# الانحدار والارتباط

## تحليل الانحدار

يعتبر الانحدار من المواضيع الأساسية في النظرية الإحصائية بشكل عام، حيث يستخدم على نطاق واسع في العديد من العلوم الطبيعية او الإدارية أو الاقتصادية، وذلك لاستكشاف القوانين التي تربط الظواهر فيما بينها، ثم وضع هذه القوانين في شكل نماذج رياضية يطلق عليها "**نماذج الانحدار**" تُسهّل عملية تقدير قيم هذه الظواهر بدلالة قيم الظواهر الأخرى.

ويعد العالم الإحصائي الانجليزي Francis Galton أول من استخدم مفهوم الانحدار في التطبيقات البيولوجية تحديدا، وذلك لاستكشاف بعض العلاقات التي تربط بعض المتغيرات البيولوجية.

يطلق على الظاهرة المراد تقدير قيمتها -وبالتالي معرفه سلوكها- **بالمتغير التابع** او **المتغير المفسَّر**، ويرمز لها عادة بالرمز $y$ ، وبالمقابل يطلق على المتغير الآخر أو بقية المتغيرات الأخرى المتحكمة فيها **بالمتغير المستقل** او **المتغير المفسِّر**، ويرمز له عادة بالرمز $x$.

وبشكل عام تعطينا دراسة الانحدار إجابة عن سؤالين رئيسين:

- ✓ **الأول**: هل توجد علاقة بين المتغيرين المدروسين؟
- ✓ **الثاني**: إذا وجدت علاقة بين هذين المتغيرين او أكثر، فما هو الشكل الرياضي لهذه العلاقة؟

يمكن تقسيم الانحدار من حيث شكلُ العلاقة الى قسمين مهمين: خطي وغير خطي.

أما الانحدار الخطي فتكون فيه العلاقة خطية بين الظاهرة المدروسة من جهة والمتغيرات المتحكمة فيها من جهة أخرى، ومعنى ذلك ان اي تغير يحصلُ في المتغير المستقل يتبعه مباشرة تغير ثابت في المتغير التابع.

وأما الانحدار غير الخطي فتكون فيه العلاقة غير خطية بين الظاهرة المدروسة والمتغيرات المؤثرة فيها كأن تكون علاقة أسية او علاقة قطع مكافئ... الخ. وسنركز في محاضرتنا على القسم الاول من الانحدار.

أ. **الانحدار الخطي البسيط :** إن الهدف الرئيس من تحليل الانحدار الخطي عموما هو تقدير العلاقة الرياضية الخطية التي تربط بين متغيرين اثنين أو أكثر، وبهذا الصدد يمكن أن نجد نوعين من الانحدار الخطي: بسيط ومتعدد.

اما **الانحدار الخطي البسيط** فهو الذي يدرس العلاقة بين متغيرين اثنين فقط، أحدهما تابع والآخر مستقل، واما **الانحدار الخطي المتعدد** فهو الذي يدرس العلاقة الخطية بين عدد من المتغيرات، أحدها تابع والبقية متغيرات مستقلة. وفي هذه المحاضرة سنصب اهتمامنا على الانحدار الخطي البسيط، أين نقوم من خلاله بتقدير العلاقة الخطية بين متغيرين فقط -كما ذكرنا- أحدهما مستقل والآخر تابع. يمكن تلخيص النموذج الرياضي الخطي للانحدار الخطي البسيط في المعادلة الأتية :

$$\hat{y} = ax + b$$

حيث:

$\hat{y}$ القيم التقديرية للمتغير التابع Y.

