SPSS

XLSTAT

تحليل المعطيات

د. خالد علي

Data analysis

تحليل المعطيات

مع التحليل باستخدام برنامجي: الحزم الإحصائية للعلوم الاجتماعية (SPSS)، برنامج XLSTAT

مطبوعة موجهة إلى طلبة السنة الثالثة اقتصاد كمي

د. خالد علي

SPSS®

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة الشهيد حمة لخضر –الواد–

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

مطبوعة بيداغوجية في مقياس تحليل المعطيات

المقياس: تحليل المعطيات.

التخصص: اقتصاد كمي.

المستوى: السنة الثالثة.

تقديم: د. خالد علي

قسم العلوم الاقتصادية

السنة الجامعية: 2023/2022

فهرست المحتويات

الفصل الثاني: الفضاء الشعاعي وحساب المصفوفات

الفصل الثالث: التحليل بالمركبات الأساسية (PCA)

الفصل الرابع: التحليل العاملي للتوفيقات (AFC)

الفصل الخامس: التصنيف التسلسلي الهرمي (AHC)

تقديم:

غالبا ما نطرح العديد من الأسئلة إزاء حياتنا ومحيطنا الذي نعيش فيه، وعن الكيفية التي نتخذ بها القرار الصائب لمواجهة التحديات التي تعترض سبيلنا، فعندما نحاول البحث عن الجواب الشافي لتلك الأسئلة سنجمع كما هائلا من البيانات والمعطيات المبعثرة التي قد لا تفيدنا في اتخاذ القرار، لأنّ النظر إلى المعطيات وفحصها بغرض الوصول إلى العلاقات التي تحكمها لا يفيدنا في عملية التعميم أو التنبؤ، وليس بالقول الفصل في تحديد المشكلة، فالارتباط لا يستلزم السببية، إذ أنك اكتشفت علاقة بين متغيرين على أساس التحليل الاستكشافي وحده.

وبناء عليه، تهدف هذه المادة التعليمية إلى تزويد الطالب بأهم المعارف النظرية والتطبيقية، والأدوات والأساليب الاحصائية والإستدلالية التي تمكنه من فهم العلاقات بين الظواهر المختلفة والتحليل والتفسير والتنبؤ بالأحداث المستقبلية. إضافة إلى تزويده بالتقنيات الإحصائية التي تستخدم لتحليل العلاقات بين المتغيرات المتعددة وتقليل الأبعاد، ما يمكنه من تحويل مجموعة كبيرة من المتغيرات الأولية إلى مجموعة صغيرة من المتغيرات المستخلصة، التي تساعده في تحليل العلاقات بين المتغيرات وفهم الأنماط الرئيسية في البيانات.

ولقد بني هذا العمل على أساس منهجي يستند على ما هو مقرر في البرنامج الوزاري لطبة السنة الثالثة، مبتعدين قدر الامكان عن البراهين الرياضية التي قد لا تفيد الطالب في هذه المرحلة بالقدر الذي قد تكون سببا في تعقيد المادة المدروسة. ولقد حاولنا في هذا العمل سلوك سبيل بسيط في تقديم المادة المدروسة، انطلاقا من عرض بعض المفاهيم الرياضية والاحصائية اللازمة لفهم الطرائق التي سيتم عرضها، مدعمين ذلك بأمثلة وتمارين، إضافة إلى ذلك لم يخلو هذا العمل من الرسومات التوضيحية التي تيسّر فهم العديد من الجوانب المتعلقة بطرائق تحليل المعطيات.

مدخل إلى تحليل المعطيات

تمهيد:

تحليل المعطيات عملية لتنظيف وتحويل ونمذجة البيانات لاكتشاف معلومات مفيدة من أجل اتخاذ قرارات الأعمال. والغرض من تحليل المعطيات هو استخراج معلومات مفيدة، واتخاذ القرار بناءً على مخرجات التحليل، فعندما نتخذ أي قرار في حياتنا اليومية يكون من خلال التفكير فيما حدث في المرة السابقة أو ما سيحدث باختيار هذا القرار المحدد، هذا ليس سوى تحليل لماضينا أو مستقبلنا واتخاذ القرارات بناءً عليه، من أجل ذلك، نجمع ذكريات ماضينا أو أحلامنا في المستقبل. وما ذلك إلاّ تحليلا للمعطيات، ونفس الشيء الذي يفعله المحلل لأغراض العمل يسمى تحليل المعطيات.

ونظرا لأهمية المناهج الكمية في تشخيص الظواهر المختلفة، أصبح الاهتمام بها من الأولويات، خاصة عندما يتعلق الأمر بعملية اتخاذ القرار. فتحليل المعطيات كجزء من هذه المناهج تساهم بشكل كبير في إعطاء تفسيرات لموضوع البحث، وذلك باستخدام الأدوات الإحصائية المختلفة، وبالتالي تعتبر وسيلة لتحقيق غاية ما.

ما المقصود بالتحليل؟

التحليل يعني التفكيك والدراسة والفحص والتقييم وتحديد الأساسي...إلخ، من أجل الفهم الجيد للعلاقات بين الكل والأجزاء بشكل أفضل.

المعطيات: ما هو معروف ومقبول منطقيا كنقطة انطلاق للبحث.

المعطيات ⟵ المعالجة ⟵ المعلومات.

مفهوم تحليل المعطيات:

يسمى أيضًا بتحليل البيانات الاستكشافية، وهو مجموعة من الأساليب الإحصائية متعددة الأبعاد والوصفية، تساعدنا في فحص البيانات وتدقيقها لتكون أكثر دقة، ومن ثمة إعادة تشكيلها وتخزينها لنحصل في النهاية على معلومات يمكن على أساسها اتخاذ القرارات.

وهي مجموعة فرعية ممّا يسمّى عمومًا بالإحصاءات متعددة المتغيرات، تساعد طرق معينة- في معظمها هندسيًا- على إبراز العلاقات التي قد توجد بين البيانات المختلفة واستخلاص المعلومات الإحصائية منها، مما يجعل من الممكن وصف المعلومات الرئيسية الواردة في هذه البيانات بشكل أكثر إيجازًا، إذ يتيح تحليل المعطيات معالجة كمية كبيرة جدًا من البيانات وتحديد الجوانب الأكثر إثارة للاهتمام في هيكلها.

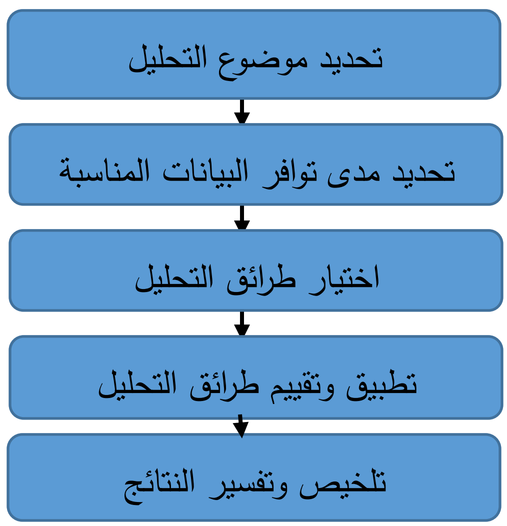
ويعود نجاح هذا التخصص في السنوات الأخيرة، إلى حد كبير إلى التمثيلات الرسومية المتقدمة، فيمكن لهذه الرسوم البيانية أن تسلط الضوء على العلاقات التي يصعب التقاطها عن طريق تحليل البيانات المباشر، ولكن قبل كل شيء، لا ترتبط هذه التمثيلات برأي مسبق حول قوانين الظواهر التي تم تحليلها، على عكس طرق الإحصاء الكلاسيكية. وقد بدأت الأسس الرياضية لتحليل المعطيات في التطور في مطلع القرن العشرين، لكن

أجهزة الكمبيوتر هي التي جعلت هذا النظام عمليًا، وجعلت منه مستخدمًا على نطاق واسع، من أجل ذلك يرتبط تحليل المعطيات بالرياضيات وعلوم الكمبيوتر ارتباطًا وثيقًا.

وبناءً على ما سبق تقديمه يتضّح أنّ تحليل المعطيات عبارة عن مجموعة من التقنيات الوصفية، الأداة الرياضية الرئيسية فيها هي جبر المصفوفات، والتي يتم التعبير عنها دون افتراض نموذج احتمالي مسبق. ولتحليل المعطيات العديد من الطرائق التي تختلف باختلاف المجال المستخدمة فيه، إذ يمكننا استخدام تحليل المعطيات في العلوم الطبيعية والعلوم الاجتماعية والمالية.

خطوات تحليل المعطيات:

تتضمن عملية تحليل المعطيات مجموعة من الخطوات التي لا بد من التعرض إليها بالتسلسل المطلوب حتى تكون النتائج التي يتحصل عليها ذات مغزى، ويمكن الاعتماد عليها في اتخاذ القرارات الصائبة. وتبدأ العملية بجمع جميع البيانات المتعلقة بالمشكلة المراد دراستها، ومن ثم معالجتها واستكشافها واستخدامها للعثور على الأنماط والرؤى الأخرى. ويمكن تلخيصها في الآتي:



طرائق تحليل المعطيات:

1. طرائق عاملية (Factorielles methods): تهدف هذه الطرائق إلى الكشف عن متغيرات كامنة لا يمكن مشاهدتها انطلاقا من المتغيرات الأصلية. وتشمل الطرائق العاملية الاختبارات الموالية.

– تحليل المكونات الرئيسية (PCA) (l'analyse en composantes principales)، المستخدم في البيانات الكمية، وطرقه المشتقة:

– (التحليل العاملي التقابلي) (CFA) (l'analyse factorielle des correspondances) المستخدم في البيانات النوعية (جدول الارتباط) والتحليل المتعدد بالمعاملات للتوفيقات AFCM أو ACM الذي يعمم السابق.

– التحليل القانوني والتحليل القانوني المعمم (L'analyse canonique et l'analyse canonique généralisée)، وهما أطر نظرية أكثر من الطرق القابلة للتطبيق بسهولة، يوسعان العديد من هذه الأساليب ويتجاوزان تقنيات الوصف. فالتحليل متعدد العوامل هو عبارة عن جداول مناسبة يتم فيها تنظيم المتغيرات في مجموعات ويمكن أن تكون كمية و/أو نوعية.

– التحليل العاملي التمييزي (DFA) أو التحليل التمييزي يجعل من الممكن تحديد المجموعات المتجانسة داخل المجتمع من وجهة نظر المتغيرات المدروسة.

– تحليل المكونات المستقلة (الأحدث) ((ICA) l'analyse en composantes indépendantes)، المشتق من فيزياء الإشارة والمعروف في البداية باسم طريقة فصل المصدر الأعمى، أقرب بشكل حدسي إلى طرق التصنيف غير الخاضعة للإشراف. وتنظم أيقونة الارتباطات الخاصة بالبيانات النوعية والكمية الارتباطات بين المتغيرات في شكل رسوم بيانية.

تحليل تاكر بين البطاريات (inter–batterie de Tucker) هو وسيط بين التحليل القانوني وتحليل المكون الرئيسي، تحليل التكرار المسمى أيضًا تحليل المكون الرئيسي في المتغيرات الآلية يشبه الانحدار نظرًا لأن متغيرات إحدى المجموعات التي تم تحليلها تعتبر تابعة، والبعض الآخر مستقل، وأنّ الوظيفة التي سيتم تعظيمها هي مجموع معاملات الارتباط بين المجموعتين.

2. طـرائق التصـنيف الأوتومـاتيكي (Automatic classification): وهو أسلوب إحصائي متقدم لتصنيف المشاهدات أو الأفراد لعدد من المجموعات بالاعتماد على مجموعة من المتغيرات ذات التأثير المعنوي. كما يستخدم في تصـنيف البيانـات إلـى فئـات متجانسـة بالتسلسل أو بدونه، مـع الأخـذ بعـين الاعتبـار المتغيـرات المدروسـة، كأن نصنف مثلا طلبة السنة الثالثة بناءً على نقاطهم في الامتحان.

الفصل الأول: مفاهيم احصائيـة

الفصل الأول: مفاهيم احصائية

يهدف هذا الفصل إلى مراجعة بعض المفاهيم الاحصائية التي يحتاج إليها الطالب في طرائق تحليل المعطيات، مثل طريقة حساب المتوسط الحسابي، الانحراف المعياري، التباين والتباين المشترك، الارتباط الخطي البسيط، القيم المعيارية، الجداول المتقاطعة، كاي تربيع...إلخ.

