جامعة محمد خيضر (بسكرة)
كلية العلوم الاقتصادية والتجارية والتسيير
قسم العلوم التجارية

السنة الجامعية: 2024/2023
المستوى: أولى ماستر تسويق
مقياس: الأساليب الكمية في التسويق 2

المحاضرة الرابعة: التحليل التمييزي

1. تمهيد:

إن كل إنسان يمارس التحليل التمييزي يومياً وبصورة لاشعورية: عندما نسير في الشارع نميز فورا بين الناس حسب النوع، ونفرزهم إلى ذكور واناث؛ كما نميز بينهم حسب العمر دون أن نسألهم إلى كبار وشباب وأطفال؛ الطبيب عندما يفحص المرضى يميز بينهم بسرعة حسب حالاتهم الصحية (مريض أو سليم)، وذلك قبل إجراء الفحوص الطبية اللازمة؛ المتجول في الغابة يميز بين أنواع الأشجار بسرعة؛ الشرطي على بوابة مطار دولي، يميز في لحظة خاطفة بين المسافرين حسب جنسياتهم، ويفرزهم إلى أوروبيين، أفارقة، هنود 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، مثلاً لا نستطيع ببساطة التمييز بين العملاء/ الموظفين ذوي الرضا العالي أو المنخفض، الشركات المتعثرة أو غير المتعثرة..الخ، وذلك لأن التمييز في مثل هذه الحالات يحتاج إلى دراسة معمقة، وتحديد المؤشرات التي تساعدنا في عملية التمييز.

وهنا نطرح **السؤال التالي**: كيف استطعنا تمييز الأفراد، وتحديد المجموعات التي ينتمون إليها وبهذه السرعة الفائقة؟ و**الإجابة هي**: استخدمنا صفات أو خواص ملازمة لهذه العناصر، وقمنا بتركيبها ضمن صيغ معينة (فورية)، ومنها حصلنا على مؤشر معين ساعدنا على فرز العناصر وتحديد المجموعات التي تنتمي إليها.

إن بعض عمليات الفرز والتصنيف قد تكون خاطئة بسبب تشابه خواص العناصر، أو الالتباس بينها، فشرطي المطار الدولي فقد يقع في خطأ ما، فالذي اعتبره اوروبياً قد يكون كنديا، والإفريقي قد يكون أمريكيا، والهندي قد يكوف باكستانيا، والصيني قد يكوف يابانياً... الخ، وهكذا نجد أن عمليات التمييز بين العناصر ليست أمرا سهلا، وهي تحتاج إلى دراسات معمقة وأدوات رياضية/احصائية متقدمة، يأتي في مقدمتها الدالة التمييزية، التي تستخدم لتصنيف عناصر العينة إلى المجموعات المناسبة لها (مع احتمال الخطأ في ذلك).

يطلق على هذه الدراسات والأدوات: **التحليل التمييزي**، حيث يستخدم في تصنيف الأفراد في مجموعات، وذلك بناء على أوزان أو نسب أو درجات يحصلون عليها في توليفة من الصفات أو الخصائص، التي تتنبأ بعضويتهم في إحدى المجموعات.

2. تعريف التحليل التمييزي :

التحليل التمييزي هو: "تقنية إحصائية رياضية، تستخدم لتوصيف عناصر مجتمع مدروس، وتوزيعها على مجموعات محددة ومنفصلة (2 أو أكثر)، وتحديد الحدود الفاصلة بينها، بهدف استخلاص قاعدة لتحديد انتماء أي عنصر جديد إلى أي منها".

التحليل التمييزي هو: "أسلوب إحصائي يهدف إلى وصف وشرح والتنبؤ بالعضوية في مجموعات محددة مسبقًا (فئات ...)، لمجموعة من الحالات (أفراد ..)، اعتمادا على سلسلة من المتغيرات التنبؤية الملائمة (مواصفات...)".

التحليل التمييزي هو: "أسلوب إحصائي يهدف إلى تصنيف مفردات المجتمع إلى عدة مجتمعات، وبناء قاعدة تساعد في تحديد المجتمع الذي تنتمي إليه مفردات جديدة مستقبلا".

يفضل استخدام التحليل التمييزي عندما يكون المتغير التابع وصفي ثنائي أو أكثر، إضافة إلى وجود مجتمعين أو أكثر، وهذه المجتمعات متشابهة ولكن منفصلة إحصائيا، ومنها تكوين قاعدة للفصل بين هذه المجتمعات، تستخدم في تصنيف مفردات جديدة غير معروف المجتمع الذي تنتمي إليه.