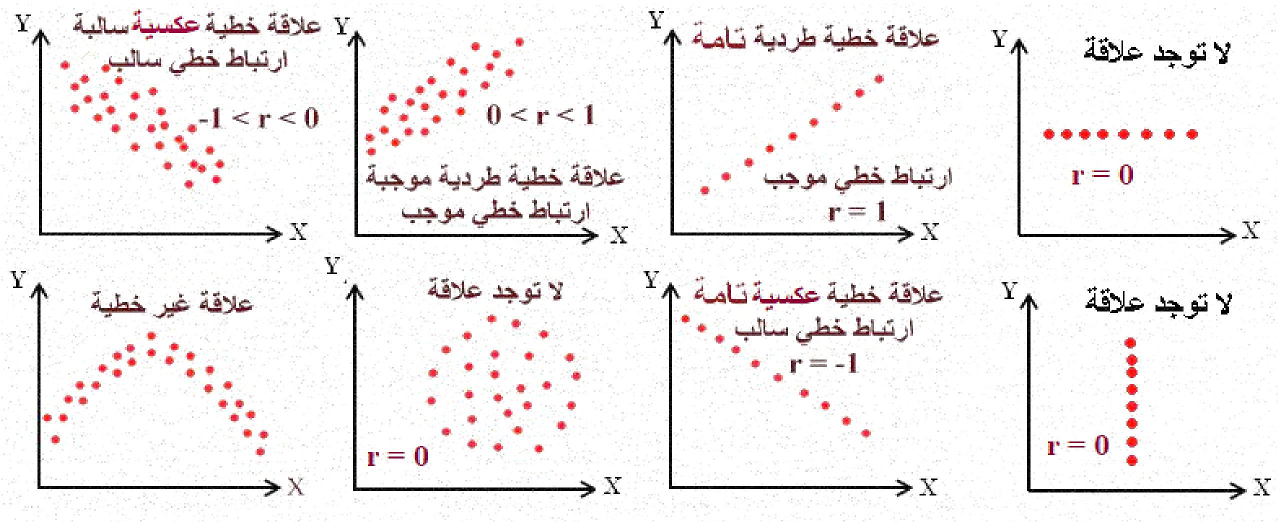
$x$ المتغير المستقل.

$a$ و $b$ ثوابت او معالم النموذج، وهي الثوابت او المعالم الواجب تحديد قيمها ليتكشف لنا النموذج كاملا.

**طرق تقدير معالم نموذج الانحدار الخطي البسيط:** إن أول خطوة يعمد إليها الباحث لتحديد معلمتي نموذج الانحدار هي رسمه لما يسمى **بلوحة الانتشار** أو **سحابة النقاط.** والتي تنتج برسم مختلف الثنائيات (x , y) في معلم متعامد، يوضع في محوره الأفقي المتغير المستقل $x$ وفي محور العمودي المتغير التابع $y$، وبحسب شكل هذه السحابة يمكن ان نحدد طبيعة العلاقة بين المتغيرين المستقل والتابع.

فإذا ظهر شكل لوحة الانتشار يقتربُ من تشكيل **خط مستقيم** نقول إن العلاقة خطية بين المتغيرين، وإذا اخذت السحابة شكلا آخر غير الشكل المستقيم، كأن تكون أسيةً مثلا، فنقول إن العلاقة بين المتغيرين المدروسين أسية، وإذا اخذت **شكلا عشوائيا** لا يشبه أي شكل رياضي معروف، نقول إنه لا توجد علاقة أصلا بين المتغيرين المدروسين. (أنظر الشكل 01 أسفله)

**الشكل رقم 01: سحابة النقاط وأشكال العلاقة بين المتغيرين X و Y.**




المصدر: الرابط: https://www.almuheet.net/post/247336

بالنسبة للانحدار الخطي البسيط فإنه يُنتظر -بعد رسم لوحة الانتشار- أن نحصُل على سحابة تقترب من تشكيل خط مستقيم، ويتوقف مدى انتشار وتشتت هذه السحابة على قوة العلاقة بين المتغيرين؛ حيث كلما قويت هذه العلاقة قل تشتت هذه السحابة، واقتربت من الوقوع على خط مستقيم، وتعتبر حينها العلاقة تامة أو وظيفية، والعكس صحيح، والأشكال السابقة تبين ذلك بوضوح.

