم في المستوى المنحني التالي :



الشكل (6-1) منحني التوزيع اللوجستي

وهناك نوعان للتحليل اللوجستي هما: الثنائي والمتعدد :

1- **التحليل اللوجستي الثنائي:** وفيه يشترط أن يكون التابع Y نوعياً ويأخذ حالتين متنافيتين (نجاح أو فشل، ربح أو خسارة، مدخن أو غير مدخن، حامل للمرض أو غير حامل له، محقق لشرط ما أو غير محقق له، ...الخ)، وأن يأخذ مقابل الحالة المرغوبة الأولى القيمة العددية (1) واحد، وأن يأخذ مقابل الحالة الثانية القيمة العددية (0) صفر. أما المتحولات X المؤثرة في التصنيف فيمكن أن تكون كمية أو نوعية أو مختلطة، وتأخذ قيمها ضمن مجالات أو فئات محددة، ولا يشترط عليها أن تحقق أية شروط مسبقة .

2- **التحليل اللوجستي المتعدد:** وفيه يشترط أن يكون التابع Y نوعياً ويأخذ عدة حالات متنافية (مستوى التعليم، حالة العمل، الحالة الاجتماعية، ...الخ) وأن يأخذ مقابل إحدى الفئات القيمة (1) ومقابل الفئات المتبقية القيمة (0) .

ولتوضيح الأساس الرياضي للنموذج اللوجستي الثنائي نعود إلى العلاقتين (3−32) و(3−33) من الفصل الثالث، اللتين تعطيانا المنطقتين R_1 و R_2 المقابلتين للمجموعتين G_1 و G_2 والاحتمالين السابقين P_1 و P_2 وللمتحولين الخاضعين للتوزيع الطبيعي X_1 و X_2 .

ولنفترض الآن أن تكاليف التصنيف الخاطئ متساوية $\left[C(1/2) = C(2/1)\right.$أي أن$\left.\right]$ فعندها نجد أن العلاقة (3−32) تأخذ الشكل التالي:

$$R_1: (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)' * \boldsymbol{S}_{\boldsymbol{P}}^{-1} * \boldsymbol{X} - \frac{\boldsymbol{1}}{\boldsymbol{2}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}_{\boldsymbol{P}}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \geq \ln\left(\frac{P_2}{P_1}\right) \qquad (2-6)$$

وبما أن: $\ln\left(\frac{P_2}{P_1}\right) = -\ln\left(\frac{P_1}{P_2}\right)$ فإنه يمكننا كتابة (6−2) كما يلي:

$$+\ln\left(\frac{P_1}{P_2}\right) \geq +\frac{\boldsymbol{1}}{\boldsymbol{2}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\,\boldsymbol{S}_{\boldsymbol{P}}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) - (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\boldsymbol{S}_{\boldsymbol{P}}^{-1} * \boldsymbol{X} \qquad (3-6)$$

وبما أن $P_2 = 1 - P_1$ فإنه يمكننا كتابة (6−3) كما يلي:

$$\ln\left(\frac{P_1}{P_2}\right) = \ln\left(\frac{P_1}{1-P_1}\right) \geq \propto + \beta X \qquad (4-6)$$

حيث أن ∝ وB تقدران من بيانات العينة وحصراً من العلاقتين :

$$\tilde{\propto} = a = +\frac{1}{2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\,\boldsymbol{S}_{\boldsymbol{P}}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \qquad (5-6)$$

$$\tilde{\beta} = b = -(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\boldsymbol{S}_{\boldsymbol{P}}^{-1} \qquad (6-6)$$

نعود إلى العلاقة (6−4) فنجد أنه يمكننا صياغتها كما يلي:

$$\frac{P_1}{1-P_1} \geq e^{a+bX} \qquad (7-6)$$

نقسم بسط ومقام الطرف الأيسر على P_1 الموجب فنجد أن:

$$\frac{1}{\frac{1}{P_1} - 1} \geq e^{a+bX}$$

ثم نأخذ مقلوب الطرفين فنجد أن:

$$\frac{1}{P_1} - 1 \leq \frac{1}{e^{a+bX}} = \bar{e}^{(a+bX)}$$

$$\frac{1}{P_1} \leq +1 + \bar{e}^{(a+bX)}$$

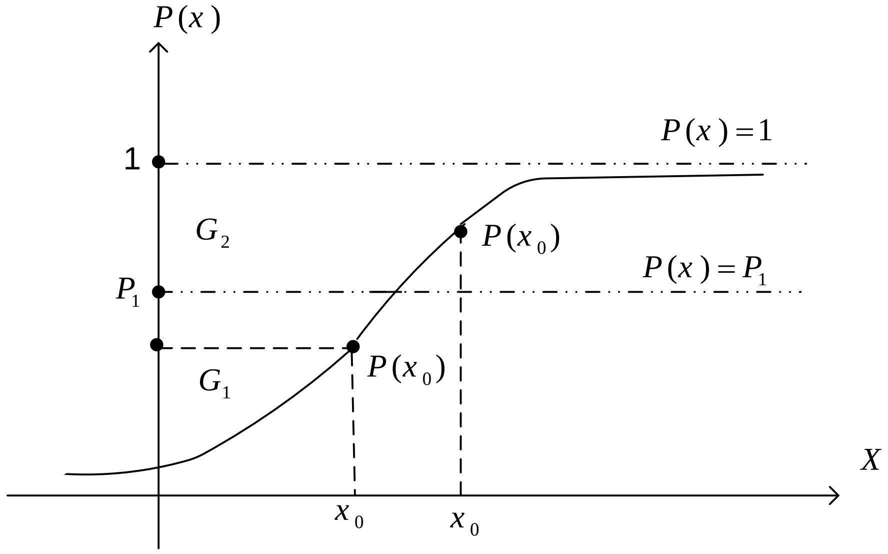
ثم نأخذ مقلوب الطرفين مرة أخرى فنجد أن:

$$\boxed{P_1 \geq \frac{1}{1 + \bar{e}^{(a+bX)}} = P(x)} \qquad (8-6)$$

وهي صيغة النموذج اللوجستي الذي يأخذ قيمه المستمرة في المجال $[0,1]$، وهكذا تصبح القاعدة (3−32) على الشكل التالي :

إذا كان لدينا x_0 عنصراً جديداً من المجتمع فإننا نصنفه في G_1 إذا كانت قيمة التابع اللوجستي : $P(x_0) = \frac{1}{1+\bar{e}^{(a+bx_0)}}$ أصغر من قيمة الاحتمال السابق P_1 (وهو معلوم لأنه عبارة عن نسبة المجموعة G_1 في المجتمع) .

أما إذا كانت قيمة $P(x_0)$ أكبر من الاحتمال P_1، فإننا نصنف x_0 في المجموعة G_2، ويمكننا تمثيل ذلك بيانياً كما يلي:



الشكل (6−2): قاعدة التصنيف للنموذج اللوجستي

ومنه نلاحظ أنه إذا كانت قيمة التابع $P(x_0)$ أصغر من P_1 فإن x_0 ينتمي إلى G_1، وإذا كانت أكبر من P_1 فإن x_0 ينتمي إلى G_2 .

مثال (6−1): لنفترض أن دراسة شملت (20) طالباً، لمعرفة علاقة متوسط عدد ساعات الدراسة يومياً X مع نتيجة اجتيازهم للامتحان Y، الذي يأخذ القيمة (1) في حالة النجاح والقيمة (0) في حالة الرسوب . وكانت نتائج الاستبيان كما في الجدول التالي :

جدول (6−1): نتائج الاستبيان لعلاقة عدد الساعات بنتيجة الامتحان [Wikipedia.org] :

i رقم الطالب	1	2	3	4	5	6	7	8	9	10
X: عدد الساعات	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50
Y: نتيجة الامتحان	0	0	0	0	0	0	1	0	1	0

يتبع

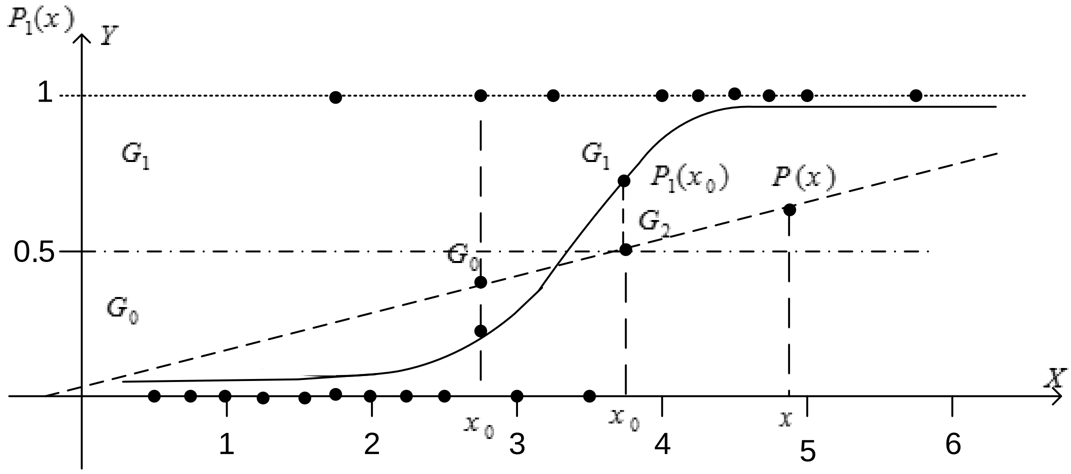
i رقم الطالب	11	12	13	14	15	16	17	18	19	20
X: عدد الساعات	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Y: نتيجة الامتحان	1	0	1	0	1	1	1	1	1	1

ونريد الآن معرفة مدى تأثير عدد ساعات الدراسة على احتمال النجاح .

ومن الجدول السابق نلاحظ أن عدد الناجحين $m = 10$، أي أن المعدل العام للنجاح = $P = \frac{10}{20} =$ 0.50، وبالمقابل نجد أن المعدل العام للرسوب $q = 0.50$.

وإذا أردنا رسم شكل الانتشار لهذه البيانات، فإننا نلاحظ أن المتحول المستقل X هو متحول مستمر ويأخذ قيمه في المجال [0 , 6] .

أما المتحول التابع Y فهو متحول منقطع وثنائي القيمة، فهو يأخذ القيمة $(Y = 1)$ في حالة النجاح، ويأخذ القيمة (0) في حالة الرسوب. وإذا قمنا برسم النقاط (x_1, y_1) على المستوى نجد أن قيم Y الصفرية تتوضع على المحور $0X$ وتميل نحو الجانب الأيسر، أما قيم Y المساوية للواحد فتتوضع على المستقيم$(Y = 1)$ وتميل نحو الجانب الأيمن . ويوجد بعض النقاط المتقابلة في المنطقة الوسطى كما هو مبين على الشكل التالي :



الشكل (6−3): شكل الانتشار للبيانات

والسؤال الآن كيف سنتعامل مع هذا الشكل العجيب؟

وللإجابة على هذا السؤال ندرس بعض خصائص الظواهر الثنائية من خلال بيانات المثال (5−1) السابق .

6−2-خواص الظواهر الثنائية :

إن الظواهر الثنائية هي عبارة عن توابع ثنائية تأخذ حالتين A و$\bar{A}$ فقط [موافق أو غير موافق، نعم أو لا، نجاح أو فشل، ربح أو خسارة، قبول أو رفض، ...الخ] . ولقد أصطلح على إعطاء التابع الثنائي Yقيمة الواحد (1) عندما تتحقق الحالة المرغوبة A، وقيمة الصفر (0) عندما تحقق الحالة غير المرغوبة $\bar{A}$ (عدم تحقق A). ولنفترض أنه عند إجراء (n=100) تجربة على أية ظاهرة ثنائية كانت نتائج تلك التجارب التي تخضع لتوزيع (برنويللي) كما في بالجدول التالي :

جدول (6−2): مخطط جدولي لتوزيع (برنوبللي) لـ 100 تجربة على Y

الحالة	$A = G_1$	$\bar{A} = G_0$	المجموع
قيمة التابع Y	1	0	----
احتمال التحقيق	p	q	$p + q = 1$
عدد التكرارات المطلقة	n_1	n_0	$n_1 + n_0 = n$
توزع عدد التجارب	60	40	$100 = n$

وعندما نكرر هذه التجربة n مرة سنحصل على عينة من قيم Y بحجم n ، وإن عناصرها تتوزع حسب الجدول (6−2) على مجموعتين هما:

- المجموعة G_1 : وهي مجموعة العناصر التي تقابل القيم$(Y = 1)$ [مجموعة الناجحين]، وتضم n_1 عنصراً، ويفترض أن يكون احتمال تحققها في كل تجربة ثابتاً ويساوي p، وإن p يقدر من $\tilde{p} = \frac{n_1}{n} = \frac{60}{100}$.
- المجموعة G_0 : وهي مجموعة العناصر التي تقابل القيم $(Y = 0)$ [مجموعة الراسبين] وتضم n_0 عنصراً، ويفترض أن يكون احتمال تحققها في كل تجربة يكون ثابتاً ويساوي $q = 1 - p$. ويقدر من: $\tilde{q} = \frac{n_0}{n} = \frac{40}{100}$.

وبناءً على نتائج هذه التجارب يمكننا تعريف عدة مؤشرات تستخدم في التحليل اللوجستي أهمها الأرجحية (*odds*) .

- **مفهوم الأرجحية (*odds*) وتعريفها:** يعود ظهور مفهوم الأرجحية إلى عمليات الرهان في الظواهر الثنائية (نجاح أو فشل) .

حيث يقال: إن إمكانية فوز اللاعب A تساوي n_1 مقابل n_0 للاعب $\bar{A}$. وإذا قام اللاعبان بإجراء $n = 100$ تجربة وفاز اللاعب A بـ $n_1 = 60$ تجربة وخسر $n_0 = 40$ تجربة منها، فإن تعريف الأرجحية لحادث فوز اللاعب A على اللاعب $\bar{A}$ يعطى بالعلاقة التالية:

$$odds(A) = \frac{n_1}{n_0} = \frac{\text{عدد مرات تحقق فوز } A}{\text{عدد مرات عدد فوز } \bar{A}} = \frac{60}{40} = \frac{3}{2} = \frac{1.5}{1} \qquad (9-6)$$

وعندها نقول أن إمكانية فوز A على $\bar{A}$ تساوي 60 مقابل 40 . وهنا يفضل اختصار الكسر $\frac{60}{40}$ إلى آخر عددين صحيحين مثل $\left(\frac{3}{2}\right)$، ونقول إن إمكانية فوز A على $\bar{A}$ تساوي 3 مقابل 2 . وإنه من الأفضل تحويل الكسر الأخير إلى نسبة عدد إلى الواحد مثل $\left(\frac{1.5}{1}\right)$، ونقول أن فوز A على $\bar{A}$ تساوي 1.5 مقابل 1 . ونكتب ذلك على الشكل 1.5 : 1 . (ويكتب بالعكس بالنسبة لـ $\bar{A}$ *1,5:1*) .

- وبطريقة مشابهة نعرف الأرجحية لفوز اللاعب $\bar{A}$ بالعلاقة :

$$odds(\bar{A}) = \frac{n_0}{n_1} = \frac{\text{عدد مرات تحقق فوز } \bar{A}}{\text{عدد مرات عدد فوز } A} = \frac{40}{60} = \frac{2}{3} = \frac{1}{1.5} \qquad (10-6)$$

- ومن التعريفين السابقين نستنتج أن:

$$odds(\bar{A}) = \frac{1}{odds(A)} \qquad (11-6)$$

$$odds(A) * odds(\bar{A}) = 1 \qquad (12-6)$$

- تعريف احتمال تحقق فوز اللاعب (A): ويعرف بالعلاقة التالية :

$$P(A) = \frac{n_1}{n_1+n_0} = \frac{n_1}{n} = p = \frac{60}{100} = 0.60 = \frac{1.5}{1+1.5} \qquad (13-6)$$

- احتمال تحقق فوز اللاعب ($\bar{A}$): ويعرف بالعلاقة التالية :

$$P(\bar{A}) = \frac{n_1}{n_1+n_0} = \frac{n_0}{n} = q = \frac{40}{100} = 0.40 = \frac{1}{1+1.5} \qquad (14-6)$$

ومن العلاقتين (6–13) و(6–14) نستخلص أن:

$$P(A) + P(\bar{A}) = p + q = 1 \qquad (15-6)$$

كما يمكننا استخلاص العلاقة التي ترتبط بين الأرجحية واحتمال تحقق حالتها، حيث نجد أنه يمكننا كتابة العلاقة (6–9) كما يلي:

$$odds(A) = \frac{n_1}{n_0} = \frac{\frac{n_1}{n}}{\frac{n_0}{n}} = \frac{P(A)}{P(\bar{A})} = \frac{p}{q} = \frac{p}{1-p} \qquad (16-6)$$

وكذلك نجد أن:

$$odds(\bar{A}) = \frac{n_0}{n_1} = \frac{\frac{n_0}{n}}{\frac{n_1}{n}} = \frac{P(\bar{A})}{P(A)} = \frac{q}{p} = \frac{1-p}{p} \qquad (17-6)$$

وسنستخدم العلاقة (6–16) في عمليات استخراج التابع اللوجستي [انظر الفقرة (6–7) في آخر هذا الفصل حول مفهوم الأرجحية وعلاقتها باحتمالات الظواهر الثنائية).

6–3–استخراج النموذج اللوجستي الثنائي :

لقد رأينا أن التابع Y (نتيجة الطالب في المثال (6–1)) هو تابع ثنائي ويأخذ إحدى القيمتين (1) للنجاح و(0) للرسوب، ومن شكل الانتشار (6–2) نلاحظ أن هذا التابع Y لا يصلح من وجهة نظر نظرية الانحدار، لأن يكون نتيجة لأي تركيب خطي (أو غير خطي) للمتحول المستقل X. لذلك يجب البحث عن بديل للتابع Y مرتبط به ويعبر عنه ويوصلنا معه .

ومن جهة أخرى نجد أن الاحتمال الشرطي لأن يأخذ التابع Y القيمة $(Y = 1)$، عند قيمة معطية x يساوي احتمال أن ينتمي العنصر المعلوم x إلى المجموعة G_1، ونكتب ذلك على الشكل التالي:

$$P(G_1/x) = P(Y = 1/x) = P_1(x) = Y \text{ الاحتمال اللاحق لـ} \qquad (18-6)$$

وهو احتمال النجاح عند أية قيمة معطية x، وهو تابع مستمر ويأخذ قيمه في المجال [1 , 0]، وهو يصلح لأن يكون بديلاً عن Y، لأنه أصبح من الممكن رياضياً دراسة علاقة $P_1(x)$ مع المتحول المستقل x .

وإذا استطعنا أن نجد العلاقة بين هذا الاحتمال $P_1(x)$ والمتحول X، فإننا نكون قد تجاوزنا المشكلة، التي واجهتنا أثناء تمثيل Y عبر X .

وهكذا نجد أنه يجب علينا الآن أن نقوم بإيجاد قيم $P_1(x)$ المقابلة لجميع قيم X، حتى نستطيع أن نقابلها مع قيم X، ثم استخلاص علاقة الانحدار بينهما دون وضع شروط مسبقة على المتحول X .

لذلك نقوم بحساب قيم الاحتمالات $P_1(x)$ اللاحقة من علاقات (بايز) (2–33) التي تأخذ الشكل التالي:

$$P_1(x) = P(G_1/x) = \frac{P * f(x/G_1)}{P * f(x/G_1) + q * f(x/G_0)} \qquad (19-56)$$

حيث أن: $f(x/G_1)$ و $f(x/G_0)$ هما التوزيعان التجريبيان لـ X ضمن المجموعتين G_1 و G_0 على الترتيب، وهما يحسبان (يعد تكرار التجربة n مرة) من التكرارات النسبية المقابلة لقيم X المختلفة كمايلي:

$$f(x/G_1) = \frac{n_1(x)}{n} \qquad (20-6)$$

$$f(x/G_0) = \frac{n_0(x)}{n} \qquad (21-6)$$

حيث أن: $n_1 + n_0 = n$

وأن: $n_1(x)$ هو عدد تكرار مرات النجاح مقابل القيمة (x) .

وأن: $n_0(x)$ هو عدد تكرار مرات الرسوب مقابل القيمة (x) .

وبعدها يمكننا أن نفترض أن العلاقة بين $P_1(x)$ وX هي علاقة انحدار خطية من الشكل التالي :

$$\tilde{P}_1(x) = \propto + \beta x \qquad (22-6)$$

ثم نقوم بحساب تقدير لـ $\propto$ و β بطريقة المربعات الصغرى أو بطريقة الإمكانية العظمى، فنحصل على مستقيم محدد يفصل بين المجموعتين G_1 و G_0. كما هو مبين على الشكل (6–3) السابق.

ومنه نحسب القيم النظرية للاحتمالات اللاحقة $\tilde{P}_1(x)$ الواقعة على ذلك المستقيم مقابل كل قيمة لـ X .

ثم نقوم بحساب الاحتمالات اللاحقة المتممة له: $\tilde{P}_0(x)$ من العلاقة :

$$\tilde{P}_0(x) = 1 - \tilde{P}_1(x) \qquad (23-5)$$

وأخيراً نقوم بمقارنة $\tilde{P}_1(x)$ مع $\tilde{P}_0(x)$ ونصنف أي عنصر جديد x وفق القاعدة التالية:

$(24-5)$ إذا $P_1(x) \geq P_0(x)$ نصنف x في المجموعة G_1 (في مجموعة الناجحين)

وإذا كان $P_1(x) < P_0(x)$ نصنف x في المجموعة G_0 (في مجموعة الراسبين)

والخط المستقيم على الشكل (6–3) يوضح ذلك .

ولكن الشكل (6–3) يظهر لنا أن جودة التمثيل لذلك المستقيم ضعيفة جداً (لأن قيمة R^2 صغيرة) . لذلك كان لابد من البحث عن حل آخر أو نموذج آخر لتمثيل العلاقة بين $P_1(x)$ وX، ومن أجل البحث

عن تلك العلاقة، سنحاول الاستفادة من شكل العلاقة (6−16) ونستبدل $P_1(x)$ بتابع مستمر جديد ومناسب، وهو متحول الأرجحية *(odds)*، والذي يعرف بدلالة الاحتمال $P_1(x)$ من خلال العلاقة (6−16) والتي تأخذ الشكل التالي :

$$odds(x) = \frac{P_1(x)}{1 - P_1(x)} = \frac{\text{احتمال تحقق } Y}{\text{احتمال عدم تحقق } Y} = \frac{n_1}{n_0} \qquad (25 - 6)$$

حيث أن: $P_1(x)$ هو احتمال أن يأخذ التابع Y القيمة (1) عند القيمة x، أو احتمال أن ينتمي العنصر x إلى المجموعة G_1، ونكتب ذلك كما يلي :

$$P_1(x) = P(Y = 1/x) = P(G_1/x) \qquad (26 - 6)$$

ولإيجاد علاقة الانحدار بين هذه الأرجحية (*odds*) والمتحول المستقل X، نفترض أنهما يرتبطان بعلاقة خطية لوغاريتمية كالعلاقة (6−4) السابقة، والتي نكتبها كما يلي:

$$ln(odds) = ln\left(\frac{P_1(x)}{1 - P_1(x)}\right) = \propto + \beta x \qquad (27 - 6)$$

ويسمى التابع اللوغاريتمي الأيسر باسم $logit\big(P_1(x)\big)$ ويكتب على الشكل التالي :

$$logit[P_1(x)] = ln\left[\frac{P_1(x)}{1 - P_1(x)}\right] = ln(odds) \qquad (28 - 6)$$

أي أن التابع $logit[P_1(x)]$ هو عبارة عن تحويل الاحتمال $P_1(x)$ المحسوب في (6−26) إلى (*odds*) ثم إلى الشكل اللوغاريتمي $ln\left[\frac{P_1(x)}{1-P_1(x)}\right]$، وهو عبارة عن تابع مستمر ويأخذ قيمه في المجال $]-\infty , +\infty[$ ، **لأنه لدينا** $0 \le P_1(x) \le 1$ ، **فيكون** $0 \le \frac{P_1(x)}{1-P_1(x)} < +\infty$، **وبالتالي فإن** $-\infty < ln\left[\frac{P_1(x)}{1-P_1(x)}\right] < +\infty$ **، ومن (6−27) و(6−28) يمكننا أن نفترض أن العلاقة بين التابع $logit[P_1(x)]$ والمتحول X هي خطية وتأخذ الشكل التالي :**

$$logit[P_1(x)] = \propto + \beta X = ln(odds) \qquad (29 - 5)$$

وبعد حساب القيم العددية لـ $logit[P_1(x)]$ من العلاقة (6−28)، يمكننا إيجاد تقديرات لـ $\propto$ و β بتطبيق طريقة المربعات الصغرى أو طريقة الامكانية العظمى .

والآن نعود إلى العلاقة (6−27) فنجد أنه يمكننا كتابتها على الشكل التالي :

$$\frac{P_1(x)}{1 - P_1(x)} = e^{\propto + \beta X} \qquad (30 - 6)$$

ومنها يمكننا أن نستخرج $P_1(x)$ كما يلي:

نقسم البسط والمقام في الطرف الأيسر على $P_1(x)$ فنجد أن:

$$\frac{1}{\frac{1}{P_1(x)} - 1} = e^{\propto + \beta X}$$

ثم نأخذ مقلوب الطرفين فنجد أن:

$$\frac{1}{P_1(x)} - 1 = \frac{1}{e^{\propto+\beta X}} = \bar{e}^{(\propto+\beta X)}$$

$$\frac{1}{P_1(x)} = 1 + \bar{e}^{(\propto+\beta X)}$$

نأخذ مقلوب الطرفين مرة أخرى ثم نضرب البسط والمقام بالحد الأسي فنجد أن:

$$P_1(x) = \frac{1}{1+\bar{e}^{(\propto+\beta X)}} = \frac{e^{\propto+\beta X}}{1+e^{\propto+\beta X}} \qquad (31-6)$$

وبناء على $(6-18)$ نجد أن:

$$P(G_1/x) = P_1(x) = \frac{1}{1+\bar{e}^{(\propto+\beta X)}} = \frac{e^{\propto+\beta X}}{1+e^{\propto+\beta X}} \qquad (32-6)$$

وهو عبارة عن منحني التابع اللوجستي المرسوم على الشكل $(3-6)$. وهنا نلاحظ أن هذا المنحني يختلف جذرياً عن المستقيم المرسوم على نفس الشكل، لأنه يقترب بطرفه الأيسر من نقاط المجموعة G_0، ويقترب بطرفه الأيمن من نقاط المجموعة G_1، وهو يعطينا بدقة أفضل، احتمال أن ينتمي x إلى G_1 مقابل كل قيمة x من قيم X .

ولحساب الاحتمال المتمم له نقوم بحساب $P_0(x)$ من العلاقة :

$$P_0(x) = 1 - P_1(x) = 1 - \frac{1}{1+\bar{e}^{(\propto+\beta X)}} = \frac{\bar{e}^{(\propto+\beta X)}}{1+\bar{e}^{(\propto+\beta X)}} \qquad (33-6)$$

وبتقسيم البسط والمقام على البسط نحصل على أن:

$$P_0(x) = P(G_0/x) = \frac{1}{1+e^{\propto+\beta X}} \qquad (34-6)$$

قاعدة: لاتخاذ قرار حول انتماء أي عنصر x لإحدى المجموعتين نطبق القاعدة التالية:

إذا كان $P_1(x) \geq P_0(x)$ نصنف x في المجموعة G_1

إذا كان $P_1(x) < P_0(x)$ نصنف x في المجموعة G_0 $\qquad (35-6)$

والمنحني المنطقي الملتوي على الشكل $(6-3)$ يوضح ذلك [مع ملاحظة أن G_1 حلت محل G_2 وأن G_0 حلت محل G_1 من الشكل $(6-2)$] .

ويمكننا تطوير أو تعديل القاعدة $(6-35)$ السابقة لتصنيف العناصر x، وذلك بأخذ نسبة الاحتمالين التالين $\frac{P_1(x)}{P_0(x)}$ فنجد أن :

$$\frac{P_1(x)}{P_0(x)} = \frac{\dfrac{1}{1+\bar{e}^{(\propto+\beta X)}}}{\dfrac{\bar{e}^{(\propto+\beta X)}}{1+\bar{e}^{(\propto+\beta X)}}} = e^{\propto+\beta X} \qquad (36-6)$$

وبذلك تصبح قاعدة التصنيف لأي عنصر x كما يلي:

قاعدة: نصنف أي عنصر x_0 إلى المجموعة G_1 إذا كانت النسبة :

$$\frac{P_1(x_0)}{P_0(x_0)} \geq 1 \qquad \Leftrightarrow \qquad e^{\propto+\beta x_0} \geq 1 \qquad (37-6)$$

ونصنف x إلى المجموعة G_0 إذا كانت النسبة :

$$\frac{P_1(x_0)}{P_0(x_0)} < 1 \qquad \Leftrightarrow \qquad e^{\propto+\beta x_0} < 1 \qquad (38-6)$$

ويمكن تحويل هذه القاعدة إلى الشكل الخطي بأخذ اللوغاريتم الطبيعي للطرفين فنحصل على القاعدة التالية .

