جامعة محمد بوضياف -المسيلة
كلية العلوم الاقتصادية والتجارية والتسيير
قسم العلوم التجارية

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

المحور الثاني: أدوات تقسيم السوق

التحليل التمييز والتحليل العنقودي

المحاضرة الأولى: تقسيم السوق والتجزئة السوقية

1.تمهيد

تعتبر تجزئة السوق أداة لصنع القرار لمدير التسويق في المهمة الحاسمة لاختيار السوق المستهدفة لمنتج معين وتصميم مزيج تسويقي مناسب، تجزئة السوق هي إحدى الركائز الأساسية للتسويق الاستراتيجي. وتجزئة السوق أمر ضروري لنجاح التسويق: فالشركات الأكثر نجاحاً تقود أعمالها على أساس التجزئة.

2. التخطيط التسويقي الاستراتيجي والتخطيط التسويقي التكتيكي

الغرض من التسويق هو التوفيق بين الاحتياجات والرغبات الحقيقية للمستهلكين وعروض الموردين المناسبة بشكل خاص لتلبية تلك الاحتياجات والرغبات. وتفيد عملية المطابقة هذه المستهلكين والموردين على حد سواء، كما أنها تقود عملية التخطيط التسويقي للمؤسسة.

التخطيط التسويقي عبارة عن تسلسل منطقي وسلسلة من الأنشطة التي تؤدي إلى تحديد الأهداف التسويقية وصياغة الخطط لتحقيقها. تتكون خطة التسويق من عنصرين: خطة تسويق استراتيجية وخطة تسويق تكتيكية. تحدد الخطة الاستراتيجية الاتجاه طويل الأجل للمؤسسة، ولكنها لا تقدم الكثير من التفاصيل حول الإجراءات التسويقية قصيرة الأجل المطلوبة للتحرك في هذا الاتجاه طويل الأجل. تقوم خطة التسويق التكتيكي بالعكس. فهي تترجم الخطة الاستراتيجية طويلة الأجل إلى تعليمات مفصلة للإجراءات التسويقية قصيرة الأجل. تنص خطة التسويق الاستراتيجية على الوجهة التي تريد المؤسسة التوجه إليها ولماذا. تحتوي خطة التسويق التكتيكي على تعليمات حول ما يجب القيام به للوصول إلى هناك. تحدد خطة التسويق الاستراتيجي عادةً احتياجات المستهلكين ورغباتهم، ونقاط القوة والضعف الداخلية للمؤسسة، والفرص والتهديدات الخارجية التي قد تواجهها المؤسسة .

ينص تحليل **SWOT** صراحةً على نقاط القوة (S) ونقاط الضعف (W) والفرص (O) والتهديدات (T) الخاصة بالمؤسسة. على هذا النحو، يحدد تحليل **SWOT** أحد جانبي عملية المطابقة: ما الذي يناسب المورد بشكل خاص ليقدمه للمستهلكين. أما الجانب الآخر من عملية المطابقة - احتياجات المستهلكين ورغباتهم - فيتم عادةً بحثه باستخدام أبحاث السوق. وعلى الرغم

من الاعتماد الكبير لأبحاث السوق على منهجية المسح، إلا أن هناك مجموعة واسعة من مصادر المعلومات المتاحة لاستكشاف ما يحتاجه المستهلكون أو يرغبون فيه واكتساب رؤية مفصلة حول ما يحتاجه المستهلكون أو يرغبون فيه، بما في ذلك الأبحاث النوعية التي تشمل مجموعات التركيز والمقابلات والبحوث القائمة على الملاحظة والتجريب.

.2. تعريف وأهمية تقسيم السوق

تعرّف تجزئة السوق بأنها النظر إلى سوق غير متجانسة (سوق تتسم بتباين الطلب) على أنها عدد من الأسواق المتجانسة الأصغر حجمًا. ومن الناحية المفاهيمية، يقع تجزئة السوق بين وجهتي النظر المتطرفتين القائلتين بأن (أ) جميع الأشياء فريدة وغير قابلة للتجزئة و(ب) أن السكان متجانسون

أحد أبسط التعريفات وأوضحها هو التعريف الذي استخدمته شركة **Grey Advertising Inc**. تجزئة السوق يعني تقسيم الأسواق إلى شرائح. من الناحية المثالية، يكون المستهلكون الذين ينتمون إلى نفس شرائح السوق - أو مجموعات من المشترين متشابهين جدًا مع بعضهم البعض فيما يتعلق بخصائص المستهلك التي تعتبرها الإدارة حاسمة. وفي الوقت نفسه، من الناحية المثالية، يكون المستهلكون الذين ينتمون إلى قطاعات السوق المختلفة مختلفين تمامًا عن بعضهم البعض فيما يتعلق بخصائص المستهلكين تلك. ويشار إلى خصائص المستهلكين التي تعتبرها الإدارة حاسمة بالنسبة لتقسيم السوق بمعايير التجزئة.

3. معايير تقسيم السوق

يعتمد اختيار المعيار المناسب لتقسيم السوق على **عدة عوامل** أهمها:

- نوع المنتج أو الخدمة التي تقدمها؛
- أهداف المؤسسة التسويقية؛
- الميزانية المتاحة (المخصصة).

