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 الأولى: أدوات تقسيم السوق -التحليل العنقودي-

1. تمهيد:

يؤدي التحليل العنقودي دورا مهما في الكثير من الميادين مثل: علم النفس، إحصاء، العلوم الاجتماعية، بيولوجيا، ... الخ، وفي التسويق تقوم الشركات بجمع بيانات كثيرة عن العملاء الحاليين والمحتملين، وهنا يساعدها التحليل العنقودي في تقسيم العملاء إلى عدد صغير من المجموعات، بهدف استخدامها في تحليلات إضافية وأنشطة التسويق، فمثلا تصنيف الأشخاص، المناطق والمنتجات إلى مجاميع عنقودية ذات صفات متشابهة، يساعد في تحديد قطاعات السوق والشركات المنافسة.

إذن التحليل العنقودي هي أداة إحصائية تستخدم في مجال الأبحاث التسويقية لوضع المستهلكين الذين يتشابهون في نفس الصفات ضمن مجموعات محددة، مما يسهل على المؤسسة تصميم المنتجات بشكل يتناسب مع كل مجموعة، كما يسهل عليها وضع الخطط التسويقية المناسبة لكل مجموعة.

كما يستخدم التحليل العنقودي لتجميع أي مجموعة من المتغيرات ضمن مجموعات متجانسة ومنفصلة، ويطبق هذا الاتجاه في مراجعة وتنقيح الاستبيانات بناءً على إجابات المستجوبين على استمارات الاستبيان، حيث أن تجميع الأسئلة حسب المتوسطات بواسطة التحليل العنقودي يساعدنا في التعرف على الأسئلة الضعيفة واعادة النظر فيها، وهذا يزيد حظوظ معدل الإجابات الجيدة بالنسبة لإجمالي أسئلة الاستبيان.

2. تعريف التحليل العنقودي:

التحليل العنقودي هو: "أحد الأساليب الاحصائية الرياضية لتقسيم عناصر المجتمع المدروس إلى عدة مجموعات متعالقة ومتجانسة داخليا (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".

التحليل العنقودي عبارة عن: "إجراءات تهدف إلى تصنيف مجموعة حالات Cases (أو متغيرات Variables) بطرق معينة وترتيبها داخل عناقيد Clusters، بحيث تكون الحالات المصنفة داخل عنقود معين متجانسة فيما يتعلق بخصائص محددة، وتختلف عن حالات أخرى موجودة في عنقود آخر".

3. أنواع التحليل العنقودي:

بصورة عامة يتفرع التحليل العنقودي إلى نوعين أساسيين، كل منهما ينقسم إلى طرق فرعية عديدة، هذين النوعين هما:

أ. التحليل العنقودي الهرمي Hierarchical

هذه الطريقة لا تتطلب المعرفة المسبقة لعدد العناقيد المسبقة التي سيتم تجميع العناصر على أساسها، وهي تناسب العينات الصغيرة نسبيا.

ب. التحليل العنقودي غير الهرمي Non- Hierarchical

في هذه الطريقة يكون عدد العناقيد محدد مسبقا أو معين كجزء من إجراءات التحليل العنقودي، ومن بين أهم أشكالها: "طريقة ك- متوسطات (K-means)، والتي تعتمد في التصنيف على تجميع العناصر المدروسة في k عنقود أولي، ثم حساب مراكز(النقطة المتوسطة) هذه العناقيد، والمسافات بين العناصر والمراكز، ثم نقل العناصر إلى العناقيد أقرب إليها.

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 عنقودا، بواسطة دمج المفردات المتقاربة ضمن مجموعات متعالقة تسمى عناقيد، وبحيث يكون العنقود الأول C_1 أصغر وأبسط العناقيد، ويكون العنقود الأخير C_m أعقدها وأشملها (لأنه يضم جميع المفردات والعنقيد الجزئية)، على أن يتألف كل عنقود من عدة

مجموعات متقاربة ومرتبطة مع بعضها بواسطة علاقات تحقق شروط التقارب المفضلة (حسب المتغير أو الصفة المدروسة).

4. أساليب التحليل العنقودي الهرمي:

يستخدم التحليل العنقودي الهرمي لعنقدة مفردات العينة أسلوبين عمليين ما:

أ. أسلوب التجميع Agglomerative Technique

يفترض 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.

ب. أسلوب التقسيم Divisive Technique

يفترض هذا الأسلوب من البداية أن جميع مفردات العينة n (مفردة) تشكل عنقوداً واحداً شاملا، 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.