العرض الجدولي للمعطيات: تُعدّ الجداول التكرارية إحدى النماذج التنظيمية لتجميع المعطيات حتى تعكس صورة الواقع الذي أخذت منه من جانب، ومن جانب آخر حتى تكون قابلة للتحليل بشكل يناسب احتياجات البحث الذي أعدت لأجله. فبمجرد الانتهاء من ترتيب المعطيات المتحصل عليها، لا بد من عرضها بكيفية أو بأخرى، ليتم تحليلها ولكي يكون لها عندئذ معنى قد نسعى إلى اختصارها وتقديمها بكيفية مرسومة أو مصورة وإقامة علاقات بينها دائما بهدف جعلها دالة بالنسبة إلى مشكلة البحث. فالجدول إذا هو شكل تقني مختصر تجمع فيه البيانات بشكل متناسب مع خصائصها البحثية التي جمعت من أجلها، كما أنه الطريق المنهجي نحو تحويل المعطيات إلى دلالاتها الإحصائية التي تنقل الظاهرة المدروسة من مستوى وقوعها إلى مستوى تفسيرها واستخلاص النتائج منها.

التحليل وحيد المتغير:

1. مقاييس النزعة المركزية:

1.1. المتوسطات:

1.1.1. المتوسط الحسابي (Mean): من أهم المقاييس الإحصائية، ويعتمد عليه بشكل كبير في إيجاد حالة من الاتزان بين جميع قيم البيانات الإحصائية، ومن أهمّ خصائص الوسط الحسابيّ ما يأتي: يأخذ بعين الاعتبار جميع القيم والمشاهدات المتوفّرة؛ يعدّ محدود التأثّر بالتقلبات العينيّة؛ لا يمكن استخدام هذا المقياس الإحصائي في حالِ وجود فئات تكراريّة مفتوحة.

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{n_x}{n}$$

مثال: يحتوي الجدول الموالي على نقاط مجموعة من الطلبة في مادتي الرياضيات والاحصاء.

الرياضيات	
12	1
11	2
9	3
8	4
12	5
48	المجموع

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{n_x}{n}=\frac{48}{5}=10.4$$

يظهر من النتائج أعلاه أنّ علامات التلاميذ كانت متوسطة في مادة الرياضات.

3.1.1. الوسط التوافقي (Harmonic Mean): وهو أحد مقاييس النزعة المركزية وهو مقلوب الوسط الحسابي لمقلوبات هذه القيم. ويفضل استخدامه على باقي المتوسطات في حالة إيجاد معدل السرعات ومعدلات التغيير، ولا يمكن استخدامه في حالة إذا كانت إحدى هذه القيم مساوية للصفر.

$$\frac{1}{H}=\frac{1}{n}\sum_{i=1}^{k}\frac{n_i}{a_i}$$

مثال: سائق سيارة يقود بسرعة 60 كلم في الساعة عند الذهاب و30 عند العودة. والمطلوب حساب متوسط السرعة.

لحسابه يدويا نقسم العدد 1 على كل قيمة من قيم المتغير، ثم نجمع النواتج ونقسم عدد القيم على الناتج الذي تحصلنا عليه. في مثالنا:

$\frac{1}{H}=\frac{1}{2}(\frac{1}{60}+\frac{1}{30})=\frac{1}{40}$

2.1.1. الوسط الهندسي (Geometric Mean): هو أحد أشكال المتوسطات، ويستخدم بكثرة في دراسة المعدلات التي تميل إلى الزيادة بنسب ثابتة، كما يستخدم في دراسة السلاسل الزمنية القائمة على مقارنة معدل ما واختلافاته باختلاف الزمن. والوسط الهندسي يمتاز بعدم تأثره بالقيم المتطرفة ولكن لا يمكن استخدامه مع القيم السالبة أو الصفر لعدم جواز الجذر للقيم السالبة ذات العدد الزوجي والقيمة صفر تلغي باقي القيم لكون الضرب في الصفر يساوي صفراً. ويعطى بالعلاقة:

$$G=\left(\prod_{i=1}^{k}a_i^{n_i}\right)^{\frac{1}{n}}$$

مثال: تتزايد أسعار المنازل في الثلاث سنوات الأخيرة بنسبة 9%، 12%، 15%. فما هو متوسط معدل الزيادة في أسعار المنازل للسنوات الثلاث.

$G=\sqrt[3]{1.09*1.12*1.15}=1.1186$

ومنه، متوسط الزيادة السنوية يساوي 11.86%.

4.1.1. الوسط المرجّح (الموزون) (Weighted Mean):

يعد هذا المقياس من المقاييس المهمة للنزعة المركزية، وهو من حيث الفكرة يماثل الوسط الحسابي الاعتيادي، غير أنّ الوسط الاعتيادي يعتبر مفردات العينة قيد الدراسة لها نفس الأهمية والتأثير في حساب أي مؤشر إحصائي، ولكن في العديد من الحالات تكون بعض المفردات أكثر أهمية من غيرها مما يستوجب استخدام مؤشر أخر لحساب المعدل مع الأخذ بعين الاعتبار أهمية كل مفردة من مفردات العينة، وهذا المؤشر هو الوسط الحسابي المرجح أو الموزون. وقيمة هذا الوسط أكثر دقة من الوسط الاعتيادي، غير أنّه أقل استخدامًا منه.

$$\bar{X}=\frac{\sum_{i=1}^{n}X_iW_i}{\sum_{i=1}^{n}W_i}$$

مثال: محل يبيع نوع من الفاكهة بأربعة أسعار مختلفة، باع من النوع الأول 5 كيلوجرام بسعر 1.2 دينار للكيلوجرام ومن النوع الثاني 8 كيلوجرام بسعر 0.8 دينار، ومن النوع الثالث باع 12 كيلوجرام بسعر 0.4 دينار. أحسب الوسط الحسابي المرجح لسعر بيع الكيلوجرام من هذه الفاكهة. لحساب الوسط المرجّح يدويا نستخدم المعادلة التالية:

$$\bar{X}= \frac{(5*1.2)+(8*0.8)+(12*0.4)}{5+8+12}= 0.68$$

5.1.1. الوسط الحسابي المشذب (Trimmed Mean %5): يشبه الوسط المقتطع (الذي يطلق عليه أحيانًا المتوسط المقطوع أو المقصوص)، لكنه يقلص أي حدود خارجية يمكن أن تؤثر القيم الخارجية على الوسط (خاصةً إذا كان هناك واحد أو اثنين فقط من القيم الكبيرة جدًا)، لذلك غالبًا ما يمكن أن يكون الوسط المقصوص مناسبًا بشكل أفضل لمجموعات البيانات ذات القيم العالية أو المنخفضة، غير المنتظمة، أو بالنسبة للتوزيعات المنحرفة للغاية. ويتم التعبير عن هذه المتوسطات بالنسب المئوية، إذ تخبرك النسبة المئوية بنوع البيانات المراد إزالتها. على سبيل المثال، مع خفض متوسط 5 ٪، يتم استبعاد أدنى 5٪ وأعلى 5٪ من البيانات، ثم يتم حساب الوسط من 90٪ المتبقية من نقاط البيانات.

ملاحظة: إذا كان الفرق بينه والمتوسط الحسابي أكبر من 5 بالمائة، يعني وجود قيم متطرفة كبيرة والبيانات لا تتبع التوزيع الطبيعي.

مثال: حساب الوسط المشذب للبيانات التالية.

x
32,00
20,00
22,00
14,00
18,00
19,00
27,00
21,00
18,00
18,00
16,00
14,00
25,00
26,00
30,00
28,00
18,00
19,00
22,00
20,00

لحسابه يدويا نقوم بترتيب البيانات من الأصغر إلى الأكبر، ثم نحذف 5% من الأعلى و5% من الأسفل، ونحسب المتوسط المشذب من 90% المتبقية من البيانات. في مثالنا 5% من 20= 1. أي سنقوم بحذف أعلى رقم وأدنى رقم (14، 32).

الوسط الحسابي لهذه البيانات قبل التشذيب يساوي 21.35 (مجموع جميع القيم مقسوما على عددها) وبعد التشذيب يساوي 21.16.

2.1. الوسيط (Median): وهو القيمة التي تتوسط البيانات الإحصائية بعد عملية ترتيبها بشكل تصاعدي أو تنازلي، ومن أهمّ خصائص الوسيط ما يلي: لا يتأثّر الوسيط بالقيم الإحصائية المتطرفة؛ يُستخدم بشكل كبير في حالات الفئات المفتوحة؛ يستخدم فيما يعرف بالتوزيعات الملتوية.

3.1. المنوال (Mode): يشير إلى القيمة الأكثر تكرارًا في البيانات الإحصائية، ومن أهمّ خصائصه: لا يمكن الاعتماد عليه في العمليات الإحصائيّة اللاحقة؛ يتأثّر بشكلٍ كبير بعامل طول الفئة.

4.1. المجموع (Sum): مجموع القيم أو إجماليها.

مثال: حسبنا المتوسط والوسيط والمنوال لدرجات الامتحان لتسعون (90) تلميذا في مجموعة من المواد التي يدرسونها، وكانت النتائج كما يلي:

Statistics

		الرياضيات	العربية	ت_اسلامية	رياضة	ت_علمية
N	Valid	96	96	96	96	94
	Missing	0	0	0	0	2
Mean		5,9063	5,0833	6,5729	6,4375	3,8191
Median		7,0000	5,0000	7,0000	6,0000	2,0000
Mode		9,00	6,00[a]	7,00	9,00	1,00

a. Multiple modes exist. The smallest value is shown

يظهر من الجدول أعلاه أنّ علامات التلاميذ كانت متوسطة في مادتي الرياضات واللغة العربية، وفوق المتوسط في مادتي التربية الإسلامية والرياضة، بينما متوسط علاماتهم في التربية العلمية كان ضعيفا (3.81)، مما يدل على ضعف تحصيلهم في هذه المادة، مقارنة ببقية المواد. كما يتضح من قيم المنوال أنّ علامات التلاميذ في مادة الرياضيات كانت أغلبها 9 من 10، وكذلك الحال بالنسبة لمادة الرياضة، وحصل أغلب التلاميذ على علامة 1 من 10 في مادة التربية العلمية. ويتبين لنا من الجدول أنّ منحنى توزيع العلامات في مادة التربية العلمية ملتوي نحو اليمين (إلتواء موجب)، وفي بقية المواد ملتوي نحو اليسار (إلتواء سالب).

ملاحظة: يكون منحنى التوزيع معتدلا في حالة تساوي الوسط والوسيط والمنوال؛ وفي حالة كان الوسط أكبر من الوسيط وأكبر من المنوال، يكون إلتواء منحنى التوزيع نحو اليمين؛ أمّا في حالة كان الوسط أقل من الوسيط والمنوال، فالتوزيع يلتوي نحو اليسار.

مثال 2: طلبنا من صديقنا تسجيل أجور 24 رجلا كانوا جالسين في مقهى، صادفناهم عشوائيا، ثم قمنا بحساب التكرارات، الوسط الحسابي، الوسيط، المنوال، والمجموع، وكانت النتائج كالتالي:

Statistics

الأجور

N	Valid	24
	Missing	0
Mean		174291.6667
Median		29000.0000
Mode		18000.00
Sum		4183000.00

يتبين من الجدول أنّ حجم العينة يساوي 24، ولا توجد لدينا قيم مفقودة، كما أنّ متوسط الأجور يساوي 174291.66 دج، ووسيطها 29000 دج، أمّا الأجر الأكثر تكرارا في العينة فهو 18000 دج، وهو المنوال، ومجموع أجور كامل أفراد العينة 4183000 دج.

نلاحظ أنّ نتيجة الوسط الحسابي تبدو للوهلة الأولى غير منطقية إن قارناها بالوسيط والمنوال، وللتأكد نفحص جدول التكرارات.

الأجور

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18000.00	7	29.2	29.2	29.2
	22000.00	3	12.5	12.5	41.7
	25000.00	1	4.2	4.2	45.8
	26000.00	1	4.2	4.2	50.0
	32000.00	4	16.7	16.7	66.7
	42000.00	4	16.7	16.7	83.3
	48000.00	3	12.5	12.5	95.8
	3500000.00	1	4.2	4.2	100.0
	Total	24	100.0	100.0	

توجد قيمة للأجر كبيرة جدا مقارنة ببقية الأجور، ربما يكون السبب خطأ في الإدخال، أو سبب آخر يمكن تبريره، بأننا صادفنا في المقهى ملياردير، وبياناتنا سليمة. هذا ما نطلق عليه بمنطق الاقتصاد رجل أعمال، وبمنطق علم الاجتماع بورجوازي، أمّا في المنطق الإحصائي فهو قيمة شاذة تسبب مشكلة في بياناتنا، علينا تبريرها أو حذفها، لكن لو لم يسجل صديقنا أجر الملياردير فلن تكون عينتنا عشوائية، ما الحل إذا؟ علينا أن نستخدم الوسيط بدل الوسط الحسابي.