لكن السؤال المطروح هنا: كيف نحدد بدقة الخط المستقيم الذي تقتربُ السحابةُ من تشكيله، وهذا لتحديد معلمتي نموذج الانحدار $a$ و $b$ السابقتين؟ للإجابة عن هذا السؤال هناك طريقتان:

**الطريقة الأولى - الطريقة البيانية**: أو التحديد بالنظر؛ حيث يقوم الباحث بوضع أقرب خط يراه مناسبا لتمثيل السحابة، وهو الخط الأقرب الى جميع النقاط كما يبدو له. لكن يعاب على هذه الطريقة أنها غير دقيقة، حيث انه يمكن لعدد من الاشخاص ان يرسم خطوطا مختلفة على السحابة نفسها، بل إن الشخص نفسه يمكن ان يرسم عدة خطوط في مناسبات مختلفة.

**الطريقة الثانية - طريقه المربعات الصغرى**:

وهي أشهر وأدق طريقة لتحديد معالم نموذج الانحدار، تقوم فكرتها الأساسية على جعل "المسافات" او "الانحرافات" بين سحابة النقاط (القيم الحقيقية للمتغير التابع y) وبين خط الانحدار (القيم التقديرية للمتغير التابع $\hat{y}$) أقل ما يمكن.

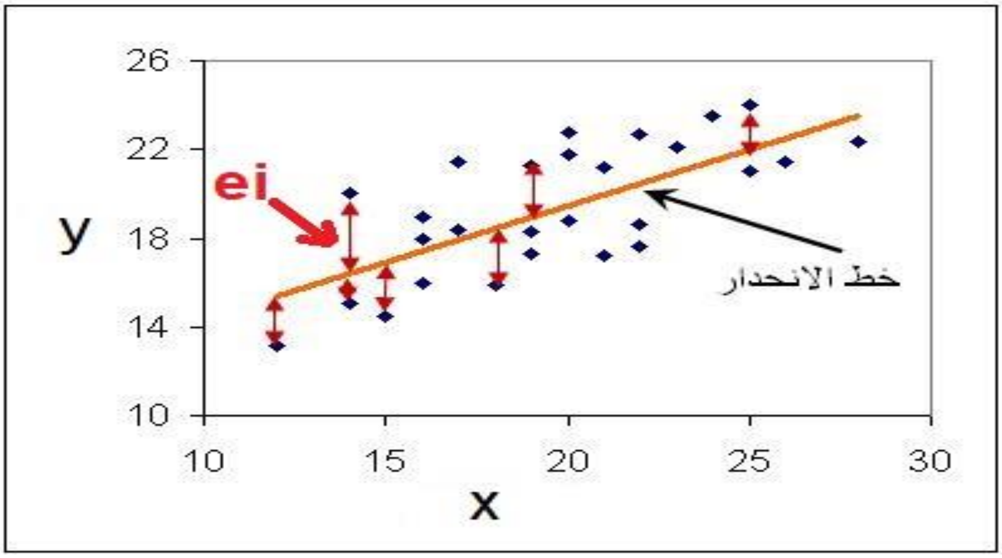
ولأن أفضل خط نبحث عنه يؤدي دورا في سحابة النقاط يشبه دور الوسط الحسابي في مجموعة القيم، فإن مجموع انحرافات القيم الحقيقية $y$ في السحابة عن القيم المقدرة $\hat{y}$ الواقعة على الخط يساوي "الصفر" حسب الخاصية رقم 01 من خصائص الوسط الحسابي التي رأيناها سابقا.

نرمز لهذه الانحرافات بالرمز $e_i$ حيث عند $\sum e_i = \sum(y_i - \hat{y}_i) = 0$

ومعنى هذا أن مجموع "مربعات" هذه الانحرافات يجب ان يكون أصغر ما يمكن عند أفضل خط نبحث عنه، وهذا حسب الخاصية رقم 02 من خصائص الوسط الحسابي التي رأيناها سابقا، وهذا هو سبب تسميتها بطريقة المربعات الصغرى، حيث أن مجموع $e_i^2$ هو دالة في المعلمتين $a$ و $b$. (أنظر الشكل 02 أسفله)

$$\sum e_i^2 = \sum(y_i - \hat{y}_i)^2 = \sum(y_i - ax - b)^2 = f(a,b)$$

الشكل رقم **02**: توضح انحراف القيمة الحقيقية عن القيمة المقدَّرة.