قاعدة: نصنف أي عنصر x_0 إلى المجموعة G_1 إذا كانت قيمة التركيب الخطي موجبة أو غير سالبة، أي إذا كان :

$$\propto +\beta x_0 \geq 0 \qquad (39-6)$$

نصنف أي عنصر x إلى المجموعة G_0 إذا كانت قيمته سالبة، أي إذا كان :

$$\propto +\beta x_0 < 0 \qquad (40-6)$$

وبذلك نحصل على توابع تمييزية فاصلة بين المجموعات بأساليب متعددة . ولكن الاختلاف بين هذه القاعدة والقواعد الخطية (في الفصول السابقة)، هو أنها تستند على النسبة بين التوزيعات الاحتمالية اللاحقة للمجموعات بينما كانت القواعد السابقة تعتمد على النسبة بين الاحتمالات السابقة P_1 و P_2 .

كما إن شكل العلاقة (6–32) لـ$P_1(x)$ يساعدنا في الحصول على تقدير لـ $\propto$ و β بطريقة الامكانية العظمى .

6-4-: تقدير معالم النموذج اللوجستي (بطريقة الامكانية العظمى MLE) بمتحول واحد:

(*Maximum Likelihood Estimation*) *X*. *[Webb P.159]* بتصرف وإضافة] .

إن تقدير معالم النموذج اللوجستي $\propto$ و β بطريقة الإمكانية العظمى *MLE* يعتمد على بيانات إحصائية معينة، ويتطلب حساب المواصفات الميدانية لتلك للبيانات، ويرتبط بتصميم المعاينة التي تطبق على المجموعتين G_1 و G_0 .

وهناك عدة تصاميم لهذه المعاينة هي :

1- المعاينة المختلطة : وتكون من توزيعات مختلطة بين المجموعتين G_1 و G_0 (توزيع مشترك) .

2- المعاينة الشرطية لـ X: تجري بحيث يكون x ثابتاً، ثم نسحب عينة أو أكثر من العناصر (التي يمكن أن تنتمي إلى G_1 أو إلى G_0) .

3- المعاينة المنفصلة من كل مجموعة على حدة : حيث تكون التوزيعات الشرطية $P(x/G_1)$ أو $P(x/G_0)$ هي التوزيعات المعتمدة .

علماً بأن طريقة الإمكانية العظمى تعطينا تقديرات للمعلم β، تكون مستقلة عن شكل تصميم المعاينة . وإن بعض تصاميم المعاينة تعطينا تقديرات أفضل من تصاميم أخرى للمعلم β_0 (تصميم المعاينة المنفصلة) .

والآن لنفترض إننا نعمل ضمن المعاينة المختلطة، التي تفترض أن العينة العشوائية مسحوبة من المجتمع المختلط للمجموعتين بحجم n، ومؤلفة من: n_1 من G_1 و n_0 من المجموعة G_0، والتي سنرمز لها لضرورات رياضية بالرمز G_2 ولعدد عناصرها بـ n_2، وعندها نجد أن تابع الإمكانية العظمى L في هاتين المجموعتين يأخذ الشكل التالي:

$$L = L_1 * L_2 = \prod_{i=1}^{n_1} P(x_{1i}/G_1) * \prod_{i=1}^{n_2} P(x_{2i}/G_2) \qquad (41-6)$$

حيث أن $i: 1\ 2\ 3 \ldots n_s$ وحيث أن: x_{si} هي المشاهدة المسحوبة من المجموعة G_s وأن: $s: 1\ 2$ ،

وبما أن التوزيع الشرطي لـ $P(x\ /G_s)$ يساوي :

$$P(x\ /G_s) = \frac{P(x) * P(G_s/x)}{P(G_s)} \qquad : s = 1\ 2 \qquad (42-6)$$

نقوم الآن بتعويض $P(x/G_s)$ من (6–42) في العلاقة (6–41) فنحصل على أن:

$$L = \prod_{i=1}^{n_1} P(G_1/x_{1i}) \frac{P(x_{1i})}{P(G_1)} * \prod_{i=1}^{n_2} P(G_2/x_{2i}) \frac{P(x_{2i})}{P(G_2)} \qquad (43-6)$$

وبإخراج التوزيعين $P(G_1)$ و $P(G_2)$ خارج الجداء لأنه ليس لهما علاقة بدليل الجداء i ، فنجد أن :

$$L = \frac{1}{P(G_1) * P(G_2)} * \prod_{i=1}^{n_1} P(x_{1i}) * P(G_1/x_{1i}) * \prod_{i=1}^{n_2} P(x_{2i}) * P(G_2/x_{2i}) \quad (\mathbf{44-6})$$

وبما أن $P(x_{1i})$ و $P(x_{2i})$ ليس لهما علاقة بالمعلمتين ∝ و β للنموذج، لذلك نكتب جداءاتهما على الشكل التالي :

$$\prod_{i=1}^{n_1} P(x_{1i}) * \prod_{i=1}^{n_2} P(x_{2i}) = \prod_{i=1}^{n} P(x_i) : \qquad (\mathbf{45-6})$$

وهنا نلاحظ أن الجداء الأخير قد أصبح مأخوذاً على كامل حجم العينة $\boldsymbol{n}$

وبذلك نجد بأن تابع الامكانية العظمى يأخذ الشكل التالي:

$$L = \frac{\prod_{i=1}^{n} P(x_i)}{P(G_1) * P(G_2)} * \prod_{i=1}^{n_1} P(G_1/x_{1i}) * \prod_{i=1}^{n_2} P(G_2/x_{2i}) \qquad (46-6)$$

وبما أن الحد $\frac{\prod P(x_i)}{P(G_1)*P(G_2)}$ ليس له علاقة بالمعلمتين ∝ و β للنموذج اللوجستي . لذلك يمكننا افتراض (كما فعل اندرسون 1967)، أن L مستقل عن الاحتمالات السابقة $P(x)$. وبذلك يمكننا اختصار التابع L إلى تابع مكافئ له L' يساوي :

$$L' = \prod_{i=1}^{n_1} P(G_1/x_{1i}) * \prod_{i=1}^{n_2} P(G_2/x_{2i}) \qquad (47-6)$$

والآن نأخذ اللوغاريتم الطبيعي للطرفين في (6–47) فنجد أن :

$$ln\, L' = \sum_{i=1}^{n_1} ln\, P(G_1/x_{1i}) + \sum_{i=1}^{n_2} ln\, P(G_2/x_{2i})$$

وبتعويض ما تحت اللوغاريتمات بما تساويها من العلاقتين (6–32) و(6–34)، نحصل على أن $ln\, L'$ يساوي :

$$ln\, L' = \sum_{i=1}^{n_1} (\propto + \beta X_{1i}) - \sum_{i=1}^{n_1} ln\left(1 + e^{\propto + \beta x_{1i}}\right) - \sum_{i=1}^{n_2} ln\left(1 + e^{\propto + \beta x_{2i}}\right) \quad (\mathbf{48-6})$$

وبعد دمج المجموعين الأخيرين وأخذ المجموع على كامل العينة n نجد أن :

$$ln\, L' = \sum_{i=1}^{n_1} (\propto + \beta X_{1i}) - \sum_{i=1}^{n} ln\left(1 + e^{\propto + \beta x_i}\right) \qquad (49-6)$$

والآن نقوم بأخذ المشتقات الجزئية لـ$(\ell n\, L')$ بالنسبة لـ $\propto$ و β ونضعها مساوية للصفر، فنحصل بعد الإصلاح والاستبدال على أن هذين المشتقين يساويان:

$$\frac{\partial\, \ell n\, L'}{\partial \propto} = n_1 - \sum_{i=1}^{n} P(G_1/x_i) = 0 \qquad (50-6)$$

$$\frac{\partial\, \ell n\, L'}{\partial \beta} = \sum_{i=1}^{n_1} x_{1i} - \sum_{i=1}^{n} (x_i) * P(G_1/x_i) = 0 \qquad (51-6)$$

مع الانتباه إلى أن المجموعين الأخيرين ($\sum_{i=1}^{n}$) مأخوذين على جميع قيم x في العينة n، ثم نقوم بحل هاتين المعادلتين فنحصل على تقدير لـ $\propto$ و β، ومنهما نحصل على النموذج اللوجستي المطلوب .

ملاحظة: إذا كان عدد المتحولات المؤثرة X يساوي p متحولاً فإننا سنرمز لها بشعاع واحد كما يلي :

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

وعندها يمكننا كتابة العلاقة (6–27) كما يلي :

$$ln\left[\frac{P_1(x)}{1 - P_1(x)}\right] = \propto + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_p X_p = \propto + \beta' X \qquad (52-6)$$

حيث أن: $\beta'\left(\beta_1 \;, \beta_2 \ldots \beta_p\right)$

وعندها أيضاً فإن صيغة النموذج اللوجستي تأخذ الشكل التالي :

$$P_1(x) = \frac{1}{1 + \bar{e}^{(\propto + \beta' X)}} = \frac{e^{(\propto + \beta' X)}}{1 + e^{(\propto + \beta' X)}} \qquad (53-6)$$

ثم نجد أن $P_0(x)$ تحسب من العلاقة:

$$P_0(x) = 1 - P_1(x) = \frac{1}{1 + e^{\propto + \beta_1' X}} \qquad (53\,a - 6)$$

وعندها فإن معادلات الإمكانية العظمى لحساب $\propto$ و β' تأخذ الشكل التالي :

$$\frac{\partial\, ln\, L'}{\partial \propto} = n_1 - \sum_{i=1}^{n} P(G_1/x_i) = 0 \qquad (54 - 6)$$

$$\frac{\partial\, ln\, L'}{\partial \beta_j} = \sum_{i=1}^{n_1} (x_{1i})_j - \sum_{i=1}^{n} (x_i)_j * P(G_1/x_i)_j = 0 \qquad (55 - 6)$$

حيث أن J تأخذ القيم $j{:}\,1\ 2\ 3 \ldots P$

ومن هذه المعادلات يمكننا حساب تقديرات لـ $\propto$ و $\beta_1\ ,\beta_2 \ldots \beta_p$، علماً بأن المعادلات (6–54) و(6–55) هي معادلات غير خطية بالنسبة لـ $\propto$ و β، وإن حلها يحتاج إلى أساليب وبرامج متقدمة .

مثال (6–2): بناءً على بيانات المثال (6–1) السابق تم تقدير معالم النموذج (6–31) بطريقة الامكانية العظمى، فحصلنا على الجدول التالي: [المصدر : *Wikipedia*] :

جدول (6–3): نتائج علاقة الانحدار لـ *logit P(X)* على *X*

البيان	قيم الأمثال	الانحراف المعياري	قيمة *Z*	قيمة *P* حسب اختبار *(Wald)*
الثابت	$\propto = -4.0777$	1.7610	−2.316	0.0206
الساعات *X*	$\beta = 1.5046$	0.6287	2.393	0.0167

نلاحظ أن هذه المخرجات تشير إلى أن عدد ساعات الدراسة *X* ترتبط معنوياً مع احتمال النجاح في الامتحان (وذلك لأن قيمة *P* حسب اختبار *Wald* تساوي *P = 0.0167*، وهي أصغر من قيمة مستوى الدلالة 0.05).

وإن معادلة العلاقة بين $logit[P_1(x)]$ و*X* تساوي :

$$logit[P_1(x)] = ln\left[\frac{P_1(x)}{1 - P_1(x)}\right] = 1.5046X - 4.0777 = 1.5046[X - 2.71]$$

وهكذا نجد أن معادلة الأرجحية مع *x* هي:

$$odds(Y) = \frac{P_1(x)}{1 - P_1(x)} = e^{1.5046(X-2.71)}$$

ومنها نجد أنه إذا كان عدد الساعات *X = 2.71* فإن: $odds(Y) = e^0 = 1$، وفي هذه الحالة نستنتج أن إمكانية النجاح ($odds$) تساوي إمكانية الرسوب، وإن احتمال كل منهما يساوي 0.50، وفي هذه الحالة (عندما *X=2.71*) يمكننا أن نقول أن : أرجحية النجاح مقابل الرسوب هي كما يلي: واحد مقابل واحد ونكتب ذلك كما يلي (1 : 1).

ولحساب الاحتمال $P_1(x)$ من النموذج نلاحظ أن:

$$P_1(x) = \frac{1}{1 + \bar{e}^{(\tilde{\alpha}+\tilde{\beta}X)}} = \frac{1}{1 + \bar{e}^{(-4.0777+1.5046X)}}$$

فعندما يكون $X = 2$ نجد أن احتمال النجاح يساوي :

$$P_1(2) = \frac{1}{1 + \bar{e}^{(-4.0777+1.5046*2)}} = 0.26 \quad \Longrightarrow \quad P_0(2) = 0.74$$

وهذا يعني أن احتمال نجاح من يدرس ساعتين ($X=2$) أصغر من احتمال رسوبه (0.74)، لذلك نصنف ذلك الطالب في المجموعة G_0 ونعتبره من مجموعة الراسبين .

أما عندما يكون $X=2.71$ فإننا نجد أن احتمال النجاح يساوي:

$$P_1(x) = \frac{1}{1 + e^0} = \frac{1}{2} = 0.50$$

وهذا ما لاحظناه سابقاً وهذا يعني أن القيمة X=2,71 هي النقطة الفاصلة بين النجاح والرسوب .

أما عندما تكون $X=4$ فإن احتمال النجاح $P_1(x)$ يساوي :

$$P_1(4) = \frac{1}{1 + \bar{e}^{(-4.0777+1.5046*4)}} = 0.87 \quad \Longrightarrow \quad P_0(4) = 0.13$$

وهذا يعني أن احتمال نجاح من يدرس ($X=4$) ساعات أكبر من احتمال رسوبه (0.13)، لذلك نصنف ذلك الطالب في المجموعة G_1 [مجموعة الناجحين] .

أما عندما يكون $X=3$ فإن احتمال النجاح يساوي :

$$P_1(3) = \frac{1}{1 + \bar{e}^{(-4.0777+1.5046X)}} = 0.61 \quad \Longrightarrow \quad P_0(3) = 0.39$$

أي أن احتمال نجاح من كان يدرس ($X=3$) ساعات أكبر من احتمال رسوبه (0.39)، ولذلك نصنف ذلك الطالب في المجموعة G_1 [الناجحين]، رغم إنه كان من بين الراسبين (انظر الجدول (6−1)).

وأخيراً يمكننا أن ننظم بعض النتائج الممكنة لهذا النموذج في جدول مناسب كالجدول التالي :

جدول (6−4): قيم التحليل اللوجستي

احتمال النجاح $P_1(x)$	قيمة الـ *odds*	قيمة *ln (odds)*	عدد ساعات الدراسة X
مؤشرات النجاح في الامتحان			
0.07	$0.078 \approx 1:13.1$	−2.57	1
0.26	$0.034 \approx 1:12.91$	−1.07	2
0.50	$1 \approx 1:1$	0.00	2.71
0.61	$1.55 \approx \cdots$	0.44	3
0.87	$6.96 \approx \cdots$	1.94	4
0.97	$31.4 \approx \cdots$	3.45	5
0.99	141.16085	4.95	6

وبالعودة إلى الجدول (6−3) نجد أن احتمال الدلالة تساوي $P = 0.0167$، علماً بأن هذه القيمة محسوبة استناداً إلى علامة اختبار *Wald –Z*. ولكن هناك طريقة أفضل من طريقة *Wald*− الطريقة المعتمدة في حساب قيمة *P* للتوابع اللوجستية – وهي طريقة اختبار نسبة الامكانية العظمى (*LRT*)، والتي تعطينا من هذه البيانات أن قيمة *P* للنموذج اللوجستي المدروس $P = 0.00006$.

6−5− التحليل اللوجستي المتعدد: [webb P. 161 بتصرف]

إن التحليل اللوجستي المتعدد هو تعميم للتحليل اللوجستي الثنائي، وهو يعالج الحالات التي يكون فيها المجتمع مؤلفاً من عدة مجموعات (أو فئات) منفصلة نرمز لها بـ :

$$G_1 \;\; G_2 \;\ldots\; G_j \;\ldots\; G_g \qquad (58-6)$$

وعندها فإننا نشكل تابع الـ *logit* لكل زوج من هذه المجموعات، وإذا أخذنا المجموعة G_j مقابل أية مجموعة أخرى ولتكن G_g فإن تابع الـ *logit* يأخذ الشكل التالي:

$$logit\left[P_j(x)\right] = ln\left[\frac{P_j(x)}{P_g(x)}\right] = \propto_j + \beta_j' X \qquad (59-6)$$

حيث أن: $j = 1\ 2\ 3 \ldots g-1$ وأن: $X'(X_1\ \ X_2\ \ldots\ X_P)$ وأن: $\beta'(\beta_1\ \ \beta_2\ \ldots\ \beta_P)$

وهي تمثل $(g-1)$ تابعاً تمييزياً تفصل بين تلك المجموعات .

وهذا يعني أن لوغاريتم نسبة الإمكانية، $\left[\frac{P_j(x)}{P_g(x)}\right]$ لأي زوج ممكن من المجموعات يرتبط مع المتحولات *X* بعلاقة خطية. وهي تشكل المستوى الفاصل بينهما .

وبطريقة مشابهة لما عرضناه في التحليل اللوجستي الثنائي يمكننا صياغة الاحتمالات اللاحقة $P_j(x)$ و $P_g(x)$ بدلالة *X* بواسطة العلاقتين التاليتين :

$$P_j(x) = P\left(G_j/x\right) = \frac{e^{\propto_j + \beta_j' X}}{1 + \sum_{j=1}^{g-1} e^{\propto_j + \beta_j' X}} \qquad (60-6)$$

حيث أن: $j = 1\ 2\ 3 \ldots g-1$ وأن: $(g-1)$ هو عدد التوابع في (6−59)، أما التابع المتمم $P_g(x)$ فيساوي :

$$P_g(x) = P\left(G_g/x\right) = \frac{1}{1 + \sum_{j=1}^{g-1} e^{\propto_j + \beta_j' X}} \qquad (61-6)$$

وهكذا نجد أن قاعدة التصنيف التمييزي تصبح تابعة للعلاقة الخطية $\left(\propto_j + \beta_j' X\right)$. وتأخذ الصيغة التالية:

قاعدة: نصنف x_0 إلى المجموعة G_k إذا كانت $\propto_k + \beta_k' x_0 > 0$ وكانت قيمته أكبر من القيم الأخرى: أي إذا كانت:

$$0 < \propto_k + \beta_k' x_0 = Max\left[\propto_j + \beta_j' X\right] \qquad (62-6)$$

حيث أن: $j = 1\ 2\ 3 \ldots g-1$ وإذا كان العكس نصنفه إلى المجموعة G_g .

وكذلك يمكننا أن نستخلص تابع الإمكانية العظمى من العلاقة:

$$L = \prod_{j=1}^{g} \prod_{i=1}^{n_i} P\left(x_{ji}/G_j\right) \qquad \mathbf{(63-6)}$$

وبإجراء نفس العمليات والمعالجات على L نحصل على التابع المكافئ لهL' ونأخذ لوغاريتمه فنجد أن:

$$ln(L') = \sum_{j=1}^{g} \sum_{i=1}^{n_j} ln\, P\left(G_j/x_{ji}\right) \qquad (64-6)$$

وباشتقاقه نحصل على المعادلات التالية:

$$\frac{\partial ln(L')}{\partial \propto} = n_j - \sum_{X \text{كل}}^{n} P\left(G_j/X\right) = 0 \qquad (65-6)$$

$$\frac{\partial ln(L')}{\partial \left(\beta_j\right)_s} = \sum_{i=1}^{n_j} \left(x_{ji}\right)_s - \sum_{X \text{كل}}^{n} x_s * P\left(G_j/X\right)_s = 0 \qquad \mathbf{(66-6)}$$

حيث أن: j: 1 2 3 ….. $g-1$

وبحل هذه المعادلات نحصل على تقديرات للمعلمات $\propto$ و β ، ولكن بما أن هذه المعادلات ليست خطية، فإنه يتم البحث عن حلول تقاربية لها (بطريقة نيوتن أو بطريقة المعاودة)، وعندها نحتاج إلى حل ابتدائي ننطلق منه لإيجاد الحلول المتتالية والمتقاربة، ويمكن أن نأخذ الحل الابتدائي المقابل للقيم الصفرية، ونضع في البداية $\propto_j = 0$ و $\beta_1 = \beta_2 = \cdots \beta_g = 0$، ثم نتابع البحث عن قيم $\propto$ و β المثالية .

6-6-: تقييم جودة النموذج اللوجستي: [Wikipedia بتصرف] :

إن عملية تقييم جودة النموذج اللوجستي تختلف عن عملية تقييم الجودة في الانحدار الخطي، ومع أن الانحدار اللوجستي يقدر معالم النموذج β_j بطريقة الإمكانية العظمى، فهو لا يعتمد على معامل التحديد R^2 لتقييم جودة التمثيل . ولكنه يعتمد في تقييم جودة التمثيل على مفاهيم جديدة هي:

- **النموذج المشبع (Saturated Model)**: وهو النموذج الذي يمثل (نظرياً) البيانات المدروسة تمثيلاً تاماً، ونرمز لتابع الإمكانية العظمى لهذا النموذج بالرمز L_S . مأخوذة من (Likelihood of the saturated model) ولكن عملية الحصول على هذا النموذج في الحالة العامة قد تكون غير ممكنة . ويبقى L_S مجهولاً .
- **النموذج الصفري (Null Model)**: وهو النموذج الذي يتوافق مع فرضية العدم $\left(H_0: \beta_j = 0\right)$. أي أنه النموذج الذي لا يتضمن أي من المتحولات X، ويأخذ قيمة ثابتة هي قيمة الثابت β_0 أو$\propto$. ونرمز لتابع الإمكانية لهذا النموذج بالرمز L_0 .
- **النموذج المقدر (Fitted Model)**: وهو النموذج الذي ينتج عن حساب وتقدير المعالم β_i بطريقة الإمكانية العظمى، وهو يتضمن بعض المتحولات X (واحد على الأقل). ونرمز لتابع الإمكانية

العظمى لهذا النموذج بالرمز L_M . علماً بأن توابع الإمكانية العظمى L_0 و L_M تحسب اعتماداً على العلاقة (6–41) أو على العلاقة (6–47) وهي تأخذ شكل جداءات توزيع (بيرنويللي) التالي:

$$L = \prod_{\substack{i=1 \\ j=1,2}}^{n} P(G_j/x_{ji}) = \prod_{i=1}^{n} \left(P_1(x_i)\right)^{y_i}[1 - P(x_i)]^{1-y_i} \qquad (67-6)$$

حيث y_i تأخذ (1) أو (0) .

وبناء على ذلك تم تعريف المؤشرات التالية:

1– **حيدان النموذج المقدر** (**Deviance Model**): ويعرف بالعلاقة التالية:

$$D_{\not f} = -2\ln\frac{(\text{قيمة تابع الإمكانية العظمى للنموذج المقدر})}{(\text{قيمة تابع الإمكانية العظمى للنموذج المشبع})} = -2\ln\left[\frac{L_M}{L_S}\right] \qquad (68-6)$$

وهو يعبر عن الاختلاف النسبي بين النموذج (بمتحول واحد أو أكثر) وبين النموذج المشبع، لقد تم وضع الإشارة السالبة قبل اللوغاريتم لأن $L_M < L_S$.

2– **حيدان النموذج الصفري**: ويعرف كما يلي:

$$D_0 = -2\ln\frac{(\text{قيمة تابع الإمكانية العظمى للنموذج الصفرى})}{(\text{قيمة تابع الإمكانية العظمى للنموذج المشبع})} = -2\ln\left[\frac{L_0}{L_S}\right] \qquad (69-6)$$

3– **الفرق** Δ : ويحسب للتخلص من L_S المجهولة، وعند حساب الفرق بين D_0 و $D_{\not f}$ نجد أن:

$$\Delta = D_0 - D_{\not f} = -2\,ln\left[\frac{L_0}{L_S}\right] + 2\,ln\left[\frac{L_M}{L_S}\right] = -2\left[ln\left(\frac{L_0}{L_S}\right) - ln\left(\frac{L_M}{L_S}\right)\right]$$

$$\Delta = D_0 - D_{\not f} = -2\ln\left[\frac{\frac{L_0}{L_S}}{\frac{L_M}{L_S}}\right] = -2\,ln\left[\frac{L_0}{L_M}\right] = -2[ln\,L_0 - ln\,L_M\,] \qquad (70-6)$$

ويستخدم هذا الفرق Δ في تقييم جودة التمثيل، لأنه من هذه العلاقات يمكننا أن نستنتج أن: $L_0 < L_M < L_S$ وأن: $D_0 > D_{\not f}$ وأنه كلما ازدادت قيمة L_M واقتربت من L_S كان التمثيل جيداً، ولكن زيادة L_M تعني زيادة $D_{\not f}$، وهذا يؤدي إلى تناقص الفرق $\Delta = D_0 - D_{\not f}$، وهذا يعني أنه كلما تناقص الفرق Δ كان التمثيل جيداً . لذلك يجب علينا عند تقدير المعالم ∝ و β البحث عن النموذج الذي يجعل Δ أصغر ما يمكن، ولهذا السبب تخصص البرامج الحاسوبية عموداً خاصاً للفرق $\Delta = -2\,ln\left[\frac{L_0}{L_M}\right]$ وتراقب تغيراته وتتخذه كمعيار للتوصل إلى الحل المثالي للمعادلات (6–54) و (6–55) بطريقة المعاودة . وذلك عندما يأخذ ذلك الفرق Δ قيمة ثابتة خلال المعاودات الأخيرة .

4- **دراسة قيم معاملات التحديد اللوجستية:** وهناك عدة معاملات تحديد لتقييم جودة النموذج اللوجستي وهي :

- معامل نسبة الإمكانية العظمى (Likelihood Ratio):

$$R_L^2 = \frac{D_0 - D_f}{D_0} = 1 - \frac{ln\left[\frac{L_M}{L_S}\right]}{ln\left[\frac{L_0}{L_S}\right]} \qquad (71-6)$$

- معامل cox & snele:

$$R_{cs}^2 = 1 - \left(\frac{L_0}{L_M}\right)^{\frac{2}{n}} = 1 - e^{\frac{2[ln\,L_0 - ln\,L_M]}{n}} \qquad (72-6)$$

- معامل Mc Fadden:

$$R_{MCF}^2 = 1 - \frac{ln(L_0)}{ln(L_M)} \qquad (73-6)$$

وإن هذه المعاملات ترتبط مع بعضها بالعلاقات التالية :

$$R_{cs}^2 = 1 - \left(\frac{1}{L_0}\right)^{\frac{2}{n}R_{MCF}^2} \qquad (74-6)$$

$$R_{MCF}^2 = \frac{n}{2}\frac{(1 - R_{cs}^2)}{ln(L_0)} \qquad (75-6)$$

5- **حساب جدول تقاطع حالات التصنيف اللوجستي مع التصنيف الفعلي** (السابق) والذي يأخذ الشكل التالي (في حالة مجموعتين) :

جدول (6-5)

اللوجستي / الفعلي	G_1	$G_2 = G_0$	المجموع
G_1	n_{11}	n_{12}	n'_1
$G_2 = G_0$	n_{21}	n_{22}	$n'_2 = n_0$
المجموع	n_1	n_2	n

ومنه يتم حساب معدل التصنيف الصحيح بحساب نسبة مجموع عناصر القطر الرئيسي على المجموع الكلي n فنجد أن:

$$R = \frac{n_{11} + n_{22}}{n} 100\% \qquad (76-6)$$

ومنه يمكن حساب معدل التصنيف الخاطئ :

$$MR = 1 - R = \frac{n_{12} + n_{21}}{n} 100\% \qquad (77-6)$$

6- **حساب المعامل** kappa من العلاقة:

$$kappa = \frac{n\sum^2 n_{ii} - \sum^2 n_i * n_i'}{n^2 - \sum^2 n_i * n_i'} 100\% \qquad (78-6)$$

وهناك مقاييس أخرى لتحليل جودة مثل هذه الجداول، وكالحساسية (SE) والخصوصية (SP) ونسبة الأرجحية (OR) وغيرها . وهي مذكورة ومعرفة في العلاقات (6–85) و(6–87) و(6–89) من الفقرة اللاحقة (6–7) .

7- **اختبار (هوسمير- ليمشو** Hosmer– Lemshow**): لجودة المطابقة:** وهو يختبر صحة فرضية العدم التالية H_0: يتساوى عدد الحالات المشاهدة مع عدد الحالات المتوقعة (أي أن النموذج يمثل البيانات بشكل صحيح). وعندها نتخذ القرار بقبول فرضية العدم H_0، إذا كان مستوى المعنوية أو احتمال الدلالة p لاختبار (كاي مربع) أكبر من مستوى الدلالة المحدد بـ $\propto$.