يمكن أن يكون معيار التجزئة خاصية واحدة من خصائص المستهلك، مثل العمر، أو الجنس، أو بلد المنشأ، أو مرحلة في دورة حياة الأسرة. وبدلاً من ذلك، يمكن أن يحتوي على مجموعة أكبر من خصائص المستهلك، مثل عدد من المزايا المطلوبة في المنتج عند شرائه، أو القيم التي يتمسك بها المستهلك فيما يتعلق بالبيئة، أو نمط الإنفاق. ويمكن تقسيم هذه المعايير كالآتي:

a) **المعيار الديمغرافي:** مثل ما يلي:

- العمر :تقسيم السوق حسب الفئات العمرية (مثلاً: جيل الألفية، جيل Z)؛
- الجنس :تقسيم السوق حسب الذكور والإناث؛
- الدخل :تقسيم السوق حسب مستوى الدخل (مثلاً: منخفض، متوسط، مرتفع)؛
- المستوى التعليمي :تقسيم السوق حسب مستوى التعليم (مثلاً: ثانوي، جامعي، دراسات عليا)؛
- المهنة :تقسيم السوق حسب نوع المهنة (مثلاً: موظفون، طلاب، ربات بيوت)؛
- الحالة الاجتماعية: تقسيم السوق حسب الحالة الاجتماعية (مثلاً: عازب، متزوج، مطلق).

b) **المعيار الجغرافي:** مثل ما يلي:

- الموقع: تقسيم السوق حسب الموقع الجغرافي (مثلاً: دولة، مدينة، منطقة)؛
- المناخ: تقسيم السوق حسب المناخ (مثلاً: حار، بارد، معتدل)؛
- الكثافة السكانية: تقسيم السوق حسب الكثافة السكانية (مثلاً: حضري، ريفي).

c) **المعيار السلوكي:** مثل ما يلي:

- سلوك الشراء: تقسيم السوق حسب سلوك الشراء (مثلاً: مستهلكون متكررون، مستهلكون مندفعون)؛
- معدل استخدام المنتج: تقسيم السوق حسب معدل استخدام المنتج (مثلاً: مستخدمون خفيفون، مستخدمون متكررون)؛
- الولاء للعلامة التجارية: تقسيم السوق حسب الولاء للعلامة التجارية (مثلاً: مخلصون لعلامة تجارية معينة، يميلون لتجربة علامات تجارية جديدة).

d) **المعيار النفسي:** مثل ما يلي:

- نمط الحياة :تقسيم السوق حسب نمط الحياة (مثلاً: رياضي، اجتماعي، تقليدي)؛
- القيم :تقسيم السوق حسب القيم الشخصية (مثلاً: دينية، بيئية، إنسانية)؛
- الشخصية :تقسيم السوق حسب السمات الشخصية (مثلاً: انطوائي، منفتح، عاطفي).

4. أهمية وخطوات تقسيم السوق

هناك أهداف مختلفة لتجزئة السوق. من بينها ما يلي:

- المنتج: يعد إنشاء منتجات ناجحة أحد الأهداف الرئيسية للمؤسسات وأحد أسباب إجرائها لأبحاث السوق. يتيح ذلك إضافة الميزات الصحيحة للمنتج، يساعد ذلك أيضًا على تقليل التكاليف لتلبية احتياجات الجمهور المستهدف؛
- السعر: هدف آخر لتجزئة السوق هو تحديد السعر المناسب للمنتجات مع تحديد الجمهور الذي سيكون على استعداد لدفع ثمنها؛
- الترويج: يساعد الترويج على استهداف أعضاء كل شريحة واختيارهم في فئات مختلفة التي تسمح بتوجيه استراتيجياتك بشكل مناسب؛
- المكان: الهدف النهائي للتجزئة هو تحديد كيفية تقديم منتج لكل مجموعة من المستهلكين وجعله ممتعًا وهاما لهم.

ومن أجل تقسيم السوق يتم اتباع خطوات لتنفيذ استراتيجية تجزئة السوق، من بينها:

- تحديد السوق المستهدف: في هذه المرحلة من التقسيم، يجب أن تركز على المؤسسة اكتشاف حجم السوق، والتي تناسب علامتها التجارية، وما إذا كان بإمكان منتجاتها القدرة على تقديم ما تعد به؛
- تقسيم السوق: تتكون هذه الخطوة من اختيار الأسواق التي تناسب أهداف المؤسسة بشكل أفضل؛

- فهم المستهلكين من السوق وسلوكهم: ويتم ذلك عن طريق الاختلاط بالعملاء وسؤالهم الأسئلة الصحيحة، حسب السوق الذي تم اختياره، فعلى المؤسسة أن تعرف جمهورها المستهدف بالتفصيل، ويساعدها بذلك استخدام الاستطلاعات عبر الإنترنت للحصول على إجاباتهم؛
- بناء شريحة العملاء الخاصة: بعد جمع الردود، تحتاج المؤسسة إلى إجراء تحليل البيانات لإنشاء شرائح ديناميكية فريدة لعلامتها التجارية؛
- اختبار الاستراتيجية: تساعد دراسة وتفسير بيانات الاستبيان الموزعة على مستهلكي الوسق المستهدف بشكل صحيح على إعادة النظر في استراتيجيات تجزئة السوق وإجراء التغييرات اللازمة .