المصدر: https://samehar.wordpress.com/2009/11/11/120909/

لتكون الدالة $f$ أصغر ما يمكن يجب أن يتوافر فيها شرطان: الأول لازم والثاني كافٍ.

- ✓ **الشرط اللازم**: يجب أن تنعدم المشتقتان الجزئيتان الأوليان بالنسبة للمعلمتين $a$ و $b$.
- ✓ **الشرط الكافي**: أن تكون المشتقتان الجزئيتان الثانيتان بالنسبة للمعلمتين $a$ و $b$ موجبتين.

الشرط اللازم:

- ➢ نشتق $f$ بالنسبة للمعلمة $a$ ونضعها تساوي الصفر:

$$\left(\frac{\delta \sum e_i^2}{\delta a} = 0\right) \Leftrightarrow \left[-2\sum (y_i - ax_i - b)x_i = 0\right] \Leftrightarrow \left[\sum (y_i - ax_i - b)x_i = 0\right]$$

$$\Leftrightarrow \left[\sum (x_i y_i - a{x_i}^2 - bx_i) = 0\right] \Leftrightarrow \left(\sum x_i y_i - a\sum {x_i}^2 - b\sum x_i\right) = 0$$

$$\Leftrightarrow \sum x_i y_i = a\sum {x_i}^2 + b\sum x_i \ldots\ldots\ldots\ldots\ldots (1)$$

- ➢ نشتق $f$ بالنسبة للمعلمة $b$ ونضعها تساوي الصفر:

$$\left(\frac{\delta \sum e_i^2}{\delta b} = 0\right) \Leftrightarrow \left[-2\sum (y_i - ax_i - b) = 0\right] \Leftrightarrow \left[\sum (y_i - ax_i - b) = 0\right]$$

$$\Leftrightarrow \left(\sum y_i - a\sum x_i - nb\right) = 0 \Leftrightarrow \sum y_i = a\sum x_i + nb \ldots\ldots\ldots (2)$$

من (1) و (2) يمكن أن نكون جملة معادلتين ذات مجهولين $a$ و $b$ كما يلي:

$$\begin{cases} \sum x_i y_i = a\sum {x_i}^2 + b\sum x_i \\ \sum y_i = a\sum x_i + nb \end{cases}$$

وبحل هذه الجملة نخلص إلى أن:

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum {x_i}^2 - \sum x_i}$$

$$b = \frac{\sum {x_i}^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum {x_i}^2 - \sum x_i} = \bar{Y} - a\bar{X}$$

**مثال**: يبين الجدول التالي البينات الخاصة بعينة عشوائية من منتجي الأدوية، حيث $X$ تمثل مصاريف البحث، و $Y$ تمثل الأرباح (بمليون دينار):

| Y | X |
|---|---|
| 2 | 1 |
| 5 | 3 |
| 9 | 5 |
| 10 | 7 |
| 14 | 9 |
| **40** | **25** |

1. أرسم لوحة انتشار البيانات.
2. ما هو شكل العلاقة بين $X$ و $Y$؟
3. أوجد معادلة خط الانحدار.
4. ما ذا تعني قيمة الثابتين $a$ و $b$؟
5. حدد القيمة التقديرية للربح إذا كانت المصاريف 10 مليون دينار.

**الحل**:

1. رسم لوحة الانتشار (سحابة النقاط) يكون في معلم متعامد أين نحدد عليه الثنائيات $(x_i, y_i)$.
2. بملاحظة لوحة الانتشار نلاحظ أن النقاط تقترب من تشكيل خط مستقيم، وعليه فشكل العلاقة بين المتغيرين خطية شكلها الرياضي: $\hat{y} = ax + b$
3. تحديد المعادلة يعني تحديد القيمتين $a$ و $b$

| $X^2$ | XY | Y | X |
|---|---|---|---|
| 1 | 2 | 2 | 1 |
| 9 | 15 | 5 | 3 |
| 25 | 45 | 9 | 5 |
| 49 | 70 | 10 | 7 |
| 81 | 126 | 14 | 9 |
| **165** | **258** | **40** | **25** |