انطلاقا من مجموعة من المتغيرات المستقلة، يحاول التحليل التمييزي إيجاد تركيبات خطية من أنسبها، والتي تسمح بالتمييز بشكل أفضل بين مجموعات الحالات المختلفة. وهنا يجب التمييز بين نوعين من المتغيرات التي تستخدم في التحليل التمييزي:

- **المتغير التابع:** متغير نوعي اسمي فئوي (عميل راضي/ غير راضي؛ مؤسسة متعثرة/ غير متعثرة، شخص مصاب / سليم ...).
- **المتغيرات التمييزية المستقلة:** تمثل الخصائص المميزة لكل مجموعة من المجموعات الداخلة في التحليل التمييزي، تستخدم في إيجاد دالة التمييز من خلالها.

3. استخدامات التحليل التمييزي:

يستخدم التحليل التمييزي في العديد من المجالات:

- في الطب: الكشف عن المجموعات المعرضة لخطر الإصابة بمرض ما، بناءً على خصائص مثل النظام الغذائي، التدخين، التاريخ العائلي...
- في القطاع المصرفي: تقييم موثوقية طالب الائتمان بناءً على دخله، القروض المستحق، تعثرات سابقة ...
- في علم الأحياء: نسبة كائن إلى عائلته بناءً على خصائصه الفيزيائية.، مثل قزحية رونالد فيشر ...
- في الحوسبة: التعرف البصري على الحروف المطبوعة بناءً على وجود أو عدم وجود التماثل، وعدد الأطراف ..
- في الإحصاء الاستكشافي: التعرف على الأنماط، التعلم الآلي، استخراج البيانات ...

في مجال التسويق:

- تصنيف الشركات (أو عملاء) إلى متعثرة/ غير متعثرة في سداد القروض، وتقدير تعثر شركة (أو عميل) بناء على توليفة نسب مالية ملائمة، وبالتالي تحديد فئة المخاطرة التي يقع فيها طالب الإئتمان.
- تصنيف العملاء إلى راضين أو غير راضين اتجاه منتجات/ خدمات شركة، وتقدير رضا عميل بناء على متغيرات ملائمة.
- تصنيف تجارب إطلاق منتجات جديدة إلى ناجحة/ فاشلة، وتقدير احتمال نجاح إطلاق منتج جديد.
- تصنيف الموظفين إلى راضين/ غير راضين عن إدارة الموارد البشرية، وتقدير رضاهم عن العمل في المؤسسة.

4. المصطلحات المتشابهة والمتداخلة:

أ. التمييز: يشترط وجود مجتمع مقسم إلى مجموعتين أو أكثر محددة مسبقاً، وسحب عينة عشوائية من كل منها، ثم دراسة خواص هذه العينات، لاستخلاص معيار مبني على بيانات العينات، يسمح بفرز عناصر المجتمع إلى تلك المجموعات، وبالتالي تحديد المجموعة التي ينتمي إليها أي عنصر جديد.

ب. التصنيف: توزيع عناصر المجتمع أو عينة منه إلى مجموعات متجانسة داخلياً ومختلفة عن بعضها أكثر ما يمكن، وينتج عنه مجموعات محددة ذات مواصفات معينة.

ج. التجزئة: تقسيم عناصر مجتمع كلي أو عينة إلى مجموعات جزئية، بغض النظر عن وجود حدود طبيعية لذلك التقسيم أو لا، مثل توزيع الطلاب عشوائياً على الشعب الدراسية بدون أية شروط مسبقة.

5. أهداف التحليل التمييزي:

- الحصول على نموذج/دالة مثالية، تفصل بين مجموعات عناصر، أو مشاهدات التي أجريت عليها عدة قياسات أو متغيرات .
- اسناد عنصر جديد إلى أحد المجتمعات بواسطة نموذج/دالة تم إيجاده اعتمادا على مجموعات مصنفة سابقا، مع نسبة خطأ أصغر ما يمكن.
- اختبار وجود فروق ذات دلالة احصائية بين المجموعات بالنسبة للمتغيرات المستقلة
- تحديد المتغيرات المستقلة التي تساهم بأكبر قدر من الاختلاف بين فئات المتغير التابع.

- تقييم (كنسبة مئوية) دقة تقسيم المجتمع الكلي إلى مجموعات متمايزة.