5. الأساليب الإحصائية المستخدمة في تقسيم السوق

يعتبر استخدام الأداة البحثية المناسبة وتحليل وتفسير نتائجها بشكل صحيح من أهم عناصر تقسيم السوق والاستهداف وذلك من اجل تقسيم السوق بطريقة تمكن مسؤول التسويق من اتخاذ القرار بشكل صحيح، يتم اعتماد مجموعة من البرمجيات والأدوات الإحصائية المساعدة على تحديد المجموعات السوقية المستهدفة، والتي تسمح بجمع مختلف البيانات حول هذه الأسواق، ومن ثم إخضاعها لمختلف الأساليب الإحصائية المناسبة، من بينها التحليل التمييزي والتحليل العنقودي.

البرمجيات المساعدة على تحديد مجموعات السوق المقسمة:

- البرامج الإحصائية: SPSS، SAS؛
- أدوات تحليل البيانات عبر الإنترنت: Google Analytics، Tableau؛
- منصات التسويق عبر الإنترنت: Facebook Ads، Google Ads؛
- أدوات استطلاعات الرأي: SurveyMonkey، Qualtrics؛
- أدوات تحليل مواقع الويب: Hotjar، Crazy Egg؛
- أدوات تحليل مواقع التواصل الاجتماعي: Sprout Social، Hootsuite؛
- أدوات تحليل الشخصية: Myers-Briggs Type Indicator (MBTI)، Enneagram؛
- أدوات تحليل القيم: Schwartz Values Survey.

ملاحظة:

- لا توجد أداة واحدة تناسب الجميع لتقسيم السوق، ويجب على المؤسسة اختيار الأدوات التي تناسب احتياجاتها وأهدافها؛
- من المهم استخدام مجموعة متنوعة من الأدوات لجمع بيانات شاملة عن السوق؛
- يجب مراجعة وتحليل البيانات بشكل منتظم لضمان أن تقسيم السوق دقيق ومُحدث.

أما عن التحليل التمييزي والتحليل العنقودي فسيتم شرحهم في القسم الموالي.

المحور الثاني: أدوات تقسيم السوق

التحليل التمييزي والتحليل العنقودي

المحاضرة الثانية: التحليل التمييزي والتحليل العنقودي

1. تمهيد

يُعدّ كل من التحليل التمييزي والتحليل العنقودي تقنيتين إحصائيتين متعددتي المتغيرات تُستخدمان لتحليل مجموعات البيانات في آن واحد، ومن بين أهم صعوبات التعامل معهما صعوبة التفريق بين استعمال كل أسلوب ووقته المناسب (أسلوب التحليل العنقودي والتحليل التمييزي)، إذ يهدف كل منهما إلى التصنيف، من خلال هذا القسم سيتم إعطاء صورة مختصرة لأوجه الاختلاف والشبه بين التحليل العنقودي والتحليل التمييزي.

2. أوجه الشبه والاختلاف بين التحليل العنقودي والتحليل التمييزي وبعض تطبيقاتهما

1.2. أوجه الشبه:

- كلاهما تقنيات إحصائية لتصنيف البيانات: يُستخدم كل من التحليل العنقودي والتحليل التمييزي لتصنيف البيانات إلى مجموعات بناءً على خصائصها؛
- كلاهما يُستخدم في تقسيم السوق: يُستخدم كل من التحليل العنقودي والتحليل التمييزي لفهم سلوك العملاء وتحديد مجموعات العملاء ذوي الخصائص واحتياجات متشابهة؛
- كلاهما يتطلب بيانات كمية: يتطلب كل من التحليل العنقودي والتحليل التمييزي بيانات كمية لكي يعمل بشكل صحيح.

2.2. أوجه الاختلاف:

يمكن التفرقة بين التحليل التمييزي والتحليل العنقودي من خلال الجدول التالي:

الميزة	التحليل التمييزي	التحليل العنقودي
الهدف	تصنيف البيانات إلى مجموعات معروفة مسبقًا	اكتشاف مجموعات متميزة في البيانات
طريقة العمل	يفترض وجود مجموعات معروفة مسبقًا ويستخدم خصائص البيانات لتصنيفها إلى هذه المجموعات	يبحث عن مجموعات متشابهة في البيانات دون أي افتراضات مسبقة
النتائج	مجموعات مسماة	مجموعات غير مسماة
متطلبات البيانات	يتطلب معرفة مسبقة بخصائص المجموعات	لا يتطلب أي افتراضات مسبقة حول البيانات
التفسير	النتائج قابلة للتفسير بسهولة	قد يكون من الصعب تفسير النتائج
التطبيقات	يُستخدم في التسويق الموجه وتحديد العملاء المحتملين	يُستخدم في أبحاث السوق وفهم سلوك العملاء

الجدول: أوجه الاختلاف بين التحليل التمييزي والتحليل العنقودي

3.2. أمثلة على استخدامات كل من التحليل العنقودي والتحليل التمييزي:

التحليل العنقودي:

- تصنيف العملاء حسب سلوكهم الشرائي؛
- تحديد مجموعات من المنتجات ذات خصائص متشابهة؛
- فهم سلوك المستهلكين في السوق.

التحليل التمييزي:

- تصنيف العملاء حسب مخاطر الائتمان؛
- تحديد العملاء المحتملين لحملة تسويقية؛
- تشخيص الأمراض الطبية.