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum {x_i}^2 - \sum x_i} = \frac{5(258) - 25 \times 40}{5(165) - 25^2} = \mathbf{1.45}$$

$$b = \frac{\sum {x_i}^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum {x_i}^2 - \sum x_i} = \frac{165 \times 40 - 25 \times 258}{5(165) - 25^2} = \mathbf{0.75}$$

ومنه المعادلة كالتالي: $\hat{y} = 1.45x + 0.75$

4. تمثل $a$ الميل ومعناها مقدار الزيادة المتوقعة في الأرباح كلما صرفنا وحدة واحدة من المصاريف، أي بعبارة أخرى أن الأرباح يتوقع ان تزيد بمقدار 1450000 دج كلما زدنا مقدار المصاريف بمبلغ 1000000 دج.
أما $b$ فقيمتها تمثل مقدار الأرباح المتوقعة إذا لم نصرف ولا دينار في البحث والتطوير، يقدر هذا الربح بمبلغ 750000 دج. (لأن وحدة القياس في التمرين هي مليون دينار)
5. تحديد القيمة التقديرية للربح إذا كانت المصاريف 10 مليون دينار:

$$\hat{y} = 1.45(10) + 0.75 = 15.25$$

أي أننا لو صرفنا 10 مليون دينار كمصاريف بحث وتطوير فسنجني أرباحا متوقعة تقدر بمبلغ 15250000 دج.

(11)

## * الارتباط :

يهتم تحليل الارتباط بقوة العلاقة بين متغيرين أو أكثر، مثل العلاقة بين مهارة العاملين والإنتاجية أو بين سعر السلعة والكمية المطلوبة، وتقاس العلاقة بمؤشر احصائي يدعى معامل الارتباط ويرمز له بالرمز (r) (1) وهناك عدة أنواع من معاملات الارتباط حسب عدد المتغيرات وطبيعتها، وسنتطرق الى أربعة ~~ثلاث~~ معاملات هي :-

أ/ معامل الارتباط الخطي البسيط (بيرسون) .

ب/ معامل ارتباط الرتب (سبيرمان) .

ج/ معامل الاقتران .

د/ التوافق

## * خصائص معامل الارتباط :

1- تتراوح قيمة معامل الارتباط بين (-1) و(+1) أي: $-1 \leq r \leq +1$ إذ يمكن الحكم عن قوة العلاقة من حيث درجة قربها أو بعدها عن (±1)

2- تكون قيمة (r) مساوية للصفر عندما يكون المتغيران مستقلان عن بعضهما تماما (لا توجد علاقة بينهما)، ويكون مساويا للواحد الصحيح عندما يكون الارتباط تاما .

3- تكون قيمة (r) موجبة عندما يكون الارتباط بين المتغيرين طرديا ويكون قويا عندما يكون معامل الارتباط قريبا من الواحد الصحيح وضعيفا عندما يكون قريبا من الصفر .

4- تكون قيمته (r) سالبة عندما يكون الارتباط بين المتغيرين عكسيا ويكون قويا عندما يكون المقدار السالب قريبا من (-1)، وضعيفا عندما يكون المقدار السالب قريبا من الصفر .

## أولاً : معامل ارتباط بيرسون :- (معامل الارتباط الخطي البسيط) .

هو عبارة عن الوسط الهندسي لمعاملي معادلتي خطي الانحدار حيث يبين قوة العلاقة الخطية بين متغيرين أحدهما تابع والآخر مستقل ، وهو مستقل عن وحدة القياس التي تقاس بها الظاهرة المدروسة ، ويرمز له بـ (r) أو ( $r_p$ ) .

فإذا كانت معادلتي خط الانحدار كالتالي :

$\hat{y} = a x + b$ ، $\hat{x} = \acute{a} y + \acute{b}$

* a و $\acute{a}$ هما معاملا خطي الانحدار ، فإن :

$$\boxed{r = \sqrt{a \acute{a}}}$$

كما يمكننا أن نعوض بقيمتي معاملي الانحدار ونجد جملة من التبسيطات يمكن الحصول على القانون التالي لـ r :