6. أنواع التحليل التمييزي:

هناك عدة أنواع للتحليل التمييزي حسب الهدف، عدد دوال التمييز، طرق إدخال المتغيرات المستقلة في التحليل.

أ. حسب الهدف:

- **التحليل التمييزي الوصفي Descriptive Discriminant analysis:** يتناول توصيف العناصر ضمن المجموعات المحددة، ووضع قواعد الانتماء إليها .
- **التحليل التمييزي التنبؤي Predictive Discriminant analysis:** يتناول تصنيف العناصر إلى مجموعات، ويتنبأ بانتماء أي عنصر جديد إلى إحدى تلك المجموعات .

ب. حسب عدد الدوال التمييزية:

- **التحليل التمييزي الأحادي Simple discriminant analysis:** يحتوي على دالة تمييزية واحدة.
- **التحليل التمييزي المتعدد Multi-discriminant analysis:** يحتوي على أكثر من دالة تمييزية واحدة.

ج. حسب طرق إدخال المتغيرات في التحليل:

- **التحليل التمييزي المباشر Direct Discriminant Analysis:** يتم إدخال جميع المتغيرات المستقلة في التحليل مرة واحدة عند إنشاء دوال التمييز، ودون استثناء أو إعطاء أي أهمية لترتيب دخولها.
- **التحليل الهرمي Hierarchical Discriminant Analysis:** يتم إدخال المتغيرات للتحليل تبعاً لما يراه الباحث من أهمية للمتغيرات المستقلة، وبالترتيب الذي يراه.
- **التحليل التمييزي التدريجي Stepwise Discriminant Analysis:** ترتيب المتغيرات فيه يكون وفقاً لمعايير إحصائية، حيث يتم البدء بالمتغير الأكثر تمييزاً بين المجموعات.

د. حسب العلاقة بين المتغيرات:

- **التحليل التمييزي الخطي linear discriminant analysis (LDA):** يعتمد على النماذج الخطية للفصل بين مجموعات المجتمع المدروس .
- **التحليل التمييزي غير الخطي Non linear discriminant analysis:** يعتمد على النماذج غير الخطية للفصل بين مجموعات المجتمع المدروس.

7. افتراضات وشروط تطبيق التحليل التمييزي:

- المجموعات في المجتمع محددة بشكل طبيعي، أي ليست عشوائية، كما أنها ومتكاملة وغير متقاطعة، وعددها لا يقل عن 2، وحجومها متقاربة أو غير مختلفة كثيرا؛
- المشاهدات مستقلة عن بعضها، أي العينات الطبقية مسحوبة عشوائيا من المجموعات.
- حجم العينة الكلية n أكبر بـ 20 مرة من عدد المتغيرات المستقلة p، ولا يقل حجمها عن $n= 30$.
- المتغيرات مستقلة عن بعضها، أي عدم وجود ارتباط عالٍ بين المتغيرات المستقلة، ولا يكون عددها كبيراً (أقل من 5% من حجم المجتمع الكلي)، وإلا يجب اختزالها حسب معايير محددة،.
- المتغيرات كمية خاضعة للتوزيع الطبيعي، أي لا توجد بيانات شاذة، وإلا نقوم بإجراء تحويلات أو نستغني عن بعض المتغيرات.
- مصفوفات التباين والتباين المشترك للمتغيرات المستقلة X داخل كل مجموعة، متشابهة (متماثلة)، أي متساوية أو متجانسة على الأقل .
- الأخطاء (البواقي) موزعة عشوائيا، أي أن يكون توقعها مساويا للصفر.

ملاحظة: عندما يزيد حجم العينة فإن الشروط السابقة تقل، وهذا يحدث عندما يكون لدينا على الأقل 20 حالة في أصغر المجموعات.

8. خطوات التحليل التمييزي:

1. تحديد نوع التحليل التمييزي المناسب: وصفي أم تنبؤي، بسيط أم متعدد، خطي أم غير خطي، كمي أو نوعي أم لوجستي .
2. تحديد المتغيرات المستقلة المناسبة واللازمة لتحقيق أهداف البحث، وجمع البيانات عنها .
3. تحديد عدد المجموعات في المجتمع التي ستستخدم في التحليل(g≥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 ، من خلال التنبؤ بانتماء وتصنيف مفردات جديدة في المجتمع

9. الدال التمييزية Function Discriminant:

يقوم التحليل التمييزي ببناء نموذج احصائي تنبؤي يسمى دوالة التمييز، انطلاقا من بيانات العينة المدروسة، تستخدم في التصنيف (تحديد انتماء فرد ما إلى مجموعة معينة)، ثم التنبؤ(تطبيق هذه الدالة على أفراد جدد).