باختصار يمكن القول بأن التحليل العنقودي والتحليل التمييزي:

- تقنيتان إحصائيتان قويتان يمكن استخدامهما لتصنيف البيانات؛
- يُستخدم التحليل العنقودي لاكتشاف مجموعات متميزة في البيانات، بينما يُستخدم التحليل التمييزي لتصنيف البيانات إلى مجموعات معروفة مسبقًا؛
- اختيار التقنية المناسبة يعتمد على احتياجات وأهداف الدراسة (المؤسسة).

3. مفهوم التحليل العنقودي والتحليل التمييزي

يعرف التحليل العنقودي بأنه عبارة عن مجموعة من الأدوات لإنشاء مجموعات(عناقيد) من بيانات متعددة المتغيرات والهدف من ذلك هو تكوين مجموعات ذات خصائص متجانسة من أصل مجموعات كبيرة غير متجانسة".

فيما يعرف التحليل التمييزي بأنه: أحد الطرائق الإحصائية في تحليل البيانات متعددة المتغيرات؛ إذ يهتم بمسألة تمييز K من المجموعات التي تكون متشابهة في كثير من الخصائص (الصفات) اعتمادا على P من المتغيرات المستقلة من خلال استخدام الدالة المميزة والتي هي عبارة عن تركيب خطي للمتغيرات المستقلة.

4. أهداف التحليل العنقودي والتحليل التمييزي

1.4. يمكن تلخيص أهم اهداف استخدام التحليل العنقودي فيما يلي:

a. وصف التصنيفات واستكشافها؛
b. تبسيط البيانات واختزالها، ليتم التعامل معها بشكل مجاميع عوضا عن الكميات الكبيرة من البيانات والتي يصعب التعامل معها؛
c. تحديد العلاقات والتي يمكن التعرف عليها بعد تشكيل القطاعات(المجاميع).

2.4. فيما يهدف التحليل التمييزي إلى:

a. تصميم التركيبات الخطية للمتغيرات الأفضل في التمييز بين فئات المتغير التابع؛

b. التحقق فيما إذا كان هناك فروق ذات دلالة إحصائية بين المجموعات فيما يتعلق بالمتغيرات؛

c. تصنيف المتغيرات التي تسهم بأكبر قدر من الاختلاف بين فئات المتغير التابع؛

d. تقييم دقة التصنيف كنسبة مئوية.

أولا: التحليل التمييزي

I. 1. خطوات التحليل التمييزي

تتمثل خطوات التحليل التمييزي فيما يلي:

A. تحديد المتغير التابع والمتغيرات المستقلة؛

B. التحقق من توافر الشروط اللازمة لإجراء التحليل؛

C. تحديد النموذج والطريقة الإحصائية المقترح استعمالها؛

D. تقدير معالم النموذج المقترح؛

E. اختبار جودة توفيق النموذج المقترح مع تحليل البواقي؛

F. اختبار معنوية مقدرة النموذج المقترح على التنبؤ باستعمال القيم المعرفة سابقا، وتفسير النتائج.

I. 2. أنواع التحليل التمييزي

توجد أنواع للتحليل التمييزي عديده سواء من عدد الدوال التمييزية، أو من الطرق المستعملة في إدخال المتغيرات المستقلة في التحليل.

I. 1.2. من حيث طرق إدخال المتغيرات المستقلة في التحليل

ينقسم التحليل التمييزي من حيث الطرق المستعملة في إدخال المتغيرات المستقلة في التحليل إلى ثلاثة أنواع وهم:

❖ التحليل التمييزي المباشر: إذ يتم إدخال جميع المتغيرات مره واحده إلى التحليل؛

❖ التحليل التمييزي الوهمي: إذ يتم إدخال المتغيرات بناء لما يراه الباحث من أهمية للمتغيرات وبالترتيب الذي يراه مناسب؛

❖ التحليل التمييزي المتدرج: إذ يتم تحديد معيار إحصائي يحدد أولوية إدخال المتغيرات في التحليل.

I. 2.2. من حيث عدد الدوال التمييزية

ينقسم التحليل التمييزي من حيث عدد الدوال التمييزية إلى نوعين وهما:

❖ التحليل التمييزي الخطي: يعتمد التحليل التمييزي الخطي على النماذج الخطية للفصل بين المجموعات، ويشترط تساوي تباينات المجتمع محل الدراسة، وينصف التميز الخطي إلى نوعين هما:

- التميز الخطي لمجموعتين؛
- التميز الخطي لأكثر من مجموعتين.

❖التحليل التمييزي غير الخطي: يعتمد التحليل التمييزي غير الخطي على النماذج غير الخطية للفصل بين المجموعات ويستعمل في حالة عدم تساوي تباينات المجتمع المدروس، ويصنف التمييز غير الخطي إلى:

- التحليل التمييزي التربيعي (Quadratic Discriminant)؛
- الشبكات العصبية متعددة الطبقات (Multi-Layer Perceptron)؛
- شجرة القرار(Decision Trees)؛
- خوارزميات الغابة العشوائية (Random Forest)؛
- خوارزمية الجار الأقرب (K-Nearest Neighbors)؛

من خلال هذا الجزء سيتم توضيح أسلوب التحليل التمييزي الخطي لمجموعتين وذلك لتوضيح خطوات، وشروط، وفرضيات تطبيق التحليل التمييزي.