5. متطلبات التحليل العنقودي التجميعي:

إن أسلوبي التجميع والتقسيم يعتمدان على عدد عناصر العينة المدروسة (n مفردة أو مشاهدة)، وعلى طبيعة المتغيرات المستقلة $x_1, x_2, \ldots x_p$ المستخدمة في عملية العنقدة، على أن تنظم بيانات مفردات العينة حسب المتغيرات النظامية (معيارية أو ثنائية)، وتوضع في جدول مناسب كما يلي:

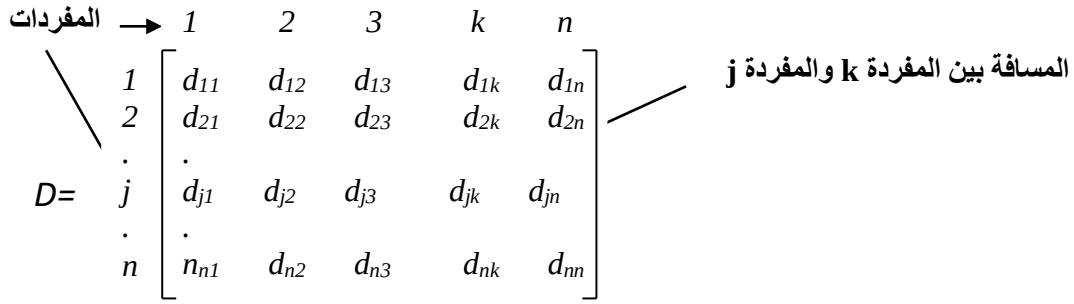
نموذج جدول البيانات اللازمة للتحليل العنقودي

المتغيرات/المفردات	X_1	X_2	X_3	X_i	X_p
1	X_{11}	X_{12}	X_{13}	$X_{1\,i}$	X_{1p}
2	X_{21}	X_{22}	X_{23}	X_{2i}	X_{2p}
.					
.					
J	X_{j1}	X_{j2}	X_{j3}	X_{ji}	X_p
.					
.					
n	X_{n1}	X_{n2}	X_{n3}	X_{ni}	X_{np}

إذا كانت المتغيرات $x_1, x_2, \ldots x_p$ كمية، فإننا نقوم بحساب عناصر مصفوفة التباعد Dissimitlarity(D) ، وهي عبارة عن المسافات التي تفصل بين مفردات العينة. أما إذا كانت المتغيرات نوعية أو مختلطة، فإننا نقوم بحساب عناصر مصفوفة أخرى مصفوفة التشابه أو التقارب Similarity(S)، وهي أوزان التكرارات التي تقابل المفردتين المتشابهتين (j, k).

5. حساب مصفوفة التباعد (حالة متغيرات كمية):

تتألف مصفوفة التباعد من قيم المسافات بين أزواج مفردات العينة، وتحسب من قيم المتغيرات المستخدمة في عملية العنقدة، لذلك نفترض أنه لدينا n مفردة هي 1 2 3 ... j...n، نريد تصنيفها ضمن عناقيد حسب قيم المتغيرات المؤثرة عليها، والتي سنرمز لها بـ $x_1, x_2, \ldots x_p$، وسنستخدم قيم هذه المتغيرات لحساب عناصر مصفوفة التباعد أو مصفوفة المسافات، والتي سنرمز لها بـ D ونكتبها كمايلي:



حيث أن: العنصر d_{jk} هو المسافة بين المفردتين (k , j) وهنا نلاحظ أن هذه المصفوفة هي مصفوفة مربعة ومتناظرة (لأن: $d_{jk}=d_{kj}$)، وأن عناصر قطرها الرئيسي تساوي أصفارا (لأن المسافة بين النقطة j ونفسها: $d_{jj}=0$)، ولهذا يمكن كتابتها اختصارا على شكل مصفوفة مثلثية عليا كما يلي:

$$D=\begin{array}{c} \\ 1 \\ 2 \\ . \\ j \\ . \\ n \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & k & n \end{array} \\ \begin{bmatrix} 0 & d_{12} & d_{13} & d_{1k} & d_{1n} \\ & 0 & d_{23} & d_{2k} & d_{2n} \\ & & 0 & & \\ & & & 0 & d_{jk} \quad d_{jn} \\ & & & & 0 \\ & & & & 0 \end{bmatrix} \end{array}$$

ولكن حساب هذه المسافات يتعلق بطبيعة المتغيرات، فهل **هي مستمرة أم منقطعة؟** كما أنها تتعلق بطبيعة المسافة المراد حسابيا، **فهل هي المسافة النظرية أم المسافة الفعلية؟** لذلك فإننا نعرف هذ المسافات حسب هذه الحالات، فعندما تكون المتغيرات X كمية (عددية)، فإن المشكلة الوحيدة التي تعترض حساب المسافة بين أي مفردتين هي وحدات القياس لتلك المتغيرات (موحدة أم مختلفة).