لاحظ أنّ وسيط الأجور هو 29000 وهو رقم معقول بالنظر إلى التكرارات، لو نحذف أجر الملياردير ونعيد الحساب.

Statistics

الأجور

N	Valid	23
	Missing	0
Mean		29695.6522
Median		26000.0000
Mode		18000.00
Sum		683000.00

نعم أصبح متوسط الأجور 29695.65، وهو تقريبا مساوي لقيمة الوسيط دون حذف القيمة الشاذة. وبالاعتماد على قيمة الوسيط يمكن القول: إنّ نصف الأفراد في المقهى أو 50 في المائة منهم يتقاضون أجورا أقل من 29000، و50 في المائة يتقاضون أجورا أعلى من 29000 دج. كما يمكننا الحكم على شكل التوزيع من خلال القيم السابقة، إذ نلاحظ أنّ قيمة المتوسط الحسابي أكبر من الوسيط والمنوال، ما معناه أنّ التوزيع ملتوي نحو اليمين.

2. مقاييس التشتت (dispersion): مقاييس التشتت مؤشرات إحصائية وصفية تستخدم لقياس مدى التباعد أو الاختلاف فيما بين قيم مجموعة بيانات أو توزيع تكراري. ويعرّف التشتت بأنه مقدار تباعد أو اختلاف قيم بيانات أو توزيع تكراري عن بعضها البعض أو عن إحدى مقاييس النزعة المركزية، وتحديدا (الوسط الحسابي، الوسيط، المنوال)، وبناءً على هذا التعريف فإنّ مقاييس التشتت تستخدم لدراسة مدى صلاحية عينة عشوائية لأغراض الدراسة، فكلما زادت قيمة مقاييس التشتت كلما زاد التباعد بين بيانات العينة، وبالتالي عدم صلاحيتها للدراسة والعكس صحيح.

ملاحظة: مقاييس التشتت المطلقة، هي مقاييس تعتمد القيم المطلقة أو المربعة في حسابها، وعليه تكون قيمها دائما موجبة.

ولتبيين أهمية مقاييس التشتت احصائيا إليك المثال التالي: لدينا بيانات عن أعمار مجموعتين من لاعبي كرة اليد.

الفريق_الأول	الفريق_الثاني
18	26
19	26
29	26
33	26
31	26

لو نقوم بحساب متوسط أعمار الفريقين سنجده كالتالي:

Statistics

		الفريق_الأول	الفريق_الثاني
N	Valid	5	5
	Missing	0	0
Mean		26.00	26.00

يظهر من مخرجات التحليل أن متوسط أعمار الفريقين متساوي، فهل نستطيع الحكم على أعمار الفريقين بأنها متساوية أو مختلفة من خلال الوسط الحسابي فقط، الجواب لا ولكن لو علمنا أن الوسط الحسابي لأعمار الفريق الثاني 26 سنة واختلافها (تشتتها) يساوي صفرا، والوسط الحسابي لأعمار الفريق الأول يبلغ 26 سنة وتشتتها يبلغ قيمة معينة غير الصفر، فنستطيع عندها إعطاء وصفا أفضل للبيانات وبالتالي مقارنة أفضل.

1.2. الانحراف المعياري:

$$\delta = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$$

مثال: لو أردنا دراسة الحالة المادية للمجتمع، وعلى ضوئها يتم اتخاذ قرارات معينة، هنا سوف يتم معرفة أين تتمركز الأغلبية وأين الشريحة أو الأقلية التي تقع في مدى المعيار أو فوق المعيار أو تحته. كيف يتم ذلك؟ إذا كان الانحراف المعياري يساوي صفرًا، فهذا يعني أنه يساوي الوسط الحسابي أي لا يوجد لا فقر ولا غنى، ولكن إن وجد انحراف يساوي مثلا 50 والمتوسط الحسابي 150 فهذا يعني أنّ الحالة المادية للمجتمع تتمركز بالمعظم بين الـ 100 والـ 200 وأنّه يوجد فقر تحت الـ 100 وغنى فوق الـ 200. وكلما زاد الانحراف المعياري عن المتوسط، فهذا يعني أن هناك تباين كبير حول المتوسط على نطاق واسع، حيث لا يوجد تجانس.

2.2. التباين:

$$V = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\delta = \sqrt{V}$$

يعتبر التباين أحد مقاييس التشتت المهمة لأنّه من جانب يأخذ جميع القيم في الاعتبار عند حسابه، ومن جانب آخر يقيس التشتت عن الوسط الحسابي للقيم، هذا بالإضافة إلى أنه تسهل معالجته رياضياً، ويدخل في تكوين عدد من المقاييس والاختبارات الإحصائية المهمة. والفكرة الأساسية للتباين هي حساب انحرافات جميع القيم عن وسطها الحسابي (أي حساب الفرق بين كل قيمة والوسط الحسابي)، وسوف نجد أن بعض القيم أكبر من الوسط فتكون الفروق (أو الانحرافات) بالموجب، والبعض الأخر أصغر من الوسط فتكون الفروق بالسالب، ودائماً يكون مجموع هذه الانحرافات مساوياً للصفر. وبتربيع هذه الانحرافات، ثم حساب متوسط الانحرافات المربعة نحصل على التباين. أي أنّ التباين هو متوسط مربعات انحرافات القيم عن وسطها الحسابي.

ملاحظة: على الطالب أن يفرّق بين التباين (Variance) والتغاير (التباين المشترك) (Covariance)، فالتغاير هو مقياس لكمية تغير متغيرين مع بعضهما (التباين هو حالة خاصة من التغاير؛ يسمى التغاير تباينا عندما يكون المتغيرين متساويين).

التحليل ثنائي المتغير(Bivariate analysis):

يهدف إلى تحليل العلاقة التي قد تكون بين متغيرين. مثل العلاقة بين عدد السكان وعدد خطوط النقل في مدينة ما.

ملاحظة: العلاقة بين متغيرين، حتى وإن كانت قوية لا تعني السببية.

الجداول المتقاطعة: تعتمد الجداول المتقاطعة على مبدأ التكرار إلا أنها تعطي معلومات أكثر عمقا ودلالة بالمقارنة مع ما تعطيه جداول التكرار. حيث تدمج الجداول المتقاطعة متغيرين أو أكثر. مثلا إذا كان أحد الأسئلة يتعلق بالجنس " ذكر، أنثى" والآخر عن المستوى التعليمي فإنه في الجداول المتقاطعة تستطيع أن تظهر عدد كل من الذكور والإناث عند كل مستوى تعليمي.

	y_1	y_2	…..	y_j	…..	y_p	**المجموع**
X_1	n_{11}	n_{12}	…..	n_{1j}	…..	n_{1p}	$n_{1.}$
X_2	n_{21}	n_{22}	…..	n_{2j}	…..	n_{2p}	$n_{2.}$
…..	…..	…..	…..	…..	…..	…..	…..
X_i	n_{i1}	n_{i2}	…..	n_{ij}	…..	n_{ip}	$n_{i.}$
…..	…..	…..	…..	…..	…..	…..	…..
X_k	n_{k1}	n_{k1}	…..	n_{kj}	…..	n_{kp}	$n_{k.}$
المجموع	$n_{.1}$	$n_{.2}$	…..	$n_{.j}$	…..	$n_{.p}$	$n_{..}$

حيث: i تتغير من واحد إلى k، j تتغير من واحد إلى p، .. المجموع.

1. المتوسط الحسابي:

$$\bar{X}= \frac{1}{n}\sum_{i=1}^{k} n_{i.}\, x_i= \sum_{i=1}^{k} f_{i.}\, x_i$$
$$\bar{y}= \frac{1}{n}\sum_{j=1}^{p} n_{.j}\, y_j= \sum_{j=1}^{p} f_{.j}\, y_j$$

2. التباين:

$$\mathrm{Var}(x):=\overline{x^2}-(\bar{x})^2$$
$$\overline{x^2} := \frac{1}{n}\sum_{i=1}^{k} n_{i.}{x_i}^2= \sum_{i=1}^{k} f_{i.}x_i^2$$
$$\mathrm{Var}(y):=\overline{y^2}-(\bar{y})^2$$
$$\overline{y^2} := \frac{1}{n}\sum_{j=1}^{p} n_{.j}{y_j}^2= \sum_{j=1}^{p} f_{.j}y_j^2$$

3. السلسلة الشرطية:

3.1. السلسلة الشرطية بالنسبة للمتغير X: نشير لها بـ X_j، ونقول بأنها السلسلة الشرطية لـ X مع العلم أنّ $Y=y_j$، ونحسب في هذه الحالة التكرار الشرطي $f_{i/j}$.

$$f_{i/j}:=\frac{n_{ij}}{n_{.j}}=\frac{f_{ij}}{f_{.j}}$$

3.2. السلسلة الشرطية بالنسبة للمتغير Y: نشير لها بـ Y_i، ونقول بأنها السلسلة الشرطية لـ Y مع العلم أنّ $X=X_i$، ونحسب في هذه الحالة التكرار الشرطي $f_{i/j}$.

$$f_{j/i}:=\frac{n_{ij}}{n_{i.}}=\frac{f_{ij}}{f_{i.}}$$

4. التباين:

$$\text{Var(X)} = \frac{1}{N}\sum_{i=1}^{n} n_{i.}x_i^2-\bar{x}^2$$

التباين المشترك:

$$\text{Cov(X,Y)}=\overline{xy}-\bar{x}\bar{y}=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{p} n_{ij}x_iy_j-\bar{x}\bar{y}$$

يعبر التباين المشترك بين المتغيرين عن نوع العلاقة بينهما، فإذا كان التباين المشترك يساوي صفرا، فالمتغيرين مستقلين عن بعضهما، وإذا كان أكبر من الصفر يرتبطان ارتباطا موجبا، وإن كان أقل من الصفر يرتبطان ارتباطا سلبيا.

5. معامل الارتباط الخطي البسيط:

$$r_{XY}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}}$$

$$r_{XY}=\frac{Cov(X,Y)}{\delta_X\delta_Y}$$

وهو تحليل ثنائي المتغير يقيس قوة الارتباط بين متغيرين واتجاه العلاقة، ومن حيث قوة العلاقة تتراوح قيمة معامل الارتباط بين +1 و −1، وتشير قيمة ± 1 إلى درجة الكمال من الارتباط بين المتغيرين، وعندما تذهب قيمة معامل الارتباط نحو 0 ستكون العلاقة بين المتغيرين أضعف، ويشار إلى اتجاه العلاقة بواسطة علامة المعامل، وتشير علامة + إلى وجود علاقة إيجابية وتشير العلامة إلى وجود علاقة سلبية.

ملاحظة: يمكن تقييم القوة من خلال هذه المبادئ العامة:

- $0.1< |r| <0.3$ .... ارتباط ضعيف/ العلاقة ضعيفة.
- $0.3< |r| <0.5$ .... ارتباط متوسط/ علاقة متوسطة.
- $0.5 < |r| <1$ ....ارتباط كبير/ علاقة قوية.

مثال: يحتوي الجدول الموالي على نتائج دراسة للعلاقة بين متوسط عدد الساعات (X) التي يقضيها الفرد يوميا في استخدام هاتفه النقال، وعدد الكتب (Y) التي يقرأها في السنة.

X \ Y	0−12	13−25	26−38	39−51	المجموع

0−3	66	54	52	49	221
3−6	58	34	22	6	120
6−9	59	9	4	2	74
المجموع	183	97	78	57	415

المطلوب:

1. حدد:

n_{23}، n_{32}، n_{12}، $n_{3.}$، $n_{.2}$، f_{31}، $f_{3.}$، $f_{i=2\,/j=1}$

2. أحسب المتوسط الحسابي.
3. أحسب التباين المشترك بين المتغيرين.

الحل:

.1

$n_{23}= 22$

22 فردا يقضون بين 3−6 ساعات يوميا وراء هواتفهم، يقرؤون في العام بين 26−38 كتابا.

$n_{32} = 9,\ n_{12} = 54$

$n_{3.} = 74$

74 فردا يقضون بين 6−9 ساعات يوميا في استخدام الهاتف.

$n_{.2} = 97$

97 فردا يقرؤون بين 13−25 كتابا في السنة.

$f_{31} = \frac{59}{415}=0.142$

14.2% من الأفراد يقضون في المتوسط بين 6 و 9 ساعات يوميا في استخدام هواتفهم. ويقرؤون في السنة بين 0 و 12 كتابا.

$F_{3.}= \frac{74}{415}= 0.178$

17.8% من الأفراد يقضون بين 6 و 9 ساعات من وقتهم في استخدام الهاتف.