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}}$$

أي :

$$r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

مثال : البيانات التالية توضح عدد أيام الغياب لخمس طلبة ونقاط (درجات) الطلاب في أحد اختبارات الرياضيات . أحسب معامل الارتباط الخطي البسيط؟

| عدد أيام الغياب x | درجة الاختبار y | $x^2$ | $y^2$ | xy |
|---|---|---|---|---|
| 0 | 27 | 0 | 729 | 0 |
| 4 | 21 | 16 | 441 | 84 |
| 4 | 18 | 16 | 324 | 72 |
| 10 | 10 | 100 | 100 | 100 |
| 6 | 16 | 36 | 256 | 96 |
| 24 | 92 | 168 | 1850 | 352 |

$$r_p = \frac{5(352) - (24)(92)}{\sqrt{(5(168) - (24)^2)(5(1850) - (92)^2)}}$$

$$\boxed{r_p = -0,98}$$

وعليه هناك ارتباط عكسي قوي بين عدد أيام الغياب ودرجات الطلبة في الرياضيات .

## ثانيا: معامل ارتباط الرتب (سبيرمان):

وهو معامل ارتباط ثنائي يصلح في المتغيرات الكمية والنوعية، وهو أقل دقة من معامل بيرسون للارتباط لأنه لا يتعامل مع القيم بل مع رتبها، ويستخدم خصوصا عندما تكون البيانات المتعلقة بالمتغيرين x، y أو بأحدهما بيانات نوعية (وصفية)، ويمكن ايجاد معامل سبيرمان حسب الخطوات التالية:

1- اعطاء رموز رقمية للبيانات (الرتب) لكل من x و y.

2- نستخرج الفرق بين رتب x ورتب y بعمود جديد هو d.

3- نقوم بتربيع الفرق بعمود آخر هو $d^2$.

4- نطبق قانون سبيرمان لإيجاد معامل الارتباط بالعلاقة التالية:

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

حيث: $r_s$: معامل سبيرمان للارتباط.

d: الفرق بين القيم المتقابلة في (x)، (y).

N: عدد أزواج القيم (x,y) في البيانات.

* ملاحظة: إذا كانت هناك بعض البيانات المتساوية (أو تأخذ نفس رتبة صفة المتغير الكيفي) ففي هذه الحالة لا يمكن ترتيبها. لذا نتبع الخطوات التالية:

1- نرتب القيم ونعطي لكل قيمة ترتيبا، (كما لو أن ليس فيها قيما متساوية) بحيث أكبر قيمة تأخذ الرتبة 1 والقيمة الأقل تأخذ الرتبة الأخيرة.

2- نوجد الوسط الحسابي لرتب كل مجموعة من البيانات المتساوية، ونعطي لكل قيمة من القيم المتساوية ترتيبا يساوي هذا الوسط.

مثال: وجد أن تقديرات مجموعة من الطلبة في مادتي ~~الاقتصاد~~ الرياضيات ($x$) والاحصاء ($y$) كالآتي:

| X | Y | رتب x | رتب y | (x-y) d | $d^2$ |
|---|---|---|---|---|---|
| امتياز | جيد جدا | 1 | 2 | -1 | 1 |
| جيد جدا | امتياز | 2 | 1 | 1 | 1 |
| جيد | متوسط | 3,5 | 4,5 | -1 | 1 |
| جيد | جيد | 3,5 | 3 | 0,5 | 0,25 |
| متوسط | متوسط | 5 | 4,5 | 0,5 | 0,25 |
| | | | | 0 | 3,5 |

المطلوب: إيجاد معامل الارتباط بينهما والتعليق عليه؟

$$r_s = 1 - \frac{6\sum d_i^2}{N(N^2-1)} = 1 - \frac{6(3,5)}{5(25-1)} = \boxed{0,82}$$

وهذا يعني أن هناك علاقة ارتباط قوية بين تحصيل الطالب في الرياضيات وتحصيله في الاحصاء، وهي علاقة طردية أي أن الطالب المتفوق في الرياضيات يكون متفوقا في الاحصاء.