فإذا كانت وحدات القياس لجميع المتغيرات X كمية موحدة (كدخل الأسرة أو نفقاتها على الغذاء، اللباس، السكن، النقل الاتصالات... الخ)، فإننا نقوم بحساب المسافة بين أي مفردتين حسب المسافة الإقليدية:

$d_{jk}^2= (x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2$: حالة متغيرين فقط X_1 وX_2 فقط، و j و k هما المفقردتان.

$d_{jk}^2= (x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2 + \dots + (x_{pj} - x_{pk})^2$: حالة p متغير، و j و k هما المفقردتان.

أما إذا كانت وحدات القياس للمتغيرات X كمية مختلفة (كالدخل وعدد أفراد الأسرة ومساحة السكن... الخ)، فإننا نقوم بتحويل هذه المتغيرات إلى متغيرات معيارية Z وفق العلاقة التالية: $z_i= (x_i-\bar{x}_i)/\sigma_i$

فنحصل على المتغيرات المعيارية: $Z_1, Z_2, \dots Z_p$، التي تتميز بأن متوسط كل منها يساوي 0، وتباينها (وانحرافها المعياري) يساوي 1، ثم نتعامل معها كالعادة.

وبعد حساب المسافات بين كل المفردات d_{jk}، وتشكيل مصفوفة التباعد D، نقوم بتجميع المفردات حسب الأقرب فالأقرب، ورياضياً تعتبر النقطة j أقرب إلى النقطة k من أية نقطة أخرى ℓ إذا كانت المسافة بينهما d_{jk} تحقق العلاقة:

$$d_{jk} < d_{j\ell} \qquad \ell = 1 \dots n,\ \ell \neq k$$

6. حساب مصفوفة التشابه أو التقارب (حالة متغيرات نوعية: اسمية أو ترتيبية):

عندما تكون المتغيرات X نوعية ترتيبية أو اسمية كمستوى التعليم أو الحالة الاجتماعية، فإنه لا يمكننا حساب المسافات بين المفردات بطريقة المسافة الإقليدية، لذا نلجأ إلى استخدام التكرارات المطلقة المقابلة لحالات أو فئات تلك المتغيرات، لأن الحالات المتشابهة تكون تكرارتها متقاربة.

- فإذا كان المتغير X ثنائيا، 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 للمتغير X أكثر من حالتين، فإننا نعتبر كل حالة مستقلة عن الحالات الأخرى، ونعرف عليها متغيرات ثنائية جديدة ($X_1', X_2' \dots X_P'$)، نفترض أن كل متغير جديد X_t' ، يأخذ القيمة (1) عندما تتحقق حالة معينة، ويأخذ القيمة (0) عندما لا تتحقق تلك الحالة(أي كل الحالات الأخرى). ولحساب نسبة التقارب S_{jk} بين أي مفردتين (j و k)نقوم بتطبيق الصيغة التالية:

$$S_{jk} = \frac{\text{عدد الأزواج المتشابهة}}{\text{عدد المتغيرات P}} = \frac{a+d}{P}$$

حيث a عدد المتغيرات X'_i التي تحصل عندها المفردتين j وk على نفس القيمة 1، وb هو عدد المتغيرات X'_i التي تحصل عندها المفردتين j وk على نفس القيمة 0 في نفس الوقت.

وأخيرا نحصل على عناصر مصفوفة التقارب S ونكتبها كما يلي:

$$S = \begin{array}{c} \\ 1 \\ 2 \\ . \\ j \\ . \\ n \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & k & n \end{array} \\ \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{1k} & S_{1n} \\ S_{21} & S_{22} & S_{23} & S_{2k} & S_{2n} \\ . & & & & \\ S_{j1} & S_{j2} & S_{j3} & S_{jk} & S_{jn} \\ . & & & & \\ S_{n1} & S_{n2} & S_{n3} & S_{nk} & d_{nn} \end{bmatrix} \end{array}$$

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 ($S_{jj}=1$)، لذلك يمكننا كتابة مصفوفة التقارب S على شكل مصفوفة مثلثية سفلى (لتمييزها عن مصفوفة التباعد D) كما يلي:

$$S = \begin{array}{c} \\ 1 \\ 2 \\ . \\ j \\ . \\ n \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & k & n \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ S_{21} & 1 & 0 & 0 & 0 \\ . & & & & \\ S_{j1} & S_{j2} & 1 & 0 & 0 \\ . & & & 1 & 0 \\ S_{n1} & S_{n2} & S_{n3} & \dots & 1 \end{bmatrix} \end{array}$$

ملاحظة (1): كل عناصر مصفوفة التقارب S تأخذ قيم كسرية في المجال [0, 1]. عكس عناصر مصفوفة التباعد D التي تأخذ قيمها في المجال [0, +∞[.