$f_{i=2\,/j=1} =\frac{n_{21}}{n_{.1}}= \frac{58}{183}= 0.317$

31.7% من الأفراد الذين يقرؤون بين 0 و 12 كتاب في السنة يقضون من 3 إلى 9 ساعات يوميا وراء هواتفهم.

2. المتوسط الحسابي.

<table>
<tr><td colspan="2">Y</td><td>0−12</td><td>13−25</td><td>26−38</td><td>39−51</td><td rowspan="2">المجموع</td></tr>
<tr><td colspan="2">X</td><td>6</td><td>19</td><td>32</td><td>45</td></tr>
<tr><td>0−3</td><td>1.5</td><td>66</td><td>54</td><td>52</td><td>49</td><td>221</td></tr>
</table>

3−6	4.5	58	34	22	6	120
6−9	7.5	59	9	4	2	74
المجموع		183	97	78	57	415

الوقت المتوسط لاستخدام الهاتف هو: 3.437، ثلاثة ساعات و26 دقيقة.

متوسط عدد الكتب المقروئة يساوي 19.3.

3. حساب التباين المشترك:

$$\text{Cov}(X,Y)=\overline{xy}-\bar{x}\bar{y}=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{p} n_{ij}x_iy_j-\bar{x}\bar{y}$$

66*6*1.5=594, 54*19*1.5=1539....

594	1539	2496	3307.5
1566	2907	3168	1215
2655	1282.5	960	675

المجموع= 22365

$\text{Cov}(X,Y)=\frac{1}{415}*22365-3.437*19.3=-12.378$

التباين المشترك بين المتغيرين سالب، ما معناه أنّ المتغيرين متعاكسين في الاتجاه، كما أنّ وقت قراءة الكتب أقل أهمية من الوقت الذي يقضيه الفرد وراء الهاتف.

4. حساب معامل الارتباط:

$$r_{XY}=\frac{Cov(X,Y)}{\delta_X\delta_Y}$$

$\text{Var}(X)=\frac{1}{N}\sum_{i=1}^{n} n_{i.}x_i^2-\bar{x}^2$

Y / X		0−12	13−25	26−38	39−51	المجموع	$n_{i.}x_i$	$n_{i.}x_i^2$
		6	19	32	45			
0−3	1.5	66	54	52	49	221	331.5	497.25
3−6	4.5	58	34	22	6	120	540	2430
6−9	7.5	59	9	4	2	74	555	4162.5
المجموع		183	97	78	57	415	1426.5	7089.75

221*1.5=331.5, 331.5*1.5^2=497.25.....

$\text{Var}(X)=\frac{1}{415}*7089.75-3.437^2=5.27$

$\delta_X=\sqrt{V_X}=\sqrt{5.27}=2.29$

$\delta_Y = 14.10$

$r_{XY} = \frac{-12.378}{2.29*14.10} = -0.38$

توجد علاقة عكسية متوسطة بين المتغيرين.

6. التوافق والاستقلالية بين متغيرين نوعيين (اختبار كاي تربيع للاستقلالية (*CHI−SQUARE*)).

يستخدم مربع كاي للاستقلالية لاختبار العلاقة بين متغيرين من النوع الإسمي أو الترتيبي. ويستخدم اختبار كاي تربيع للترابط لاختبار العلاقة بين متغيرين فئويين (اسمي أو ترتيبي أو مزيج من الاثنين)، ويشار إليه أيضًا باسم اختبار مربع كاي للاستقلالية. وكلاهما اسمان صحيحان للاختبار، حيث ينبع الاختلاف في المصطلحات فيما إذا كانت الفرضية الصفرية صحيحة (أي الاستقلالية) والفرضية البديلة (أي الارتباط).

شروطه:

– يجب قياس المتغيرين على مستوى ترتيبي أو اسمي (أي البيانات الفئوية).

– يجب أن يتكون المتغيران لديك من مجموعتين فئتين مستقلتين أو أكثر. المتغيرات المستقلة التي تلبي هذا المعيار تشمل الجنس (مجموعتان: ذكور وإناث)، المهنة (على سبيل المثال، 5 مجموعات: الجراح، الطبيب، الممرضة، طبيب الأسنان، المعالج) وهكذا.

الفرضية الصفرية للاختبار: لا توجد علاقة بين المتغيرين في المجتمع. ويتضمن اختبار ما إذا كانت تكرارات الخلية المشاهدة (أو الاحتمالات المشتركة) في البيانات تختلف اختلافًا كبيرًا عن تلك التي يمكن توقعها، إذا لم تكن هناك علاقة (أي الاستقلال) بين المتغيرات داخل المجتمع.

ويمكننا ترجمة هذا الكلام في ظل فرضية الاستقلالية H_0.

P(X=i et Y=j)= P(X=i) x P(Y=j).

وعلى العكس، الفرضية البديلة H_1

P(X=i et Y=j) # P(X=i) x P(Y=j).

بعبارة أخرى سنقوم باختبار فرضية العدم:

$H_0 ; f_{ij} = f_{i.} * f_{.j}$

في مقابل الفرضية البديلة:

$H_1 ; f_{ij} \# f_{i.} * f_{.j}$

وعند دراسة ظاهرة كيفية متعلقة بمتغيرتين نوعيين، يمكن البحث عن وجود اعتمادية أو استقلالية باستخدام اختبار كاي تربيع للاستقلالية، المعطى بالعلاقة:

$$X^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(nij-(ni.*n.j/n))^2}{ni.*n.j/n} = \frac{(fij-fi.*f.j)^2}{fi.*f.j}$$

ملاحظة: يستخدم التحليل التقابلي جدول الاحتمالات، لذلك فهو لا يقول أي شيء عن الدلالة الإحصائية، فقط يتصور طبيعة الارتباط بين المتغيرين.

ويتم حساب قيمته ومن ثم مقارنته مع القيمة الجدولية الحرجة بدرجة حرية،

$$r=(p-1)(q-1).$$

((عدد الأسطر−1)(عدد الأعمدة−1)).

الفرضية الصفرية للاختبار (عند مستوى دلالة 0.05 فأقل) المتغيرين مستقلين، أمّا الفرضية البديلة فتقول بأنّهما غير مستقلين عن بعضهما. ولحساب كاي تربيع، نحسب التكرار المتوقع لكل خلية في الجدول المتقاطع حسب العلاقة الموالية.

$$\text{Expected frequency } E_{r,c} = \frac{(\text{Sum of row } r) \times (\text{Sum of column } c)}{\text{Sample size}}.$$

نضرب مجموع عناصر السطر في مجموع عناصر العمود الذي تنتمي له الخلية، ونقسم الناتج على المجموع الكلي. أو عن طريق العلاقة في معادلة حساب كاي تربيع $n_{i.}*n_{.j}/n_{..}$.

مثال: نحن مهتمون بالعلاقة بين لون العينين ولون الشعر لمجموعة من 484 من الإناث. وجدول الاقتران أو التقاطع الموالي يبين تقاطع المتغيرين الوصفيين، لون العينين في الأسطر ولون الشعر في الأعمدة. حيث يشمل متغير لون العينين أربعة خصائص للون العينين، هي: كستنائي، عسلي، أخضر، أزرق؛ في حين يشمل متغير لون الشعر ثلاثة خصائص من لون الشعر، هي: بني، أحمر، أشقر.

	بني	أحمر	أشقر
كستنائي	119	26	7
عسلي	54	14	10
أخضر	29	14	16
أزرق	84	17	94

المطلوب: حساب كاي تربيع وتفسير النتيجة.

الحل:

سنحسب في الخطوة الأولى جدول التكرارات المتوقعة بالاعتماد على جدول التقاطع.

نضرب مجموع عناصر السطر في مجموع عناصر العمود الذي تنتمي إليه الخلية، ونقسم الناتج على المجموع الكلي. أو عن طريق العلاقة في معادلة حساب كاي تربيع.

$$n_{i.}*n_{.j}/n_{..}$$

جدول التكرارات النظرية (المتوقعة)

	بني	أحمر	أشقر
كستنائي	89.818	22.297	39.884
عسلي	46.090	11.442	20.466
أخضر	34.863	8.654	15.481
أزرق	115.227	28.605	51.167

مثال: التكرار المتوقع للخلية الناتجة من تقاطع السطر الأول مع العمود الأول:

$$n_{11}=\frac{152*286}{484}= 89.818$$

جدول حساب كاي تربيع

	بني	أحمر	أشقر	
كستنائي	9.481	0.614	27.112	37.207
عسلي	1.357	0.571	5.352	7.28
أخضر	0.985	3.302	0.017	4.304
أزرق	8.462	4.708	35.856	49.026
	20.285	9.195	68.337	97.817

$$X^2_{11}=\frac{(89.818-(119))^2}{89.818}= 9.481$$

نحسب درجات الحرية من العلاقة:

$$(r-1)(c-1)= (4-1)(3-1)=6$$

من الجد ... عند مستوى

دلالة 5 ... يع المحسوبة

تقع ضمن ... ر والأعمدة،

أي بين ...

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
	Probability of a larger value of x²								
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48

مصطلحات

Terminology

1. جدول البيانات الكمية: مصفوفة مستطيلة الشكل مكونة من صفوف وأعمدة. كل صف يمثل فرد وكل عمود يمثل متغير. ويطلق عليه أيضا اسم مصفوفة البيانات، ويرمز له بالرمز X.

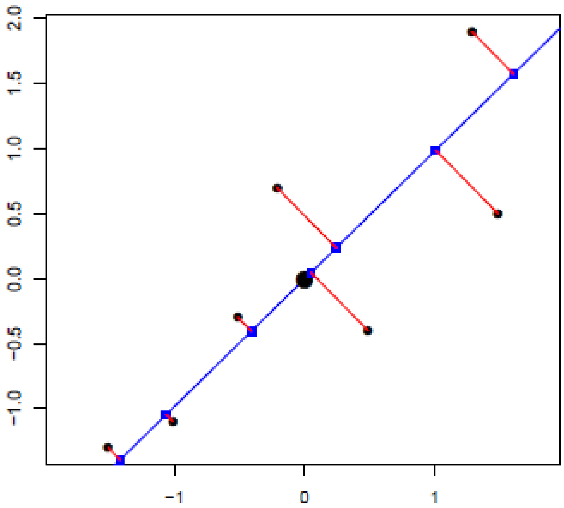
2. الأفراد: كل فرد يقابل صفًا في جدول البيانات، وتسمية فرد تطلق على أي شيء، شخص، حيوان، بلد، منتج، شركة...إلخ.

3. الخاصية: الخصائص أو الفئات، هي القيم التي يتخذها متغير اسمي. على سبيل المثال، متغير الجنس له خاصيتان: ذكر وأنثى. ويُعرف أيضًا باسم المجموعة أو الفئة.

4. التحليل العاملي: التصور الأمثل (بمعنى معين) لسحابة من النقاط في فضاء متعدد الأبعاد.

5. المحاور العاملية: المحاور، مرتبة حسب أهميتها، والتي نستخدمها لتصور سحابة النقاط في تحليل العاملي. ويتم تحديدها من خلال اتجاهات التمدد الأكبر (القصور الذاتي) لسحابة النقاط.

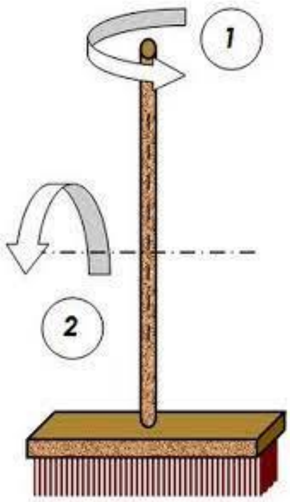
6. سحابة النقط (Cloud of Points): نسمي الرسم البياني المستخدم لتمثيل مجتمع إحصائي مكون من حرفين (X, Y) بسحابة النقط، حيث تشير كل نقطة إلى موضع المشاهدة وفقًا للحرفين. وفي معظم الأحيان لا نلاحظ فقط موقع النقاط، بل تكون مصحوبة بخط مستقيم يلخص السحابة بشكل أو بآخر، هذا الخط المستقيم تمثل معادلته الرابط الذي قد يكون موجودًا بين الحرفين.



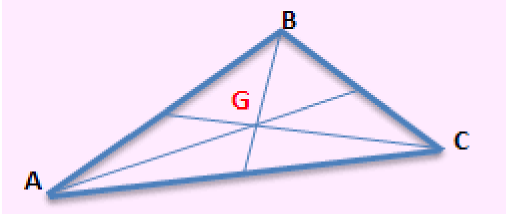
7. المسافة: المسافة بين نقطتين في سحابة معينة يمكن حسابها بناءً على إحداثيات النقاط وفقا لنظرية فيتاغورس (Pythogorean) الشهيرة.