دالة التمييز هي تركيبات خطية لمجموعات من المتغيرات المستقلة الملائمة (متغيرات التمايز)، حيث تعمل على تعظيم الفروق بين متوسطات المجموعات، وتقليل التشابه في أخطاء التصنيف في الوقت ذاته.

وعدد دوال التمييز الممكنة لتحليل ما هو= عدد المجموعات -1، فمثلا إذا كان لدينا 3 مجموعات و4 متغيرات كمية، فإن عدد دوال التمييز = 2.

تأخذ دالة تمييز الصيغة التالية:

$$Z= \alpha +\beta_1 X_{i1}+ \beta_2 X_{i2}+\ldots\ldots\ldots\ldots+\beta_n X_{in}$$

Z: القيمة التمييزية؛ α: تمثل عدد ثابت؛ X_i: تمثل المتغيرات؛ β_i: تمثل معاملات التمميزية للمتغيرات التفسيرية.

10. خطوات بناء النموذج تنبؤي (دالة التمييز):

أ. تحديد حجم العينة:

يجب أن يكون حجم العينة كبيراً نسبياً، بحيث لا يقل عن 20 مشاهدة، وذلك لإضفاء نوع من الاستقرار والصلاحية على نتائج دوال التمييز. كما يجب تجنب الأحجام الكبيرة جداً، لأنها قد تسفر عن نتائج ذات دلالات إحصائية غير صحيحة. وعند استخدام عينات تضم 50 حالة أو أقل، و20 حالة كحد أدنى، فإن التحليل التمييزي يحقق قدرة التنبؤية أفضل، وعموما هناك اختلافا وتبايناً في تحديد حجم العينة المناسب.

ب. تحديد المتغيرات المستقلة للنموذج:

يتم اقتراح قائمة أولية من المتغيرات المستقلة، التي يرى أنها تفسر المتغير التابع، ثم اختيار من بينها الأكثر قدرة على التمييز بين المجموعات، وفق أسلوب التحليل المتدرج Stepwise Analysis، والتي تدخل في بناء النموذج الذي سيستخدم مستقبلاً للتنبؤ، حيث يعمل التحليل المتدرج على اختيار أفضل المتغيرات التي تعمل على التمييز بين المجموعات الجزئية، وذلك بتخفيض عدد المتغيرات المستقلة التي ستدخل في تكوين النموذج التنبؤي بشكل متدرج ومرحلي .

وينطلق التحليل المتدرج من عدم وجود أي متغير مستقل في النموذج، ثم يبدأ كخطوة أولى باختيار أفضل متغير مستقل يميز بين المجموعات الجزئية، يعطي أكبر معامل إرتباط مع المتغير التابع، ويحتفظ به كمتغير أول يدخل في النموذج مع حذفه من القائمة الكلية للمتغيرات المستقلة المقترحة، ليتم بعد ذلك اختيار المتغير الثاني الذي يمثل أفضل المتغيرات المستقلة المتبقية، ويحتفظ به كمتغير ثاني يدخل في النموذج مع حذفه أيضاً من القائمة الكلية للمتغيرات المستقلة المقترحة، ويتم الإستمرار إلى أن يتم تحديد جميع المتغيرات التمييزية .

ج. تحديد الصيغة التامة لنموذج التحليل التمييزي:

يتم إيجاد المعاملات التمييزية للنموذج من أجل احتساب القيمة التمييزية للمتغيرات المستقلة، وذلك بضرب تلك المعاملات بالقيم الفعلية للمتغيرات المستقلة. ويستفاد من المعاملات التمييزية في تحديد أهمية المتغيرات المستقلة في تكوين النموذج، فإذا كانت القيمة المطلقة لمعامل تمييزي لمتغير مستقل ما كبيرة، فإنه يعني أن المتغير يساهم بشكل كبير في بناء النموذج، وأنه ذو أثر كبير في تصنيف الأفراد الجدد. كما تدل إشارة المعامل على اتجاه التمييز، فإذا كانت + فإنه يساهم في تصنيفها الأفراد إلى مجموعة، وإذا كان - فإنه يزيد من حظوظ تصنيفها إلى المجموعة الأخرى.