I. 3. التحليل التمييزي الخطي لمجموعتين

تتمثل أهم خطوات حساب الدالة التمييزية فيما يلي:

1.حساب متوسطات المتغيرات في كل مجموعة وإيجاد الفرق بين المتوسطات

-حساب متوسطات المتغيرات في المجموعة الأولى:

$$\bar{x}_{i(1)} = \begin{bmatrix} \bar{x}_{1(1)} \\ \bar{x}_{2(1)} \\ . \\ . \\ . \\ \bar{x}_{k(1)} \end{bmatrix}$$

-حساب متوسطات المتغيرات في المجموعة الثانية:

$$\bar{x}_{i(2)} = \begin{bmatrix} \bar{x}_{1(2)} \\ \bar{x}_{2(2)} \\ . \\ . \\ . \\ \bar{x}_{k(2)} \end{bmatrix}$$

إذا أن kعدد المتغيرات المستقلة.

حساب الفرق بين متوسط المتغيرات في المجموعتين:

$$d_i = \bar{x}_{i(1)} - \bar{x}_{i(2)} = \begin{bmatrix} \bar{x}_{11} - \bar{x}_{12} \\ \bar{x}_{21} - \bar{x}_{22} \\ \cdot \quad - \quad \cdot \\ \cdot \quad - \quad \cdot \\ \bar{x}_{k(1)} - \bar{x}_{k(2)} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{bmatrix}$$

حيث d هي المسافة بين متوسط المتغيرات.

2. إيجاد التباين والتغاير المشترك بين المجموعتين:

التباين المشترك:

$$v_{ii} = \frac{s_{ii(1)} + s_{ii(2)}}{n_1 + n_2 - 2}$$

التغاير المشترك:

$$v_{ij} = \frac{s_{ij(1)} + s_{ij(2)}}{n_1 + n_2 - 2}$$

إذ يمكن إيجاد S_{ii} وS_{ij} من الصيغ التالية:

$$s_{ii} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$s_{ij} = \sum x_i\, x_j - \frac{\sum x_i \sum x_j}{n}$$

مصفوفة التباين والتغاير المشترك بين المجموعتين:

$$v = \begin{bmatrix} v_{11} & v_{12} & v_{13} & \ldots . v_{1k} \\ v_{21} & v_{22} & v_{23} & v_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ v_{k1} & v_{k2} & v_{k3} & v_{kk} \end{bmatrix}$$

إذ يمثل التباين المشترك بين عناصر القطر الرئيسي للمصفوفة والعناصر المتبقية تمثل التغاير المشترك.

3. بناء الدالة التمييزية:

تعد الدالة التمييزية الخطية تركيب خطي من المتغيرات بناء على الصيغة التالية:

$$\hat{l} = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$$

حيث:

α: معاملات النموذج وتستخدم في عملية التصنيف؛

: عدد المتغيرات؛

x: متجه المتغيرات.

علما أن: $$\alpha = v^{-1} d$$

حيث:

v^{-1}: معكوس مصفوفة التباين والتغاير المشترك؛

d: مصفوفة المسافة بين متوسط المتغيرات من كل المجموعتين.

4. الأهمية النسبية للعوامل المؤثرة (المتغيرات المستقلة)

تحدد هذه المرحلة الأهمية النسبية للمتغيرات المستقلة المؤثرة في عملية التميز والفصل بين المجموعات؛ إذ تعد المعاملات المعيارية ذات اعلى قيمة هي الأكثر أهمية، وتحدد نسبة مساهمة المتغير في عملية التميز بواسطة عامل الارتباط القانوني.

5.اختبار قدرة الدالة على التميز بين المجموعتين

توجد مجموعة من الاختبارات لقياس قدرة الدالة على التميز والفصل بين المجموعتين وهي:

- ♣ اختبار F (F test): ويتم ذلك عن طريق اختبار الفرضية التالية:

الفرضية الصفرية H_0: الدالة ليس لديها قدرة على التميز؛

ويعتمد هذا الاختبار على تكوين جدول تحليل التباين:

Source	SS	DF	MS	F
بين المجموعات	*SSB*	*K-1*	M_{SB}	$\frac{M_{SB}}{M_{SE}}$
الخطأ	*SSE*	*n-K*	M_{SE}	
الكلي	*SST*	*n-1*		

حيث:

$$SSE = D^2 = \widehat{\alpha_1} d_1 + \widehat{\alpha_2} d_2 + \cdots + \widehat{\alpha_k} d_k$$

$$SSB = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)} \times (D^2)^2$$

$$SST = SSE + SSB$$

ويتم حساب F المحسوبة من خلال الصيغة التالية:

$$F = \frac{M_{sB}}{M_{sE}}$$

ولحساب القيمة الجدولية لـ F تستخدم الصيغة التالية:

$$F\ (k\text{-}1,\ n\text{-}K)$$

إذا كانت قيمة F المحسوبة أكبر من القيمة الجدولية، يتم رفض الفرض الصفري (العدمي)، ويتم قبول الفرض البديل، وهو أن الدالة لديها القدرة على التمييز بين المجموعتين.