8- **دراسة معنوية المتحولات الداخلة فيه:** وذلك من خلال مقارنة قيم احتمال الدلالة $(P = sig)$ مع مستوى الدلالة $\propto = 0.05$. فإذا كانت قيمة $\mathrm{P_j}$ المقابلة للمتحول X_j أصغر من 0.05 ، تكون علاقة ذلك المتحول مع النموذج معنوية والعكس بالعكس . ولكن دراسة هذه المعنوية تعتمد على اختبار جديد هو اختبار (*Wald*) وهو اختبار شبيه بالاختبار الطبيعي ويعرف بالعلاقة التالية: $Wald = \left(\frac{\beta_i}{SE(\beta_i)}\right)^2$.

9- **دراسة قوة تأثير كل متحول X_j على النموذج:** وذلك من خلال حساب نسبة الأرجحية له OR (Odds Ratio) والتي يمكن تعريفها وحسابها من العلاقة التالية :

$$\mathrm{OR} = \frac{odds(x_j+1)}{odds(x_j)} = \frac{e^{\tilde{\propto}+\tilde{\beta}(x_j+1)}}{e^{\tilde{\propto}+\tilde{\beta}(x_j)}} = e^{\beta} \qquad (78\,a-6)$$

لذلك يقوم البرنامج الحاسوبي بحساب القوة e^{β}، لجميع المتحولات الداخلة في النموذج . لأنها تدل على قوة تأثير المتحول X_j عندما يزداد بمقدار واحد (1) على النموذج ككل . وهو يعبر عن مرونة المتحول X_j بالنسبة للتابع (odds) .

مثال (6–3): في دراسة نشرها [شاهين 2014] حول تطبيق التحليل اللوجستي على مرضى سرطان الدم (اللوكيميا) في محافظة البصرة، لاحظ أولاً أن الجهات الطبية تصنف هؤلاء المرضى إلى نوعين أو مجموعتين هما: مرض سرطان الدم النخاعي الحاد (AME) ومرض سرطان الدم اللمفاوي (ALL) .

لذلك قام الباحث بسحب عينيتين عشوائيتين بحجمين متساويين $n_1 = n_2 = 80$ من ملفات هاتين المجموعتين، ثم قام بدراستها وتحليلها وتحديد المتحولات التي تفسر عملية الإصابة بهذين المرضين. فكانت ثمانية متحولات مختلطة (كمية ونوعية)، ثم قام بتحويلها إلى متحولات ثنائية وصنف قيم كل منها ضمن فئتين محددتين كما يلي:

X_1 – جنس المريض وصنفه إلى (ذكر = 1 ، انثى = 2) .

X_2 – عمر المريض وصنفه إلى $[(X_2 \le 50) = 1\ ,\ (X_2 > 50) = 2\ ]$.

X_3 – وزن المريض وصنفه إلى $[\ (X_3 \le 40) = 1\ ,\ (X_3 > 40) = 2\ ]$.

X_4 – نسبة الكريات الحمراء (P,C,V) وصنفه إلى $[\ (X_4 \le 0.35) = 1\ ,\ (X_4 > 0.35) = 2\ ]$.

X_5 – هيموغلوبين الدم (H,B) وصنفه إلى $[\ (X_5 \le 11.5) = 1\ ,\ (X_5 > 11.5) = 2\ ]$.

X_6 – معدل الكريات البيضاء (W,B,C) وصنفه إلى $[\ (X_6 \le 7.5) = 1\ ,\ (X_6 > 7.5) = 2\ ]$.

X_7– **سرعة ترسب الكريات الحمراء** (E,S,Q) **وصنفه إلى** $[(X_7 \le 22.5) = 1\ ,(X_7 > 22.5) = 2]$

X_8 – عدد الصفائح الدموية (P,C) وصنفه إلى $[\ (X_8 \le 150) = 1\ ,\ (X_8 > 150) = 2\ ]$.

كما قام بتسمية تابع الاستجابة الثنائي Y، الذي يعبر عن نوع الإصابة بمرض سرطان الدم وصنف حالتيه (النخاعي واللمفاوي) كما يلي:

إذا كان المريض مصاب بسرطان الدم النخاعي فإن قيمة $Y = 0$.

إذا كان المريض مصاب بسرطان الدم اللمفاوي فإن قيمة $Y = 1$.

ثم قام بجمع البيانات اللازمة له من ملفات عناصر العينة المسحوبة من المجموعتين، وعمل على تحليلها وتحويلها إلى متحولات ثنائية، ووضعها في جداول منظمة، ثم قام بتطبيق برنامج التحليل اللوجستي على المتحولات الثنائية التالية :

$$Y: X_1\ ,X_2\ ,X_3\ ,X_4\ ,X_5\ ,X_6\ ,X_7\ ,X_8$$

واستخدم لذلك طريقة (enter)، واختار طريقة (نيوتن) لإجراء المعاودة لإيجاد الحلول التقريبية للمعادلات غير الخطية الواردة في العلاقات (6–54) و(6–55)...الخ . وذلك من أجل الحصول على معالم النموذج اللوجستي التالي :

$$\ln\left(\frac{P_1(x)}{1-P_1(x)}\right) = \propto +\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_8 X_8$$

فكانت النتائج بعد إجراء (6) دورات للمعاودة كما يلي:

جدول (6–6): قيم الأمثال للتابع اللوجستي خلال دورات المعاودة:

رقم الدورة	$\Delta = -2\,\ell n\left[\frac{L_0}{L_M}\right]$	الثابت C	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	162.878	0.758	−2.268	−0.207	−0.871	−1.080	0.557	−0.359	−0.195	0.949
2	156984	0.901	−3.387	−0.236	−1.184	−1.370	0.669	−0.474	−0.194	1.246
3	156.243	0.913	−3.998	−0.228	−1.258	−1.130	0.679	−0.497	−0.185	1.212
4	156.200	0.912	−4.188	−0.227	−1.264	−1.416	0.679	−0.499	−0.183	1.316
5	156.200	0.912	−4.206	−0.227	−1.264	−1.416	0.679	−0.499	−0.183	1.316
6	156.200	0.912	−4.206	−0.227	−1.264	−1.416	0.679	−0.499	−0.183	1.316

وهنا نلاحظ أن قيم أمثال التابع اللوجستي قد استقرت تماماً بعد إجراء خمس أو ست دورات لمعاودة الحسابات .

كما إن قيمة مؤشر الجودة للنموذج المتمثل بالعمود الثاني الذي يتضمن قيم الفرق $\left(D_0 - D_f\right)$، والتي استقرت عند القيمة 156,200 في الدورة الرابعة ومابعدها. لذلك توقف عن الحسابات عند الدورة السادسة . ثم انتقل الباحث إلى تقدير معنوية معالم النموذج اللوجستي المتعدد فحصل على النتائج التالية:

جدول (6–7): نتائج الانحدار اللوجستي (أمثال النموذج اللوجستي وخواصها):

المقدرات المتحولات	β	S. E	Z Wald	df	Sig	Exp (B)	Lower of Ic 0.95	Upper of Ic 0.95
Constant	0.912	0.625	2.133	1	0.144	2.490	–	–
X_1 – جنس المريض	−4.206	1.071	15.426	1	0.000	0.015	0.002	0.122
X_2 – عمر المريض	−0.227	0.430	0.278	1	0.598	0.797	0.343	1.853
X_3 – وزن المريض	−1.264	0.435	8.461	1	0.004	0.283	0.121	0.662
X_4 – نسبة الكريات الحمراء	−1.416	0.878	2.603	1	0.107	0.243	0.043	1.356
X_5 – هيموغلوبين الدم	0.679	0.852	0.635	1	0.425	1.973	0.371	10.494
X_6 – معدل الكريات البيضاء	−0.499	0.459	1.179	1	0.278	0.607	0.247	1.494
X_7– سرعة ترسب	−0.183	0.407	0.202	1	0.653	0.833	0.375	1.850
X_8 – عدد الصفائح	1.316	0.404	10620	1	0.001	3.730	1.690	8.233

وعند دراسة هذا الجدول نجد أن عناصر β في العمود الأول هي نفسها العناصر التي استقرت عليها الحلول في السطر الأخير من الجدول (6–6) السابق. وهذا يجعلنا نكتب النموذج على شكل العلاقة (6–52) كما يلي:

$$ln(odds) = ln\left(\frac{P_1(X)}{1 - P_1(X)}\right) = 0.912 - 4.206X_1 - 0.227X_2 - 1.264X_3 - 1.416X_4 + \\ +0.679X_5 - 0.499X_6 - 1.183X_7 - 1.316X_8$$

ومنها يمكننا حساب الاحتمال $P_1(X)$ وهو احتمال الانتماء إلى G_1 وكتابته حسب العلاقة (6–53) كمايلي:

$$P_1(X) = \frac{1}{1 + e^{-(0.912 - 4.206X_1 - 0.227X_2 - \cdots \ldots + 1.316X_8)}}$$

ومنها أيضاً نقوم بحساب احتمال الانتماء إلى G_0 من العلاقة:

$$P_0(X) = 1 - P_1(X) = \frac{1}{1 + e^{+(0.912 - 4.206X_1 - 0.227X_2 - \cdots \ldots + 1.316X_8)}}$$

ولاختبار جودة التوفيق استخدم معيار نسبة الامكانية العظمى ورمز لها بـ $\chi^2 = -2\, ln\left[\frac{L_0}{L_1}\right]$ ، وهي تتبع تقاربياً للتوزيع χ^2 وبدرجة حرية $(p_1 - p_0)$. حيث: p_0 و p_1 هما أبعاد L_0 و L_1 على الترتيب.

وحيث أن L_0 هي قيمة دالة الامكانية العظمى عند الفرضية H_0 في G_0

وإن L_1 هي قيمة دالة الامكانية العظمى عند الفرضية H_1 في G_1 . فنجد أن :

جدول (6–8): اختبار χ^2 للتوفيق:

	$\chi^2 = -2\, ln\left[\frac{L_0}{L_1}\right]$	df	Sig
Model	65.607	8	0.008

وهذا يدل على معنوية النموذج بشكل عام (لأن قيمة sig أقل بكثير من 0.05)، ونلاحظ أن الجدول (6–7) السابق يعطينا عموداً خاصاً باسم Wald، وهو عبارة عن مؤشر Wald لتعريف معنوية ومصداقية تقدير كل من الأمثال β_i وهو يحسب من العلاقة التالية:

$$Wald_i = \left(\frac{\tilde{\beta}_i}{S.E(\tilde{\beta}_i)}\right)^2$$

ومن خلال عمود sig نلاحظ أن هناك ثلاثة متحولات فقط ، ذات تأثير معنوي وهي X_1 و X_3 و X_8، لأن قيم sig المقابلة لها أقل من 0.05، أما بقية المتغيرات فليس لها تأثيرات ذات أهمية أو معنوية .

كما نلاحظ أن العمود الذي يضم $EXP(\beta_1) = e^{\beta_i}$ فهو يعبر عن مقدار زيادة تابع الاستجابة Y أو التابع $logit\ (P)$ عندما يزداد المتحول المرافق لـ β_i بمقدار واحد، فمثلاً نجد أن :

$$eXP(\beta_1) = e^{-4.206} = 0.015$$

وهو يعني أن تابع الاستجابة Y سيزداد بمقدار 0.015 إذا تعبر المتحول الأول X_1 بمقدار (1): أي إذا تغير نوع الجنس من ذكر إلى أنثى .

وبناء على صيغة النموذج الأخيرة والمحددة نقوم بحساب الاحتمالات $P_1(X)$ لكل مفردة i من مفردات العينة ثم نقوم بحساب الاحتمالات المكملة $P_0(X)$ لكل مفردة i من العلاقة التالية:

$P_0(X) = 1 - P_1(X)$، ثم نقوم بمقارنة هذين الاحتمالين لكل مفردة i . ونصنف المفردة i كما يلي:

القاعدة: فإذا كانت $P_{1i}(X) \geq P_{0i}(X)$ فإننا ننسب تلك المفردة i إلى G_1 .

أما إذا كانت $P_{1i}(X) < P_{0i}(X)$ فإننا ننسب تلك المفردة i إلى G_0 .

وذلك كما فعلنا على الشكل (6–3) السابق.

وبعد إجراء كل هذه العمليات وتصنيف كل مفردات العينة في إحدى المجموعتين G_1 أو G_0، نقوم من جديد بتبويب النتائج الجديدة للتصنيف مع نتائج التصنيف الأصلي المستخدم في الإدارة . وعند إجراء ذلك التبويب حاسوبياً حصل الباحث على الجدول التالي:

جدول (6-9):

		التصنيف المستنبط من النموذج		النسبة المئوية	مجموع الأعداد
		G_0	G_1	%	n_i'
التصنيف الأصلي الإداري	G_0	57	23	71.3	80
	G_1	21	59	73.8	80
	–	49.75	51.25	72.5	–
مجموع الأعداد	n_i	78	82	–	160

ومن هذا الجدول نستنتج أن احتمال التصنيف الصحيح في المجموعة G_0 فقط كان يساوي 71.3% وفي المجموعة G_1 كان يساوي 73.8% ولكن الاحتمال الاجمالي للتصنيف الصحيح كان يساوي 72.5% وهذا يعني أن المعدل الاجمالي للتصنيف الخاطئ يساوي 27.5% . ولتقدير جودة التصنيف نحسب المؤشر kappa فنجد أن:

$$kappa = \frac{n * (\sum n_{ii}) - \sum n_i n_i'}{n^2 - \sum n_i n_i'} = \frac{160(57+59) - [(80*78) + (80*82)]}{(160)^2 - [(80*78) + (80*82)]}$$
$$= 0.45$$

وهي قيمة ضعيفة نسبياً . وتدل على جودة ضعيفة لعملية التصنيف المجراة في ذلك البحث .

6-7-: إضافات رياضية عن الأرجحية (odds):

لتوضيح مفهوم الأرجحية وعلاقتها بالاحتمالات للظواهر الثنائية، نفترض إننا نريد معرفة معدلات الإصابة بإحدى الأمراض (كالسكري مثلاً) بين الأشخاص المعرضين له (وراثياً وصحياً وسلوكياً)، فأخذنا عينة عشوائية مؤلفة من (1000) شخص من مجتمع المعرضين لذلك المرض، وأجرينا عليهم الفحوصات المخبرية والسريرية، ثم قمنا بتبويب نتائج هذه الاختبارات حسب حالة المريض الفعلية (D^+= مصاب فعلاً و D^-= غير مصاب)، وحسب نتيجة الاختبار الإيجابية (T^+= مصاب مخبرياً) والسلبية (T^-= غير مصاب مخبرياً)، وضعناها في الجدول التالي:

جدول (6-10): نتائج تبويب المرضى حسب حالة المريض الفعلية ونتيجة الاختبار (فرضية)

حالة المريض / نتيجة الاختبار	حالة المريض الفعلية		المجموع
	D^+= مصاب بالمرض	D^-= غير مصاب	
T^+ = الإصابة إيجابية	$a = 200$	$b = 85$	$n_1' = 285$
T^- = عدم إصابة	$c = 15$	$d = 700$	$n_2' = 715$
المجموع	$n_1 = 215$	$n_2 = 785$	$n = 1000$

واعتماداً على الجدول السابق (6-10) نلاحظ أن حجم العينة $n = 1000$ شخص وهي تتألف من مجموعتين: مجموعة المصابين فعلاً (D^+) وحجمها $n_1 = 215$ شخصاً، ومجموعة غير المصابين (D^-) وحجمها $n_2 = 785$ شخصاً .

ولكن نتائج الاختبارات أظهرت لنا أن عدد المصابين مخبرياً T^+ يساوي $(n_1' = 285)$ شخصاً، وعدد غير المصابين مخبرياً $n_2' = 715$ شخصاً .

وبناءً على ذلك يمكننا أن نعرف الاحتمالات التالية (وهي تحسب من هوامش الجدول):

احتمال أن يكون الشخص مصاباً : $P(D^+) = \frac{n_1}{n} = \frac{215}{1000} = 0.215 = \tilde{P}$

احتمال أن يكون الشخص غير مصاب : $P(D^-) = \frac{n_2}{n} = \frac{785}{1000} = 0.785 = \tilde{q}$

احتمال أن تكون نتيجة الاختبار إيجابية : $P(T^+) = \frac{n_1'}{n} = \frac{285}{1000} = 0.285$ **(79 − 6)**

احتمال أن تكون نتيجة الاختبار سلبية : $P(T^-) = \frac{n_2'}{n} = \frac{715}{1000} = 0.715$

وهنا نلاحظ أن: $P(D^+) + P(D^-) = 1$ وأن: $P(T^+) + P(T^-) = 1$

وبناء على ذلك يمكننا أن نعرف الأرجحيات (odds) المختلفة. لذلك نضع التعريف العام لأرجحية تحقق حادث (A) خل n تجربة عليه كما يلي:

$$\text{قيمة الأرجحية للحادث}(A) = \frac{\text{عدد مرات تحقق الحادث}(A)\text{ خلال التجربة}}{\text{عدد مرات عدم تحقق الحادث}(A)\text{ خلال التجربة}} = \frac{n_1(A)}{n_2(\bar{A})}$$

ونرمز لها بالرمز odds (A) ونكتبها رياضياً كما يلي:

$$odds(A) = \frac{n_1(A)}{n_2(\bar{A})} \quad : \quad \left[\text{ونكتب ذلك كمايلي: } n_1(A) : n_2(\bar{A})\right] \qquad (80-6)$$

حيث أن: $n_1(A) + n_2(\bar{A}) = n$

وبناء على ذلك يمكننا أن نحسب قيم الأرجحيات المختلفة لحالات المرضى ولنتائج الاختبارات من بيانات الجدول (6−10) كما يلي:

يوجد **43** مريضاً مقابل كل **157** غير مريض : $odds(D^+) = \frac{n_1}{n_2} = \frac{215}{785} = \frac{43}{157}$

يوجد **157** غير مريض مقابل كل **43** مريضاً : $odds(D^-) = \frac{n_2}{n_1} = \frac{785}{215} = \frac{157}{43}$

هناك **57** نتيجة إيجابية مقابل كل **143** نتيجة سلبية : $odds(T^+) = \frac{n_1'}{n_2'} = \frac{285}{715} = \frac{57}{143}$ **(81 − 6)**

هناك **143** نتيجة سلبية مقابل كل **57** نتيجة إيجابية : $odds(T^-) = \frac{n_2'}{n_1'} = \frac{715}{285} = \frac{143}{57}$

كما يمكننا ببساطة استخراج العلاقات التي تربط هذه الأرجحيات بالاحتمالات (6−79) السابقة. وذلك بتقسيم بسط ومقام كل أرجحية على حجم العينة (n) فنجد أن:

$$odds(D^+) = \frac{\frac{n_1}{n}}{\frac{n_2}{n}} = \frac{P(D^+)}{P(D^-)} = \frac{p}{q} = \frac{p}{1-p} = \frac{0.215}{0.785} = 0.274$$

أي أنه يوجد مقابل كل 274 مصاب بالمرض 1000 شخص غير مصاب به .

$$odds(D^-) = \frac{\frac{n_2}{n}}{\frac{n_1}{n}} = \frac{P(D^-)}{P(D^+)} = \frac{q}{p} = \frac{1-p}{p} = \frac{1}{odds(D^+)} = 3.65 \qquad (82-6)$$

أي أنه يوجد مقابل كل 365 غير مصاب يوجد 100 مصاب .

$$odds(T^+) = \frac{\frac{n'_1}{n}}{\frac{n'_2}{n}} = \frac{P(T^+)}{P(T^-)} = \frac{p'}{q'} = \frac{p'}{1-p'} = \frac{0.285}{0.715} = 0.363$$

أي أنه مقابل كل 363 نتيجة إيجابية يوجد 1000 نتيجة سلبية .

$$odds(T^-) = \frac{n'_1/n}{n'_2/n} = \frac{P(T^-)}{P(T^+)} = \frac{q}{p'} = \frac{1-p'}{p'} = \frac{1}{odds(T^+)} = 2.75$$

أي أنه مقابل كل 275 نتيجة سلبية يوجد 100 نتيجة إيجابية .

ومما سبق يمكننا استنتاج أن أرجحية أي حادث (A) ترتبط باحتمال تحققه وعدم تحققه وفق العلاقة التالية:

$$odds(A) = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1-P(A)} \qquad (83-6)$$

وأن نستنتج أن: $odds(\bar{A}) = \frac{1}{odds\ (A)}$ وأن: $odds(\bar{A}) = \frac{P(\bar{A})}{P(A)}$

$$odds(A) * odds(\bar{A}) = 1 \qquad (84-6)$$

كما يمكننا أن نعرّف عدداً من المؤشرات الاحصائية اعتماداً على التكرارات الداخلية والتكرارات الهامشية والمبينة في الجدول (6–10) السابق. وهذه المؤشرات هي الاحتمالات الشرطية التالية:

1- الحساسية (sensibility): وهي نسبة عدد المرضى (D^+) الذين اعتبرتهم الاختبارات إنهم مصابين بالمرض (أي كانت نتيجة الاختبار T^+ متطابقة مع الحالة الفعلية للمريض D^+) ونرمز لها بالرمز $P(T^+/D^+)$ ونحسبها من الاحتمال الشرطي التالي:

$$P(T^+/D^+) = \frac{a}{a+c} = \frac{a}{n_1} = \frac{200}{215} = 0.93023 \qquad (85-6)$$

وهو احتمال أن تكون نتيجة اختبار المريض متطابقة مع حالته المرضية (أي أن تكون نتيجة الاختبارات الإيجابية صحيحة) .

2- الخصوصية أو النوعية (specificity): وهي نسبة عدد غير المرضى (D^-) الذين اعتبرتهم الاختبارات إنهم غير مصابين (أي كانت نتيجة الاختبارات السلبية T^- متطابقة مع الحالة الفعلية للمريض D^-)، ونرمز لها بـ $P(T^-/D^-)$ ونحسبه من الاحتمال الشرطي التالي:

$$P(T^-/D^-) = \frac{d}{b+d} = \frac{d}{n_2} = \frac{700}{785} = 0.89172 \qquad (86-6)$$

وهو احتمال أن تكون نتيجة اختبار غير المريض متطابقة مع حالته الصحية (أي أن تكون نتيجة الاختبارات السلبية الصحيحة) .

3- نسبة الاختبارات الإيجابية الكاذبة (غير الصحيحة): وهي نسبة عدد غير المرضى (D^-) الذين اعتبرتهم الاختبارات مصابين بالمرض (T^+) ونرمز لها بالرمز $P(T^+/D^-)$ ونحسبها من الاحتمال الشرطي التالي:

$$P(T^+/D^-) = \frac{b}{b+d} = \frac{85}{285} = 0.10828 \qquad (87-6)$$

وهو احتمال أن تكون نتيجة اختبار غير المريض إيجابية (النتيجة غير صحيحة).

4- نسبة الاختبارات السلبية الكاذبة (غير الصحيحة): وهي نسبة عدد غير المرضى (D^+) الذين اعتبرتهم الاختبارات أنهم غير مصابين (T^-) ونرمز لها بالرمز $P(T^-/D^+)$ وتحسب من الاحتمال الشرطي التالي:

$$P(T^-/D^+) = \frac{c}{a+c} = \frac{15}{215} = 0.06977 \qquad (88-6)$$

وهو احتمال أن تكون نتيجة اختبار المريض سلبية (النتيجة غير صحيحة).

5- مصداقية الاختبارات: وهي نسبة كل الاختبارات الصحيحة (الإيجابية والسلبية) إلى المجموع الكلي للاختبارات (n) ويُرمز لها بـ AT وتحسب من العلاقة:

$$AT = \frac{a+d}{n} = \frac{a+d}{a+b+c+d} = \frac{200+700}{1000} = 0.90 \qquad (89-6)$$

6- معدل التصنيف الخاطئ: وهو نسبة كل الاختبارات غير الصحيحة إلى المجموع الكلي للاختبارات (n) ويرمز له بـ MR وتحسب من العلاقة:

$$MR = \frac{b+c}{n} = \frac{b+c}{a+b+c+d} = \frac{85+15}{1000} = 0.10 \qquad (90-6)$$

وهنا نلاحظ أن: $MR = 1 - AT$

كما يمكننا تعريف عدداً من الأرجحيات الشرطية. بناء على التكرارات الداخلية فقط المبينة في الجدول (6–10) السابق وهذه الأرجحيات هي:

$$odds(T^+/D^+) = \frac{a}{c} = \frac{200}{15} = \frac{40}{3} \qquad (91-6)$$

أي أنه يوجد 40 نتيجة إيجابية مقابل كل 3 نتائج سلبية وذلك عند مجموعة المرضى D^+

$$odds(T^+/D^-) = \frac{b}{d} = \frac{85}{700} = \frac{17}{140}$$

أي أنه يوجد 17 نتيجة إيجابية مقابل كل 140 نتيجة سلبية عند مجموعة غير المرضى D^- .

$$odds(T^-/D^+) = \frac{c}{a} = \frac{15}{200} = \frac{3}{40}$$

أي يوجد 3 نتائج سلبية مقابل كل 40 نتيجة إيجابية عند مجموعة المرضى D^+ .

يوجد 140 سلبية مقابل كل 7 إيجابية عند غير المرضى $odds(T^-/D^-) = \frac{d}{b} = \frac{700}{85} = \frac{140}{7}$

يوجد 40 مريض مقابل كل 17 في التحاليل الإيجابية $odds(D^+/T^+) = \frac{a}{b} = \frac{200}{85} = \frac{40}{17}$

يوجد 3 مريض مقابل كل 140 في التحاليل السلبية $odds(D^+/T^-) = \frac{c}{d} = \frac{15}{700} = \frac{3}{140}$

يوجد 17 غير مريض مقابل كل 40 في التحاليل الإيجابية $odds(D^-/T^+) = \frac{b}{a} = \frac{85}{200} = \frac{17}{40}$

يوجد 140 غير مريض مقابل كل 3 في التحاليل السلبية $odds(D^-/T^-) = \frac{d}{c} = \frac{700}{15} = \frac{140}{3}$

وأخيراً نعرف المؤشرات الهامة التالية:

7- نسبة الأرجحية (odds Ratio): وتعرف للمرضى ككل ونرمز لها بـ OR وتحسب من العلاقة التالية:

$$OR = \frac{odds(D^+/T^+)}{odds(D^+/T^-)} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a*d}{b*c} = \frac{200*700}{85*15} = 109.80 \qquad (92-6)$$

أما نسبة الأرجحية لغير المرضى فتحسب من العلاقة :

$$\overline{OR} = \frac{odds(D^-/T^+)}{odds(D^-/T^-)} = \frac{\frac{b}{a}}{\frac{d}{c}} = \frac{b*c}{a*d} = \frac{1}{OR} = 0{,}009107 \qquad (93-6)$$

8- نسبة المخاطرة (Risk Ratio): وهي نسبة احتمال الحادث المطلوب على احتمال الحادث المتمم له ونرمز لها بـ RR وتحسب لاختبارات المرضى من العلاقة :

$$RR = \frac{P(D^+/T^+)}{P(D^+/T^-)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{ac+ad}{ac+bc} = \frac{200*15+200*700}{200*15+85*15} = 33.45 \qquad (94-6)$$

وعندما يكون الحد $(a*c)$ صغيراً بالنسبة لـ $(a*d)$ فيمكن اهماله وتصبح نسبة المخاطرة المقربة كما يلي:

$$\widetilde{RR} \approx \frac{ad}{bc} \approx OR \qquad (95-6)$$

9- المخاطرة المطلقة لاختبارات المرضى وتحسب من العلاقة :

$$AR = P(D^+/T^+) - P(D^+/T^-) = \left[\frac{a}{a+b} - \frac{c}{c+d}\right] = 0.68078 \qquad (96-6)$$

الفصل السابع

التحليل العنقودي Cluster Analysis

7-1 تمهيد:

إن التحليل العنقودي هو مجموعة من الطرائق الرياضية لاستكشاف الخواص الهيكلية للبيانات الاحصائية لمفردات العينة المسحوبة من المجتمع، وذلك من خلال تصنيفها إلى مجموعات (ضمن عناقيد)، بحيث تكون المفردات داخل كل مجموعة متشابهة مع بعضها (وذلك بالنسبة للمتحولات أو الصفات المعتمدة لذلك)، وبحيث تكون المجموعات مختلفة عن بعضها البعض، وبعبارة أخرى أن هدف التحليل العنقودي هو تجميع مفردات العينة وتصنيفها ضمن مجموعات متجانسة داخلياً ومتباينة خارجياً بين بعضها البعض .