ملاحظة (2): يمكن حساب عناصر مصفوفة التقارب S_{jk} انطلاقا من عناصر مصفوفة التباعد d_{jk} من خلال العلاقة:

$$S_{jk} \quad \frac{1}{1+ d_{jk}} \qquad 0 \leq S_{jk} \leq 1$$

ملاحظة (3):

يمكن معالجة المتغيرات النوعية أو المختلطة دون تحويلها إلى متغيرات ثنائية، حيث نستبدلها بأرقامها وندخلها إلى الجدول الأساسي لبيانات العينة، مثلا: النوع X_1 (0 أنثى، 1 ذكر)؛ X_2 الدخل (1 منخفض، 2 متوسط، 3 مرتفع)؛ X_3 الحالة الاجتماعية (1 أعزب، 2 متزوج، 3 مطلق، 4 أرمل)؛ X_4 العمر بالسنوات (1 أقل من 30، 2 لـ [30,50[، 3 لـ 50 فأكبر)؛ ...

ثم نعرف على بيانات المفردتين (j وk) متغير جديد Z، يأخذ القيمة (1) إذا كانت إجابة المفردة j مساوية لإجابة المفردة k، ويأخذ القيمة (0) إذا كانت إجابة j مختلفة عن إجابة k، ثم نقوم بحساب عناصر مصفوفة التقارب s_{jk} من خلال حساب المتوسط الحسابي لقيم المتغير Z بالمنسبة لكل مفردتين i و k بتطبيق العلاقة التالية:

$$s_{jk}= \frac{\sum_i z_{jk}}{p} \qquad i= 1 \ldots P ; \quad J, k= 1.. n$$

7. خطوات التحليل العنقودي الهرمي التجميعي:

إن عملية العنقدة الهرمية التجميعية لـ n مفردة (مشاهدات أو متغيرات)، 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 تشابها من بين العناقيد المدروسة، وذلك من أجل دمجهما وتشكيل عنقود جديد منهما.

3- ندمج العنقودين u وv في عنقود واحد، ونرمز له بـ: (uv)، ثم نقوم بتحديث العناصر المتقاطعة في مصفوفة المسافات D كما يلي:

أ. نحذف العمودين والسطرين المقابلين للعنقودين u وv من المصفوفة.

ب. نضيف سطرا وعمودا جديدين، ونضعهما في مكان u 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) المتبقية في المصفوفة، ثم نضعها في العمود والسطر المخصصين لـ (uv) في المصفوفة، علما 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).

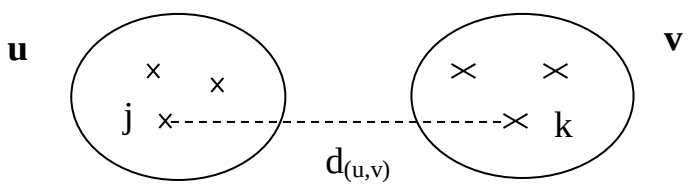
د. نكرر الخطوات (2) و(3) حتى (n-1) مرة حتى تصبح جميع المفردات في عنقود واحد وفي كل مرحلة نقوم بتسجيل مواصفات وخواص العناقيد المدمجة ومستوياتها الجديدة (عناصر المسافات أو التشابه) في الأماكن المخصصة للدمج.

هـ. نقوم برسم مخطط الأغصان لكل مرحلة من هذه المراحل، وأخيرا ندمجها في المخطط العنقودي العام.

أ. طريقة الربط المنفرد Single Linkage

تبدأ من اعتبار كل مفردة من المفردات تشكل عنقوداً خاصا، وتعتمد على مصفوفة المسافات D (أو التشابه S) بين أزواج المفردات المدروسة، ويتم تشكيل العناقيد فيها من دمج العناقيد الأكثر تقاربا (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 (u) و(v) بأصغرهما (بأقربهما)، أي نطبق العلاقة: $d_{(u,v)= \min[d_{jk}]} \quad j\in u, k\in v$؛ حيث j ينتمي إلى u و k ينتمي إلى v.



ولحساب شعاع المسافات بين العنقود الجديد وأي عنقود آخر(أو مفردة w) نطبق العلاقة التالية:

$$d_{(u,v)w} = \min [d_{uw} , d_{vw}]$$

حيث أن: d_{uw} و d_{vw} هما المسافتان بين العنقود الأكثر تقاربا u مع العنقود w، والعنقود الأكثر تقاربا v مع العنقود w على الترتيب.