- ♣ اختبار ويلكس لمدا(A)Wilk's lambda:

ويتم ذلك عن طريق اختبار الفرضية التالية:

الدالة ليس لديها قدرة على التمييز$\mu 1 = \mu 2$:0H

ويحسب من العلاقة:

$$\Lambda = \sum_{i=1}^{k} \frac{1}{1 + \lambda_i}$$

حيث:

λ_i: الجذر الكامن Eigen values لكل المتغيرات؛

k: عدد المتغيرات.

وتنحصر قيمة Λ بين الصفر والواحد الصحيح : $0 \leq \Lambda \leq 1$.

حيث إذا كانت:

- $\Lambda = 1$ ويقصد بها تساوي متوسطات المجوعتين (عدم مقدرة الدالة على التمييز)؛
- $\Lambda = 0$ ويقصد بها عدم تساوي متوسطات المجوعتين (مقدرة الدالة على التمييز) ؛
- في حالة اقتراب قيمة Λ من الواحد الصحيح فذلك يدل على عدم مقدرة الدالة على التمييز، أما في الحالة العكس (اقتراب القيمة من الصفر) دل ذلك على قدرة الدالة على التمييز.

- ♣ اختبار هوتلنج Hotelling-Lawely test (T^2):

ويأخذ الصيغة التالية:

$$T^2 = \sum_{i=1}^{s} \lambda_i$$

حيث:

S عدد المتغيرات؛

ويمكن حساب قيمة F من اختبار هوتلنج باستعمال الصيغة الآتية:

$$F = \frac{n_1 + n_2 - k - 1}{(n_1 + n_2 - 2)k} \times T^2$$

والقيمة الجدولية:

$$F_\alpha = (k - 1, n_1 + n_2 - k - 1)$$

إذا كانت F المحسوبة أكبر من F الجدولية فإن الدالة تملك قدرة عالية على التمييز.

♣ نقطة الفصل Cut Off Point:

تستخدم نقطة الفصل بين مجموعتين غرض تصنيف مفردة معينة إلى المجموعة الأقرب لها، فإذا كانت قيمة الدالة بعد تعويض قيم المفردة فيها أكبر من هذه النقطة فالمفردة تعود إلى المجموعة الأولى، أما إذا كانت قيمة الدالة أكبر من هذه النقطة فالمفردة تعود إلى المجموعة الثانية، ويتم ذلك من حساب الصيغة الآتية:

$$\bar{\bar{l}} = \frac{\bar{l}_{(1)} + \bar{l}_{(2)}}{2}$$

حيث:

$\bar{\bar{l}}$ نقطة الفصل؛

$\bar{l}_{(1)}$: متوسط القيم التميزية للمجموعة الأولى؛

$\bar{l}_{(2)}$: متوسط القيم التميزية للمجموعة الثانية.

يتم تصنيف المفردة إلى احدى المجموعات من اتباع قاعدة التصنيف وهي:

إذا كان $\bar{L}_1 > \bar{L}_2$

❖ إذا كانت القيمة التميزية للمفردة الجديدة أكبر من نقطة الفصل : تصنف المفردة الجديدة ضمن المجموعة الأولى؛

❖ إذا كانت القيمة التميزية للمفردة الجديدة أقل من نقطة الفصل: تصنف المفردة الجديدة ضمن المجموعة الثانية؛

إذا تساوت القيمة التميزية للمفردة الجديدة ونقطة الفصل: تصنف المفردة الجديدة عشوائيا ضمن أي مجموعة من المجموعتين.

إذا كان: $\bar{L}_1 < \bar{L}_2$

❖ إذا كانت القيمة التميزية للمفردة الجديدة أعلى من نقطة الفصل: تصنف المفردة الجديدة ضمن المجموعة الثانية؛

❖ إذا كانت القيمة التميزية للمفردة الجديدة أقل من نقطة الفصل: تصنف المفردة الجديدة ضمن المجموعة الأولى؛

❖ إذا تساوت القيمة التميزية للمفردة الجديدة مع نقطة الفصل: تصنف المفردة الجديدة عشوائيا ضمن أي مجموعة من المجموعتين.

I. 4. أخطاء التصنيف

يقصد بأخطاء التصنيف حدوث خطأ في تصنيف البيانات، حيث يتم تصنيفها في فئات خاطئة، وذلك بوضع المفردة في مجموعة غير المجموعة التي تنتمي إليها، ويعتبر خطأ التصنيف عامل مهم عند الحكم على كفاءة الدالة التمييزية.

يوجد نوعان من أخطاء التصنيف وهي:

- احتمال خطأ التصنيف $P12$: ويقصد به احتمال تصنيف المفردة إلى المجموعة الثانية فيما تنتمي إلى المجموعة الأولى؛
- احتمال خطأ التصنيف $P21$: ويقصد به احتمال تصنيف المفردة إلى المجموعة الأولى وهي تنتمي إلى المجموعة الثانية.

ويمكن تقدير احتمال خطأ التصنيف من الصيغة الآتية:

$$P_{12} = P_{21} = \Phi\left(\frac{-D}{2}\right)$$

حيث:

Φ: تمثل دالة التوزيع الطبيعي القياسي؛

D: يمثل مقياس مهالانوبيس(Mahalanobis Distance.)