وخلافاً لما ذكرناه حول مسألة التصنيف في التحليل التمييزي، حيث يكون انتماء كل مشاهدة إلى إحدى المجموعات معروفاً، فإن التحليل العنقودي يهدف إلى التنبؤ بالمجموعة التي سينتمي إليها أي عنصر جديد، ويحاول اكتشاف عدد أو تركيبات هذه المجموعات .

كما يستخدم التحليل العنقودي لتجميع أي مجموعة من المتحولات X ضمن مجموعات متجانسة ومنفصلة. ويطبق هذا الاتجاه في مراجعة وتنقيح الاستبيانات بناءً على إجابات المستجوبين على مسودات الاستبيان، حيث أن تجميع الأسئلة حسب المتوسطات بواسطة التحليل العنقودي يساعدنا في التعرف على الأسئلة الضعيفة وإعادة النظر فيها . وهذا يزيد حظوظ معدل الإجابات الجيدة بالنسبة لإجمالي أسئلة الاستبيان .

وأخيراً يمكننا أن نعرف التحليل العنقودي بما يلي: هو أحد الأساليب الاحصائية الرياضية لتقسيم عناصر المجتمع المدروس إلى عدة مجموعات متعالقة ومتجانسة داخلياً (متشابهة) ومتباينة خارجياً عن بعضها البعض . أي أنه يهدف إلى جعل تباين العناصر داخل كل مجموعة أصغر ما يمكن، وجعل التباين بين المجموعات (بين مراكزها) أكبر ما يمكن . وبصورة عامة يتفرع التحليل العنقودي إلى نوعين أساسيين هما:

- التحليل العنقودي الهرمي (Hierarchical) .
- التحليل العنقودي غير الهرمي (Non– Hierarchical)

ويعتبر أسلوب التحليل العنقودي الهرمي من الأساليب المفضلة في التحليل العنقودي، لأنه يعتمد على أسس بسيطة، ويعمل على عنقدة مفردات العينة (n مفردة). وبشكل متتالي، ضمن m عنقوداً، بواسطة دمج المفردات المتقاربة ضمن مجموعات متعالقة تسمى عناقيد، وبحيث يكون العنقود الأول C_1 أبسط

العناقيد، ويكون العنقود الأخير أعقدها وأشملها (لأنه يضم جميع مفردات العينة)، وبحيث يتألف كل عنقود من عدة مجموعات متقاربة ومرتبطة مع بعضها بواسطة علاقات تحقق شروط التقارب المفضلة (حسب المتحول أو الصفة المدروسة) .

ويستخدم التحليل العنقودي الهرمي لعنقدة مفردات العينة أسلوبين عمليين هما:

1- أسلوب التجميع The Agglomerative Technique :

ويفترض هذا الأسلوب من البداية أن كل مفردة من مفردات العينة تشكل عنقوداً خاصاً بها، ثم يتم دمج أي مفردتين متقاربتين في عنقود خاص (أول)، ثم نضيف إليهما أي مفردة ثالثة متقاربة مع ذلك العنقود فيتشكل لدينا عنقود ثانٍ، وهكذا نتابع إضاقة المفردات واحدة بعد الأخرى إلى بعضها أو إلى العناقيد السابقة، مع تحديد العلاقات بينها ضمن العناقيد، حتى نحصل على العنقود الأخير، الذي يضم جميع مفردات العينة (n مفردة) مع العلاقات التي ترتبط بينها. ويعتمد هذا الأسلوب على مصفوفة التقارب بين مفردات العينة حسب المسافات المحسوبة .

2- أسلوب التجزئة أو التقسيم The Divisive Technique :

ويفترض هذا الأسلوب من البداية أن جميع مفردات العينة (n مفردة) تشكل عنقوداً واحداً شاملاً . ثم تتم تجزئته إلى عناقيد جزئية متباينة تتضمن عدداً أقل من المفردات . وبعد فرز هذه العناقيد وتحديد العلاقات بينها تتم تجزئتها إلى عناقيد أصغر فأصغر، وتتابع هذه العملية حتى يتكون عنقود خاص لكل مفردة من مفردات العينة أو نتوقف عن التقسيم عند حد معين .

وأخيراً نشير إلى أن هذين الأسلوبين يعتمدان على بيانات العينة المدروسة . وعلى طبيعة المتحولات المستقلة $X_1\ X_2\ X_3\ ....\ X_P$ المستخدمة في عملية العنقدة .

فإذا كانت المتحولات $X_1\ X_2\ ....\ X_P$ متحولات كمية . فإننا نقوم بحساب عناصر المصفوفة D التي تسمى مصفوفة التباعد Dissimitlarity وهي عبارة عن المسافات التي تفصل بين مفردات العينة .

أما إذا كانت المتحولات $X_1\ X_2\ ....\ X_P$ نوعية أو مختلطة فإننا نقوم بحساب عناصر مصفوفة أخرى S والتي تسمى بمصفوفة التشابه أو التقارب Similarity، وهي عبارة عن أوزان التكرارات التي تقابل الأزواج المتشابهة (j, k)، ولهذا فإننا سنقوم بتقديم كيفية حساب عناصر هاتين المصفوفتين حسب المتحولات المؤثرة على مفردات العينة: أي حسب المتحولات الكمية أو النوعية أو حسب المتحولات المختلطة من هذين النوعين .

كما نشير إلى أن بيانات العينة لـ n مفردة أو مشاهدة تُنظم حسب المتحولات النظامية (المعيارية أو الثنائية) $X_1\ X_2\ X_3\ ....\ X_P$ المعتمدة في عملية العنقدة وتُوضع في جدول مناسب كما يلي:

جدول (7−1) نموذج جدول البيانات اللازمة للتحليل العنقودي :

المتحولات / رقم المفردة	X_1	X_2	X_3	…..	X_i	….	X_P
1	x_{11}	x_{12}	x_{13}	…	x_{1i}	…	x_{1P}
2	x_{21}	x_{22}	x_{23}	…	x_{2i}	…	x_{2P}
3	x_{31}	x_{32}	x_{33}	…	x_{3i}	…	x_{3P}
j	…..	…..	…..	….	x_{ji}	….	…..
n	x_{n1}	x_{n2}	x_{n3}	…	x_{ni}	…	x_{nP}

7−2 حساب مصفوفة التباعد D (Dissimilarity) للمتحولات الكمية

تتألف مصفوفة التباعد من قيم المسافات بين أزواج مفردات العينة وتحسب من قيم المتحولات المستخدمة في عملية العنقدة . لذلك نفترض أنه لدينا n مفردة هي: 1 2 3 … j…n ونريد تصنيفها ضمن عناقيد حسب قيم المتحولات المؤثرة عليها، والتي سنرمز لها بـ $X_1\ X_2\ X_3 \ldots . X_P$ ، وسنستخدم قيم هذه المتحولات لحساب عناصر مصفوفة التباعد أو مصفوفة المسافات، والتي سنرمز لها بـ D ونكتبها كمايلي:

$$D = \begin{array}{c c} \text{المفردات} & \begin{array}{ccccc} 1 & 2 & 3 & k & n \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \ldots \\ j \\ \vdots \\ n \end{array} & \begin{bmatrix} d_{11} & d_{12} & d_{13} & \ldots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \ldots & d_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & d_{jk} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ d_{n1} & d_{n2} & d_{n3} & \ldots & d_{nn} \end{bmatrix} \end{array} \qquad (1-7)$$

حيث أن: العنصر d_{jk} هو المسافة بين المفردتين (k , j). وهنا نلاحظ أن هذه المصفوفة هي مصفوفة مربعة ومتناظرة (لأن المسافات: $(d_{jk} = d_{kj}$ وان عناصر قطرها الرئيسي تساوي أصفاراً (لأن المسافة بين النقطة *j* ونفسها $(d_{jj} = 0$. ولهذا فإن معظم المراجع العلمية تكتبها اختصاراً على شكل مصفوفة مثلثية عليا كما يلي:

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \ldots & d_{1n} \\ & 0 & d_{23} & \ldots & d_{2n} \\ & & 0 & d_{jk} & d_{jn} \\ & & & 0 & \\ & & & & 0 \end{bmatrix} \qquad (2-7)$$

ولكن حساب هذه المسافات يتعلق بطبيعة المتحولات $X_1\ X_2\ X_3 \ldots . X_P$. فهل هي مستمرة أم منقطعة؟ كما أنها تتعلق بطبيعة المسافة المراد حسابها. فهل هي المسافة النظرية أم المسافة الفعلية ؟

لذلك فإننا نعرف هذه المسافات حسب هذه الحالات، فعندما تكون المتحولات X كمية (عددية)، فإن المشكلة الوحيدة التي تعترض حساب المسافة بين أي مفردتين هي واحدات القياس لتلك المتحولات .
فإذا كانت واحدات القياس لجميع المتحولات X موحدة (كدخل الأسرة أو نفقاتها على الغذاء أو الكساء أو السكن أو النقل والاتصالات ...الخ)، فإننا نقوم بحساب المسافة بين أي مفردتين حسب الصيغ اللاحقة .
أما إذا كانت واحدات القياس لـ X مختلفة (كالدخل وعدد أفراد الأسرة ومساحة السكن ...الخ) فإننا نقوم بتحويل هذه المتحولات إلى متحولات معيارية Z وفق العلاقة التالية:

$$Z_i = \frac{X_i - \bar{X}_i}{\sigma_i} \qquad (3-7)$$

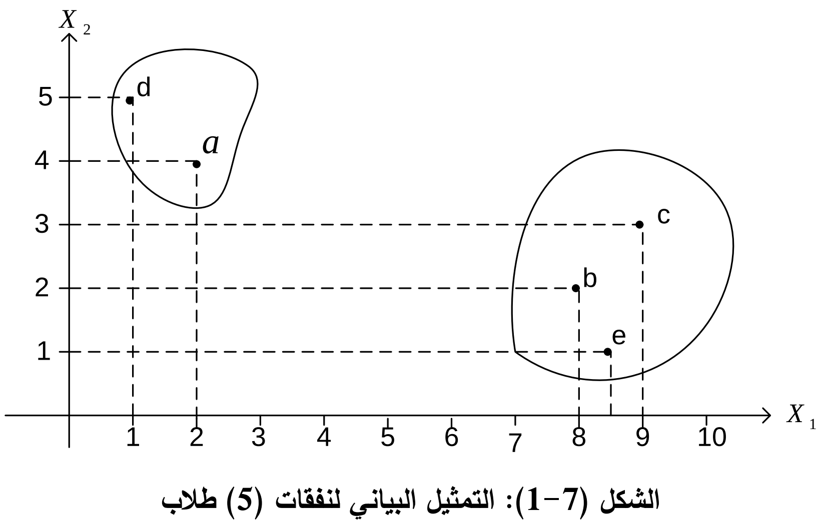
فنحصل على المتحولات المعيارية: $Z_1\ Z_2\ Z_3 \ldots . Z_P$ ، التي تتميز بأن متوسط كل منها يساوي الصفر $(\bar{Z}_i = 0)$ وأن تباينه يساوي الواحد $\left(\sigma_i^2 = 1\right)$. ثم نتعامل معها كالعادة .

مثال (7−1): لنفترض أننا نريد تصنيف (5) طلاب ضمن عناقيد حسب متحولين فقط هما: X_1 نفقات الطالب على الغذاء، وX_2 نفقات الطالب على الاتصالات، وذلك حسب البيانات المبينة في الجدول التالي:

جدول (7−2): بيانات نفقات (5) طلاب (الف ليرة والأرقام فرضية)

رمز الطالب أو رقمه	X_1 = نفقاته على الغذاء	X_2 = نفقاته على الاتصالات
1 = a	2	4
2 = b	8	2
3 = c	9	3
4 = d	1	5
5 = e	8.5	1

والآن نقوم برسم مواقع هؤلاء الطلاب على المستوى $X_1 0 X_2$ حسب إحداثيات كل منها $(X_1\ X_2)$ فنحصل على الشكل التالي:



الشكل (7−1): التمثيل البياني لنفقات (5) طلاب

ومن الشكل (7−1) نلاحظ أن مواقع هؤلاء الطلاب الخمسة تشكل حسب قيم نفقاتهم على الغذاء والاتصالات مجموعتين منفصلتين هما: $G_1(a\ ,d)$ و $G_2(e\ ,b\ ,c)$.

والسؤال الآن هل يمكن تصنيف هؤلاء الطلاب حسب X_1 و X_2، ضمن عناقيد متشابهة داخلياً ومتباينة خارجياً ؟

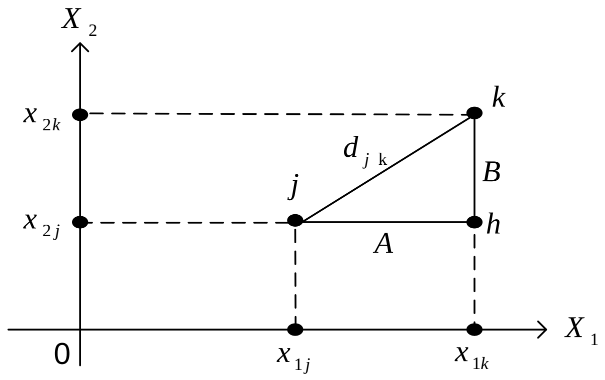
وحتى نجيب على هذا السؤال علينا أن نتبع منهجية العنقدة، والتي تقتضي حساب المسافات المختلفة بين مواقع هؤلاء الطلاب للحصول على مصفوفة التباعد D، ثم تجميع الطلاب حسب الأقرب فالأقرب . ورياضياً تعتبر النقطة j أقرب إلى النقطة k من أية نقطة أخرى ℓ إذا كانت المسافة بينهما d_{jk} تحقق العلاقة:

$$d_{jk} < d_{j\ell} \qquad \ell = 1\ 2 \dots n \qquad \ell \neq k \qquad (4-7)$$

وإن أهم مقياس للمسافات بين النقطتين j و k في المستوى هو المسافة الاقليدية، والتي تتمثل في الضلع الوتر في المثلث القائم $(j\ \ h\ \ k)$، لذلك نحسب مربع ذلك الوتر حسب نظرية (فيثاغورث) وبدلالة إحداثيات النقطتين $k(x_{1k}\ x_{2k})$ و $j(x_{1j}\ x_{2j})$ كمايلي:

$$d_{jk}^2 = A^2 + B^2 = \left(x_{1j} - x_{1k}\right)^2 + \left(x_{2j} - x_{2k}\right)^2 \qquad (5-7)$$

والشكل التالي يوضح ذلك:



الشكل (7−2): تمثيل المسافة الاقليدية d_{jk}

ومن العلاقة (7−5) نجد أن المسافة الاقليدية d_{jk} بين النقطتين j و k تساوي :

$$d_{jk} = +\sqrt{\left(x_{1j} - x_{1k}\right)^2 + \left(x_{2j} - x_{2k}\right)^2} \qquad (6-7)$$

ويمكن تعميم هذه العلاقة على عدة متحولات $X_1\ X_2\ X_3 \dots .X_P$ وفي الفضاء R^P، فنجد أن المسافة الاقليدية بين أي نقطتين j و k من الفضاء R^P تعطى بالعلاقة :

$$d_{jk} = \sqrt{\left(x_{1j} - x_{1k}\right)^2 + \left(x_{2j} - x_{2k}\right)^2 + \left(x_{3j} - x_{3k}\right)^2 + \dots + \left(x_{Pj} - x_{Pk}\right)^2} \qquad (7-7)$$

وهي أقصر مسافة ممكنة بين k و j لذلك تسمى بالمسافة النظرية .

ولكن الحياة العملية لا تعتمد كثيراً على هذه المسافات، فمثلاً لا يمكننا الذهاب من شارع لآخر دون الالتفاف حول بعض المباني التي بينهما، وبناءً على ذلك تم استنباط مقياس آخر للمسافة يسمى مسافة المقاطع (City Block Distance) ويعرف في المستوى بالعلاقة التالية (انظر الشكل 7−2) :

$$d_{jk}^{b} = |A| + |B| = \left|\left(x_{1j} - x_{1k}\right)\right| + \left|\left(x_{2j} - x_{2k}\right)\right| \qquad (8-7)$$

ويمكن تعميم (7−8) على p متحولاً في الفضاء R^{P} بالعلاقة التالية:

$$d_{jk}^{b} = \sum_{i=1}^{P} \left|\left(x_{ij} - x_{ik}\right)\right| \qquad (9-7)$$

وأخيراً نشير إلى أن حساب المسافات من العلاقات (7−6) (7−7) (7−8) (7−9) السابقة، يشترط أن تكون المتحولات $X_1\ X_2\ X_3 \ldots .X_P$ متحولات معيارية، أو تكون ذات واحدات قياس موحدة .

ونورد فيما يلي جدولاً بأسماء وتعاريف أهم المقاييس المستخدمة لحساب المسافات للمتحولات العددية الموحدة أو المعيارية . وذلك حسب طبيعة وشروط كل مسألة أو كل قضية بحثية وفي الفضاء R^{P} .

جدول (7−3): مقاييس حساب المسافات بين النقطتين (j وk) في R^{P} [من Webb P. 420] :

ملاحظات	الصيغة الرياضية للمقياس	اسم المقياس
جذر مجموع مربعات الفروقات	$d_e = \left[\sum_{i=1}^{P} \left(x_{ij} - x_{ik}\right)^2\right]^{\frac{1}{2}}$	المسافة الاقليدية Euclidean
مجموع القيم المطلقة للفروقات	$d_{cb} = \sum_{i=1}^{P} \left\|x_{ij} - x_{ik}\right\|$	مسافة المقاطع City Block
أكبر القيم المطلقة للفروقات	$d_{ch} = Max\ \left\|x_{ij} - X_{ik}\right\|$	مسافة (تشيبتشيف) Chebychev
الجذر الـ (m) لمجموع الفروقات من المرتبة m' – تعميم الاقليدية	$d_m = \left[\sum_{i=1}^{P} \left(x_{ij} - x_{ik}\right)^m\right]^{\frac{1}{m}}$	مسافة (مينكوفسكي) Minkowski
متناظرة: S^{-1}	$d_q = \sum \left(X_{ij} - X_{ik}\right)^2 S^{-1}\left(X_j' - X_k\right)$	المسافة التربيعية Q_s Mahalanobis

مثال 7−2: لنأخذ بيانات المثال (7−1) ونقوم بحساب عناصر مصفوفة التباعد D بواسطة استخدام المسافة الاقليدية المعرفة بالعلاقة (7−6) .

ولنفترض أولاً أن كل من الطلاب الخمسة يشكل لوحده عنقوداً خاصاً، ثم نقوم بحساب المسافات الاقليدية بين كل زوج منهم (j , k) من العلاقة:

$$d_{jk} = \sqrt{\left(x_{1j} - x_{1k}\right)^2 + \left(x_{2j} - x_{2k}\right)^2}$$

وهكذا نجد أن المسافة بين الطالبين (2 ،1) تساوي:

$$d_{12} = \sqrt{(2-8)^2 + (4-2)^2} = 6.325$$

وكذلك نجد أن المسافة بين الطالبين (3 ،1) تساوي:

$$d_{13} = \sqrt{(2-9)^2 + (4-3)^2} = 7.071$$

وبمتابعة حساب هذه المسافات للأزواج المختلفة الأخرى، نحصل على مصفوفة التباعد D لهؤلاء الطلاب التالية:

$$D = \begin{array}{c} \text{الطلاب} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 6{,}325 & 7.071 & 1.414 & 7.159 \\ & 0 & 1.414 & 7.616 & 1.116 \\ & & 0 & 8.246 & 2.062 \\ & & & 0 & 8.500 \\ & & & & 0 \end{bmatrix} \end{array}$$

ومنها نلاحظ أن أصغر عناصر هذه المصفوفة هو (1.116) وهو يقابل الطالبين 2 و5، لأنهما الأكثر تشابهاً في النفقات، لذلك يمكن أن ننشأ منهما العنقود الأول كما سنرى لاحقاً .

3-7 حساب مصفوفة التشابه أو التقارب (Similarity) في حالة المتحولات النوعية (المرتبة والاسمية) .

عندما تكون المتحولات X نوعية– مرتبة أو اسمية– فلا يكون لها قيم عددية، بل يكون لها حالات مرتبة أو فئات مختلفة (كمستوى التعليم أو الحالة الاجتماعية). لذلك لا يمكننا حساب المسافات بين المفردات من العلاقات الكمية المذكورة في الجدول (7-3) .

وفي مثل هذه الحالات نلجأ إلى استخدام التكرارات المطلقة المقابلة لحالات أو فئات تلك المتحولات . لأن الحالات المتشابهة تكون تكرارتها متقاربة .

فإذا كان المتحول X ثنائياً – أي يتألف من حالتين فقط (نجاح وفشل)– فإننا نفترض أنه يأخذ القيمة (1) إذا تحققت حالة النجاح، ويأخذ القيمة (0) إذا تحققت حالة الفشل .

أما إذا كان المتحول X أكثر من حالتين، فإننا نعتبر كل حالة مستقلة عن الحالات الأخرى، ونعرف عليها متحولات ثنائية جديدة ($X'_1\ X'_2\ ...\ X'_s$)، ثم نفترض أن كل متحول جديد X'_t يأخذ القيمة (1) عندما تتحقق الحالة المقابلة له، ويأخذ القيمة (0) عندما لا تتحقق تلك الحالة .

أي أنه لحساب الاختلافات في حالة المتحولات النوعية، يجب أن تكون (أو أن نجعل) تلك المتحولات متحولات ثنائية وتأخذ إحدى القيمتين قيمة (1) عند التحقق وقيمة (0) عند عدم التحقق .

وعندها فإن بيانات العينة المأخوذة من n مفردة والمعتمدة على p متحولاً ثنائياً هي $X_1\ X_2\ X_3\ ....X_P$، تنظم وتوضع في جدول كالتالي:

جدول (7–4): بيانات المتحولات النوعية (فرضية)

المتحولات / المفردات	X_1	X_2	X_3	....	X_i	....	X_P
1	1	0	1	....	0	....	1
2	0	1	1	....	0	....	0
3	1	0	0	....	1	....	1
I	....	....	....	....	X_{ik}	....	....
N	1	1	0	....	1	....	0

ونلاحظ من الجدول (7–4) أن كل مفردة j من العينة تعطينا مقابل كل متحول من المتحولات $X_1\ X_2\ X_3 \ldots X_i \ldots X_P$ إحدى القيمتين (1) أو (0) [وهي القيم التي في السطر المقابل للمفردة j] .

كما نلاحظ من جهة أخرى أن كل متحول X_i يأخذ إحدى القيمتين (1) و (0) مقابل مفردات العينة [وهي القيم التي في العمود المقابل لـ X_i] .

ولدراسة التقارب بين أي مفردتين (j و k) نقوم بإيجاد جدول التوافق بينهما، ونحسب التكرارات المتقابلة لتوافقهما ولتعارضهما في السطرين (j و k) من الجدول (7–4) فنحصل على الجدول التالي :

جدول (7–5): جدول التوافق للمفردتين (j و k) فقط

البيان		قيم المفردة K: 1	قيم المفردة K: 0	المجموع
قيم المفردة j	1	$a\ (1,1)$	$b\ (1,0)$	$a + b$
	0	$c\ (0,1)$	$d\ (0,0)$	$c + d$
المجموع		a + c	$b + d$	P = a + b + c + d

إن الرموز في الجدول (7–5) تعني أن:

a : هو عدد تكرارات الأزواج (1–1) ، b: هو عدد تكرارات الأزواج (0 و 1) .

c : هو عدد تكرارات الأزواج (1–0) ، d: هو عدد تكرارات الأزواج (0 و 0) .

ولتقدير عناصر مصفوفة التقارب S بين جميع مفردات العينة. يجب علينا أن نقوم بإيجاد جميع جداول التوافق لجميع الأزواج المختلفة، التي يمكن تشكيلها من العينة ذات الحجم n، والتي يبلغ عددها $C_n^2 = \frac{n(n-1)}{2}$ زوجاً مختلفاً . وهناك عدة مقاييس للتقارب بين هذه الأزواج تحسب من القاعدة التالية:

$$S_{jk} = \frac{\text{عدد الأزواج المتشابهة}}{\text{عدد المتحولات المؤثرة } P} = \frac{a+d}{P} \qquad (10-7)$$

وبعد إيجاد تلك الجداول التوافقية لكل مفردتين (j و k) نقوم بحساب عناصر مصفوفة التقارب S_{jk} في حالة المتحولات الثنائية باستخدام أحد المقاييس المعرفة في الجدول التالي:

جدول (7–6): مقاييس التقارب للمتحولات الثنائية (من (Jonhson,Wichern. P. 549)

ملاحظات	المقاييس الرياضية لـ S_{jk}	رقم المقياس
نسبة تكرارات الأزواج المتشابهة (1 و1) و(0 و0) بأوزان متساوية	$\frac{a+d}{P}$	1
بمضاعفة أوزان الأزواج المتشابهة (1 و1) و(0 و0)	$\frac{2(a+d)}{2(a+d)+b+c}$	2
بمضاعفة أوزان الأزواج غير المتشابهة (0 و1) و(1 و0)	$\frac{a+d}{a+d+2(b+c)}$	3
بحذف تكرار الأزواج (0 و0) من البسط: Russel + Rao	$\frac{a}{P}$	4
بحذف تكرار الأزواج (0 و0) من البسط والمقام Jaccard (لأنها خارج الموضوع)	$\frac{a}{a+b+c}$	5
بحذف تكرار الأزواج (0 و0) من البسط والمقام ومضاعفة تكرارات الأزواج (1 و1) czekanowcki	$\frac{2a}{2a+b+c}$	6
بحذف تكرارات الأزواج (0 و0) من البسط والمقام، ومضاعفة تكرار الأزواج غير المتشابهة	$\frac{a}{a+2(b+c)}$	7
نسبة الأزواج المتشابهة من الشكل (1 و1) إلى الأزواج غير المتشابهة (1 و0) و(0 و1) مع حذف الأزواج (0 و0)	$\frac{a}{b+c}$	8

ولكن استخدام هذه المقاييس يتعلق بطبيعة المسألة المطروحة وبالهدف الذي يقصده الباحث، علماً بأن هذه المقاييس تعطينا قيماً مختلفة للعناصر غير القطرية S_{jk} في مصفوفة التقارب S، وهذا قد يؤدي بنا إلى الحصول على نتائج مختلفة .

وباستخدام أحد هذه المقاييس نحصل على عناصر مصفوفة التقارب S ونكتبها كما يلي :

$$S_{P*P} = \begin{array}{c} \text{المفردات} \\ 1 \\ 2 \\ 3 \\ \cdots \\ n \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & \ldots & n \end{array} \\ \begin{bmatrix} S_{11} & S_{12} & S_{13} & \ldots & S_{1n} \\ S_{21} & S_{22} & S_{23} & \ldots & S_{2n} \\ S_{31} & S_{32} & S_{33} & \ldots & S_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ S_{n1} & S_{n2} & S_{n3} & \ldots & S_{nn} \end{bmatrix} \end{array} \qquad (11-7)$$

حيث أن العناصر S_{jk} تحسب من أحد المقاييس المذكورة في الجدول (7–6) السابق للمفردتين j وk.

ومن خواص هذه المصفوفة إنها مصفوفة مربعة من المرتبة $n * n$، ومتناظرة لأن المقاييس المستخدمة لحسابها متناظرة (أي أن $S_{jk} = S_{kj}$) كما أن عناصر القطر الرئيسي فيها $(S_{jj} = 1)$ لأن جدول التوافق للمفردة *j* مع نفسها يكون من الشكل التالي:

j	1	0	1	0	1	0
j	1	0	1	0	1	0

$\Rightarrow$

j	1	0
1	3	0
0	0	3

ومنه نجد أن قيمة العنصر S_{jj} يساوي حسب المقياس (1) ما يلي:

$$S_{jj}=\frac{a+d}{a+b+c+d}=\frac{3+3}{3+0+0+3}=1$$

وهكذا نجد أن معظم المقاييس الأخرى تعطينا نفس النتيجة (ماعدا المقياسين (4) و(8)).

لذلك يمكننا كتابة مصفوفة التقارب S على شكل مصفوفة مثلثية سفلى (لتمييزها عن مصفوفة التباعد D) كما يلي:

المفردات

$$S=\begin{array}{c} 1\\2\\3\\j\\n \end{array}\begin{bmatrix} 1 & & & & \\ S_{21} & 1 & & & \\ S_{31} & S_{32} & 1 & & \\ & & & & \\ S_{n1} & S_{n2} & S_{n3} & \dots & 1 \end{bmatrix} \qquad (12-7)$$

ملاحظة هامة: إذا كانت المتحولات $X_1\ X_2\ \dots X_i\dots X_P$ مختلفة (كمية ونوعية) فإننا نقوم بتحويل المتحولات الكمية فيها (مثل X_1 وX_2) إلى متحولات ثنائية، وذلك بتقسيم مجال تحول كل منها إلى مجالين فقط . وننشأ منها متحولات ثنائية جديدة معرفة على هذين المجالين كما يلي:

$$X_1'=\begin{cases}1: & X_1\le x_{01}\\ 0: & X_1>x_{01}\end{cases} \qquad X_2'=\begin{cases}1: & X_2\le x_{02}\\ 0: & X_2>x_{02}\end{cases} \qquad (13-7)$$

حيث أن x_{01} وx_{02} هما النقطتان الفاصلتان بين هذين المجالين .