د. إنشاء قاعدة التصنيف:

هي درجات لدالة التمييز Z ستخدم في الفصل الإحصائي بين المجتمعات الجزئية محل الدراسة، وذلك بناءا على بيانات عينة من المفردات المأخوذة من هذه المجتمعات، مثلا تحسب قاعدة الفصل والتمييز كما يلي:

في حالة عدم تساوي المجموعات: $Z^* = (n_1 Z_1 + n_2 Z_2)/(n_1 + n_2)$

في حالة تساوي المجموعات: $Z^* = (Z_1 + Z_2)/2$

حيث: n_1 عدد مفردات في المجموعة الأولى؛ n_2 عدد مفردات في المجموعة الثانية؛

Z_1 متوسط درجة التمييز المعيارية للمجموعة الأولى؛ Z_2 متوسط درجة التمييز المعيارية للمجموعة الثانية؛

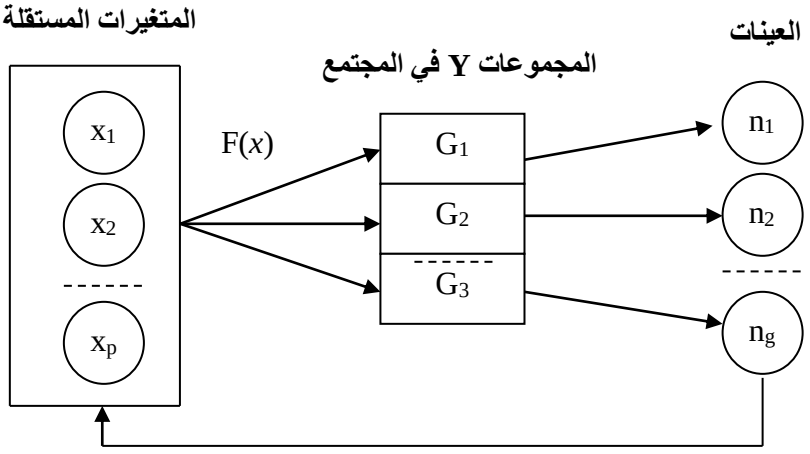
هـ. اختبار قدرة الدالة على التمييز:

قبل وضع النموذج الذي تم بناؤه وفق أسلوب التحليل التمييزي في التصنيف والتنبؤ، يجب التأكد من صلاحيته، وذلك باختبار دقته في التصنيف بشكل صحيح، من خلال تطبيقه على مفردات العينة المدروسة، فكل مفردة ذات قيمة تمييزية أكبر من نقطة الفصل المثلى تنتمي للمجموعة الأولى، وكل مفردة لها قيمة تمييزية أقل من نقطة الفصل المثلى، فإنها تنتمي مممجموعة الثانية.

11. الإطار الرياضي للتحليل التمييزي:

يمكن تمثل التأثير المشترك للمتغيرات المستقلة X على المجموعات التابعة Y بيانياً على الشكل التالي :

الشكل (01): آلية عمل التحليل التمييزي



إن الشكل (01) يبين أن جملة المتغيرات المستقلة $(X_1, X_2 .. X_p)$ تؤثر على كل مجموعة من المجموعات التي تتألف منها الدالة (المتغير التابع) Y، وذلك من خلال التراكيب الخطية (أو غير الخطية)، المعرفة عليها $F_1(x)$, $F_2(x)$ .. $F_g(x)$، وأن كل عنصر من عناصر هذه المجموعات يعطينا جملة من القياسات لهذه المتغيرات وعند تعويضها في التركيب الخطي المقابل لتلك المجموعة نحصل على قيمة محددة للدالة $F_J(x)$، تميز ذلك العنصر عن غيره في المجموعة المذكورة .

إن تطبيق التحليل التمييزي يتطلب أن نقوم بسحب عينات عشوائية طبقية من كل مجموعة من مجموعات المجتمع بحجوم متساوية أو مختلفة: $n_1, n_2 .. n_g$، فتتشكل لدينا عينة كلية حجمها: $n= \sum n_J \quad j= 1 .. g$، ويمكن تفريغ بيانات هذه العينات في جداول مناسبة وحساب متوسطاتها وتبايناتها كما يلي :

الجدول رقم (01): بيانات العينات المسحوبة من المجموعات (حالة مجموعتين)