## ثالثا: معامل الاقتران:

هناك حالات لا يمكن فيها ترتيب بيانات الظاهرة أو اعطاء رتب لها، لذلك نستخدم ما يعرف بمعامل الاقتران لمتغيرين لكل منهما حالتين كمدخن وغير مدخن، أو متعلم وغير متعلم أو ذكر و أنثى، (2x2) ولتوضيح ذلك نقول لدينا متغيرين $x$ و $y$، وهناك صفتان للمتغير $x$ هما $X_1$ و $X_2$ وصفتان للمتغير $y$ هما $Y_1$ و $Y_2$، فيكون لدينا الجدول التالي:-

|  $x_i$ \ $y_i$ | $y_1$ | $y_2$ |
|---|---|---|
| $x_1$ | $n_{11}$ | $n_{12}$ |
| $x_2$ | $n_{21}$ | $n_{22}$ |

حيث: $n_{11}$: تمثل التكرارات المشتركة في الصفة لكل من $x_1$، $y_1$
وهكذا الباقي...

وقد قام يول بوضع معامل الاقتران حسب العلاقة الرياضية الآتية:

$$r_a = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{11}n_{22} + n_{21}n_{12}} \qquad r_a \frac{ad - bc}{ad + bc}$$

حيث أن قيمة معامل الاقتران تقع بين الصفر والواحد أي:

$$0 \leq r_a \leq 1$$

مثال: الجدول الآتي يبين بيانات الاقتران بين الجنس والحل في أحد الفنادق، أحسب معامل الاقتران؟.

| الحل \ الجنس | ذكر | أنثى |
|---|---|---|
| يحل | 8 | 6 |
| لا يحل. | 3 | 7 |

$$r_a = \frac{ad - bc}{ad + bc} = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{11}n_{22} - n_{21}n_{12}}$$

$$= \frac{8(7) - 6(3)}{8(7) + 6(3)} = \boxed{0{,}51}$$

إذا العلاقة طردية ومقبولة بين المتغيرين (متوسطة).

-5-

مثال 2: قام أحد الباحثين بدراسة أثر التدخين على أمراض الصدر، فجمع بيانات عن 335 حالة فكان توزيعهم حسب الصفات على النحو التالي:

| | مدخن | غير مدخن | المجموع |
|---|---|---|---|
| مصاب | 120 | 30 | 150 |
| غير مصاب | 25 | 160 | 185 |
| المجموع | 145 | 190 | 335 |

احسب معامل الاقتران؟

$$r_a = \frac{ad - bc}{ad + bc} = \frac{(120)(160) - (25)(30)}{(120)(160) + (25)(30)} = 0,92$$

وهي علاقة طردية قوية جدًا. فكلما زاد عدد المدخنين كلما زاد احتمال الاصابة بأمراض الصدر.

رابعًا: معامل التوافق:

إذا كان للمتغيرين (أحدهم على الأقل) أكثر من صفتين، فيعرف معامل الاقتران في هذه الحالة بمعامل التوافق ويرمزله بالرمز $R_c$، ويقاس الارتباط من الصيغة الآتية والتي تعتمد على حساب معامل ($x^2$ كاي تربيع والذي سوف نتطرق له لاحقا) والذي يبين وجود العلاقة بين المتغيرين أو لا.

$$R_c = \sqrt{\frac{x^2}{x^2 + n}}$$

حيث أن: $x^2$: معامل كاي تربيع.

$n$: حجم العينة.

## اختبار مربع كاي $\chi^2$ (كاي تربيع) :

هناك استخدامات عديدة لكاي تربيع وهي حالة التوافق بين ما نتوقعه وبين ما نشاهده، فوجد [اختبار كاي مربع] مقياس لمدى التفاوت الموجود بين التكرارات المتوقعة والتكرارات المشاهدة وبعبارة أخرى هو مقياس التعارض أو التناقض، فرضيته تصاغ حسب طبيعة الظاهرة أو المشكلة المدروسة مثلا :

صيغته الرياضية

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

حيث أن :

$O_{ij}$ : القيمة المشاهدة أو الحقيقية.

$E_{ij}$ : القيمة المتوقعة.

$R$ : عدد الصفوف.

$C$ : عدد الأعمدة.