أما المتحولات النوعية المتبقية فنقوم بتحويلها أيضاً إلى متحولات ثنائية كما وضحنا ذلك أعلاه . وبعد إجراء هذه التحويلات نقوم بجمع البيانات ووضعها في جدول كالجدول (7–4) . ثم نقوم بإيجاد جداول التوافق ككل زوج من المفردات (j , k)، فنحصل على $\frac{n(n-1)}{2}$ جدولاً مشابهاً للجدول (7–5) السابق .

وبعد كل ذلك نقوم بحساب عناصر مصفوفة التقارب S_{jk} وذلك باستخدام أحد المقاييس المعرفة في الجدول (7–6) السابق .

مثال 7–3: لنفترض أننا نريد تصنيف (5) طلاب ضمن عناقيد متشابهة وذلك حسب (6) متحولات مختلطة هي: الوزن X_1 ، الطول X_2 ، لون العينين X_3 ، لون الشعر X_4 ، اليد المستخدمة X_5 ، والجنس X_6 ، وبعد استجوابهم حصلنا منهم على البيانات التالية:

جدول (7−7): البيانات الأولية لـ (5) طلاب حسب (6) متحولات (Johnson,Wichern. P. 550)

المتحولات / رقم الطالب	الوزن كغ X_1	الطول سم X_2	لون العينين X_3	لون الشعر X_4	اليد المستخدمة X_5	الجنس X_6
1	68	140	أخضر	أشقر	اليمنى	أنثى
2	73	185	بني	أسود	اليمنى	ذكر
3	67	165	أزرق	أشقر	اليمنى	ذكر
4	64	120	بني	أسود	اليمنى	أنثى
5	76	210	بني	أسود	اليسرى	ذكر

للاستفادة من هذه البيانات نقوم بتحويل كل من هذه المتحولات إلى متحولات ثنائية، ونعرف منها المتحولات الثنائية التالية: $X_1', X_2', X_3', X_4', X_5'\ , X_6'$ كما يلي:

$$X_1' = \begin{cases} 1 : X_1 \geq 72 \\ 0 : X_1 < 72 \end{cases} \qquad X_4' = \begin{cases} 1 : X_4 = \text{أشقر} \\ 0 : X_4 = \text{غير ذلك} \end{cases}$$

$$X_2' = \begin{cases} 1 : X_2 \geq 150 \\ 0 : X_2 < 150 \end{cases} \qquad X_5' = \begin{cases} 1 : X_5 = \text{اليمنى} \\ 0 : X_5 = \text{اليسرى} \end{cases}$$

$$X_3' = \begin{cases} 1 : X_3 = \text{بني} \\ 0 : X_3 = \text{غير ذلك} \end{cases} \qquad X_6' = \begin{cases} 1 : X_6 = \text{أنثى} \\ 0 : X_6 = \text{ذكر} \end{cases}$$

وبعد تفريغ قيم هذه المتحولات في جدولٍ كالجدول (7−4) السابق، نحصل على جدول جديد للمتحولات الثنائية يأخذ الشكل التالي :

جدول (7−8) نتائج استجواب (5) طلاب بدلالة المتحولات الثنائية:

المتحولات / رقم الطالب	X_1'	X_2'	X_3'	X_4'	X_5'	X_6'
1	0	0	0	1	1	1
2	1	1	1	0	1	0
3	0	1	0	1	1	0
4	0	0	1	0	1	1
5	1	1	1	0	0	0

ولإنشاء مصفوفة التقارب S لهؤلاء الطلاب، علينا أولاً أن نقوم بإيجاد جداول التوافق لكل زوج منهم (وعددها $C_5^2 = 10$ أزواج) . و بناءً على بيانات الجدول (7−8) نجد مثلاً أن جدول التوافق للطالبين (1) و(2) يأخذ الشكل التالي:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">قيم الطالب (2)</td><td rowspan="2">المجموع</td></tr>
<tr><td>1</td><td>0</td></tr>
<tr><td rowspan="2">قيم الطالب (1)</td><td>1</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>0</td><td>3</td><td>0</td><td>3</td></tr>
<tr><td colspan="2">المجموع</td><td>4</td><td>2</td><td>6</td></tr>
</table>

ومنه نحسب مقياس التقارب بينهما، وذلك من خلال المقياس الأول المعرف في الجدول (7−6) بالعلاقة التالية:

$$S_{12} = \frac{a+d}{P} = \frac{1+0}{6} = \frac{1}{6}$$

وكذلك نجد أن جدول التوافق للطالبين (1) و(3) يأخذ الشكل التالي:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">قيم الطالب (3)</td><td rowspan="2">المجموع</td></tr>
<tr><td>1</td><td>0</td></tr>
<tr><td rowspan="2">قيم الطالب (1)</td><td>1</td><td>2</td><td>1</td><td>3</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td colspan="2">المجموع</td><td>3</td><td>3</td><td>6</td></tr>
</table>

ومنه نحسب عنصر التقارب بين الطالبين (1) و(3) وذلك من خلال نفس المقياس الأول من الجدول (7−6) المعرف بالعلاقة :

$$S_{13} = \frac{a+d}{P} = \frac{2+2}{6} = \frac{4}{6}$$

وبمتابعة حساب بقية عناصر مصفوفة التقارب S_{jk} بين هؤلاء الطلاب وباستخدام نفس المقياس، نحصل على مصفوفة متناظرة من المرتبة 5 * 5 وتأخذ الشكل التالي:

$$S = \begin{array}{c} \text{الطلاب} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & & & & \\ \frac{1}{6} & 1 & & & \\ \frac{4}{6} & \frac{3}{6} & 1 & & \\ \frac{4}{6} & \frac{3}{6} & \frac{2}{6} & 1 & \\ 0 & \frac{5}{6} & \frac{2}{6} & \frac{2}{6} & 1 \end{bmatrix} \end{array}$$

ومن خلال هذه المصفوفة نلاحظ مباشرة أن أكبر عنصر في المصفوفة S هو العنصر التكراري $\frac{5}{6}$، الذي يقابل الطالبين (2) و(5)، وهذا يعني أن هذين الطالبين أقرب إلى بعضهما (حسب المتحولات المستخدمة) من أي طالبين آخرين، لذلك يمكننا أن نشكل منهما مجموعة أولى تمثل العنقود الأول .

كما نلاحظ أن أبعد طالبين عن بعضهما هما الطالبين (1) و(5) لأنهما يقابلان أصغر عنصر في المصفوفة هو $(S_{15} = 0)$، وهناك أزواج تقع بين هاتين الحالتين .

وإذا أردنا تقسيم الطلاب إلى مجموعتين جزئيتين متجانستين نسبياً بناءً على بيانات مصفوفة التقارب، يمكننا أن نشكل مجموعتين جزئيتين مؤلفتين من هؤلاء الطلاب : $G_1 = (2\,,5)$ و $G_2 = (1\,,3\,,4)$.

ملاحظة هامة: إن عناصر مصفوفة التقارب S_{jk} ترتبط مع عناصر مصفوفة التباعد d_{jk} من خلال العلاقة التالية [Johnson,Wichern, P. 551]:

$$\tilde{S}_{jk} = \frac{1}{1+d_{jk}} \qquad : \qquad 0 < \tilde{S}_{jk} \le 1 \qquad\qquad (14-7)$$

ولكن حساب d_{jk} من S_{jk} يحتاج إلى توفر شرط (Gower) وهو أن تكون المصفوفة S غير سالبة التحديد $(X' * S * X \ge 0)$. وعندما يكون التشابه أعظمياً ومميزاً بـ $(\bar{S}_{ii} = 1)$، فإن المسافة d_{jk} ترتبط مع S_{jk} بالعلاقة التالية :

$$d_{jk} = \sqrt{2(1-\bar{S}_{jk})} \qquad\qquad (15-7)$$

7-4 أسلوب (جور Gower) لمعالجة بيانات المتحولات النوعية أو المختلطة دون تحويلها إلى متحولات ثنائية:

لنفترض إننا نريد تصنيف عينة من الموظفين إلى عناقيد متشابهة حسب المتحولات النوعية والكمية المرفقة حسب حالاتها المختلفة وأرقامها الرمزية التالية:

X_1 – المستوى الاقتصادي: منخفض= (1)، متوسط= (2)، مرتفع= (3) .

X_2 – الحالة الاجتماعية: أعزب= (1)، متزوج= (2)، مطلق= (3)، أرمل= (4) .

X_3 – لون العينين: سوداء= (1)، خضراء= (2)، زرقاء= (3) .

X_4 – العمر (بالسنوات): أقل من 30 =(1)، من(30–50) = (2)، أكبر من 50 = (3) .

X_5 – الجنس: ذكر= (1)، أنثى= (2) .

نعالج حالات هذه المتحولات ونستبدلها بأرقامها وندخلها إلى الجدول الأساسي لبيانات العينة، ولنفترض أن بيانات مفردتين منه مثل (j وk) على هذه الحالات كانت كما يلي:

المتحولات / رمز الموظف	X_1	X_2	X_3	X_4	X_5	
J	2	2	3	1	2	
K	3	2	3	1	1	
$S_i(jk)$	0	1	1	1	0	$\bar{S}_{jk}$

ثم نعرف على إجابات المفردتين (j وk) متحول جديد $S_{i(jk)}$ ونضعه في السطر الثالث، ونجعله يأخذ القيمة (1) إذا كانت إجابة المفردة j متوافقة مع إجابة المفردة k (مساوية لها)، ويأخذ القيمة (0) إذا كانت إجابة j مختلفة عن إجابة k . فنحصل على السطر الأخير في الجدول السابق .

ثم نقوم بحساب عناصر مصفوفة التقارب S_{jk} من خلال حساب المتوسط الحسابي لقيم المتحول $S_{i(jk)}$ ونرمز له بالرمز التالي:

$$\bar{S}_{jk} = \frac{\sum_{i=1}^{P} S_{i(jk)}}{P} \qquad j,k = 1\ 2\ 3 \ldots n \qquad (16-7)$$

ولكن (جور Gower) اقترح تثقيل هذا المتوسط بأوزان مناسبة مع أهمية المتحولات $X_1\ X_2\ X_3 \ldots X_P$، ورمز لها بالرموز المقابلة لها: $W_1\ W_2\ W_3 \ldots W_P$، ثم حساب تقدير عنصر مصفوفة التقارب S_{jk} من العلاقة التالية:

$$\bar{S}_{jk} = \frac{\sum_{i=1}^{P} W_i S_{i(jk)}}{\sum_{i=1}^{P} W_i} \qquad (17-7)$$

وهكذا نجد أن هذا الأسلوب يعطينا قيماً تقديرية لعناصر مصفوفة التقارب $\tilde{S}_{jk}$ ، وتأخذ قيمها المختلفة بين (0) و(1). وإنها تأخذ القيمة (0) عندما تكون جميع إجابات j مختلفة عن إجابات k (لأنه عندها تكون جميع $S_{i(jk)} = 0$)، وبالتالي فإن متوسطها $\bar{S}_{jk} = 0$)، وتأخذ القيمة (1) عندما تكون جميع إجابات j متوافقة مع إجابات k (لأنه عندها تكون جميع $S_{i(jk)} = 1$، وبالتالي يكون متوسطاتها $\bar{S}_{jk} = 1$) .

وإن قيم $\bar{S}_{jk}$ الأخرى تأخذ قيماً عددية تقع في المجال [0 , 1] أي أن يكون لدينا: $0 \le \bar{S}_{jk} \le 1$.

أي أن أسلوب (*Gawer*) يحافظ على أن تأخذ عناصر المصفوفة *S* قيماً كسرية في المجال [0 , 1] .

7-5 تجميع المتحولات X:

إن التحليل العنقودي يستخدم- بين الحين والآخر- لتجميع المتحولات X بالاعتماد على المشاهدات. وهذه الحالة مطلوبة في تصميم الاستبيانات، حيث أن مسودة الاستبيان غالباً ما تتضمن بعض الأسئلة التي تحرص على تأمين معدل جيد للإجابات، وعندما يتم اختبار الاستبيان على عدد قليل من المستجوبين، يمكننا مباشرة أن نلاحظ أن الاجابات على مجموعات الأسئلة المتشابهة تكون مرتبطة بشدة، ولكن أهم تطبيق للتحليل العنقودي يمكن أن يتم على مجموعات أخرى من الأسئلة هي الأسئلة المتباعدة أو الغامضة حيث تكون أجوبتها غير مرتبطة، وبذلك تكون هذه الأسئلة نقطة ضعف في الاستبيان . وبعد تحليل الإجابات نقوم بتجميع أسئلة الاستبيان ضمن مجموعات متشابهة وعناقيد متباينة، ثم نقوم بإعادة صياغة الاستبيان . فنقوم بدمج بعض الأسئلة المتشابهة في سؤال واحد . ونعيد صياغة بعضها الآخر ونوضح معانيه ودلالاته .

مثال (7−4): لتوضيح ذلك نأخذ إجابات (5) أشخاص على (3) أسئلة في مسودة أحد الاستبيانات والتي كانت كما يلي:

جدول (7−9): بيانات المثال فرضية

الأسئلة / المستجوبين	Q_1	Q_2	Q_3
a	10	5.0	3.00
b	30	7.5	3.10
c	20	6.0	2.90
d	40	8.0	2.95
e	50	9.0	3.80

إن أفضل مقياس لتقارب الأسئلة $Q_1\ ,Q_2,Q_3$ هو معاملات الارتباط الخطي بين كل زوج منها $Q_i\ ,Q_j$ مأخوذة بالقيمة المطلقة، وبعد حسابها نحصل على مصفوفة التقارب التالية:

$$S = \begin{array}{c c} \text{الأسئلة} & \begin{array}{ccc} Q_1 & Q_2 & Q_3 \end{array} \\ \begin{array}{c} Q_1 \\ Q_2 \\ Q_3 \end{array} & \begin{bmatrix} 1 & |r_{12}| & |r_{13}| \\ |r_{12}| & 1 & |r_{23}| \\ |r_{13}| & |r_{23}| & 1 \end{bmatrix} \end{array} = \begin{bmatrix} 1 & & \\ 0.984 & 1 & \\ 0.076 & 0.230 & 1 \end{bmatrix} \qquad (18-7)$$

وإن مقياس التباعد (حسب المسافات) بين أي سؤالين Q_i و Q_j يمكن أن يحسب من التحويل التالي :

$$d_{ij} = 1 - |r_{ij}|$$

وهكذا نحصل على المصفوفة الأولى للمسافات بين كل زوج من الأسئلة، والتي تأخذ الشكل التالي :

$$D = \begin{array}{c c} \text{الأسئلة} & \begin{array}{ccc} Q_1 & Q_2 & Q_3 \end{array} \\ \begin{array}{c} Q_1 \\ Q_2 \\ Q_3 \end{array} & \begin{bmatrix} 0 & 0.016 & 0.924 \\ & 0 & 0.770 \\ & & 0 \end{bmatrix} \end{array} \qquad (19-7)$$

وبذلك يمكننا أن نطبق على هذه المصفوفة أية طريقة من طرائق العنقدة وباستخدام الإجراءات المعتادة . فمثلاً نجد أن أصغر قيمة في D هي: $d_{12} = 0.016$ وهذا يعني أن السؤالين Q_1 و Q_2 يشكلان عنقوداً واحداً، لأنهما متشابهان جداً، لذلك يمكن أن نختار أحدهما أو أن نصيغ منهما سؤالاً ثالثاً يعبر عنهما معاً. أما إذا كانت الأسئلة ثنائية (محولة من متحولات نوعية أو كمية) فإنه يمكننا تصنيف بياناتها المؤلفة من (1) و(0) ضمن جداول التوافق . وذلك باستبدال المفردات بالأسئلة Q . وبذلك نحصل مقابل كل زوج من الأسئلة $(Q_j\ ,Q_k)$ على جدول للتوافق لهما يأخذ الشكل التالي:

جدول (7−10): جدول التوافق للسؤالين Q_j و Q_k

<table>
<tr><th colspan="2" rowspan="2">البيان</th><th colspan="3">السؤال Q_k</th></tr>
<tr><th>1</th><th>0</th><th>المجموع</th></tr>
<tr><td rowspan="2">السؤال Q_j</td><td>1</td><td>a</td><td>b</td><td>$a + b$</td></tr>
<tr><td>0</td><td>c</td><td>d</td><td>$c + d$</td></tr>
<tr><td></td><td>المجموع</td><td>$a + c$</td><td>$b + d$</td><td>$n = a + b + c + d$</td></tr>
</table>

حيث أن: الرموز $d\ ,c\ ,b\ ,a$ هي عبارة عن أعداد التكرارات المطلقة المتقابلة للأزواج $(1,1)$ و$(0,1)$ و$(1,0)$ و$(0,0)$ على الترتيب، وإن أفضل مقياس للتقارب بين هذين السؤالين $\left(Q_j\ ,Q_k\right)$ هو معامل الاقتران أومعامل ارتباط المتحولات الثنائية المعرف بالعلاقة التالية:

$$r_{jk} = \frac{a * d - b * c}{(a + b)(c + d)(a + c)(b + d)} \qquad (20 - 7)$$

وبعد حساب هذه المعاملات لجميع الأزواج، نأخذ القيمة المطلقة لها، ونضعها في مصفوفة خاصة للتقارب بين الأسئلة المدروسة ونرمز لها بـ

$$S = \begin{bmatrix} 1 & 0 & 0 \\ |r_{12}| & 1 & 0 \\ |r_{13}| & |r_{23}| & 1 \end{bmatrix} \implies \qquad (21 - 7)$$

ومنها يمكننا أن نحصل على مصفوفة التباعد D وتطبيق الإجراءات اللازمة عليها .

علماً بأن معاملات الاقتران أو الارتباط r_{ij} المعرفة في (7−20) ترتبط مع المتحول χ^2 بواسطة العلاقة $\left(r = \frac{\chi^2}{n}\right)$، وتستخدم هذه الخاصة عند اختبار الاستقلال لأي متحولين نوعيين . ومن أجل قيمة ثابتة لـ n فإن التشابه الكبير (أو الارتباط القوي) يتوافق مع انعدام الاستقلال .

وأخيراً نشير إلى أن بيانات الجدول (7−10)، ومقاييس الاقتران (التقارب) المعرفة في (7−20) هي مشابهة تماماً للمقاييس المعرفة في الجدول (7−6)، والفرق الوحيد بينها هو استبدال عدد المتحولات p بعدد المشاهدات n . ومما سبق نستنتج أنه يوجد عدد كبير من مقاييس التقارب بين الموضوعات، ولكن المقاييس الأكثر تطبيقاً هي مقاييس المسافة المعرفة للمتحولات الكمية في الجدول (7−3) ومقاييس العلاقات النوعية المعرفة في الجدول (7−6)، والتي يتم استخدامها لعنقدة المفردات أو لتجميع المتحولات X على الترتيب، ولتنفيذ ذلك تستخدم خوارزميات حاسوبية خاصة .

7−6 : طرائق التحليل العنقودي الهرمي التجميعي:

لقد أشرنا سابقاً إلى أن التحليل العنقودي الهرمي يعمل على تصنيف مفردات العينة باتباع أحد الأسلوبين التاليين:

- **أسلوب التجميع (Agglomerative)**: وهو الذي يبدأ من المفردات، ويعتبر كل منها تشكل عنقوداً مستقلاً، ثم يقوم بدمج المفردات الأكثر تشابهاً ويشكل منها عناقيد متشابهة، ثم يقوم بدمج العناقيد المتشابهة في عناقيد أكبر منها، حتى يتم الحصول على عنقود واحد يتضمن جميع مفردات العينة.

- **أسلوب التجزئة أو التقسيم (Divisive):** وهو يبدأ من اعتبار جميع المفردات تشكل عنقوداً واحداً، ثم يتم تقسيمها إلى عناقيد متشابهة، ثم تقسيم تلك العناقيد إلى عناقيد أصغر منها، حتى نحصل على عناقيد مقابل كل مفردة من مفردات العينة أو نتوقف عن عملية التقسيم .

وإن نتائج هذين الأسلوبين تمثل بياناً في المستوى على شكل عناقيد متعالقة ضمن عنقود واحد. ويسمى هذا الشكل مخطط الأغصان (Dendogram) .

ويستخدم التحليل العنقودي الهرمي التجميعي عدة طرائق تسمى طرائق الربط وهي:

- طريقة الربط المنفرد (Single Linkage) .
- طريقة الربط التام (Complete Linkage) .
- طريقة الربط المتوسط (Average Linkage) .

وتوجد طريقتان أخريتان هما:

- طريقة مجموع المربعات، وتسمى طريقة (وارد) (Ward's Method) .
- طريقة المراكز (Centroid Method) .

علماً بأن كل طريقة من هذه الطرائق تعتمد عند دمج المفردات أو العناقيد (المؤلفة من مفردات) على معيار معين لقياس التشابه بين الجانبين وتتبع خوارزمية موحدة .

إن خوارزمية العنقدة الهرمية التجميعية لـ n مفردة (مشاهدات أو متحولات)، تتألف من الخطوات التالية:

1- تبدأ العنقدة الهرمية التجميعية من اعتبار كل مفردة تشكل عنقوداً خاصاً، ومن مصفوفة متناظرة للمسافات (أو للتشابه)، نرمز لها بـ $D = [d_{jk}]$.

2- نبحث في مصفوفة المسافات D عن أصغر عنصر فيها (أصغر مسافة) ومنه نحدد العنقودين الأكثر قرباً أو تشابهاً من بين العناقيد المدروسة، وذلك من أجل دمجهما وتشكيل عنقود جديد منهما.

3- ندمج العنقودين v , u في عنقود واحد . ونرمز له بـ (uv) . ثم نقوم بتحديث العناصر المتقاطعة في مصفوفة المسافات D كما يلي :

أ- نحذف العمودين والسطرين المقابلين للعنقودين v , u من المصفوفة .

ب- نضيف سطراً وعموداً جديدين. ونضعهما في مكانu أو v، أو في آخر أو أول المصفوفة D.

ج- نقوم بحساب شعاع عناصر العنقود الجديد (uv) من عناصر العنقودين u وv، ونرمز لذلك الشعاع بـ $d_{(uv)}$، وهو عبارة عن المسافات المتقاطعة بين العنقود الجديد (uv) والعناقيد المتبقية في المصفوفة، ثم نضعها في العمود والسطر المخصصين لـ (uv) في المصفوفة، علماً بأن عناصر العنقود الجديد (uv) تحسب باستبدال كل عنصرين متقابلين في العنقودين u وv بأصغرهما أو بأكبرهما أو بمتوسطهما، وذلك حسب الربط المستخدم في عملية العنقدة المنفرد أو التام أو المتوسط على الترتيب، وسنشرح ذلك من خلال الأمثلة القادمة .

4- نكرر الخطوات (2) و(3) حتى $(n-1)$ مرة (حتى تصبح جميع المفردات في عنقود واحد) وفي كل مرحلة نقوم بتسجيل مواصفات وخواص العناقيد المدمجة ومستوياتها الجديدة (عناصر المسافات أو التشابه) في الأماكن المخصصة للدمج .

كما نقوم برسم مخطط الأغصان لكل مرحلة من هذه المراحل، وأخيراً ندمجها في المخطط العنقودي العام وسنوضح ذلك من خلال الأمثلة القادمة .

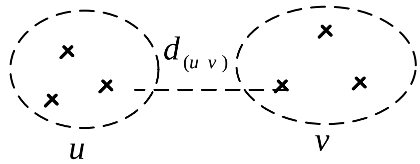
7-6-1 طريقة الربط المنفرد (Single Linkage):

إن هذه الطريقة تبدأ من اعتبار كل مفردة من المفردات تشكل عنقوداً خاصاً، وتعتمد على مصفوفة المسافات D (أو التشابه S) بين أزواج المفردات المدروسة، ويتم تشكيل العناقيد فيها من دمج العناقيد الأكثر تقارباً (الجوار الأقرب) .

ولتحديد العنقودين الأكثر تقارباً نقوم بدراسة عناصر المصفوفة D، ونحدد أصغر عنصر فيها . ولنفترض أنه كان يقابل العنقودين u وv، لذلك نقوم بدمج هذين العنقودين في عنقود جديد ونرمز له بـ $(u\,,v)$، ولحساب شعاع المسافات للعنقود الجديد $(u\,,v)$ نستبدل كل عنصرين متقابلين من (v) و(u) بأصغرهما (بأقربهما)، أي نطبق العلاقة:

$$d_{(u\,,v)} = \min_{j\in u\,,k\in v}\left[d_{jk}\right] \qquad (22-7)$$

حيث j ينتمي إلى u وk ينتمي إلى v .



ولحساب شعاع المسافات بين العنقود الجديد $(u\,,v)$ وأي عنقود آخر (أو مفردة) W نطبق العلاقة التالية:

$$d_{(u\,,v)w} = \min[d_{uw}\,,d_{vw}] \qquad (23-7)$$

حيث أن: d_{uw} و d_{vw} هما المسافتان بين العنقود الأكثر تقارباً u مع العنقود w . والعنقود الأكثر تقارباً v مع العنقود w على الترتيب .