8. القصور الذاتي (**Inertia**): مصطلح مستعار من الميكانيك (في الفيزياء)، يكافئ المفهوم الإحصائي للتباين. ويعطي القصور الذاتي فكرة عن الانتشار في سحابة من النقاط الموزونة. فإذا كان للأفراد نفس الوزن (نفس الأهمية)، فإنّ أكبر اتجاه للقصور الذاتي لسحابة النقاط يكون مع اتجاه محورها الرئيسي.

فإذا كنت لم تدرس الفيزياء أو الميكانيك، فأنت لا تعرف أن القصور الذاتي يقيس مقاومة الجسم لتغيير حركته الدورانية حول نقطة ما. خذ مثلا المكنسة، إذا أردنا تدويرها حول مركز ثقلها، ونظرًا لتوزيع مادتها، يسهل تدويرها حول محور المقبض (الحركة 1) بدلاً من تدويرها مثل العصى (الحركة 2). لذلك، يكون القصور الذاتي أقل على المحور الذي يعبر المقبض على طوله بالكامل مقارنة بالمحور الذي يخترق قطره (أعلى بقليل من الفرشاة).



ويتم قياس القصور الذاتي الكلي كمجموع مربعات مسافات النقاط من مركز ثقل (G) سحابة النقط.



ففي حالة عدم إعطاء كل النقاط نفس الوزن (في الميكانيك، نقول الكتلة)، فلا بد من أن يتم وزن مربعات المسافات، لذلك إذا لاحظنا أن d مسافة نقطة من مركز ثقل السحابة، فسيكون لدينا عدد n من النقاط المتأثرة بالوزن p.

$$I= \sum_{i=1}^{n} p_i d_i^2$$

ويترتب على ذلك أنه كلما زاد عدد النقاط التي نضيفها، زاد القصور الذاتي (من الصعب مناورة سفينة الشحن أكثر من سفينة الصيد) ونلاحظ أيضًا أنه بقدر ما نضيف مربعات من المسافات، فإن النقطة البعيدة عن مركز الثقل يكون لها تأثير أكبر على القصور الذاتي الكلي من النقطة المتوسطة، عند تساوي الوزن. لذلك لا يتم تعريف القصور فيما يتعلق بالمحور ولكن فيما يتعلق بالنقطة. ففي الحالة التي تكون فيها المتغيرات كمية، إذا تم تعيين نفس الوزن لجميع المشاهدات، فإن القصور الذاتي يندمج مع مؤشر التشتت. لذلك، يكون القصور الذاتي = عدد المشاهدات × التباين، أو في حالة توفر عدة متغيرات، فالقصور الذاتي يساوي عدد الأفراد في مجموع تباينات المتغيرات، هذا المجموع هو أثر مصفوفة التباين والتباين المشترك (Trace (V)).

وبالتالي، إذا سعى المرء إلى محور يمر بمركز الثقل O والذي يجب أن "يمتص" أقصى قدر من القصور الذاتي، فمن الضروري تقليل مسافات الإسقاطات على المحور. وبالتالي، تسعى طريقة التحليل بالمركبات الرئيسية إلى تحديد المحاور التي تمتص أقصى قدر ممكن من القصور الذاتي. فإذا كانت لدينا سحابة من النقاط على شكل مكنسة، فإن اتجاه "القصور الذاتي الأقصى المتوقع" (أو الحد الأدنى من التشوه المتوقع) سوف يندمج مع المقبض، وسيشمل معظم القصور الذاتي الكلي، ويتم امتصاص الباقي بواسطة المحور الثاني الذي يتقاطع بشكل عمودي مع الأول فوق الفرشاة .

الفصل الثاني:

الفضاء الشعاعي وحساب المصفوفات

أولا: الفضاء الشعاعي.

المصفوفة $X_{(nxp)}$ هي المصفوفة الأولوية للمعطيات، تحتوي على n فرد ممثلين في الأسطر و p متغير ممثلين في الأعمدة.

$$X_{(nxp)}=\begin{bmatrix} n_{11} & n_{12} & n_{1j} & n_{1p} \\ n_{21} & n_{22} & n_{2j} & n_{2p} \\ n_{i1} & n_{i2} & n_{ij} & n_{ip} \\ .. & .. & .. & .. \\ n_{n1} & n_{n2} & n_{nj} & n_{np} \end{bmatrix}$$

حيث أنّ n_{ij} هو العنصر المولّد للمصفوفة X.

المتجه (الشعاع) e في الفضاء R^n.

$$u := \begin{bmatrix} x_1 \\ .. \\ x_i \\ \dots \\ x_n \end{bmatrix} \in X(n*1)$$

شعاع الوحدة في الفضاء R^n.

$$1_n=\begin{bmatrix} 1 \\ .. \\ 1 \\ \dots \\ 1 \end{bmatrix} \in X(n*1)$$

الشعاع الذاتي: نقول عن الشعاع غير الصفري U_α أنّه شعاع ذاتي للمصفوفة $X_{(pxp)}$ إذا وفقط إذا كان:

$$(X_{(pxp)}-\lambda_\alpha I_p)U_\alpha=0_p$$

وبما أنّ هذا الشعاع موجود في فضاء ذو بعد p، فإنّه يملك p مركبة.

الشعاع الذاتي الوحدوي: نقول عن الشعاع الذاتي U_α أنّه شعاع الوحدة إذا كانت طويلته تساوي الواحد.

$$||U_\alpha||=1$$

طويلة الشعاع U_α:

$$||U_\alpha||=\sqrt{x_1^2+\cdots+x_p^2}$$

وفي حالة كانت طوية الشعاع لا تساوي الواحد، يمكننا تعديل ذلك بقسمة مركبات الشعاع على طويلته.

$$U^*=\begin{pmatrix} u_1/||U_\alpha|| \\ u_2/||U_\alpha|| \\ \dots \\ u_p/||U_\alpha|| \end{pmatrix}$$

الاستقلال الخطي: نقول عن سلسلة الأشعة $(U_1,U_2,\dots U_P)$ أنها مستقلة خطيا، إذا وفقط إذا كان:

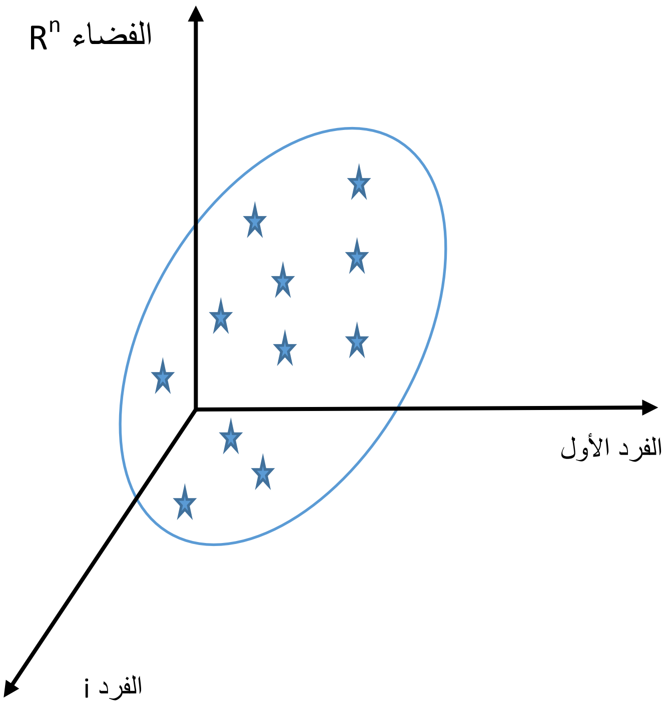
$\alpha_1U_1+\alpha_2U_2+\ldots\ldots.+\alpha_pU_P=0_p \rightarrow \alpha_1=\alpha_2=\ldots..=\alpha_p=0$

أمّا بيانيا، فيفسر الاستقلال الخطي على أساس التعامد، فسلسلة الأشعة $(U_1,U_2,\ldots U_P)$ تكون مستقلة خطيا، إذا كانت متعامدة مثنى مثنى. كما أنّ هذه الأشعة تكون أساسا للفضاء الشعاعي الجزئي E_P إذا كانت مستقلة خطيا.

التحليل في الفضاء R^n (تحليل المتغيرات في فضاء الأفراد): نعمل في هذا الفضاء على تحليل p متغير في الفضاء ذو البعد n، فننظر إلى الأفراد كمحاور والمتغيرات كنقاط. فتكون مثلا، إحداثية المتغير الأول في هذا الفضاء:

$n^1=(n_{11},n_{21},\ldots\ldots,n_{i1},\ldots.n_{n1})$

بمعنى أنّ احداثية المتغير الأول على المحور الأول هي n_{11} وعلى المحور الثاني هي n_{21} وعلى المحور n هي n_{n1}. ويتم تمثيل المتغيرات في فضاء الأفراد كما يلي:



ولحساب المسافة بين الفردين i و $í$ في الفضاء R^p:

$n_j=(n_{1j}\ldots.,n_{ij},..n_{nj})$, $n_{j'}=(n_{1j'}\ldots.,n_{ij'},\ldots..n_{nj'})$

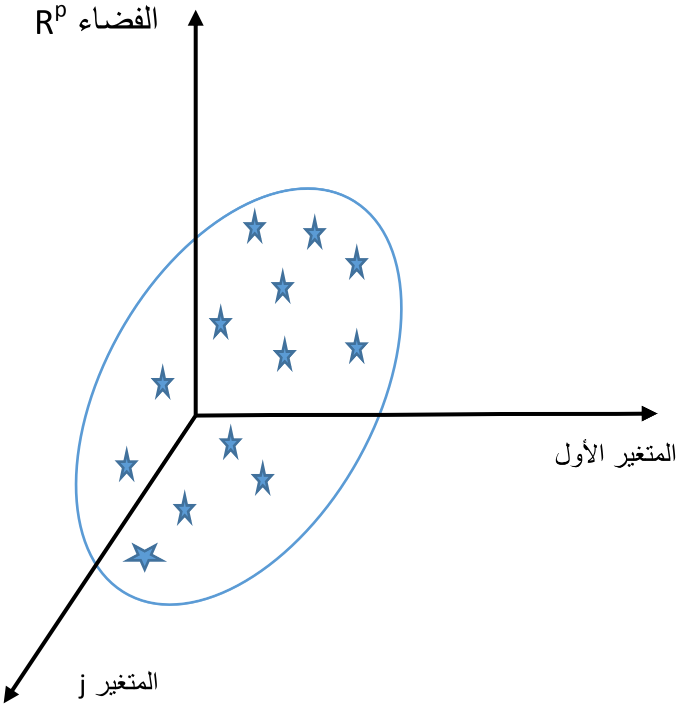
$d(j,j')= \|j-j'\|=\sqrt{(n_{1j}-n_{1j\prime})^2+\cdots+(n_{nj}-n_{nj\prime})^2}= \sqrt{\sum_{i=1}^{n}(n_{ij}-n_{ij\prime})^2}$

بشرط أن تكون المتغيرات متجانسة (لها وحدة القياس نفسها).

التحليل في الفضاء R^p (تحليل الأفراد في فضاء المتغيرات): نعمل في هذا الفضاء على تحليل n فرد في الفضاء ذو البعد p، فننظر إلى المتغيرات كمحاور والأفراد كنقاط. فتكون مثلا، إحداثية الفرد الأول في هذا الفضاء:

$n^1=(n_{11},n_{12},\ldots\ldots,n_{1j},\ldots.n_{1p})$

بمعنى أنّ احداثية الفرد الأول على المحور الأول هي n_{11} وعلى المحور الثاني هي n_{12} وعلى المحور p هي n_{1p}. ويتم تمثيل الأفراد في فضاء المتغيرات كما يلي:



ولحساب المسافة بين الفردين i و $\acute{i}$ في الفضاء R^p:

$n_i=(n_{i1}....,n_{ij},..n_{ip})$, $n_{i'}=(n_{i'1}....,n_{i'j},.....n_{i'p})$

$$d(i,i')=\|i-i'\|=\sqrt{(n_{i1}-n_{i'1})^2+\cdots.+(n_{ip}-n_{i'p})^2}=\sqrt{\sum_{j=1}^{p}(n_{ij}-n_{i'j})^2}$$

ثانيا: الجبر الخطي وحساب المصفوفات.

يتم تفسير الطرق الاحصائية متعددة المتغيرات باستخدام مفاهيم جبر المصفوفات، وعليه، لا بد على من يريد التعامل مع تحليل المعطيات أن يكون على دراية بجبر المصفوفات. وفيما يلي أهم العمليات التي يتم اجرائها على المصفوفات، ويحتاجها الذي يريد العمل على تحليل المعطيات.