مثال توضيحي : لدينا متغيرين كل متغير له عدة تقسيمات أو مستويات فإن الجدول المشاهد له يكون :

الجدول المشاهد $O_{ij}$ :

| المتغير 2 / المتغير 1 | $C_1 : O_{i1}$ | $C_2 : O_{i2}$ | - - - | المجموع |
|---|---|---|---|---|
| $R_1 : O_{1j}$ | $O_{11}$ | $O_{12}$ | | $\sum R_1$ |
| $R_2 : O_{2j}$ | $O_{21}$ | $O_{22}$ | | $\sum R_2$ |
| ⋮ | | | | |
| المجموع | $C_1$ | $C_2$ | | $\sum R_i = \sum T_{ij} = \sum C_j$ |

-7-

والجدول المتوقع : $E_{ij}$

| المتغير1 / المتغير2 | $C_1 : E_{i1}$ | $C_2 : E_{i2}$ | ... | المجموع |
|---|---|---|---|---|
| $R_1 : E_{1j}$ | $E_{11} = \frac{(R_1)(C_1)}{T_{ij}}$ | $E_{12} = \frac{(R_1)(C_2)}{T_{ij}}$ | --- | $\Sigma R_1$ |
| $R_2 : E_{2j}$ | $E_{21} = \frac{(R_2)(C_1)}{T_{ij}}$ | $E_{22} = \frac{(R_2)(C_2)}{T_{ij}}$ | -- | $\Sigma R_2$ |
| ⋮ | ⋮ | ⋮ | -- | ⋮ |
| المجموع. | $C_1$ | $C_2$ | -- | $\Sigma R_i = \Sigma T_{ij} = C_j$ |

بعد استخراج الجدول المتوقع بالاعتماد على الجدول المشاهد، نستخرج قيمة $\chi^2$ المحتسبة من القانون:

$$\chi^2 = \Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}} =$$

$$= \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \cdots + \frac{(O_{1n} - E_{1n})^2}{E_{1n}}$$

$$+ \frac{(O_{21} - E_{21})^2}{E_{21}} + \frac{(O_{22} - E_{22})^2}{E_{22}} + \cdots + \frac{(O_{2n} - E_{2n})^2}{E_{2n}}$$

$$+ \frac{(O_{n1} - E_{n1})^2}{E_{n1}} + \frac{(O_{n2} - E_{n2})^2}{E_{n2}} + \cdots + \frac{(O_{nn} - E_{nn})^2}{E_{nn}}$$

مثال: الجدول الآتي يبين عدد الطلاب الذين نجحوا ورسبوا في ثلاث مواد دراسية في قسم السياحة.

المطلوب:

1- استخرج قيمة $\chi^2$؟

2- حساب معامل التوافق؟

الحل:

← جدول التكرارات الحقيقية [من التمرين] اي $O_{ij}$

| | اقتصاد | سياحة | احصاء | Σ |
|---|---|---|---|---|
| ناجح | 56 | 47 | 50 | 153 |
| راسب | 8 | 14 | 5 | 27 |
| Σ | 64 | 61 | 55 | 180 |

ننشئ الآن جدول التكرارات النظرية او المتوقعة $E_{ij}$

| | اقتصاد | سياحة | احصاء | Σ |
|---|---|---|---|---|
| ناجح | 54,4 | 51,85 | 46,75 | 153 |
| راسب | 9,6 | 9,15 | 8,25 | 27 |
| Σ | 64 | 61 | 55 | 180 |

مثلا الخانة الأولى: (ناجح/اقتصاد)

$$E_{11}=\frac{(153)\times(64)}{180} = 54,4$$

وهكذا مع البقية،

نحسب الآن قيمة الاحصاءة $\chi^2$:

$$\chi^2=\sum\left(\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right)=\frac{(50-46,75)^2}{46,75}+\frac{(47-51,85)^2}{51,85}+\dots$$

$$=\dots+\frac{(8-9,6)^2}{9,6}=4,85$$

وأخيرا نحسب معامل التوافق $R_c$

$$R_c=\sqrt{\frac{\chi^2}{\chi^2+n}}=\sqrt{\frac{4,85}{4,85+180}}=0,16.$$

وبذلك يمكن القول ان التوافق منخفض بين المتغيرين وبالتالي فالعلاقة بينهما ضعيفة.

-9-