بيانات/ متغيرات	بيانات المجموعة الأولى 1 2 3 n_1	المتوسطات $\bar{x}_{ij}$	التباينات σ_{ij}	بيانات المجموعة الثانية 1 2 3 n_2	المتوسطات $\bar{x}_{11}$	التباينات σ_{ij}
X_1	x_{111} x_{112} x_{113} x_{11n1}	$\bar{x}_{11}$	σ_{11}	x_{121} x_{122} x_{123} x_{12n1}	$\bar{x}_{12}$	σ_{12}
X_2	x_{211} x_{212} x_{213} x_{21n1}	$\bar{x}_{21}$	σ_{21}	x_{221} x_{222} x_{213} x_{22n1}	$\bar{x}_{22}$	σ_{22}
X_3	x_{311} x_{312} x_{313} x_{31n1}	$\bar{x}_{31}$	σ_{31}	x_{321} x_{322} x_{323} x_{32n1}	$\bar{x}_{32}$	σ_{32}
..	..	..		..	..	
X_p	x_{p11} x_{p12} x_{p13} x_{p1n1}	$\bar{x}_{p1}$	σ_{p1}	X_{p21} x_{p22} x_{p13} x_{p2n1}	$\bar{x}_{p2}$	σ_{p2}

نرمز لقيم المتغيرات بـ x_{ijk} حيث: i رمز المتغير X_i؛ j رمز المجموعة G_j، وk رمز المشاهدة k من n_j

في حالة التمييز الخطي المتعدد، يأخذ النموذج أو الدالة التمييزية الشكل التالي:

$$F(x)= b_0+ b_1x_1+ b_2x_2 + \dots + b_px_p + \varepsilon$$

حيث:

$b_1, b_2, .. b_p$ الأمثال (المعاملات) التمييزية في الدالة $F(x)$؛ $(x_1, x_2 .. x_p)$ المتغيرات المستقلة الكمية؛

ε متغير الخطأ العشوائي في النموذج؛

$F(x)$ دالة كامنة وتركيب خطي للمتغيرات x المؤثرة في الدالة $F(x)$، تسمى الدالة التمييزية الفاصلة بين المجموعات.

12. الإختبارات الإحصائية المستخدمة في التحليل التمييزي:

أ. معيار ولكس- لمدا Wilks's lambda:

طريقة إحصائية تأخذ في الإعتبار الاختلاف والشمولية بين المجموعات، ويقصد بالشمولية درجة تجمع الحالات قرب مركز المجموعة، أي أنه يقيس درجة التباعد بين المجموعات، وهو يساوي نسبة مجموع مربعات الإنحرافات داخل المجتمعات الجزئية إلى مجموع مربعات الإنحرافات الكلية، ومنه يكون محصور بين 0 و1، فإذا اقترب من 0 دل على أن المتغير المستقل يحقق تمييزا عاليا بين المجتمعات الجزئية، وإذا اقترب من 1 دل على العكس. أي كلما كانت قيمته صغيرة كلما كانت أقوى في تمييز الفروق .

يعتبر **معيار ولكس- لمدا** أفضل طريقة في اختيار المتغيرات المستقلة، حيث المتغيرات المستقلة التي تدخل في النموذج هي ذات أعلى قيمة لـ F (معدل مساهمة المتغير المستقل في التمييز بين المجموعتين)، بعد الأخذ بعين الإعتبار التغيرات التي تحدثها بقية المتغيرات التمييزية، حيث يقيس الفروقات بين المراكز المتوسطة للمجتمعين.

ب. احصائية RAO'S:

يرمز لها بـ v، تبنى على مسافة Mahalanobis، تقيس الانفصال الكلي للمجموعات (مراكزها) عن بعضها، تطبق على أي عدد من المجموعات، لا يرتبط بذاته بالشمولية في المجموعات، ولذلك فإن المتغير الذي يتم اختياره على أساس v يمكن أن يكون زيادة في تماسك المجموعة، بالإضافة إلى أن المسافات التي يتم قياسها بواسطة v تكون من خلال مركز كل مجموعة متجهاً إلى المركز الأكبر الذي يتم وزنه باستخدام حجم المجموع، وبالتالي فإن v لا تؤكد الحد الأقصى من الانفصال بين كل زوجين من المجموعات.

ج. مسافة Mahalanobis:

يرمز لها بالرمز D^2، وهي إحدى الإحصائيات التي تعد مقياسا مباشراً يعطي وزنا مكافئاً لكل زوجين من المجموعات، حيث يمكن حساب مسافة Mahalanobis المربعة للمفردات عن مراكز المجموعات، والتالي يمكن من توزيع المفردة للمجموعة الأقرب لها.