1. تعريف المصفوفة: هي مجموعة مستطيلة من الأعداد أو من الرموز أو من التعبيرات منتظمة بشكل أعمدة وصفوف. يُدعى كل عنصر من هذه المجموعة بعنصر أو مدخل للمصفوفة. ويستدل عادة على أي مدخل في مصفوفة ما باسم المصفوفة بحرف لاتيني صغير وأسفله رقمين صغيرين بحيث يمثل العدد الأول رقم الصف والثاني رقم العمود .

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

تدعى الخطوط الأفقية في المصفوفة بالأسطر بينما تدعى الخطوط العمودية باسم عمود. أما الأعداد فتدعى مدخلات المصفوفة أو عناصر المصفوفة. ترمز إلى مصفوفة بحرف لاتيني كبير وتحته عددين طبيعيين على شكل جداء هما m و n حيث m هو عدد الصفوف و n عدد الأعمدة. وبالتالي تعرف المصفوفة بعدد الصفوف والأعمدة، وتعرف m و n برتبة المصفوفة. فرتبة المصفوفة A أعلاه هي 3*4 أي 4 أسطر و3 أعمدة .

$$\mathbf{A} = \begin{bmatrix} 9 & 8 & 6 \\ 1 & 2 & 7 \\ 4 & 9 & 2 \\ 6 & 0 & 5 \end{bmatrix}$$

وفيما يلي، على سبيل المثال، مصفوفة تحتوي على صفين وثلاثة أعمدة:

$$\begin{bmatrix} 1 & 9 & 13 \\ 20 & 55 & 4 \end{bmatrix}$$

أما المصفوفة ذات العمود الواحد، فتحدد بالشكل $1 \times m$ وتعرف باسم متجه عمودي. بينما المصفوفة المؤلفة من صف وحيد و n عمود تحدد بالشكل $1 \times n$ وتعرف باسم متجه صفي. كما أنّ المصفوفة هي جدول من العناصر قد تكون أعدادا حقيقية أو أعدادا مركبة وقد تكون دوالا وهي صورة رياضية لوضع الأعداد في جدول.

ويمكن إجراء عمليتي الجمع والطرح على المصفوفات المتساوية القياس. كما يمكن ضرب المصفوفات بانسجام معين في القياس. ولهذه العمليات العديد من خصائص الحساب العادي، باستثناء أن ضرب المصفوفات ليس بعملية تبديلية، وبشكل عام يمكن أن نقول إن A.B لا يساوي B.A.

2. أنواع المصفوفات:

1.2. **المصفوفة المربعة:** هي المصفوفة التي يكون فيها عدد الصفوف مساويا لعدد الاعمدة (nxn).

$$A_{(2*2)}=\begin{bmatrix}1 & 2\\4 & 0\end{bmatrix}$$

2.2. **المصفوفة القطرية:** مصفوفة مربّعة جميع عناصرها معدومة، ما عدى عناصر القطر الرئيسي.

$$D=\begin{bmatrix}d_1 & 0 & 0\\0 & d_2 & 0\\0 & 0 & d_3\end{bmatrix}$$

فإذا كانت كل عناصر القطر الرئيسي متساوية وباقي العناصر معدومة، سميت بالمصفوفة السلمية. وتكون المصفوفة القطرية قابلة للعكس إذا وفقط إذا كانت جميع عناصر القطر الرئيسي غير صفرية.

$$D^{-1}=\begin{bmatrix}\frac{1}{d_1} & 0 & 0\\0 & \frac{2}{d_2} & 0\\0 & 0 & \frac{3}{d_3}\end{bmatrix}$$

3.2. **المصفوفة المتناظرة:** مصفوفة مربّعة جميع عناصرها المتناظرة بالنسبة للقطر الرئيسي متساوية.

$$\begin{matrix}3 & 1 & 2\\1 & 3 & 6\\2 & 6 & 4\end{matrix}$$

4.2. **مصفوفة الوحدة:** مصفوفة قطرية، وجميع عناصر القطر الرئيسي تساوي الواحد.

$$I_d=\begin{matrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{matrix}$$

5.2. **المصفوفة المعدومة:** مصفوفة كل عناصرها معدومة.

6.2. **المصفوفة الشاذة:** مصفوفة مربّعة محددها معدوم (det (A)=0).

7.2. **المصفوفة النظامية:** مصفوفة مربعة محددها غير معدوم.

8.2. **المصفوفة المثلثية العلوية:** مصفوفة تكون فيها جميع العناصر الواقعة تحت القطر الرئيسي معدومة.

$$\begin{matrix}1 & 7 & 8 & 2\\0 & 2 & 3 & 6\\0 & 0 & 4 & 5\\0 & 0 & 0 & 5\end{matrix}$$

9.2. **المصفوفة المثلثية السفلية:** عكس المصفوفة المثلثية العلوية.

10.2. **منقول المصفوفة (Transposition):** وهي المصفوفة الناتجة عن قلب الأسطر أعمدة والأعمدة أسطر، مع المحافظة على الترتيب.

$$A=\begin{bmatrix}1 & 5 & 5\\3 & 4 & 2\\5 & 8 & 9\end{bmatrix},\ A^T=\begin{bmatrix}1 & 3 & 5\\5 & 4 & 8\\5 & 2 & 9\end{bmatrix}$$

ملاحظة: منقول منقول المصفوفة A هي المصفوفة A. وإذا كانت المصفوفة A متناظرة فهي تساوي منقولها.

11.2. **رتبة المصفوفة:** هي عدد الأسطر (الأعمدة) المستقلة خطيا فيها، وتساوي عدد الأسطر (الأعمدة) في المصفوفة المكافئة.

ملاحظة: نعني بمصفوفتين متكافئتين إذا أمكن الحصول على إحداهما من الأخرى بإجراء عمليات أولية متتابعة، حيث نتوقف عن إجراء العمليات الأولية عند الحصول على مصفوفة مكافئة مثلثية، وذلك لأن العمليات الأولية على الأسطر أو الأعمدة لا تؤثر على رتبة المصفوفة.

12.2. **المصفوفة المتماثلة:** نقول عن المصفوفة A أنّها متماثلة، إذا وفقط إذا كان منقولها A^T يساويها.

$$A=\begin{bmatrix}1&4&5\\4&3&0\\5&0&7\end{bmatrix}\text{--------}\rightarrow A^T=\begin{bmatrix}1&4&5\\4&3&0\\5&0&7\end{bmatrix}$$

نقول عن المصفوفتين A و b أنهما مصفوفتان متماثلتان إذا وجدت مصفوفة P، حيث:

$B= P^{-1}AP,\ A=PBP^{-1}$

3. عمليات على المصفوفات:

1.3. **حساب محدد مصفوفة:** في الجبر الخطي، المُحَدِّد (Determinant) لمصفوفة مربعة n×n، هو عدد يمكن أن يحسب من خلال مداخل المصفوفة المربعة، يحدد عددا من خصائص التحويل الخطي الذي تصفه هذه المصفوفة. ويكون الُمحَدِّد مساوِيا للصفر إذا وفقط إذا كانت المصفوفة غير معكوسة. ويرمز عادة لمحدد مصفوفة ما (A)، det (A). وقيمته تحدد قابلية المصفوفة للعكس، كما تحدد رتبة المصفوفة وحلول جملة المعادلات الخطية.

ملاحظة: لا يمكن حساب المحدد إلاّ لمصفوفة مربّعة. ويمكن حساب المحدد حسب أي سطر أو عمود.

1.1.3. محدد مصفوفة 2*2:

$$\begin{vmatrix}a&b\\c&d\end{vmatrix}=ad-bc$$

$$\begin{vmatrix}1&2\\1&3\end{vmatrix}=1\cdot 3-2\cdot 1=1$$

2.1.3. محدد مصفوفة 3*3: يمكن حساب محدد مصفوف 3*3 كما يلي:

$$\det(A)=\begin{bmatrix}a&d&g\\b&e&h\\c&f&i\end{bmatrix}=\left[\begin{matrix}a&d&g\\b&e&h\\c&f&i\end{matrix}\right]\begin{matrix}a&d\\b&e\\c&f\end{matrix}= (a.e.i + d.h.c + g.b.f) - (d.b.i + a.h.f + g.e.c).$$

$$\det(A)=\begin{bmatrix}2&1&0\\3&3&1\\4&2&5\end{bmatrix}\begin{matrix}2&1\\3&3\\4&2\end{matrix} = (2*3*5+1*1*4+0*3*2)(1*3*5+2*1*2+0*3*4)=(34)-(19)=15.$$

أو بالطريقة التالية:

det (A)= $\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$= a$\begin{bmatrix} e & h \\ f & i \end{bmatrix}$−d$\begin{bmatrix} b & h \\ c & i \end{bmatrix}$+g$\begin{bmatrix} b & e \\ c & f \end{bmatrix}$= a(i*e−h*f)−d(i*b−c*h)+g(b*f−e*c).

A= $\begin{bmatrix} 2 & 1 & 0 \\ 3 & 3 & 1 \\ 4 & 2 & 5 \end{bmatrix}$

det (A)=2 $\begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix}$ − 1$\begin{bmatrix} 3 & 1 \\ 4 & 5 \end{bmatrix}$ + 0$\begin{bmatrix} 3 & 3 \\ 4 & 2 \end{bmatrix}$= 2*(5*3−1*2)−1*(5*3−1*4)+0=26−11= 15.

3.1.3. محدد مصفوفة 4*4:

A= $\begin{bmatrix} 0 & 4 & 0 & -3 \\ 1 & 1 & 5 & 2 \\ 1 & -2 & 0 & 6 \\ 3 & 0 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix}$

نبحث عن العمود الذي يحتوي على أكبر عدد من الأصفار، في مثالنا العمود الثالث.

det (A)= −5 $\begin{bmatrix} 0 & 4 & -3 \\ 1 & -2 & 6 \\ 3 & 0 & 1 \end{bmatrix}$.

في حالة عدم وجود أصفار نستخدم حذف كاوس لتصفير عمود من الأعمدة ونتبع الطريقة السابقة.

2.3. خصائص المحدد:

أ. إذا أخذت جميع عناصر أي صف أو أي عمود (أو أكثر) القيمة صفر في محدد ما، فإن قيمة هذا المحدد تتلاشى أي تساوي صفر.

det$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 2 \\ 3 & 2 & 4 \end{bmatrix}$= ((0*1*4)+(0*2*3)+(0*1*2)−((0*1*4)+(0*2*2)+(0*1*3)=0

ب. إذا تطابقت (قيمة وإشارة) العناصر المتناظرة في أي صفين أو أي عمودين في محدد ما، فإن قيمة هذا المحدد تتلاشى أي تأخذ القيمة صفر.

det$\begin{bmatrix} 1 & -2 & 8 \\ 2 & 4 & 5 \\ 1 & -2 & 8 \end{bmatrix}$= ((1*4*8)+(−2*5*1)+(8*2*−2))−((−2*2*8)+(1*5*−2)+(8*4*1))=(32−10−32)−(−32−10+32)=−10+10=0.

ج. إذا كانت جميع عناصر محدد ما كل منها تأخذ القيمة (صفر) ما عدا العناصر التي تقع على القطر الرئيسي، فإن قيمة هذا المحدد نحصل عليها بضرب عناصر القطر الرئيسي.

det $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$= 1*5*3=15

د. محدد مصفوفة قطرية أو مصفوفة مثلثية علوية أو سفلية يساوي جداء عناصر القطر الرئيسي.

هـ. لا تتغير قيمة المحدد إذا أضيف إلى أي عمود (سطر) من أعمدة المصفوفة مزج خطي لبقية الأعمدة (الأسطر).

4. أثر المصفوفة (Trace): أثر المصفوفة المربعة هو مجموع عناصر القطر الرئيسي، أي مجموع قيمها الذاتية (Valeurs Propres).

$$A=\begin{bmatrix}1 & 5 & 5\\ 3 & 4 & 2\\ 5 & 8 & 9\end{bmatrix},\ \text{Trace }(A)=1+4+9=14.$$

5. جمع وطرح مصفوفتين: لجمع وطرح مصفوفتين يجب أن تكونا من نفس الدرجة.

$$\begin{bmatrix}5 & 3\\ 1 & -1\end{bmatrix}-\begin{bmatrix}2 & 4\\ 1 & -3\end{bmatrix}=\begin{bmatrix}1 & -1\\ 0 & -4\end{bmatrix}$$

لا يمكن جمع أو طرح مصفوفتين غير متساويتين في الدرجة. كما أنّ جمع المصفوفات عملية تبديلية وتجميعية، والمصفوفة المعدومة حيادية في الجمع.