مثال (7-5): لنفترض أن مصفوفة المسافات لـ (6) طلاب (مفردات) حسب نفقاتهم على الأغذية والاتصالات كانت كما يلي :

$$D = \begin{array}{c|c} \text{الطلاب} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 0 & 4 & 13 & 24 & 12 & 8 \\ & 0 & 10 & 22 & 11 & 10 \\ & & 0 & 7 & \mathbf{3} & 9 \\ & & & 0 & 6 & 18 \\ & & & & 0 & 8.5 \\ & & & & & 0 \end{bmatrix} \end{array}$$

ولتصنيف هؤلاء الطلاب ضمن عناقيد متشابهة نقوم بتطبيق خطوات الخوارزمية السابقة فنجد أن:

1- نعتبر أن كل طالب (مفردة) يشكل عنقوداً مستقلاً، ثم نقوم بدراسة عناصر المصفوفة D، فنجد أن أصغر عنصر في المصفوفة D هو العنصر (3) المقابل للطالبين (3) و(5)، لذلك نقوم بدمج هذين الطالبين في عنقود واحد، ونرمز له بـ (3، 5). ونحذف العمودين (3) و(5) والسطرين (3) و(5) من المصفوفة D، ثم نضيف عموداً خاصاً وسطراً خاصاً للعنقود الجديد (3، 5)، ونضعه مكان العمود (3) والسطر (3)، فنحصل على المصفوفة التالية:

$$D = \begin{array}{c} \text{العناقيد} \\ 1 \\ 2 \\ (3,5) \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & (3,5) & 4 & 6 \end{array} \\ \begin{bmatrix} 0 & 4 & d_{1(3,5)} & 24 & 8 \\ & 0 & d_{2(3,5)} & 22 & 10 \\ & & 0 & d_{(3,5)4} & d_{(3,5)6} \\ & & & 0 & 18 \\ & & & & 0 \end{bmatrix} \end{array}$$

وهنا نلاحظ أن عملية الدمج لا تؤثر على المفردات أو العناقيد الأخرى، ويتم حساب عناصر العمود (3,5) ثم عناصر السطر (3,5) من المصفوفة الأساسية D (علماً بأن $d_{jk} = d_{kj}$) وفق العلاقات التالية:

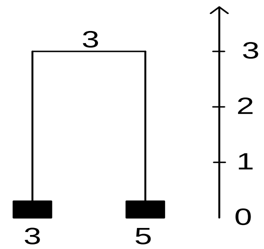
$$d_{1(3,5)} = \min(d_{13}, d_{15}) = \min(13,12) = 12$$
$$d_{2(3,5)} = \min(d_{23}, d_{25}) = \min(10,11) = 10$$
$$d_{(3,5)4} = \min(d_{34}, d_{54}) = \min(d_{34}, d_{45}) = \min(7,6) = 6$$
$$d_{(3,5)6} = \min(d_{36}, d_{56}) = \min(9,8.5) = 8.5$$

وبذلك نحصل على مصفوفة المسافات الجديدة D_1 التالية :

$D_1 =$

X_k / X_j	1	2	(3,5)	4	6
1	0	**4**	12	24	8
2		0	10	22	10
(3,5)			0	6	8.5
4				0	18
6					0

(c)



ولقد رسمنا العنقود (3,5) إلى جانب المصفوفة D_1 للتوضيح . ولمتابعة إجراء العنقدة نقوم بتطبيق الخوارزمية مرة أخرى. فنلاحظ أن أصغر عنصر في المصفوفة الأخيرة D_1 هو العنصر (4) المقابل للمفردتين (1) و(2). لذلك نقوم بدمج المفردتين (1) و(2) ضمن عنقود آخر نرمز له بـ (1,2) ونحذف العمودين (1) و(2) والسطرين (1) و(2) . ثم نخصص عموداً واحداً (1,2) وسطراً (1,2) للعنقود الجديد ونضيفهما في مكان العمود(1) والسطر (1)، فنحصل على المصفوفة التالية:

X_k / X_j	(1,2)	(3,5)	4	6
(1,2)	0	$d_{(1,2)(3,5)}$	$d_{(1,2)4}$	$d_{(1,2)6}$
(3,5)		0	6	8.5
4			0	18
6				0

ولحساب عناصر العنقود (1,2) الواقعة في السطر (1,2) نحسبها من المصفوفة D_1 وذلك بتطبيق العلاقات التالية:

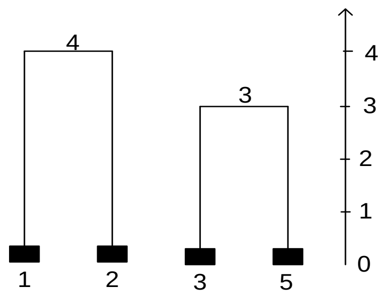
$$d_{(1,2)(3,5)} = \min[d_{1(3,5)}, d_{2(3,5}] = \min[12,10] = 10$$
$$d_{(1,2)4} = \min[d_{14}, d_{24}] = \min[24,22] = 22$$
$$d_{(1,2)6} = \min[d_{16}, d_{26}] = \min(8,10) = 8$$

وبذلك نحصل على مصفوفة المسافات D_2 التالية:

$D_2 =$

X_k / X_j	(1,2)	(3,5)	4	6
(1,2)	0	10	22	8
(3,5)		0	**6**	8.5
4			0	18
6				0

(e)



ولمتابعة عملية العنقدة نقوم بتطبيق الخوارزمية السابقة من جديد. فنجد أن أصغر عنصر في المصفوفة D_2 هو(6) المقابل للمفردة (4) والعنقود (3,5)، لذلك ندمجها في عنقود واحد ونرمز له بـ $[(3,5), 4]$، ونحذف العمودين (4) و(3,5) والسطرين (4) و(3,5)، ثم نضيف عموداً جديداً $[(3,5), 4]$ وسطراً جديداً $[(3,5), 4]$ ونضعهما مكان العمود والسطر (3,5) المحذوفين فنحصل على المصفوفة التالية:

X_k / X_j	(1,2)	[(3,5)4]	6
(1,2)	0	$d_{(1,2)[(3,5),4]}$	8
$[(3,5), 4]$		0	$d_{[(3,5)4],6}$
6			0

ولحساب عناصر العنقود الجديد $[(3,5), 4]$ نجد من المصفوفة D_2 أن :

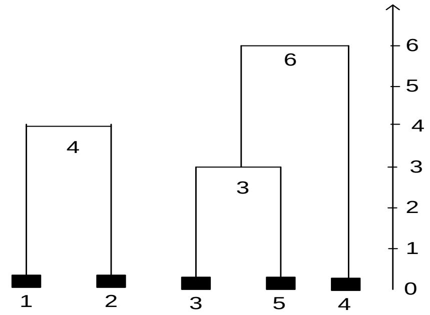
$$d_{(1,2)[(3,5),4]} = \min[d_{(1,2)(3,5)}, d_{(1,2)4}] = \min[10,22] = 10$$
$$d_{[(3,5)4],6} = \min[d_{(3,5)6}, d_{46}] = \min[8.5,18] = 8.5$$

وبذلك نحصل على المصفوفة التالية:

$D_3 =$

$X_j \backslash X_k$	(1,2)	[(3,5), 4]	6
(1,2)	0	10	**8**
[(3,5), 4]		0	8.5
6			0

(g)



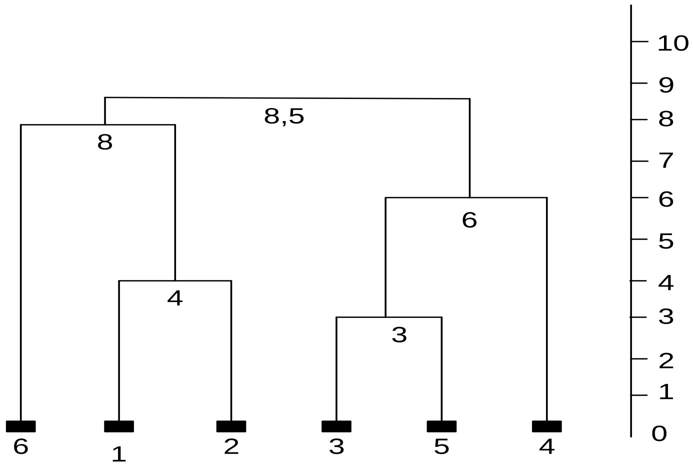
ولمتابعة عملية العنقدة نلاحظ أن أصغر عنصر في المصفوفة الأخيرة (D_3) هو (8) المقابل للمفردة (6) وللعنقود (1,2)، لذلك ندمجها في عنقود جديد نرمز له بـ $[(1,2), 6]$، وبعد إجراء العمليات والحسابات اللازمة نحصل على المصفوفة النهائية التالية:

$X_j \backslash X_k$	[(1,2), 6]	[(3,5), 4]
[(1,2), 6]	0	8.5
[(3,5), 4]		0

وذلك لأن العنصر الجديد يحسب من D_3 كما يلي:

$$d_{[(1,2),6][(3,5),4]} = \min\left[d_{(1,2)[(3,5)4]}, d_{6[(3,5)4]}\right] = \min[10,8.5] = 8.5$$

وأخيراً نجمع هذه العناقيد في عنقود واحد ورسم واحد كما يلي:



الشكل (7–3) التحليل العنقودي للمثال (7–5)

7–6–1–1: اسلوب المقارنة لحساب عناصر العنقود الجديد:

وهناك أسلوب آخر نسميه أسلوب المقارنة لحساب عناصر العنقود الجديد. ونلخص خطوات تطبيقه على المثال (7–5) كما يلي:

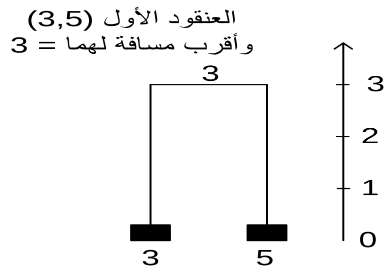
مثال(7−6): نأخذ المصفوفة D كاملة ونكتبها كما يلي:

$$
D = \begin{array}{c} \text{الطلاب} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array}
\begin{array}{c}
\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\
\begin{bmatrix}
0 & 4 & 13 & 24 & 12 & 8 \\
4 & 0 & 10 & 22 & 11 & 10 \\
13 & 10 & 0 & 7 & \mathbf{3} & 9 \\
24 & 22 & 7 & 0 & 6 & 18 \\
12 & 11 & 3 & 6 & 0 & 8.5 \\
8 & 10 & 9 & 18 & 8.5 & 0
\end{bmatrix}
\end{array}
$$

2− نبحث عن أصغر عنصر فيها فنجد أنه (3) . ويقابل الطالبين (3) و(5)، لذلك نشكل منهما عنقود جديد ونرمز له بـ (3,5)، ثم نقوم بحذف العمودين (3) و(5) والسطرين (3) و(5) من المصفوفة D، ثم نضيف عموداً جديداً على يمين المصفوفة الجديدة ونضيف سطراً جديداً إلى أسفلها نرمز له بـ (3,5) فنحصل على المصفوفة D_1 التالية:

$D_1 =$

	1	2	4	6	(3,5)
1	0	**4**	24	8	12
2	4	0	22	10	10
4	24	22	0	18	6
6	8	10	18	0	8.5
(3,5)	12	10	6	8.5	0



3− نقوم بحساب عناصر العمود (3,5) بأخذ أصغر العنصرين المتقابلين من العمودين (3) و(5) في الأسطر غير المحذوفة من المصفوفة الأساسية D فنجد أن:

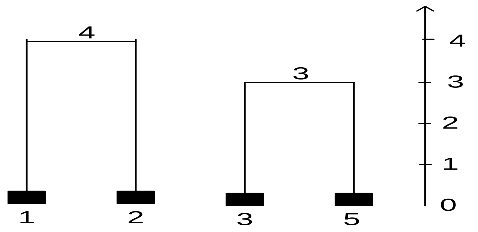
$$d_{1(3,5)} = \min[d_{13}, d_{15}] = \min[13,12] = 12$$
$$d_{2(3,5)} = \min[d_{23}, d_{25}] = \min[10,11] = 10$$
$$d_{4(3,5)} = \min[d_{43}, d_{45}] = \min[7,6] = 6$$
$$d_{6(3,5)} = \min[d_{63}, d_{35}] = \min[9,8.5] = 8.5$$

ثم نضع هذه القيم في العمود (3,5)، وكذلك في السطر (3,5)، فنحصل على المصفوفة D_1 السابقة، علماً بأن بقية العناصر في المربع القديم لا تتغير .

4− نبحث في المصفوفة D_1 عن أصغر عنصر فيها فنجد أنه الـ (4) المقابل للطالبين (1) و(2)، لذلك نحذف العمودين والسطرين (1) و(2)، ثم نشكل عنقوداً جديداً من العنقودين (1) و(2) ونرمز له بـ (1,2) ونضعه على يمين وأسفل المصفوفة الجديدة التي تأخذ الشكل التالي:

$D_2 =$

	4	6	(3,5)	(1,2)
4	0	18	**6**	22
6	18	0	8,5	8
(3,5)	6	8,5	0	10
(1,2)	22	8	10	0

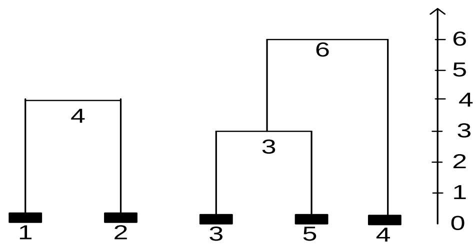


لقد تم حساب العمود (1,2) بأخذ أصغر العنصرين المتقابلين في العمودين (1) و(2) في المصفوفة D_1 والمقابلين للأسطر غير المحذوفة (4) و(6) و(3,5) . أما بقية العناصر ضمن المربع فلا تتغير .

5- نبحث في المصفوفة D_2 عن أصغر عنصر فيها، فنجد أنه الـ (6) المقابل للطالب (4) وللعنقود (3,5)، ولذلك ندمجهما في عنقود جديد ونرمز له بـ [(3,5),4]، وبعد إجراء العمليات والحساب اللازمة، نحصل على المصفوفة التالية:

$D_3 =$

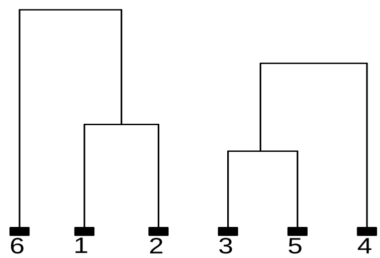
	6	(1,2)	[(3,5),4]
6	0	**8**	8.5
(1,2)	8	0	10
[(3,5),4]	8.5	10	0



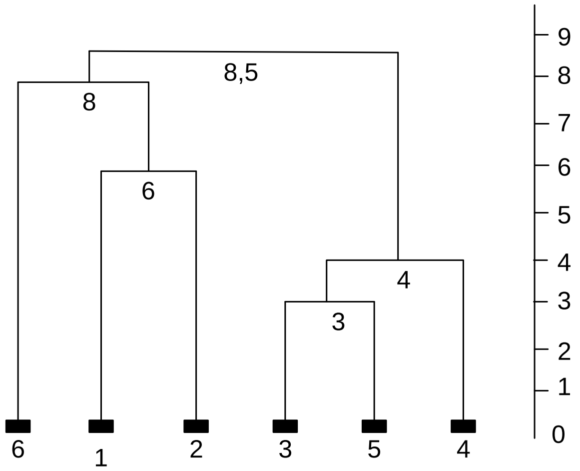
6- نبحث في المصفوفة D_3 عن أصغر عنصر فيها فنجد أنه الـ (8) المقابل للطالب (6) وللعنقود (1,2)، لذلك نقوم بدمجهما في عنقود جديد ونرمز له بـ [(1,2),6]، وبعد إجراء العمليات والحسابات اللازمة، نحصل على المصفوفة D_4 التالية:

$D_4 =$

	[(3,5),4]	[(1,2),6]
[(3,5),4]	0	8.5
[(1,2),6]	8.5	0



وبذلك نحصل على مصفوفة تتألف من عنقودين هما: [(3,5),4] و[(1,2),6]، نقوم بدمجهما لتشكيل العنقود الأخير فنحصل على المخطط التالي: (لاحظ أن أطوال الأغصان هي عبارة عن المسافات بين العناقيد) .



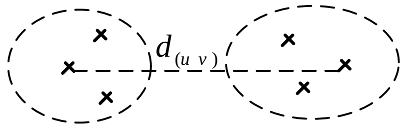
الشكل (7−4): الشكل العنقودي للمثال (7−5)

7−6−2 طريقة الربط التام (Complete Linkage)

وتبدأ هذه الطريقة أيضاً من اعتبار كل مفردة من المفردات تشكل عنقوداً خاصاً . وتعتمد على مصفوفة المسافات D (أو التشابه S)، بين أزواج المفردات المدروسة . كما إن عملية تشكيل العناقيد فيها، تتم من خلال دمج العناقيد الأكثر تقارباً وذلك حتى نحافظ على التشابه داخلياً ضمن العناقيد الجديد. ولكن عملية حساب عناصر شعاع المسافات للعنقود الجديد، تتم بواسطة استبدال كل عنصرين متقابلين من العنقودين المدمجين بأكبرهما (لجوار الأبعد) . ولهذا فإن عملية الحساب تبدأ بالبحث عن أصغر عنصر في المصفوفة D، ثم تحديد العنقودين الأكثر تقارباً. ونرمز لهما بـ u و v ، ثم نقوم بدمج هذين العنقودين في عنقود جديد نرمز له بـ $(u\,,v)$، ولحساب شعاع المسافات للعنقود الجديد $(u\,,v)$ نستبدل كل عنصرين متقابلين في u و v بأكبرهما (بأبعدهما)، أي نطبق العلاقة التالية:

$$d_{(u,v)} = \max_{j\in u\,,k\in v}\left[d_{jk}\right] \qquad (24-7)$$

حيث أن: j ينتمي إلى u و k ينتمي إلى v.



ثم نقوم بحساب شعاع المسافة بين العنقود الجديد $(u\,,v)$ وأي عنقود (أو مفردة) W من العلاقة التالية:

$$d_{(u,v)w} = \max[d_{uw}\,, d_{vw}] \qquad (25-7)$$

وعند إجراء الحسابات نتبع نفس الإجراءات، التي أتبعناها في طريقة الربط المنفرد، مع اختلاف وحيد هو طريقة حساب المسافات الجديدة، حيث أن مسافات العنقود الجديد $d_{(u,v)}$، تحسب بأخذ أكبر أي عنصرين متقابلين من عناصر u و v . ويمكن تطبيق العلاقتين السابقتين على المصفوفة المثلثية D، كما يمكن تطبيق أسلوب المقارنة على المصفوفة الأساسية كما فعلنا في حالة الربط المنفرد .

مثال (7–7): لنأخذ مصفوفة المسافات الواردة في المثال (7–5) السابق لـ (6) طلاب حسب النفقات والتي كانت كما يلي:

$$D = \begin{array}{c} \text{الطلاب} \\ \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 4 & 13 & 24 & 12 & 8 \\ 4 & 0 & 10 & 22 & 11 & 10 \\ 13 & 10 & 0 & 7 & \mathbf{3} & 9 \\ 24 & 22 & 7 & 0 & 6 & 18 \\ 12 & 11 & 3 & 6 & 0 & 8.5 \\ 8 & 10 & 9 & 18 & 8.5 & 0 \end{bmatrix} \end{array}$$

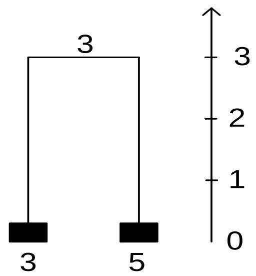
والمطلوب تصنيف هؤلاء الطلاب حسب طريقة الربط التام (الجوار الأبعد) وبأسلوب المقارنة:

حتى نصنف هؤلاء الطلاب ضمن عناقيد متشابهة نطبق الخوارزمية السابقة– حسب أسلوب المقارنة :

1– نبحث عن أصغر عنصر في D فنجد أنه الـ (3) المقابل للطالبين (3) أو(5). لذلك ندمجهما في عنقود جديد نرمز له بـ(3,5)، ونحذف العمودين والسطرين (3) و(5)، ونضيف إلى يمين المصفوفة وأسفلها عموداً وسطراً لـ (3,5)، ولحساب عناصر شعاع المسافات بين العنقود (3,5) وبقية العناقيد المتقاطعة معه، نأخذ أكبر العنصرين المتقابلين في العمودين (3) و(5) ونضعهما في العمود والسطر المخصص لـ (3,5)، فنحصل على المصفوفة D_1 التالية :

$D_1 =$

	1	2	4	6	(3,5)
1	0	**4**	24	8	13
2	4	0	22	10	11
4	24	22	0	18	7
6	8	10	18	0	9
(3,5)	13	11	7	9	0



لتوضيح كيفية حساب عناصر العمود (3,5)، قمنا بأخذ أكبر عنصر لأي عنصرين متقابلين في العمودين (3) و(5) [بعد حذف السطرين (3) و(5) من المصفوفة D] كما يلي:

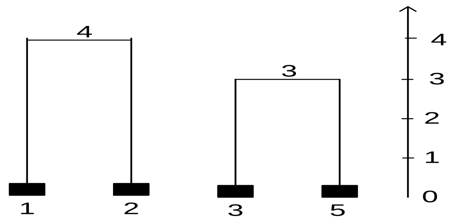
$$\begin{aligned} d_{1(3,5)} &= \max[d_{13}, d_{15}] = \max[13,12] = 13 \\ d_{2(3,5)} &= \max[d_{23}, d_{25}] = \max[10,11] = 11 \\ d_{4(3,5)} &= \max[d_{43}, d_{45}] = \max[7,6] = 7 \\ d_{6(3,5)} &= \max[d_{63}, d_{35}] = \max[9,8.5] = 9 \end{aligned}$$

أما بقية العناصر في المربع القديم فلا تتغير .

2– نبحث في المصفوفة D_1 عن أصغر عنصر فيها، فنجد أنه الـ (4) المقابل للطالبين (1) و(2)، لذلك ندمجهما في عنقود جديد نرمز له بـ (1,2). ثم نقوم بإجراء العمليات والحسابات اللازمة فنحصل على المصفوفة التالية:

$D_2 =$

	4	6	(3,5)	(1,2)
4	0	18	**7**	24
6	18	0	9	10
(3,5)	7	9	0	13
(1,2)	24	10	13	0

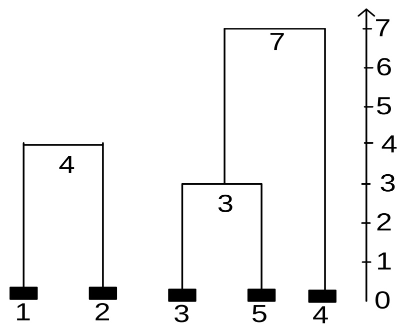


لقد تم حساب عناصر العمود (1,2) بأخذ أكبر العنصرين المتقابلين في العمودين (1) و(2). [بعد حذف السطرين (1) و(2)] فحصلنا على المصفوفة D_2 السابقة .

3- نبحث في المصفوفة D_2 عن أصغر عنصر فيها، فنجد أنه الـ (7) المقابل للطالب (4) وللعنقود (3,5)، لذلك ندمجهما في عنقود جديد ونرمز له بـ [(3,5),4]، وبعد إجراء العمليات والحسابات اللازمة نحصل على المصفوفة التالية:

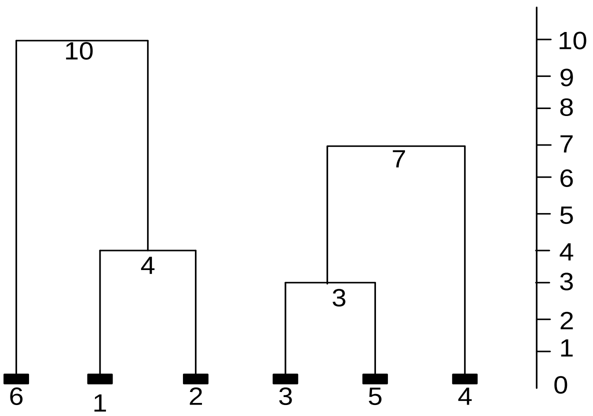
$D_3 =$

	6	(1,2)	[(3,5), 6]
6	0	**10**	18
(1,2)	**10**	0	24
[(3,5), 6]	18	24	0

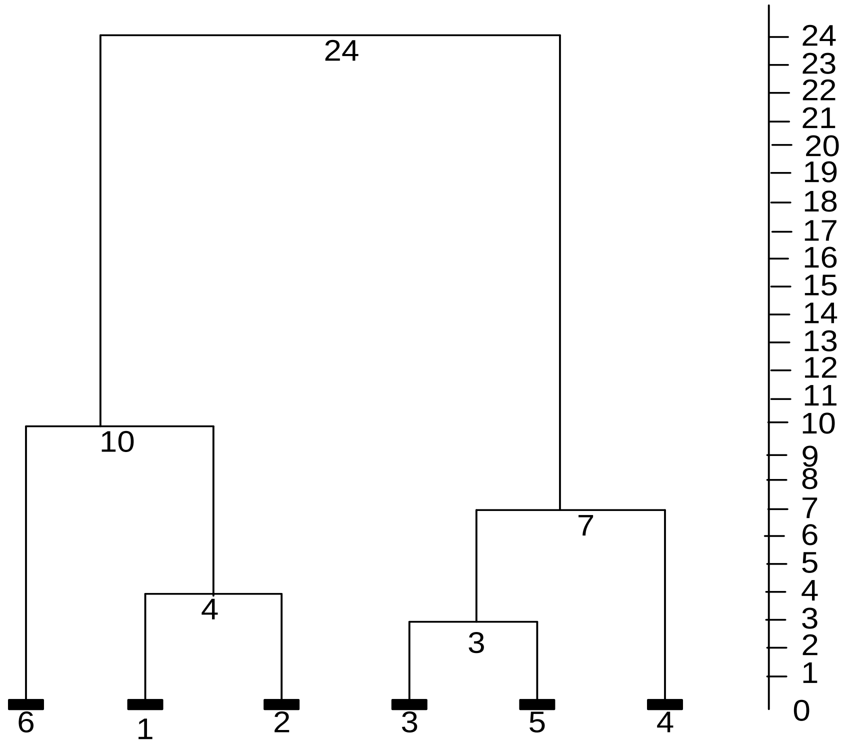


4- نبحث في المصفوفة D_3 عن أصغر عنصر فيها، فنجد أنه الـ (10) المقابل للطالب (6) وللعنقود (1,2)، لذلك ندمجهما في عنقود جديد ونرمز له بـ [(1,2)6]. وبعد إجراء العمليات والحسابات اللازمة نحصل على المصفوفة التالية:

	[(3,5), 6]	[(1,2), 6]
[(3,5), 6]	0	24
[(1,2), 6]	24	0



ثم نقوم بدمج العنقودين الأخيرين بعنقود واحد فنحصل على التصنيف العنقودي لهؤلاء الطلاب الستة حسب نفقاتهم ، وإنه سيأخذ الشكل التالي:



الشكل (7−5): التصنيف العنقودي للطلال الستة

وهنا نلاحظ أن هذا التصنيف حسب طريقة الربط التام (الجوار الأبعد) يتوافق من حيث الجوهر مع التصنيف الذي حصلنا عليه بطريقة الربط المنفرد . ولكنه يختلف عنه في أن أطوال الأغصان النهائية بين العناقيد تختلف عن بعضها البعض حيث أصبحت (24) بدلاً من (8.5) في الربط المنفرد .

ملاحظة: إن استبدال كل عنصرين متقابلين من u وv بأكبرهما في هذه الطريقة يجعل العنقود الجديد أكثر بعداً عن العناقيد الأخرى. وهذا ما يترك المجال مفتوحاً أمام العناقيد الأولية لتندمج مع بعضها قبل أن تندمج مع ذلك العنقود الجديد، وذلك خلال عمليات الاندماج والتجميع المتتالية. ولكن ذلك قد يؤدي لاحقاً إلى دمج بعض العناقيد غير المتشابهة ويجعل بعض العناقيد غير متجانسة داخلياً. ولهذا تم اللجوء إلى أسلوب آخر كحل وسطي هو أسلوب الربط المتوسط التالية.

7−6−3 طريقة الربط المتوسط: (Average Linkage)

إن هذه الطريقة تبدأ أيضاً من اعتبار كل مفردة من المفردات تشكل عنقوداً خاصاً، وتعتمد على مصفوفة المسافات D (أو التشابه S) بين أزواج المفردات المدروسة. كما إن عملية تشكيل العناقيد فيها تتم من خلال دمج العناقيد الأكثر تقارباً. ولكن عملية حساب عناصر شعاع المسافات للعنقود الجديد تتم بواسطة أخذ المتوسط الحسابي لكل عنصرين متقابلين من العنقودين المدمجين .

ولتطبيق هذه الطريقة ندرس أولاً عناصر المصفوفة D ونحدد أصغر عنصر فيها، ولنفترض أنه يقابل العنقودين u وv . لذلك علينا أن نقوم بدمج العنقودين u وv في عنقود جديد ونرمز له بـ (u,v)، ولحساب شعاع المسافات المتوسطة للعنقود الجديد (u,v) نأخذ متوسط كل عنصرين متقابلين في

العمودين u و v ونضعهما في عمود وسطر مخصص للعنقود الجديد $(u\,,v)$. أي نقوم بحساب المسافات المتوسطة للعنقود $(u\,,v)$ من العلاقة:

$$d_{(u,v)_j} = \frac{1}{2}\left[d_{uj} + d_{vj}\right] \qquad (26-7)$$

كما يمكن حساب شعاع المسافة بين العنقود الجديد $(u\,,v)$ وأي عنقود آخر W (أو مفردة) من العلاقة:

$$d_{(u,v)w} = \frac{\sum_{j=1}\sum_{k=1} d_{jk}}{n_{(uv)} * n_{(w)}} \qquad (27-7)$$

حيث أن: $n_{(uv)}$ هو عدد المفردات الموجودة في العنقود $(u\,,v)$

وأن: $n_{(w)}$ هو عدد المفردات الموجودة في العنقود (w) .

وأن: d_{jk} هي المسافة بين العنصر j من $(u\,,v)$ والعنصر k من (w)

وهناك صيغ أخرى لحساب هذه المسافات أهمها صيغة (Ward) التالية [Webb P. 368]:

$$d_{(i+j)k} = \frac{n_k + n_i}{n_k + n_i + n_j} d_{ik} + \frac{n_k + n_j}{n_k + n_i + n_j} d_{jk} - \frac{n_k}{n_k + n_i + n_j} d_{ij} \qquad (28-7)$$

حيث أن: n_k و n_j و n_i هي أعداد المفردات في المجموعات $(i\,,j\,,k)$ على الترتيب .

وحيث أن كل من d_{jk} و d_{ik} هي المسافة الاقليدية بين المفردتين $(j\,,k)$ و $(i\,,k)$ على الترتيب .

أما d_{ij} فهو مربع المسافة الاقليدية بين المفردتين $(i\,,j)$.

مثال (7−8): لنتابع البيانات المثال (7−5) لتصنيف (6) طلاب حسب نفقاتهم التي كان لها الشكل التالي:

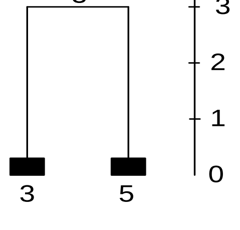
$$D = \begin{array}{c} \text{الطلاب} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 4 & 13 & 24 & 12 & 8 \\ 4 & 0 & 10 & 22 & 11 & 10 \\ 13 & 10 & 0 & 7 & \mathbf{3} & 9 \\ 24 & 22 & 7 & 0 & 6 & 18 \\ 12 & 11 & 3 & 6 & 0 & 8.5 \\ 8 & 10 & 9 & 18 & 8.5 & 0 \end{bmatrix} \end{array}$$

والمطلوب تصنيف هؤلاء الطلاب ضمن عناقيد حسب الربط المتوسط .