ثانيا: التحليل العنقودي

II. 1. خطوات التعنقد (Clustering Steps):

تتمثل أهم خطوات تطبيق التحليل العنقودي فيما يلي:

G. حساب مصفوفة التباعد أو مصفوفة التشابه؛

H. ربط العنصرين اللذين تكون المسافة بينهما أقصر المسافات ضمن المصفوفة، لتشكيل العناقيد الأولية؛

I. حساب مصفوفة المسافة الجديدة بعد تشكيل العناقيد الأولية والاستمرار بعملية ربط العناصر؛ اعتمادا على المسافة بينهما إلى أن يربط العنقودان الآخران في نهاية التحليل.

والتي سيتم ربطها مع قسمي التحليل العنقودي تاليا (سيتم تقسيم هذه الخطوات تاليا ضمن نوعي التحليل العنقودي).

II. 2. أنواع التحليل العنقودي

يتفرع التحليل العنقودي عموما إلى نوعين هما:

II. 1.2. التحليل العنقودي الهرمي(Hierarchical)

عرف بأنه: "لا يتطلب التحليل العنقودي الهرمي المعرفة المسبقة بعدد العناقيد التي سيتم تصنيف الحالات على أساسها حيثً يناسب التحليل العنقودي الهرمي العينات الصغيرة نسبيا".

ويقسم هذا التحليل على:

- التحليل العنقودي الهرمي للحالات(المفردات)؛
- التحليل العنقودي الهرمي للمتغيرات.

كما يتم الإشارة إلى وجود أسلوبين علميين لعنقدة مفردات العينة باستعمال التحليل الهرمي وهما:

- أسلوب التجميع (The Agglomerative Technique):

يبدأ هذا الأسلوب من التحليل بعنقود واحد لكل حالة ثم تجمع العناقيد المتشابهة تدريجيا حتى يصل إلى العدد المطوب.

- أسلوب التقسيم (The Divisive Technique):

يبدأ هذا الأسلوب بافتراض أن جميع الحالات تتجمع في عنقود واحد ثم يتم تصنيف الحالات في عناقيد أصغر فأصغر.

A. حساب مصفوفة التباعد أو مصفوفة التشابه

A. 1. حساب مصفوفة التباعد (Dissimilarity) للمتغيرات الكمية:

عرفت مصفوفة التباعد بتألفها من قيم المسافات بين أفراد العينة، ويرمز لها بالرمز D وهي كالآتي:

$$D=\begin{bmatrix} 0 & d12 & d13 \dots\dots & d1n \\ & 0 & d23 \dots\dots & d2n \\ & & 0 \quad djk & djn \\ & & & 0 \\ & & & & 0 \end{bmatrix}$$

حيث إن العنصر d_{jk} هو المسافة بين المفردتين k وj، يلاحظ أن هذه المصفوفة هي مصفوفة مثلثية عليا (تمثلك شكل مثلث في النصف العلوي منها) وعناصر قطرها الرئيسي يساوي أصفارا (صف الأصفر من أعلى اليسار إلى أدنى اليمين) لأن المسافة بين المفردة ونفسها تساوي صفرا (المسافة بين المفردة j_1 وَj_0 تساوي الصفر).

يتم حساب المسافة باستخدام المسافة الإقليدية وفق الصيغة الآتية:

$$D_e = \sqrt{\sum_{i=1}^{d} (x_{ij} - x_{ik})^2}$$

A. 2. حساب مصفوفة التشابه أو التقارب (Similarity):

لحساب مصفوفة التقارب يتم اتباع الخطوات التالية:

أولا: تحويل المتغيرات النوعية (رتبيه، اسمية) إلى متغيرات ثنائية وتأخذ احدى القيمتين واحد أو صفر، (1) عند التحقق و(0) عند عدم التحقق؛

ثانيا: حساب التقارب بين كل مفردتين من إنشاء جدول التوافق؛

ثالثا: حساب عناصر مصفوفة التقارب ويرمز لها S، من استعمال المقياس الرياضي الآتي:

$$s_{ik} = \frac{a+d}{p} = \frac{\text{عدد أزواج المتشابهة}}{\text{عدد المتغيرات المؤثرة}}$$

من استعمال المقياس تنتج مصفوفة التقارب وهي كالآتي:

$$S = \begin{vmatrix} 1 & & & \\ s_{21} & 1 & & \\ s_{31} & s_{32} & 1 & \\ s_n & s_{n2} & \ldots\ldots\ldots & 1 \end{vmatrix}$$

يلاحظ أن المصفوفة مثلثية سفلى من الرتبة **n*n** كما أن عناصر قطرها الرئيسي تساوي 1.

B. طرق التعنقد الهرمية (Hierarchical Clustering Method):

يطلق عليها بقوانين الربط، وهي كالآتي:

B. 1. طريقة الربط المنفرد (Linkage Single):

وتسمى أيضا بطريقة الجوار الأقرب (Neighbor Nearest)، حسب هذه الطريقة كل مفردة تشكل عنقودا خاصا ومن ثم تضاف الترابطات الأقوى بين المفردات لتجميع العناصر وتشكيل العناقيد، وتعتمد على حساب مصفوفة التباعد أو مصفوفة التقارب، ومن ثم تتم عملية الربط بالاعتماد على أقل مسافة بين أزواج المفردات وربطها معا بحسب الصيغة الآتية:

$$(A, B) = \text{Min}\ \{d\ (y_i , y_j)\ , y_i \in A\ , y_j \in B\}$$

إذ **i,j** تمثل العناصر في العناقيد، **$d\ (y_i , y_j)$** هي المسافة المحسوبة من مصفوفة التباعد باستعمال أحدى المقاييس الرياضية .