6. ضرب المصفوفات:

1.6. ضرب مصفوفة في عدد: إذا تم ضرب مصفوفة في عدد معين (x)، فإنّ جميع عناصر المصفوفة تضرب في هذا العدد.

$$5*\begin{bmatrix}3 & 2 & 1\\ 4 & -1 & 0\\ 2 & 6 & 5\end{bmatrix}=\begin{bmatrix}15 & 10 & 5\\ 20 & -5 & 0\\ 10 & 30 & 25\end{bmatrix}$$

2.6. ضرب مصفوفة في متجه (شعاع): عند ضرب شعاع في مصفوفة لا بد أن يكون عدد أعمدة الشعاع يساوي عدد أسطر المصفوفة.

$$U_{13}=[1\quad 2\quad 3].\ A_{33}=\begin{bmatrix}3 & 2 & 1\\ 4 & -1 & 0\\ 2 & 6 & 5\end{bmatrix},\ U_{13}*A_{33}=B_{13}$$

$$[1\quad 2\quad 3]x\begin{bmatrix}3 & 2 & 1\\ 4 & -1 & 0\\ 2 & 6 & 5\end{bmatrix}=[17\quad 18\quad 16]\ /(1*3+2*4+3*2),(1*2+2*(-1)+3*6),(1*1+2*0+3*5).$$

المصفوفة الناتجة B تكون رتبتها بعدد صفوف الشعاع C وأعمدة المصفوفة A.

3.6. ضرب مصفوفة في مصفوفة: حتى يمكن ضرب مصفوفتين يجب أن يكون عدد أعمدة الأولى يساوي عدد صفوف الثانية. لذلك فالضرب غير تبديلي في المصفوفات.

$$\begin{bmatrix}2 & 1 & 3\\ 4 & 0 & 1\\ 0 & 2 & 1\end{bmatrix}X\begin{bmatrix}1 & 2\\ 0 & 2\\ 1 & 3\end{bmatrix}=\begin{bmatrix}5 & 15\\ 5 & 20\\ 1 & 7\end{bmatrix}$$

المصفوفة الناتجة تكون رتبتها بعدد صفوف الأولى وأعمدة الثانية. ولإجراء عملية الضرب نقوم بتوزيع صفوف الأولى على أعمدة الثانية بالترتيب.

7. معكوس المصفوفة (inverse) (A^{-1}):

لا يمكن حساب معكوس أي مصفوفة إلا إذا كانت مربعة.

Inverse (A)= $\frac{1}{\det A} adj\ (A)$= A^{-1}.adj(A).

نحسب المحدد كما رأينا سابقا، ثم نحسب المرافق:

adj(A)= $\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$=

a= (e*i–h*f), d=(b*i–h*c), g=(b*f–e*c), b=(d*i–g*f), e=(a*i–g*c), h=(a*f–d*c), c=(d*h–g*e), f=(a*h–g*b), i=(a*e–d*b).

A = $\begin{bmatrix} 1 & 0 & -3 \\ 2 & -2 & 1 \\ 0 & -1 & 3 \end{bmatrix}$

det (A)= $(-5)-0+(-3)(-2-0))= 1$

adj (A)= $\begin{bmatrix} -5 & 6 & -2 \\ -3 & 3 & -1 \\ -6 & 7 & -2 \end{bmatrix}$

نضرب إشارة كل عنصر في المصفوفة في الإشارة المرافقة له.

adj (A)= $\begin{bmatrix} -5 & -6 & -2 \\ 3 & 3 & 1 \\ -6 & -7 & -2 \end{bmatrix}$

ثم نترك القطر الرئيسي كما هو ونبدل أماكن العناصر المتناظرة بالنسبة للقطر الرئيسي.

adj (A)= $\begin{bmatrix} -5 & 3 & -6 \\ -6 & 3 & -7 \\ -2 & 1 & -2 \end{bmatrix}$

ومنه

$A^{-1}= \frac{1}{1}* \begin{bmatrix} -5 & 3 & -6 \\ -6 & 3 & -7 \\ -2 & 1 & -2 \end{bmatrix} = \begin{bmatrix} -5 & 3 & -6 \\ -6 & 3 & -7 \\ -2 & 1 & -2 \end{bmatrix}$

ملاحظة: معكوس A^{-1} يساوي المصفوفة الأصلية.

8. القيم الذاتية:

تعريف: لتكن A مصفوفة من الدرجة n يسمى العدد الحقيقي λ قيمة ذاتية لمصفوفة A إذا وجد متجه غير صفري U بحيث يكون: AU=λU . يسمى U المتجه الذاتي للمصفوفة A.

ملاحظة: للنظام 0= A –I λ)U) حل غير تافه إذا وفقط إذا كانت المصفوفة A– λI ليس لها معكوس، حيث I هي مصفوفة الوحدة .

أي إذا وفقط إذا كان $\det(A-\lambda I)=0$ فإذا كان λ مجهول فإنّ $\det(A-\lambda I)$ عبارة عن كثيرة حدود في المجهول λ تسمى كثيرة الحدود المميزة للمصفوفة A. ولحساب القيم والمتجهات الذاتية لمصفوفة مربعة نتبع ما يلي:

أولاً: نحسب كثيرة الحدود المميز $h(\lambda)=\det(A-\lambda I)$.

ثانيا: نحسب القيم المميزة للمصفوفة A وهي عبارة عن حلول المعادلة المميزة $h(\lambda)=0$.

ثالثاً: لكل قيمة ذاتية λi نحسب المتجه الذاتي المقابل لها، وذلك بحل نظام المعادلات المتجانس:

$$(A-\lambda iI)\,U=0$$

مثال: لتكن المصفوفة الموالية:

$$A=\begin{pmatrix}1 & -1\\ 2 & 4\end{pmatrix}$$

إذا:

$$h\quad(\lambda)=\det(A-\lambda I)=\begin{vmatrix}1-\lambda & 1\\ -2 & 4-\lambda\end{vmatrix}=0$$

ومنها نجد:

$$(\lambda-1)(\lambda-4)+2=\lambda^2-5\lambda+6=0$$

حصلنا على كثير حدود من الدرجة الثانية، لإيجاد قيم λ، نحسب مميّز كثير الحدود.

$$ax^2+bx+c=0$$

مميز كثير الحدود من الدرجة 2:

$$\Delta=b^2-4ac=(-5)^2-4(1*6)=25-24=1$$

حيث: b هو معامل λ، a معامل ^{2} ، c الثابت.

وبما أنّ قيمة المميز أكبر من الصفر، فهي تقبل حلاّن:

$$\lambda=\frac{-b\pm\sqrt{\Delta}}{2a}$$

$$\lambda1=\frac{-(-5)-\sqrt{1}}{2}=2,\ \lambda2=\frac{-(-5)+\sqrt{1}}{2}=3$$

الآن نعين المتجه المميز للقيمة $\lambda1=2$ فنعوض عنها في: $(A-\lambda I)\,U=0$

ونحصل على:

$$\begin{pmatrix}\lambda-1 & 1\\ -2 & \lambda-4\end{pmatrix}\begin{pmatrix}X_1\\ X_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

$$\begin{pmatrix}1 & 1\\ -2 & -2\end{pmatrix}\begin{pmatrix}X_1\\ X_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

$$X1+X2=0$$

$$-2X1-2X2=0$$

أي أن: $X1=-X2$

إذا للنظام عدد لانهائي من الحلول، ولذلك فإنّه يوجد عدد لا نهائي من المتجهات الذاتية المقابلة للقيمة $\lambda 1=2$.

فإذا حددنا $X2=1$ فإنّ $X1=-1$ ونحصل على $X=\begin{pmatrix}-1\\1\end{pmatrix}$.

وبنفس الطريقة نجد أنّ $X=\begin{pmatrix}-1\\2\end{pmatrix}$ مقابل القيمة $\lambda 2=3$.

مثال2: عين القيم الذاتية للمصفوفة الموالية:

$$\begin{matrix}1 & 0 & -1\\ 1 & 0 & 0\\ -2 & 2 & 1\end{matrix}$$

حساب القيم الذاتية:

$$Det(A-\lambda I)=0$$

$$\det\left(\begin{bmatrix}1 & 0 & -1\\ 1 & 0 & 0\\ -2 & 2 & 1\end{bmatrix}-\lambda*\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\right)=0$$

$$\det\begin{bmatrix}1-\lambda & 0 & -1\\ 1 & -\lambda & 0\\ -2 & 2 & 1-\lambda\end{bmatrix}=(1-\lambda)\begin{bmatrix}-\lambda & 0\\ 2 & 1-\lambda\end{bmatrix}-0\begin{bmatrix}1 & 0\\ -2 & 1-\lambda\end{bmatrix}+(-1)\begin{bmatrix}1 & -\lambda\\ 2 & 2\end{bmatrix}=$$

$$((1-\lambda)(-\lambda)(1-\lambda)-(1-\lambda)*0*2-0+(-2+2\lambda)=(1-\lambda)(-\lambda+\lambda^2)+2\lambda-2=$$

$$-\lambda+\lambda^2+\lambda^2-\lambda^3+2\lambda-2=-\lambda^3+2\lambda^2+\lambda-2$$

حصلنا على كثير حدود من الدرجة الثالثة. ولحل كثير الحدود يمكننا استخدام ثلاثة طرق: طريقة المطابقة، طريقة القسمة الإقليدية، خوارزمية هورنر.

1. طريقة المطابقة: علينا تبسيط كثير الحدود:

علينا إيجاد جذر كثير الحدود لدينا، الذي يحقق $P(\lambda)=0$.

نجرب 1، -1، 2، -2، وهكذا...

نجرب مع -1:

$$P(-1)=(-1)^3-2(-1)^2-(-1)+2=-1-2+1+2=0$$

إذا -1 جذر لكثير الحدود، ويمكننا كتابة كثير الحدود على الشكل:

$$P(x)=(x-\alpha)(ax^2+bx+c)$$

حيث: $\alpha=-1$.

$$P(x)=(x+1)(ax^2+bx+c)$$

ننشر كثير الحدود:

$$P(x)=ax^3+bx^2+cx+ax^2+bx+c$$

$$P(x)=ax^3+(b+a)x^2+(c+b)x+c$$

نقوم الآن بعملية المطابقة:

$$P(x)=ax^3+(b+a)x^2+(c+b)x+c$$

$P(\lambda)= \lambda^3-2\lambda^2-\lambda+2$

$a=1,\ b+a=-2 \Rightarrow b=-3,\ c=2$

نعوض قيم المعاملات في كثير الحدود لدينا، فنجد:

$(\lambda+1)(\lambda^2-3\lambda+2)=0$

هذا يستلزم أنّ:

$\lambda+1=0 \Rightarrow \lambda=-1$ أو $\lambda^2-3\lambda+2=0$

نحسب المميز لكثير الحدود من الدرجة الثانية:

$\Delta=b^2-4ac=(-3)^2-4(1*2)=1$

وبما أنّ قيمة المميز أكبر من الصفر، فهي تقبل حلاّن:

$$\lambda=\frac{-b\pm\sqrt{\Delta}}{2a}$$

$$\lambda 2=\frac{-(-3)-\sqrt{1}}{2}=1,\ \lambda 3=\frac{-(-3)+\sqrt{1}}{2}=2$$

2. الطريقة الثانية، القسمة الإقليدية:

$-\lambda^3+2\lambda^2+\lambda-2 \quad / \ \lambda+1$

$-\lambda^3 \quad -\lambda^2 \quad (-) \qquad -\lambda^2+3\lambda-2$

$0 \quad 3\lambda^2+\lambda-2$

$-3\lambda^2-3\lambda$

$0 \quad -2\lambda-2$

$+2\lambda+2$

$0 \quad 0$

$(\lambda+1)(\lambda^2-3\lambda+2)$

3. الطريقة الثالثة، خوارزمية هورنر:

وجدنا أنّ -1 جذر للمعادلة ننقله للخانة الملونة بالأحمر معكوس الإشارة، وفي الخانات باللون الأصفر نكتب المعاملات بإشارتهم، ثم في الخانة الملونة بالأخضر نضع صفر، ونضرب افقيا ونجمع عموديا.