1− نبحث عن أصغر عنصر في المصفوفة D فنجد أنه الـ (3) المقابل للطالبين (3) و(5). لذلك ندمج الطالبين (3) و(5) في عنقود نسميه (3,5). ثم نقوم بإجراء العمليات والحسابات اللازمة فنحصل على المصفوفة التالية :

$D_1 =$

	1	2	4	6	متوسط (3,5)
1	0	**4**	24	8	12.5
2	4	0	22	10	10.5
4	24	22	0	18	6.5
6	8	10	18	0	8.75
(3,5)	12.5	10.5	6.5	8.75	0

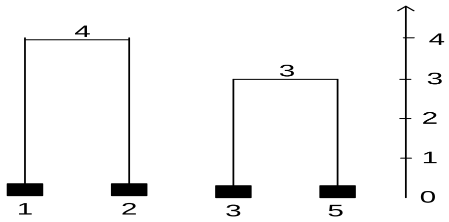


لقد تم حساب عناصر العمود (السطر) (3,5) بأخذ متوسطات عناصر العمودين (3) و(5) في المصفوفة الأساسية D . أما بقية العناصر في المربع فلا تتغير .

2- نبحث في المصفوفة D_1 عن أصغر عنصر فيها فنجد أنه الـ (4) المقابل للطالبين (1) و(2). لذلك ندمج هذين الطالبين في عنقود واحد ونرمز له بـ (1,2) ثم نقوم بإجراء العمليات والحسابات اللازمة فنحصل على المصفوفة التالية:

$D_2 =$

	4	6	(3,5)	(1,2)
4	0	18	**6.5**	23
6	18	0	8.75	9
(3,5)	6.5	8.75	0	11,5
(1,2)	23	9	11.5	0

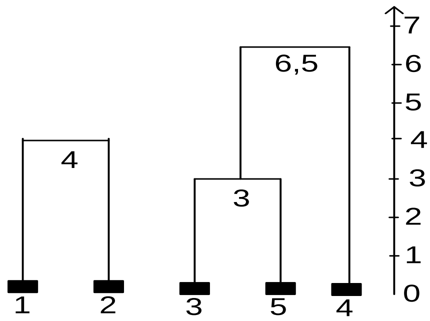


لقد تم حساب عناصر العمود (1,2) بأخذ متوسطات عناصر العمودين (1) و(2) في المصفوفة السابقة D_1 .

3- نبحث عن أصغر عنصر في المصفوفة D_2 فنجد أنه الـ (6.5) المقابل للطالب (4) وللعنقود (3,5) لذلك ندمجهما في عنقود جديد ونرمز له بـ [4(3,5)] ، وبعد إجراء العمليات والحسابات اللازمة نحصل على المصفوفة التالية:

$D_3 =$

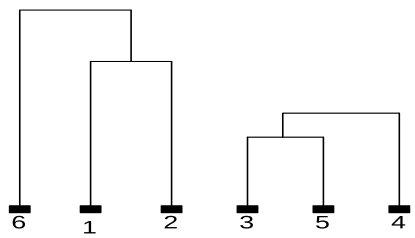
	6	(1,2)	[(3,5), 4]
6	0	**9**	13.375
(1,2)	9	0	17.25
[(3,5), 4]	13.375	17.25	0



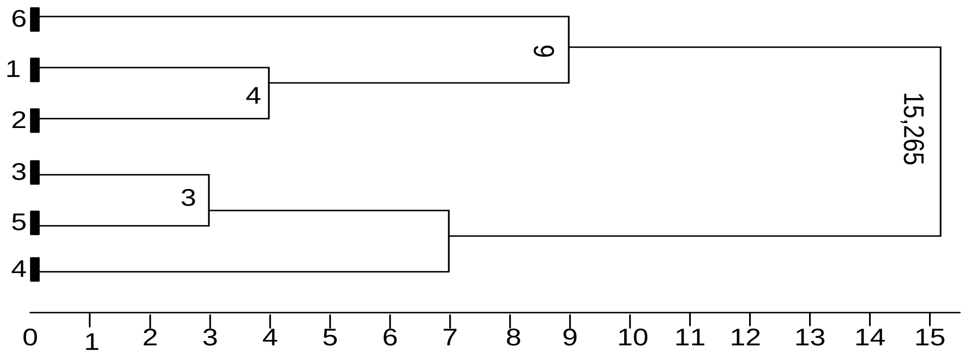
لقد تم حساب عناصر العنقود [(3,5), 4] من متوسطات العمودين (4) و(3,5) في المصفوفة D_2

4- نبحث عن أصغر عنصر في المصفوفة D_3 فنجد أنه الـ (9) المقابل للطالب (6) وللعنقود (1,2)، لذلك ندمجهما في عنقود جديد ونرمز له بـ [(1,2), 6] ، وبعد إجراء العمليات والحسابات اللازمة نحصل على المصفوفة النهائية التالية :

	[(3,5), 4]	[(1,2), 6]
[(3,5), 4]	0	15.265
[(1,2), 6]	15.265	0



نقوم بدمج العنقودين الأخيرين فنحصل على المخطط العنقودي نفسه. ولكن بأبعاد مختلفة لأطوال الأغصان والتي يمكن أن نمثله أفقياً كما يلي :



الشكل (7−6) الشكل العنقودي للمثال (7−8)

7−7 : التحليل العنقودي غير الهرمي (Non hierarchical)

يطبق التحليل العنقودي غير الهرمي على العينات الكبيرة، وهو يصلح لتجميع بيانات المفردات الكثيرة (أكثر مما يصلح للمتحولات X)، وذلك ضمن k عنقوداً محدداً. وإن عدد العناقيد k يمكن تحديده بشكل شخصي، أو من خلال أسلوب العنقدة .

ويعتمد التحليل العنقودي غير الهرمي على أسلوب التقسيم ويفترض ما يلي:

1− وجود أو إيجاد تجزئة أولية للمفردات ضمن مجموعات محددة بأي شكل كان .

2− اعتبار أن المجموعات الأولية تشكل عناقيد أولية .

3− إعادة تقسيم المجموعات الأولية وتشكيل عناقيد أصغر فأصغر حتى مرحلة التوقف .

إن الاختيار الجيد لترتيبات البداية يجب أن يكون مستقلاً عن أي تحيز. وإن أحد أساليب الاختيار، هو الأسلوب العشوائي لاختيار مفردات العناقيد الأولية من بين المفردات، أو بإجراء تجزئة عشوائية للمفردات ضمن مجموعات أولية. وتسمى هذه التجزئة بالتجزئة السريعة (Quick) . ويستخدم في التحليل العنقودي غير الهرمي عدة طرائق رياضية لفرز المفردات إلى مجموعات أو عناقيد متشابهة، وتستند هذه الطرائق على مبدأ تصغير مجموع مربعات الانحرافات داخل العناقيد المشكلة في كل مرحلة من مراحل التشكيل، وإن أهم هذه الطرائق هي:

- طريقة الـ K متوسطاً (K− means) .
- طريقة التجزئة السريعة− طريقة النمذجة المختلطة− طريقة النمذجة اللاخطية .
- طريقة الـ K متوسطاً الغامضة(**Fuzzy k− means**).

ويتم تطبيق هذه الطرائق ضمن معايير محددة، تسمى معايير العنقدة، وهي تضمن لنا درجة معينة من التشابه بين مفردات كل عنقود . وسنستعرض ذلك وفق ما يلي:

1-7-7 معايير العنقدة :

لنفترض أنه لدينا n مفردة، تشكل g عنقوداً أولياً (أو مرحلياً) وحصلنا منها على البيانات العددية لـ p متحولاً مؤثراً على عملية العنقدة هي: $X_1\ X_2\ X_3\ ...\ X_P$. وبناءً على ذلك يمكننا حساب أو تقدير المصفوفات التالية:

- مصفوفة التباين المشترك وتحسب من العلاقة التالية :

$$\tilde{V} = S = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})' \qquad (29-7)$$

- مصفوفة مجموع مربعات الانحرافات داخل within العناقيد ونرمز لها بـ S_W . أو مصفوفة المجموع المدمج لها S_p، والتي تحسب من العلاقة التالية:

$$S_w = \frac{1}{n}\sum_{j=1}^{g}\sum_{i=1}^{n_j} Z_{ji}(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{\boldsymbol{j}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{\boldsymbol{j}})' \qquad (30-7)$$

حيث أن: $Z_{ji} = 1$ إذا كانت x_i تنتمي إلى المجموعة j، و$Z_{ji} = 0$ إذا كانت x_i لا تنتمي إلى المجموعة j، وأن $\bar{x}_j$ هو متوسط المجموعة j، و n_j هو عدد المفردات في العنقود أو المجموعة j .

- مصفوفة مجموع مربعات الانحرافات بين تلك العناقيد (Between) ونرمز لها بـ S_B، التي تحسب من العلاقة التالية:

$$S_B = S - S_w = \sum_{j=1}^{g}\frac{n_j}{n}(\overline{\boldsymbol{x}}_{\boldsymbol{j}} - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_{\boldsymbol{j}} - \overline{\boldsymbol{x}})' \qquad (31-7)$$

- مصفوفة التباين العام والتي تحسب من العلاقة:

$$S = S_W + S_B \qquad (31\,a-7)$$

ويستفاد من هذه المصفوفات في تعريف معايير العنقدة، التي تسعى إلى جعل العناقيد متشابهة داخلياً ومتباينة خارجياً . ولكننا هنا لا نعلم ما هو عدد تلك العناقيد . علماً بأنه سوف يكون لدينا بيانات من عدد معلوم من المجموعات أو الفئات .

والآن سنبحث عن التحويلات التي تحقق لنا ذلك الهدف، والتي تسمى معايير العنقدة، وأهمها المعايير التالية:

1- تصغير مجموع المربعات داخل المجموعات المشكلة S_W :

يعتبر هذا المعيار أساساً للمعايير الأخرى، لأنه يسعى إلى تصغير مجموع المربعات الداخلية S_W المعرف بالعلاقة (7-30)، وبالتالي إلى تعظيم مجموع المربعات البينية S_B المعرف بالعلاقة(7-31).

ويعتمد هذا المعيار على مفهوم المسافة الاقليدية المعرف بين أية نقطة $x(x_1\ x_2\ x_3\ ...)$ ومركز المجموعة j $\bar{x}_j(\bar{x}_{1j}\ \bar{x}_{2j}\ \bar{x}_{3j}\ ...)$ بالعلاقة التالية:

$$D_{x/j} = \sqrt{(x_1 - \bar{x}_{1j})^2 + (x_2 - \bar{x}_{2j})^2 + (x_3 - \bar{x}_{3j})^2 + \cdots} \qquad (32-7)$$

ولتطبيق هذا المعيار علينا أن نقوم بحساب هذه المسافات لتحديد بعد المفردة x عن جميع مراكز المجموعات j المشكلة في كل مرحلة من مراحل التجزئة. ثم نحدد أصغر تلك المسافات، وبعدها نقوم بتنسيب المفردة x إلى المجموعة k المقابلة لتلك المسافة الصغرى. أي أن قاعدة التنسيب ستكون على الشكل التالي:

نقوم بتنسيب x إلى المجموعة k إذا كانت

$$D_{x/k} = \min\left[D_{x/j}\right] \qquad j = 1\ 2\ 3 \ldots .\, g \qquad (33-7)$$

وهكذا يتم تشكيل المجموعات (العناقيد) بشكل متتالٍ حتى ننتهي من تنسيب جميع المفردات المدروسة إلى مجموعات محددة، وتتشكل لدينا العناقيد المتشابهة داخلياً والمتباينة خارجياً .

ولكن عندما يكون عدد المفردات n كبيراً فإن المتطلبات الحسابية المعقدة تعيق عملية الحصول على حلول مثالية أو على عناقيد مثالية . لذلك نلجأ في بعض الأحيان إلى معايير شبه مثالية وإجراء تجزئات شبه مثالية. وهذا قد يقودنا إلى الحصول على عناقيد شبه مثالية .

2- تصغير أثر المصفوفة S_W :

إن أثر trace المصفوفة S_W هو مجموع عناصرها القطرية ويرمز له بـ Tr وهو يساوي :

$$Tr(S_W) = \frac{1}{n}\sum_{j=1}^{g}\sum_{i=1}^{n} Z_{ji}\left(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{\boldsymbol{j}}\right)^2 = \frac{1}{n}\sum_{j=1}^{g} S_j \qquad (34-7)$$

حيث رمزنا بـ S_j للمجموع الثاني :

$$S_j = \sum_{i=1}^{n} Z_{ji}\left(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{\boldsymbol{j}}\right)^2 \qquad (35-7)$$

وهو عبارة عن مجموع المربعات داخل المجموعة المحددة j . حيث $j = 1\ 2\ 3 \ldots .\, g$

وهكذا نجد أن تصغير هذا الأثر $Tr(S_W)$ لكل المجموعاتg، يكافئ تصغير المجموع الكلي لمربعات الانحرافات S_W داخل المجموعات g حول مراكزها $\bar{x}_j$.

وإن طرائق العنقدة التي تصغر هذه الكمية تكون أحياناً أفضل من تصغير مجموع المربعات الداخلي S_W، علماً بأن هذا المعيار لا يتأثر بواحدات المحاور ولا بمعيرة المتحولات، وكذلك فإنه لا يتأثر بالتحويلات الخطية التي تُجرى على البيانات .

3- تصغير النسبة $\frac{|S_W|}{|S|}$:

إن هذا المعيار أيضاً لا يتأثر بالتحويلات الخطية النظامية للبيانات، وهو من أجل مجموعة بيانات محددة يكافئ تجزئة البيانات التي تصغر محدد المصفوفة $|S_W|$ (لأن المصفوفة S ثابتة ومستقلة عن أشكال التجزئة) .

4– تعظيم أثر جداء المصفوفتين $Tr(S_W^{-1} * S_B)$:

وهو تعميم لطريقة مجموع المربعات، التي لا تكون العناقيد فيها على شكل مجسمات كروية، بل على شكل مجسمات قطوع ناقصة. وهو يكافئ تصغير مجموع المربعات ضمن قياسات المسافات حسب (مهالانوبيس Mahalanobis)، وهو أيضاً لا يتأثر بالتحويلات النظامية للبيانات .

5– تصغير أثر جداء المصفوفتين $Tr(S^{-1} * S_W)$:

وهو معيار يتطابق مع تصغير مجموع المربعات للبيانات، التي لها الصيغة المعيارية، وتجعل المصفوفة الكلية تساوي مصفوفة واحدية .

وهكذا نجد أن هدفنا أصبح البحث عن تجزئة حقيقية (غير مكررة) لـ n مفردة ضمن k مجموعة حتى يأخذ المعيار المختار قيمته المثلى. ولإيجاد التجزئة المثالية يجب أن نقوم بفحص كل تجزئة ممكنة . علماً بأن عدد التجزئات الحقيقية (غير المكررة) لـ n عنصراً ضمن مجموعتين فقط $(k = 2)$ يساوي حسب مفكوك ثنائي الحدين ما يلي:

$$M = \sum_{i=1}^{n} C_n^i = 2^{n-1} - 1$$

وكمثال على ذلك نجد أن عدد التجزئات الممكنة لـ 60 عنصراً ضمن مجموعتين فقط يساوي:

(طريقة التجزئة) $M = 2^{59-1} - 1 \approx 6.10^{17}$

وهذا يقدم لنا استعراضاً شاملاً لجميع المجموعات الجزئية الممكنة، ويلقي علينا أعباءً كبيرة في عمليات التصنيف، ويتطلب منا إجراءات وحسابات معقدة، لا يمكن إنجازها إلا بواسطة الحواسيب وباستخدام برامج خاصة لذلك . والآن ننتقل إلى استعراض الطرائق المستخدمة في التحليل العنقودي غير الهرمي وهي:

7–7–2 : طريقة الـ k متوسطاً (K– means) :

إن هدف هذه الطريقة هو تجزئة بيانات العينة ضمن k عنقوداً، بحيث يكون مجموع المربعات S_W داخل المجموعات أصغر ما يمكن (المعيار (1))، وإن خوارزمية تطبيق طريقة *k* متوسطاً تتألف من الخطوات التالية:

1– تحديد عدد العناقيد المطلوبة *k*. وإجراء تجزئة أولية عشوائية أو عمدية لمفردات العينة، إلى *k* مجموعة، مع فرز عناصر كل مجموعة على حدة. وتسمى هذه المجموعات بالعناقيد الأولية.

2– حساب مراكز مجموعات العناقيد الأولية، ثم حساب المسافات بين كل مفردة *x* ومراكز تلك العناقيد، ثم تحديد أصغرها للقيام بتنسيب المفردة *x* من جديد إلى العنقود الذي تكون المفردة أقرب إلى مركزه من أي عنقود آخر. أي نقوم بإضافة *x* إلى ذلك العنقود وسحبه من المجموعة التي كان فيها. أي نقوم بإعادة تنسيبه إلى العنقود الأقرب إليه .

3- نكرر الخطوة (2) حتى تشمل جميع المفردات n، ونعيد تنسيب أية مفردة إلى العنقود الأقرب إليها أو نثبتها في العنقود التي كانت فيه (عندما يكون الأقرب لها) .

4- إذا كان عدد العناقيد k غير مرتبط بدرجة دقة معينة للتوقف عن التجزئة، نقوم بتحديد عدد آخر للعناقيد (مثل $k' = k + 1$) ونكرر الخطوات (1) و(2) و(3)، ثم نقارن نتائج الحالتين ونقيّمها ونختار الحالة الأفضل .

مثال (7-9): لنأخذ بيانات المثال (7-1) والمذكورة في الجدول (7-2). ولنفترض إننا نريد تجميع هؤلاء الطلاب حسب نفقاتهم ضمن عنقودين فقط (أي نضع $k = 2$). فعندها نجد أن عدد التجزئات الممكنة يساوي: (طريقة التجزئة) $M = 2^{5-1} - 1 = 15$

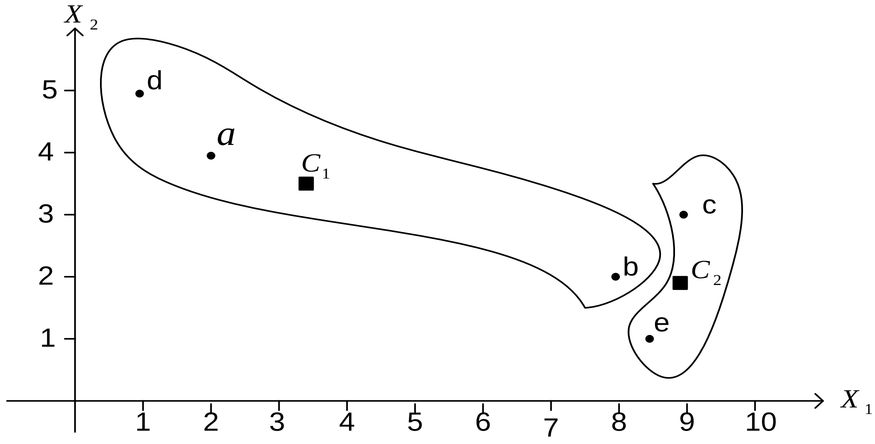
ولكننا سوف لن نحتاج إلى جميع هذه التجزئات إذا استخدمنا أحد معايير العنقدة المذكورة، وسنقوم بإجراء التجزئة اعتماداً على المعيار(1) الذي يعتمد على المسافات الاقليدية، ولنبدأ بتطبيق الخطوات كما يلي:

1- نقوم بإجراء أي تجزئة عشوائية أو عمدية، ولتكن كما يلي: نعتبر المفردات a و b و d في العنقود (1) الأولي ونعتبر المفردتين c و e في العنقود (2) الأولي .

2- ثم نقوم بحساب مركزي هذين العنقودين، اعتماداً على البيانات الواردة في الجدول (7-2) ونضعها بعد إجراء التجزئة السابقة عليها في جدول خاص كما يلي :

جدول (7-11): متوسطات العناقيد الأولية:

العنقود (1) = a b d			العنقود (2) = c e		
المفردات	X_1	X_2	المفردات	X_1	X_2
a	2	4	C	9	3
b	8	2	E	8.5	1
d	1	5			
متوسط العنقود(1)	3.67	3.67	متوسط العنقود (2)	8.75	2



الشكل (7-7) التجزئة الأولية

ومن الجدول(7−11) السابق نلاحظ أن إحداثيات مركز العنقود(1) تساوي= $C_1(X_1 = 3.67, X_2$ (3.67، وإحداثيات مركز العنقود (2) تساوي $C_2(X_1 = 8.75, X_2 = 2)$، حيث رمزنا لها على الشكل (7−7) بالرمزين C_1 و C_2 على الترتيب، علماً بأن مركز العنقود يعتبر مركز المفردات التي فيه . والآن نقوم بحساب مسافة المفردة a عن مركزي هذين العنقودين فنجد أن:

$$d_{a,dbd} = \sqrt{(2-3{,}67)^2 + (4-3{,}67)^2} = 1{,}702$$

$$d_{a,ce} = \sqrt{(2-8{,}75)^2 + (4-2)^2} = 7{,}040$$

ثم نقارن هاتين المسافتين فنجد أن: $d_{a,dbd} < d_{a,ce}$، أي أن المفردة a أقرب إلى مركز العنقود (1) منه إلى مركز العنقود (2)، لذلك نعتبر انتسابها إلى العنقود(1) انتساباً صحيحاً ونتركها في ذلك العنقود.

ثم نقوم بحساب مسافة المفردة b عن مركزي العنقودين (1) و(2) فنجد أن:

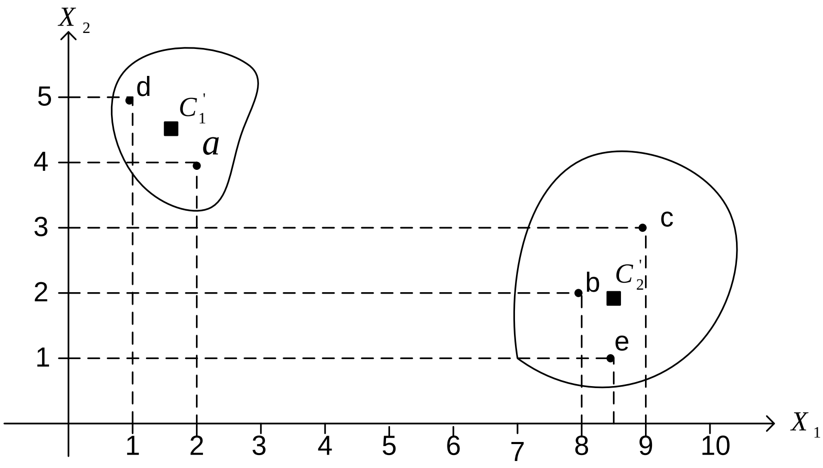
$$d_{b,abd} = \sqrt{(8-3{,}67)^2 + (2-3{,}67)^2} = 4{,}641$$

$$d_{b,ce} = \sqrt{(8-8{,}75)^2 + (2-2)^2} = 0{,}750$$

وعند مقارنة هاتين المسافتين نجد أن: $D_{(b,ce)} < D_{(b,abd)}$، أي أن المفردة b هي أقرب إلى العنقود (ce). لذلك ننقلها من العنقود (abd) إلى العنقود (ce)، فتصبح لدينا مكونات العنقودين الجديدين كمايلي: العنقود (1) ويتألف من (ad) والعنقود (2) ويتألف من (ceb). نقوم الآن بوضع إحداثيات مفردات هذين العنقودين في جدول جديد، ونحسب مركزي العنقودين الجديدين فنحصل على الجدول التالي :

جدول (7−12): نتائج حسابات مراكز العناقيد الجديدة

العنقود (2)			العنقود (1)		
X_2	X_1	المفردات	X_2	X_1	المفردات
3	9	C	4	2	a
1	8.5	E	2	8	d
2	8	B			
2	8.5	المتوسط	4.5	1.5	المتوسط



الشكل (7−8) شكل العنقودين الجديدين

ومن الجدول (7−12) نلاحظ أن إحداثيات مركز العنقود (1) الجديد أصبحت تساوي: $C_1'(X_1 = 1{,}5, X_2 = 4{,}5)$ وإحداثيات مركز العنقود(2) الجديد أصبحت تساوي: $C_2'(X_1 = 8{,}5, X_2 = 2)$ كما هو موضح على الشكل (7−8) .

ولمتابعة التحليل نعود ونحسب من جديد مسافات المفردات الأصلية عن مركزي هذين العنقودين الجديدين، فنجد أن مسافات a تساوي:

$$D_{a,ad} = \sqrt{(2-1{,}5)^2 + (4-4{,}5)^2} = 0{,}707$$

$$D_{a,ceb} = \sqrt{(2-8{,}5)^2 + (4-2)^2} = 6{,}801$$

وبالمقارنة نجد أن: $D_{a,ad} < D_{a,ceb}$، لذلك نبقي المفردة a في العنقود (1) كما كانت عليه .

أما مسافات b فتساوي :

$$D_{b,ad} = \sqrt{(18-1{,}5)^2 + (2-4{,}5)^2} = 6{,}964$$

$$D_{b,ceb} = \sqrt{(8-8{,}5)^2 + (2-2)^2} = 0{,}500$$

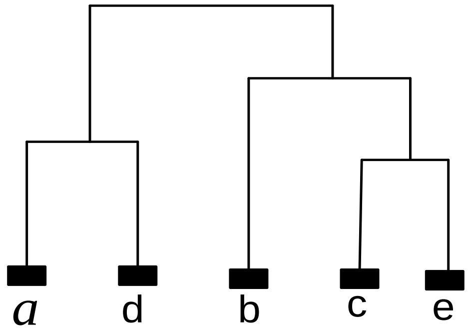
وبالمقارنة نجد أن: $D_{b,ceb} < D_{b,ad}$، لذلك نبقي المفرد b في العنقود (2) كما كانت قد أصبحت عليه في المرحلة السابقة .

وهكذا نتابع الحسابات لبقية المفردات، فنحصل على النتائج المبينة في الجدول التالي:

جدول (7−13): مسافات المفردات عن مركزي العنقودين

المفردات	عن مركز العنقود (1) $ad =$	عن مركز العنقود (2) $ceb =$	نتيجة المقارنة
a	0,707*	0,801	a. تنتمي إلى العنقود (1)
b	6,964	0,500*	b. تنتمي إلى العنقود (2)
c	7,649	1,118*	c. تنتمي إلى العنقود (3)
d	0,707*	8,078	d. تنتمي إلى العنقود (4)
e	7,826	1,000*	e. تنتمي إلى العنقود (5)

ومن نتائج الجدول (7−13) نلاحظ أن كل من المفردات a وb وc وd وe ، قد استقر في العنقود الأقرب له، ولم يعد هناك حاجة لمتابعة التحليل والعنقدة، وبذلك نحصل على الشكل العنقودي التالي:



الشكل (7−8): الشكل العنقودي للمثال (7−1)

ملاحظة: هناك عدة أشكال لتطبيق طريقة الـ (k− متوسط)، فبعضها يتطلب أن يتم تحديد مراكز لعناقيد الأولية (والتي تسمى عادة بالبذور seeds)، وهذه البذور يمكن أن تكون المفردات نفسها، وعندها فإن المفردات الداخلية التي تحقق شرط المسافة الأقرب عن مركز العنقود يتم تنسيبها إلى ذلك العنقود، أما عندما تكون المفردة الأولى أقرب إلى مركز عنقود آخر نقوم بإعادة تنسيبها إليه .

ونقوم بإعادة إجراء حساب مراكز العناقيد الجديدة، ثم نكرر هذه العمليات حتى يتم فحص جميع المفردات، ثم نعيد التكرار حتى يتم التوقف عند حد معين، (وفق المعيار المعتمد)، إن إسلوب هذه العنقدة والتي تدعي (بالسريعة أو الطازجة) يستخدم في البرامج الحاسوبية المعروفة SPSS وSAS وR وMatlab وغيرهم، ويقدم بدائل جديدة لطريقة العنقدة بواسطة الـ (k− متوسط). لذلك سنلخصها في الفقرة التالية:

7−7−3 : طرائق التجزئة السريعة Quick partition :

لقد لاحظنا أن طريقة (k− متوسط) تحتاج إلى تحديد العدد k وإجراء تجزئة أولية لمفردات العينة، وخلال التطبيقات ظهرت عدة أشكال سريعة لطريقة (k− متوسط)، فبعضها يبدأ بتحديد مراكز العناقيد الأولية (وأطلق عليها اسم البذور seeds أو النقاط النووية). وبعضها يعتبر بعض هذه المفردات بذوراً لتشكيل العناقيد الأخرى، ويتم تنسيب المفردات الأخرى إلى العناقيد الممكنة حسب أقرب مسافة لها عن أحد تلك البذور. وإن أكثر طرائق التجزئة السريعة تطبيقاً هي:

1− الاختيار العشوائي لـk مفردة: بما أننا نرغب في الحصول على k شعاعاً مختلفاً (للمتوسطات)، فإنه من الأفضل أن نسحبها عشوائياً وبدون إعادة من البيانات الكلية . فنسحب العنصر الأول من العينة ذات الـ n مفردة، والثاني من $(n-1)$ مفردة، والثالث من $(n-2)$ مفردة، .... حتى نحصل على k مفردة ثم نعتبر هذه المفردات عناقيد أولية ونحسب مراكزها فنحصل على k مركزاً للعناقيد الأولية. ثم نقوم بعملية تنسيب المفردات الأخرى إلى تلك العناقيد حسب قربها من مراكز تلك العناقيد، ثم نحسب متوسطات تلك العناقيد فنحصل على k شعاعاً (بذرة) منها .