B. 2. طريقة الربط الشامل (Complete linkage):

ويطلق عليها أيضا الجوار الأبعد (Furthest Neighbor)، تعمل هذه الطريقة بشكل معاكس للطريقة السابقة؛ إذ تحدد المسافات بين العناقيد بأكبر مسافة بين أي عنصرين ضمن العناقيد المختلفة (أبعد جوار)، ولهذا فعملية الحساب تبدأ بالبحث عن أصغر عنصر في المصفوفة **D** ثم يتم تحديد العنقودان الأكثر تقاربا من الصيغة الآتية:

$$D(A, B) = \text{Max}\ \{d(y_i , y_j)\ , y_i \in A\ , y_j \in B\}$$

B. 3. طريقة الربط المتوسط (Average Linkage):

في هذه الطريقة يتم تحديد المسافة بين عنقودين باستعمال معدل التقاربات (المسافة) الزوجية بين كل أزواج العناصر في العناقيد المختلفة، ويمثل هذا أسلوب متوسط بين طريقتي ربطMinوMax" ويُعبر عن ذلك بالمعادلة الآتية:

$$proximity(s_1, s_2) = \frac{\sum_{x_1 \in s_1} \sum_{x_2 \in s_2} proximity(x_1, x_2)}{size(s_1) * size(s_2)}$$

B. 4. طريقة الربط المركزية (Centroid):

تعرف المسافة بين العنقودين في الطريقة المركزية على أنها المسافة الإقليدية بين متجهي الوسط الحسابي للعنقودين.

B. 5. طريقة الوسيط (Median):

تستخدم في حالة كانت عدد مفردات أحد العناقيد أكبر من الأخرى.

B. 6. طريقة وورد (Ward's method):

تعتبر من أكثر الطرق استعمالا؛ إذ تتبع سلسلة من خطوات التجميع التي تبدأ بالعناقيد ويحوي كل منها عنصرا واحدا، وتنتهي بمجموعة واحدة تشمل العناصر جميعا، كما يطلق عليها طريقة أصغر تباين (Minimum Variance Method):

لأنها تستعمل أسلوب تحليل التباين ومؤشرات مجموع المربعات بين العناقيد.، ويكون التعبير عنها بالصيغ الآتية:

$$SSE_A = \sum_{i=1}^{nA} (y_i - \bar{y}_A)'(y_i - \bar{y}_A)$$

حيث SSE_A هي مجموع مربعات الخطأ في العنقودA،nA هي عدد المشاهدات الكلية في العنقود.

$$SSE_B = \sum_{i=1}^{nB} (y_i - \bar{y}_B)'(y_i - \bar{y}_B)$$

حيث SSE_B هي مجموع مربعات الخطأ في العنقودB،nB هي عدد المشاهدات الكلية في العنقود.

$$SSE_{AB} = \sum_{i=1}^{nAB} (y_i - \bar{y}_{AB})'(y_i - \bar{y}_{AB})$$

حيثSSE_{AB} هي مجموع مربعات الخطأ للعنقود الناتج من ربط العنقودينAوB، nAB هي عدد المشاهدات الكلية في العنقود.

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$$

حيث I_{AB} هو مقدار الزيادة الناتج عن ربط العنقودين بحيث يقلل الزيادة في مربع المسافات.

II. 2.2. التحليل العنقودي غير الهرمي (Non-Hierarchical):

يطبق التحليل العنقودي غير الهرمي يطبق على العينات الكبيرة، ويستعمل لتجميع بيانات المفردات الكثيرة أكثر مما يستعمل لتجميع المتغيرات ضمن عنقود محدد؛ إذ يعتمد على أسلوب التقسيم، ويمكن تحديد عدد العناقيد سابقا، شخصيًّا، أو من أسلوب العنقدة.

وتوجد طرائق عديدة للتحليل العنقودي غير الهرمي، إذ تعد طريقة K متوسط أفضلها وذات فعالية عالية للتعامل مع البيانات الضخمة.

❖ طريقة K متوسط (K-means):

تقوم هذه الطريقة على أساس تصنيف الحالات (المفردات) في مجموعات متجانسة من حيث الخصائص والصفات، وذلك باستخدام خوارزميات يمكنها معالجة عدد كبير من الحالات، وتسمى هذه الطريقة بطريقة التحليل العنقودي السريع لأنها تقوم بعملية التحليل والتصنيف في وقت قصير نسبيا، وتمثل الخطوات التالية خطوات تطبيق أسلوب المتوسطات:

1. تحديد عدد العناقيد المطلوب لإجراء عملية التجميع (العنقدة)؛
2. حساب البعد بين كل مفردة وبين جميع المراكز باستخدام المسافة الإقليدية؛
3. ضم كل مفردة إلى العنقود الأقرب اليها؛
4. تحديد مراكز العناقيد من إيجاد مراكز المفردات الموجودة في كل عنقود، وتحدد المراكز من حساب متوسط المفردات التي تنتمي للعنقود؛
5. تكرار الخطوات من2 إلى4حتى الوصول إلى الاستقرار (أي عدم توافر مفردات تنتقل ضمن العناقيد) أو عند عدد معين من التكرارات.