2- تحديد متحول التقسيم (Variable Division): نختار أحد المتحولات الكمية X لاستخدامه في تقسيم المفردات إلى k مجموعة كما يلي:
نقسم مجال تحول X إلى k مجالاً متساوياً متصلاً وغير متقاطع، ثم نقوم بوضع رموز مرتبة لهذه المجالات الجزئية من $j: 1\ 2\ 3 \dots k$، وبعدها نقوم بتوزيع المفردات حسب إحداثياتها على المجالات الجزئية لـX . فإذا وقعت المفردة في المجال j نعتبرها تنتمي إلى المجموعة J التي يتشكل منها العنقودj ، ثم نحسب متوسطات هذه العناقيد فنحصل على k شعاعاً (k بذرة) منها.

وحتى نفحص استقرار عملية العنقدة نقوم بإعادة التجارب بإجراء تجزئة جديدة. فعندما تكون العناقيد محددة، فإن التفسير البديهي لظهورها يساعد على إعادة ترتيب قائمة المفردات كما كانت، فالعنقود الأول يظهر أولاً والثاني يظهر ثانياً ...الخ . وإن جدول مراكز العناقيد (متوسطاتها) وتباينات العناقيد الداخلية تساعد على رسم صورة عن الفروقات بين المجموعات المتعنقدة .

4-7-7 : طريقة النماذج المختلطة (Mixture Models)

تعتمد هذه الطريقة على افتراض أن مفردات المجموعات المختلفة G_j في المجتمع تخضع لتوزيعات احتمالية مختلفة $f(x, Q_j)$، وإن هذه التوزيعات يمكن أن تكون من نفس العائلة، ولكن بمعالم مختلفة (متوسط وتباين)، ولنمذجة الآثار المختلفة لها (باعتبارها إشارة أو ضجة) نقوم بمقارنة مجاميع التراكيب المختلفة للتوزيعات الاحتمالية. وبذلك يمكننا توصيف المجتمع بواسطة التوزيع المختلط لهذه المجموعات والذي يعرف بالعلاقة:

$$P(x) = \sum_{j=1}^{g} P_j * f(x, \theta_j) \qquad (36-7)$$

حيث أن: P_j هو الاحتمال السابق للمجموعة G_i (وإن $\sum_{j=1}^{g} P_j = 1$)، وإن $f(x, \theta_j)$ هي التوزيعات الاحتمالية لمفردات المجموعة G_i في المجتمع والمتعلقة بشعاع المعالم θ_j .

وهنا نلاحظ أنه يوجد ثلاث أنواع من المعالم تحتاج للتقدير هي:

- القيم العددية للاحتمالات السابقة P_j .
- مركبات شعاع المعالم في التوزيعات θ_i (المتوسط والتباين) .
- قيمة العدد g الذي هو عدد المجموعات في المجتمع .

وهناك أساليب متعددة لتقدير هذه المعالم. وذلك حسب صيغ التوزيعات الاحتمالية المختلطة للمجموعات. فإذا كانت التوزيعات المختلطة توصف متحولات مستمرة، فإن التوزيع الطبيعي المختلط لها يأخذ الشكل التالي:

$$P(x) = \sum_{j=1}^{g} P_j * f(x, \boldsymbol{\mu_j}, \boldsymbol{V_j}) \qquad (37-7)$$

حيث أن $\boldsymbol{\mu_j}$ و $\boldsymbol{V_j}$ هما شعاع المتوسط ومصفوفة التباين المشترك للتوزيع الطبيعي المتعدد، والذي يساوي في الفضاء R^P ما يلي:

$$f(x, \boldsymbol{\mu_j}, \boldsymbol{V_j}) = \frac{1}{(2\pi)^{\frac{p}{2}} * |V_j|^{\frac{1}{2}}} * eXP\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu_j})' * \boldsymbol{V_j^{-1}} * (\boldsymbol{X}-\boldsymbol{\mu_j})\right] \quad (38-7)$$

أما بالنسبة للتوزيعات الثنائية [التي تأخذ 1أو 0]، فإن التوزيع المختلط لها يأخذ الشكل التالي:

$$P(x) = \sum_{j=1}^{g} P_j * f(x, \theta_j) \qquad (39-7)$$

حيث أن $f(x, \theta_i)$ هو التوزيع الاحتمالي المشترك للمتحول X الخاضع لتوزيع (برنوييللي) ويعرف في الفضاء R^P بالعلاقة:

$$f(x, \theta_j) = \prod_{\ell=1}^{P} \theta_{j\ell}^{X\ell}(1-\theta_{j\ell})^{1-X\ell} \qquad (40-7)$$

حيث أن المعلمة $\theta_{j\ell}$ هي احتمال أن يأخذ المتحول $X\ell$ في المجموعة j القيمة (1) .

وإن مسألة تقدير هذه المعالم بواسطة طريقة الامكانية العظمى أو غيرها معروضة في كثير من المراجع المختصة مثل [Hand 1981] . ولكن علينا منذ البداية تحديد صيغة التوزيع $f(x, \theta_i)$، وبعدها يمكن أن نفترض أي عدد لـg .

وعندما تصبح معالم النموذج مقدرة وتصبح المفردات مفروزة إلى مجموعات (g مجموعة). وذلك بناء على تقديرات الاحتمالات اللاحقة لعناصر المجموعة. فإن قاعدة التصنيف تصبح كما يلي:

القاعدة: ننسب المفردة x إلى المجموعة k إذا كان :

$$P_k\, f(x, \theta_k) \le P_j\, f(x, \theta_j) \qquad (41-7)$$

حيث أن: $k \neq j: 1\ 2\ 3 \ldots .\ g$

ولكن المشكلة الأساسية في تطبيق هذه الطريقة تنحصر في تحديد عدد المجموعات g . ولكن معظم الخوارزميات تطلب تحديد العدد g قبل تقدير بقية المعالم .

وهناك مشكلة أخرى هي أن عدداً من القيم الصغرى لتابع الامكانية العظمى والمجسمات الداخلية المتعددة، يمكن أن تحدث قبل أن تتحقق العنقدة المثالية . وعندها علينا أن نقوم بإجراء عدة دورات لتحقيق العنقدة المثالية، وكلما كانت نتائج التصنيف متوافقة كانت الحلول صحيحة .

7-7-5 : طريقة النماذج اللاخطية:

تعتمد هذه الطريقة على المعيار (2) الذي يعتمد على أثر مصفوفة مجموع مربعات الانحرافات داخل المجموعات (S_W)، والذي يمكن كتابته على الشكل التالي [من Webb P.379 بتصرف]:

$$Tr(S_W) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{g} Z_{ji} \sum_{k=1}^{P}(x_{ik} - m_{jk})^2 \qquad (42-7)$$

حيث أن: x_{ik} هي الاحداثي k للمفردة أو النقطة i . وأن $(i: 1\ 2\ 3 \ldots n)(k: 1\ 2\ 3 \ldots P)$.

وأن m_{jk} هي الإحداثي k لمتوسط المجموعة j ، وأن $(Z_{ji} = 1)$ إذا كانت المفردة أو النقطة i ، تنتمي إلى المجموعة j ، وأن $(Z_{ji} = 0)$ إذا كانت المفردة i لا تنتمي إلى المجموعة j .

وإن إحداثيات متوسطات المجموعات m_{jk} تحسب من العلاقة التالية:

$$m_{jk} = \frac{\sum_{i=1}^{n} Z_{ji} x_{ik}}{\sum_{i=1}^{n} Z_{ji}} \qquad (43-7)$$

وحتى نحصل على تجزئة مثالية للبيانات، يجب علينا أن نحدد قيم المتحول Z_{ji} الذي يأخذ إحدى القيمتين (1 أو 0)، والذي يجعل الأثر المعرف في (7−42) أصغر ما يمكن .

لنرمز لمصفوفة العناصر (Z_{ji}) بالرمز Z ونكتبها كما يلي:

$$Z_{g*n} = \begin{array}{c c} j\ i: & \begin{matrix} 1 & 2 & 3 & \cdots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ g \end{matrix} & \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \ldots & Z_{1n} \\ Z_{21} & Z_{22} & Z_{23} & \ldots & Z_{2n} \\ Z_{31} & Z_{32} & Z_{33} & \ldots & Z_{3n} \\ \vdots & \vdots & \vdots & \vdots & \ldots \\ Z_{g1} & Z_{g2} & Z_{g3} & \ldots & Z_{gn} \end{bmatrix} \end{array} \qquad (44-7)$$

هي عبارة عن مصفوفة مؤلفة من العناصر Z_{ji} التي تساوي (1 أو 0) ومن أهم خواصها أن عناصرها تحقق الخاصتين التاليتين:

$$\sum_{j=1}^{g} Z_{ji} = 1 \qquad , \quad Z_{ji} \geq 0 \qquad (45-7)$$

أي أن كل عمود فيها لا يتضمن إلا عنصراً واحداً يساوي الواحد والباقي تكون أصفاراً. وعلينا الآن تحديد أي هذه العناصر سيكون مساوياً للواحد لتحديد انتماء العنصر i إلى المجموعة j .

ولتصغير المقدار $Tr(S_W)$ المعرف بالعلاقة (7−42)، علينا الأخذ بعين الاعتبار العناصر (Z_{ji}) والتي تخضع للشرط $\left(\sum_{j=1}^{g} Z_{ji} = 1\right)$، وعندها نحصل على حل نهائي لعناصر المصفوفة Z والمؤلفة من العنصرين (1) و(0). وبعدها نحصل على تجزئة للبيانات التي تصغر المقدار (7−42) مع وجود الشرط (7−45) عليه، وتُفرز المفردات إلى المجموعات حسب قيم العناصر (Z_{ji})، وهكذا يكون المتوسط m_{jk} غير مساوٍ لمتوسط المجموعة j حتى تتحقق عملية التقارب بينهما خلال المعاودة .

ملاحظة: يمكن تحويل هذه المسألة إلى عملية غير مشروطة، وذلك من خلال تحويل العناصر Z_{ji} كما يلي :

$$Z_{ji} = \frac{eXP(\nu_{ji})}{\sum_{j=1}^{g} eXP(\nu_{ji})} \quad : \quad \begin{bmatrix} j = 1\ 2\ 3 \ldots g \\ i = 1\ 2\ 3 \ldots n \end{bmatrix} \qquad (46-7)$$

وهذا التحويل يجعل الأثر $Tr(S_W)$ يأخذ شكل تابع غير خطي لمعالم (ν_{ji})، ويصبح هدفنا تصغير التابع $Tr[S_W(\nu_{ji})]$ وحساب قيمته الصغرى ثم حساب Z_{ji} ، وهناك صيغ أخرى لتحويل المسألة إلى

عملية تدرج غير مشروطة، وهي أن تأخذ شكلاً مبسطاً للمقدار $Tr[S_w(v)]$ وعندها نكتب مشتقة الجزئي بالنسبة (v_{ab}) كما يلي:

$$\frac{\partial Tr[S_w(v)]}{\partial v_{ab}} = \frac{1}{n}\sum_{j=1}^{g} Z_{jb}(\sigma_{ja} - Z_{ab})(X_b - \mu_j)^2 \qquad (47-7)$$

حيث أن: σ_{ja} هو رمز (كرونكر) وهو يساوي (0) إذا كانت $(j \neq a)$ ويساوي (1) إذا كانت $(j = a)$.
وبإجراء الحسابات اللازمة لحل المعادلات (7–47) نحصل على قيم العناصر (v_{ji}) ومنها نحسب Z_{ji} ونحدد انتماء العناصر إلى المجموعات على التوالي .
وهناك طريقة أخرى لحساب (v_{ji}) وهي أن نختار العناصر Z_{ji} عشوائياً حسب التوزيع المنتظم في المجال $[1,1+a]$، ثم القيام بمعيرتها ليكون مجموعها مساوياً للواحد . وبعدها نحسب v_{ji} من العلاقة:

$$v_{ji} = \log(Z_{ji})$$

6–7–7 : طريقة (k– متوسط) الغامضة: (Fuzzy k– means)

تعتمد هذه الطريقة على افتراض أن عناصر الأصول (Paterns) يمكن أن تنتمي إلى كل العناقيد وبدرجات مختلفة لعضويتها . وتسعى خوارزمية هذه الطريقة إلى إيجاد حل للمعالم الوسيطة (y_{ji}) التي تصغر المقدار التالي [من Webb. P.380 بتصرف]:

$$J_r = \sum_{i=1}^{n}\sum_{j=1}^{g} y_{ji}^{r}(x_i - m_j)^2 \qquad (48-7)$$

حيث أن: $j: 1\ 2\ 3 \ldots g$ و $i: 1\ 2\ 3 \ldots n$

وبشرط أن يكون:

$$\sum_{j=1}^{g} y_{ji} = 1 \quad , \quad y_{ji} \geq 0 \qquad (49-7)$$

وحيث أن المعلم الوسيط y_{ji} هو متحول مستمر يعبر عن درجة الارتباط أو الاقتران، ويسمى بتابع العضوية للأصل أو المفردة i في المجموعة j . وإن الأس r في العلاقة (7–48) هو عدد سلمي يرمز لوزن أو لدرجة التشدد في العناقيد الناتجة. علماً بأن $(r \geq 1)$، أما m_j فهو مركز المجموعة j ويحسب من العلاقة المتشددة التالية :

$$m_j = \frac{\sum_{i=1}^{n} y_{ji}^{r} x_i}{\sum_{i=1}^{n} y_{ji}^{r}} \qquad (50-7)$$

وعندما تكون $(r = 1)$ فإن ذلك يعطينا حالة الطريقة اللاخطية، وإنه في هذه الحالة نجد أن عملية تصغير العلاقة (7–48) تعطينا أن قيم y_{ji} تساوي أما (0) أو (1) .
إن الخوارزمية الأساسية لهذه الطريقة تتلخص بما يلي:

1- نختار r بحيث يكون $1 < r < \infty$، اللازم لحساب قيم تابع العضوية y_{ji} (حيث أن: $j = 1\ 2\ 3 \dots g$ وأن $i = 1\ 2\ 3 \dots n$) .

2- نحسب مراكز العناقيد m_j (حيث $j = 1\ 2\ 3 \dots g$) وفق العلاقة (7-50) .

3- نحسب المسافات d_{ji} من العلاقة : $d_{ji} = |x_i - m_j|$ وهو مقياس القيمة المطلقة للمسافة بين النقطتين x_i و m_j .

4- نحسب تابع العضوية كما يلي:

إذا كانت $d_{\ell i} = 0$ من أجل بعض العناقيد ℓ . فإن قيم تابع العضوية $y_{\ell i} = 1$ وإن المفردة i تنتمي إلى العنقود ℓ . وعندها فإن $y_{\ell i} = 0$ إذا كانت $\ell \neq i$.

أما إذا كانت $d_{\ell i} \neq 0$ فإن $y_{\ell i} \neq 1$ ونحسبه من العلاقة التالية:

$$y_{\ell i} = \frac{1}{\sum_{j=1}^{g} \left(\frac{d_{\ell i}}{d_{ji}}\right)^{\frac{2}{r-1}}} \qquad (51-7)$$

وهو يعبر عن درجة عضوية المفردة *i* في العنقود ℓ .

5- إذا لم يتحقق التقارب نعود إلى الخطوة (2) ونكرر ذلك حتى يتحقق التقارب .

ملاحظة: عندما يقترب أو يسعى r إلى الواحد $(r \to 1)$ فإن الخوارزمية تتحول إلى خوارزمية الـ (k-متوسط) العادية .

وهناك عدة قواعد للتوقف عن عمليات المعاودة، وأهمها هو: أن نراقب التغيرات النسبية في قيم المراكز حتى تصبح تلك التغيرات مهملة (أو أقل من مقدار محدد من الباحث)، وإن هذا يحدث عندما تتحقق المعادلة التالية:

$$D_s^{\Delta} = \left\{\sum_{j=1}^{g} \left[m_j(k) - m_j(k-1)\right]^2\right\}^{\frac{1}{2}} < \varepsilon \qquad (52-7)$$

حيث أن $m_j(k)$ هو قيمة المتوسط j في الدورة k، وأن ε هو العدد الذي يحدده الباحث ليكون الحد الأعلى للخطأ، وهناك قاعدة أخرى لتوقيف عمليات المعاودة تستند على التغيرات في قيم تابع العضوية y_{ji} أو في تابع التكاليف J_j، ويشترط أن يتحقق فيها الشرط التالي:

$$\max_{1 \le i \le n} \propto_i < \varepsilon \qquad (53-7)$$

حيث أن $\propto_i$ تحسب من العلاقة :

$$\propto_i = \max_{1 \le j \le g} y_{ji}^{r-1} \left|x_i - m_j\right|^2 - \max_{1 \le j \le g} y_{ji} \left|x_i - m_j\right|^2 \qquad (54-7)$$

وحيث أن أصغر قيمة مكانية لها هي: $\propto_i = 0$ وأن: $i = 1\ 2\ 3 \dots n$

الخلاصة :

1- إن التحليل العنقودي الهرمي يتألف من عدة طرائق تهدف إلى تجميع المشاهدات (المفردات) أو المتحولات (أو الصفات) إلى عناقيد متجانسة داخلياً ومتباينة خارجياً، وذلك اعتماداً على مقاييس معينة للتقارب .

2- إذا كان التجميع يعتمد على المتحولات الكمية فإن مقاييس التقارب (التشابه) بين المشاهدات هو مجموع مربعات المسافات الاقليدية، أو مسافات المقاطع، وهي تطبق على المتحولات الكمية ذات الواحدات الموحدة أو على البيانات الممعيرة لها أو المثقلة منها .

أما إذا كان التجميع يعتمد على المتحولات النوعية. فإن مقاييس التقارب بين أي مشاهدتين هو نسبة الأزواج المتشابهة على عدد المتحولات المستخدمة. وهناك مقاييس أخرى .

3- إن طريقة الجوار الأقرب (الربط المنفرد) أو طريقة الجوار الأبعد (الربط التام) أو طريقة الربط المتوسط . وهي أمثلة عن أساليب التجميع في التحليل الهرمي، وإن هذه الطرائق تبدأ من عدة عناقيد (عنقود لكل مشاهدة) وتنتهي بعقود واحد يحتوي كل المشاهدات . وإن جميع العناقيد التي تشكل بواسطة هذه الطرائق تنتج عن دمج العناقيد المشكلة سابقاً .

وهناك طريقة أخرى للعنقدة هي طريقة التقسيم أو التجزئة، وهي تبدأ من عنقود واحد يضم كل المشاهدات وتنتهي بعدة عناقيد (بعدد المشاهدات) .

وهناك طرائق أخرى للتحليل غير الهرمي مثل طريقة (k- متوسط) والتي تطبق عادة في حالات (العنقدة السريعة) وباستخدام بعض البرامج الاحصائية .

إن طرائق العنقدة تستطيع أيضاً تصنيف المتحولات المؤثرة X بشكل أفضل من تصنيفها للمشاهدات. وإن ذلك يطبق في حالة عمليات تصميم الاستبيانات، وإن طرائق هذا الأسلوب تعتمد غالباً على مصفوفة معاملات الارتباط الخطي بين المتحولات X .

المراجع الأساسية:

1- Anderson, T. W. (1984). An Introduction to Multivariate Analysis (2nd ed). New york John Wiley 1984.

2- Breiman, L. Friedman Rt. (1984). Classification and Regression Trees. Wardworth Intermatial Group Belmout. SA.

3- Johson A. Richard, Wichern. W. Dean (1988) Applied multivariate statistical Analysis sec. ed. Prentice Hall. Internatial, Inc. 1988.

4- Kendall M. G. Stuart, A. (1976) The Advanced theory of Statistics, charless Griffin. Com London, Russian Trans. Nauka, Moscow .

5- Kendall, M. (1975). Multivariate Analysis, Gniffin, London.

6- Friedman J.H. (1991) Multivaraite Adoptive Regression splines, Anauls of statistics.. 19(1): 1-141 .

7- Webb, Andrew, R, (2002) Statistical pattern Recignition, Quint Q,Ltd, Malven u.k 2002 .

8- Wilks's .(1967) Mathematical statistics, Russian Trans. Nauka, Moscow.

المراجع المنشورة في المجلات العلمية وعلى الشبكة الالكترونية:

1- رضا، صبح منفي: ذياب، وسام سرحان (2008)، استخدام بعض الطرائق الاحصائية والتصنيف الشجيري في التصنيف والتنبؤ بإفلاس الشركات مالياً. مجلة العلوم الاقتصادية والإدارية- المجلد14 العدد 94- جامعة المستنصرية- البصرة- العراق .

2- شاهين، حمزة اسماعيل (2014). دراسة مقارنة لبعض طرائق التصنيف الخطية مع تطبيق عملي. مجلة العلوم الاقتصادية والإدارية، المجلد2- العدد 80- لعام 2014- جامعة المستنصرية- العراق.

3- العلي، إبراهيم محمد، علوش، ياسر محمد. (2018)، بناء نموذج رياضي للتنبؤ بتصنيف المشروعات الصغيرة باستخدام التحليل التمييزي في بعض المحافظات السورية- مجلة جامعة تشرين- العدد (2)، المجلد (49)، لعام 2018 . سلسلة العلوم الاقتصادية والقانونية .

4- علي، عمر عبد المحسن، أحمد، سهاد أحمد (2016). الأسلوب البيزي في تصنيف الانحدار الشجيري لتقدير نموذج تجميعي ومقارنته بالنموذج اللوجستي مع التطبيق. مجلة الإدارة والاقتصاد- المجلد 30- العدد 109.

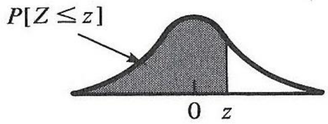
5- Mihaela, S. (2017)- classification and Regression Tree- Department of Engineering science, University of Oxfort .

6- Tryfos, P. (2001) cluster Analysis, chapter 15 .

7- Wikipedia (2018), Logistic Analysis- aspects .

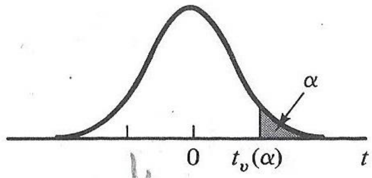
المـلاحـق: الجداول الاحصائية

TABLE 1 STANDARD NORMAL PROBABILITIES



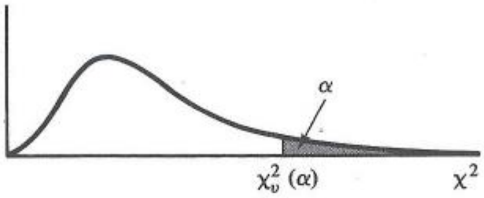
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

TABLE 2 STUDENT'S *t*-DISTRIBUTION CRITICAL POINTS



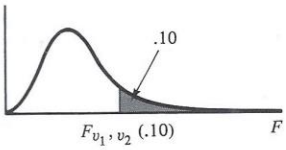
d.f.	α							
ν	.250	.100	.050	.025	.010	.00833	.00625	.005
1	1.000	3.078	6.314	12.706	31.821	38.190	50.923	63.657
2	.816	1.886	2.920	4.303	6.965	7.649	8.860	9.925
3	.765	1.638	2.353	3.182	4.541	4.857	5.392	5.841
4	.741	1.533	2.132	2.776	3.747	3.961	4.315	4.604
5	.727	1.476	2.015	2.571	3.365	3.534	3.810	4.032
6	.718	1.440	1.943	2.447	3.143	3.287	3.521	3.707
7	.711	1.415	1.895	2.365	2.998	3.128	3.335	3.499
8	.706	1.397	1.860	2.306	2.896	3.016	3.206	3.355
9	.703	1.383	1.833	2.262	2.821	2.933	3.111	3.250
10	.700	1.372	1.812	2.228	2.764	2.870	3.038	3.169
11	.697	1.363	1.796	2.201	2.718	2.820	2.981	3.106
12	.695	1.356	1.782	2.179	2.681	2.779	2.934	3.055
13	.694	1.350	1.771	2.160	2.650	2.746	2.896	3.012
14	.692	1.345	1.761	2.145	2.624	2.718	2.864	2.977
15	.691	1.341	1.753	2.131	2.602	2.694	2.837	2.947
16	.690	1.337	1.746	2.120	2.583	2.673	2.813	2.921
17	.689	1.333	1.740	2.110	2.567	2.655	2.793	2.898
18	.688	1.330	1.734	2.101	2.552	2.639	2.775	2.878
19	.688	1.328	1.729	2.093	2.539	2.625	2.759	2.861
20	.687	1.325	1.725	2.086	2.528	2.613	2.744	2.845
21	.686	1.323	1.721	2.080	2.518	2.601	2.732	2.831
22	.686	1.321	1.717	2.074	2.508	2.591	2.720	2.819
23	.685	1.319	1.714	2.069	2.500	2.582	2.710	2.807
24	.685	1.318	1.711	2.064	2.492	2.574	2.700	2.797
25	.684	1.316	1.708	2.060	2.485	2.566	2.692	2.787
26	.684	1.315	1.706	2.056	2.479	2.559	2.684	2.779
27	.684	1.314	1.703	2.052	2.473	2.552	2.676	2.771
28	.683	1.313	1.701	2.048	2.467	2.546	2.669	2.763
29	.683	1.311	1.699	2.045	2.462	2.541	2.663	2.756
30	.683	1.310	1.697	2.042	2.457	2.536	2.657	2.750
40	.681	1.303	1.684	2.021	2.423	2.499	2.616	2.704
60	.679	1.296	1.671	2.000	2.390	2.463	2.575	2.660
120	.677	1.289	1.658	1.980	2.358	2.428	2.536	2.617
∞	.674	1.282	1.645	1.960	2.326	2.394	2.498	2.576

TABLE 3 χ^2 CRITICAL POINTS



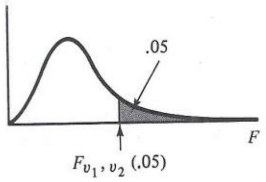
d.f. ν	.990	.950	.900	α .500	.100	.050	.025	.010	.005
1	.0002	.004	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.02	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.11	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.30	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.55	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.87	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	1.24	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.65	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.95
9	2.09	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.56	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	3.05	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.57	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	4.11	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.66	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	5.23	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.81	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	6.41	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	7.01	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	7.63	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	8.26	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.90	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	9.54	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	10.20	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	10.86	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	11.52	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	12.20	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	12.88	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	13.56	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	14.26	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	14.95	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	22.16	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	29.71	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	37.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	45.44	51.74	55.33	69.33	85.53	90.53	95.02	100.43	104.21
80	53.54	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	61.75	69.13	73.29	89.33	107.57	113.15	118.14	124.12	128.30
100	70.06	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

TABLE 4 F-DISTRIBUTION CRITICAL POINTS ($\alpha = .10$)



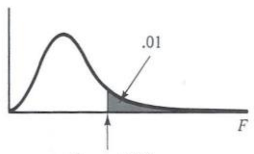
v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.79
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.76
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.51
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.21
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.11
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.96
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.86
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.82
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.78
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.75
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.80	1.78	1.75	1.72
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.78	1.76	1.73	1.70
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.76	1.74	1.71	1.68
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.74	1.72	1.69	1.66
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.71	1.69	1.66	1.62
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.68	1.66	1.63	1.59
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.67	1.65	1.61	1.58
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.66	1.64	1.60	1.57
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.65	1.63	1.59	1.56
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.64	1.62	1.58	1.55
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.63	1.61	1.57	1.54
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.50	1.48	1.44	1.40
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24

TABLE 5 F-DISTRIBUTION CRITICAL POINTS ($\alpha = .05$)



v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.3	250.1	251.1	252.2
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.38
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.80
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.79
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.75
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.43
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.51	1.46	1.39	1.32

TABLE 6 F-DISTRIBUTION CRITICAL POINTS ($\alpha = .01$)



$F_{v_1, v_2}(.01)$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6023.	6056.	6106.	6157.	6209.	6240.	6261.	6287.	6313.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.32
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.65
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.20
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.06
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.82
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.03
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.48
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.08
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.78
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.54
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.34
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.18
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.05
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.93
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.83
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.75
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.67
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.61
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.55
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.50
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.45
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.40
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.36
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.57	2.50	2.42	2.33
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.54	2.47	2.38	2.29
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.51	2.44	2.35	2.26
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.48	2.41	2.33	2.23
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.21
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.02
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.84
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.93	1.86	1.76	1.66
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.78	1.70	1.59	1.47