	λ^3	$-2\lambda^2$	$-\lambda$	$+2$
	1 +	-2 +	-1 +	2 +
-1	0	-1	3	-2
	1	-3	2	0

$(\lambda+1)(\lambda^2-3\lambda+2)$

2.4. حساب الاشعة الذاتية:

$(A-\lambda I)U=0$

حساب الشعاع الذاتي المرتبط بـ $\lambda_1=-1$:

لدينا: $\lambda_1=-1$

$$\begin{bmatrix}1-\lambda & 0 & -1\\ 1 & -\lambda & 0\\ -2 & 2 & 1-\lambda\end{bmatrix}\begin{bmatrix}X1\\X2\\X3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \Rightarrow \begin{bmatrix}2 & 0 & -1\\ 1 & 1 & 0\\ -2 & 2 & 2\end{bmatrix}\begin{bmatrix}X1\\X2\\X3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

حتى نصل على الحل ببساطة نستخدم حذف كاوس (تعتمد على تصفير المثلث أسفل القطر الرئيسي):

$$\begin{array}{c} R_1\\ R_2\\ \\ \end{array}\begin{bmatrix}2 & 0 & -1\\ 1 & 1 & 0\\ -2 & 2 & 2\end{bmatrix} \xrightarrow{*1/2} \begin{bmatrix}1 & 0 & -1/2\\ 1 & 1 & 0\\ -2 & 2 & 2\end{bmatrix} \quad r_2=R_2-1*R_1$$

نجري العملية فتصبح المصفوفة كما يلي:

$$\begin{bmatrix}1 & 0 & -1/2\\ 0 & 1 & 1/2\\ -2 & 2 & 2\end{bmatrix} *2 \quad \boxed{r_3=R_3+2*R_1} \Rightarrow \begin{bmatrix}1 & 0 & -1/2\\ 0 & 1 & 1/2\\ 0 & 2 & 1\end{bmatrix} *-2 \quad \boxed{r_3=R_3-2R_2}$$

نجري العمليات فتصبح المصفوفة الجديدة كما يلي:

$$\begin{bmatrix}1 & 0 & -1/2\\ 0 & 1 & 1/2\\ 0 & 0 & 0\end{bmatrix}$$

يمكننا الآن إيجاد المتجهات الذاتية:

$X_1-1/2X_3=0 \Rightarrow X_1=1/2X_3$

$X_2+1/2X_3=0 \Rightarrow X_2=-1/2X_3$

$X_3=X_3$

وعليه،

$$V_1=\begin{bmatrix}1/2X3\\ -1/2X3\\ X3\end{bmatrix}=X_3\begin{bmatrix}1/2\\ -1/2\\ 1\end{bmatrix}=\begin{bmatrix}1/2\\ -1/2\\ 1\end{bmatrix}.$$

الشعاع الذاتي المرافق للقيمة الذاتية $\lambda_2=1$:

$$\begin{matrix}1-\lambda & 0 & -1\\ 1 & -\lambda & 0\\ -2 & 2 & 1-\lambda\end{matrix} \Rightarrow \begin{matrix}0 & 0 & -1\\ 1 & -1 & 0\\ -2 & 2 & 0\end{matrix}$$

نبدل السطرين الأول والثاني في مكان بعضهما فتصبح المصفوفة كما يلي:

$$\begin{matrix}1 & -1 & 0\\ 0 & 0 & -1\\ -2 & 2 & 0\end{matrix} *2 \quad \boxed{r_3=R_3+2*R_1} \Rightarrow \begin{matrix}1 & -1 & 0\\ 0 & 0 & -1\\ 0 & 0 & 0\end{matrix} \quad /\text{-}1 \Rightarrow \begin{matrix}1 & -1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{matrix}$$

$X_1 - X_2 = 0 \Rightarrow X_1 = X_2$

$X_3 = 0$

$V_2 = \begin{bmatrix} x2 \\ x2 \\ 0 \end{bmatrix}$

$V_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

الشعاع الذاتي المرافق للقيمة الذاتية $\lambda_3=2$:

$$\begin{matrix} -1 & 0 & -1 \\ 1 & -2 & 0 \\ -2 & 2 & -1 \end{matrix}$$

$V_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$

مثال 2: احسب القيم والمتجهات الذاتية للمصفوفة:

$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & -4 & 2 \end{bmatrix}$

$\det(A - \lambda I) = 0$

$$\det\left(\begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & -4 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = 0$$

$$\det\left(\begin{bmatrix} 1-\lambda & 2 & 0 \\ 0 & 3-\lambda & 0 \\ 2 & -4 & 2-\lambda \end{bmatrix}\right) = (1-\lambda)\begin{bmatrix} 3-\lambda & 0 \\ -4 & 2-\lambda \end{bmatrix} - 2\begin{bmatrix} 0 & 0 \\ 2 & 2-\lambda \end{bmatrix} + 0\begin{bmatrix} 0 & 3-\lambda \\ 2 & -4 \end{bmatrix} =$$

$(1-\lambda)[(3-\lambda)(2-\lambda) - (-4)(0)] - 2[(2-\lambda)(0) - (2)(0)] =$

$(1-\lambda)[(3-\lambda)(2-\lambda)]$--------→

$\lambda_1=3,\ \lambda_2=2,\ \lambda_3=1$

$D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$(A - \lambda I)U = 0$

$$\left(\begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & -4 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\lambda_1=3$

$$\begin{bmatrix} -2 & 2 & 0 \\ 0 & 0 & 0 \\ 2 & -4 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

إذاً لدينا نظام متجانس من المعادلات الخطية، نقوم بحلها بواسطة حذف كاوس:

$$\begin{bmatrix}-2 & 2 & 0\\ 0 & 0 & 0\\ 2 & -4 & -1\end{bmatrix} \boxed{\longrightarrow} \begin{bmatrix}-2 & 2 & 0\\ 2 & -4 & -1\\ 0 & 0 & 0\end{bmatrix} \boxed{\begin{matrix}*1\\ +\end{matrix}} \begin{bmatrix}-2 & 2 & 0\\ 0 & -2 & -1\\ 0 & 0 & 0\end{bmatrix}$$

$$\begin{bmatrix}-2 & 2 & 0\\ 0 & -2 & -1\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}u_1\\ u_2\\ u_3\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$$

$-2u_1+2u_2=0 \dashrightarrow u_1=u_2$

$-2u_2-u_3=0 \dashrightarrow u_2=-\frac{1}{2}u_3$

$u_3=u_3$

$$U_1=\begin{bmatrix}u_1=-\frac{1}{2}u_3\\ u_2=-\frac{1}{2}u_3\\ u_3=u_3\end{bmatrix}\ldots u_3=1 \dashrightarrow U_1=\begin{bmatrix}-\frac{1}{2}\\ -\frac{1}{2}\\ 1\end{bmatrix}$$

$$U_2=\begin{bmatrix}0\\ 0\\ 1\end{bmatrix},\ U_3=\begin{bmatrix}-\frac{1}{2}\\ 0\\ 1\end{bmatrix}$$

$$U=\begin{bmatrix}-\frac{1}{2} & 0 & -\frac{1}{2}\\ -\frac{1}{2} & 0 & 0\\ 1 & 1 & 1\end{bmatrix}$$

الاستقطار (diagonalization): هو تحويل المصفوفة إلى مصفوفة قطرية مماثلة لها.

$$A=PDP^{-1}$$

$$A=\begin{bmatrix}1 & 2 & 0\\ 0 & 3 & 0\\ 2 & -4 & 2\end{bmatrix},\ U^{-1}=\begin{bmatrix}0 & -2 & 0\\ 2 & 0 & 1\\ -2 & 2 & 0\end{bmatrix}$$

$$A=\begin{bmatrix}-\frac{1}{2} & 0 & -\frac{1}{2}\\ -\frac{1}{2} & 0 & 0\\ 1 & 1 & 1\end{bmatrix}*\begin{bmatrix}3 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 1\end{bmatrix}*\begin{bmatrix}0 & -2 & 0\\ 2 & 0 & 1\\ -2 & 2 & 0\end{bmatrix}=\begin{bmatrix}1 & 2 & 0\\ 0 & 3 & 0\\ 2 & -4 & 2\end{bmatrix}$$

الفصل الثالث:

التحليل بالمركبات الأساسية

PCA

التحليل بالمركبات الأساسية

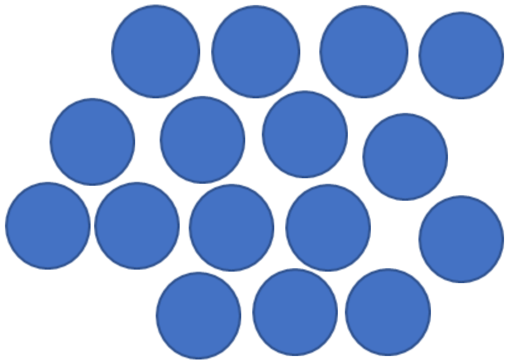
Principal component analysis

هو أسلوب رياضي يقوم على أساس تحويل مجموعة من المتغيرات المترابطة فيما بينها إلى مجموعة جديدة من المتغيرات غير المترابطة (أو المتعامدة – Orthogonal) تدعى بالمركبات الرئيسية. فكل مركبة رئيسية هي عبارة عن توليفة خطية تضم جميع المتغيرات الأصلية. والهدف من هذه الطريقة هو التمثيل البياني لأكبر قدر ممكن من المعلومات الموجودة في جدول البيانات الكمية، الذي يحتوي على n فرد و p متغير، فكل فرد يمثل في فضاء ذو p بعد، وكل متغير يمثل في فضاء ذو n بعد.

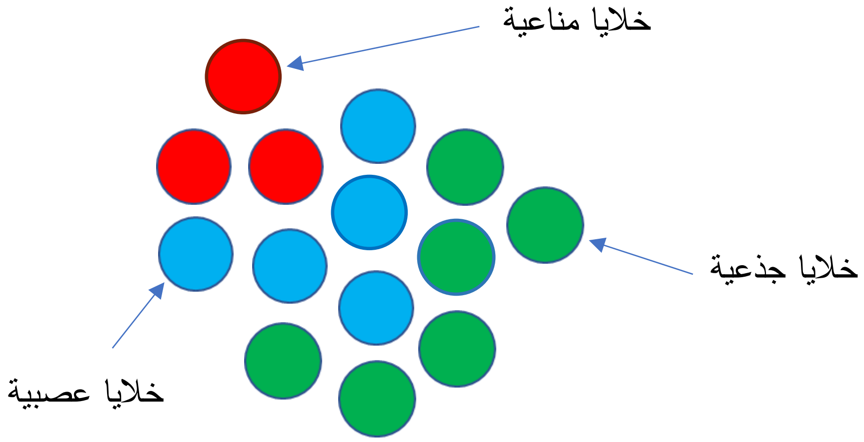
وهو أداة مرنة للغاية تسمح بتحليل مجموعات البيانات التي قد تحتوي، على سبيل المثال، الخطية المتعددة، القيم المفقودة، البيانات الفئوية، والقياسات غير الدقيقة. والهدف هو استخراج المعلومات المهمة من البيانات والتعبير عن هذه المعلومات كمجموعة من المؤشرات الموجزة تسمى المكونات الرئيسية.

فعلى سبيل المثال، يمكننا عن طريق استخدام التحليل بالمركبات الأساسية بناء مؤشر اقتصادي يقيس القدرة الاقتصادية لمجموعة من الأفراد. ويمكن استخدامه أيضا كخطوة وسيطة تكون فيها مخرجاتها جزءًا من تحليل لاحق، مثل: الانحدار أو التجميع أو التصنيف، كما يمكن استخدامه في تقليل البيانات.

لماذا التحليل بالمركبات الأساسية؟ لنفترض أنه لدينا مجموعة من الخلايا:



لاحظ أنّه من غير الممكن ملاحظة أي اختلافات بين الخلايا أعلاه. سنقوم بتحليل الحمض الريبي النووي الرسول (RNA) لتحديد الجينات النشطة في كل خلية، حتى نتمكن من تحديد وظيفتها.

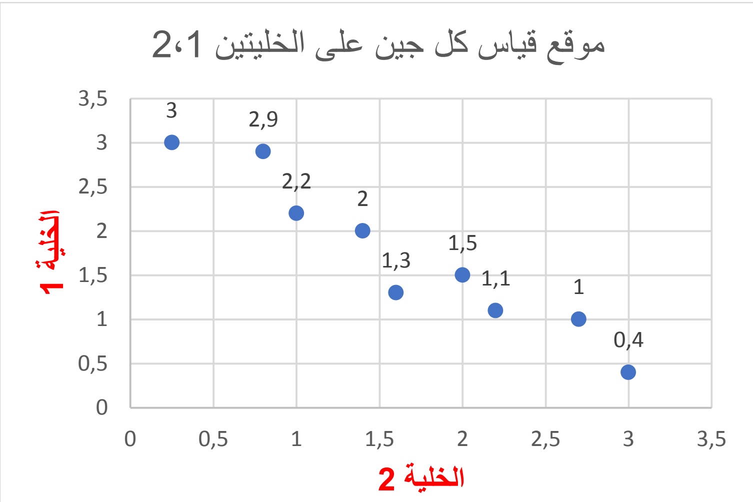


والبيانات في الجدول الموالي، تمثل عدد الجينات المنسوخة في كل خلية.

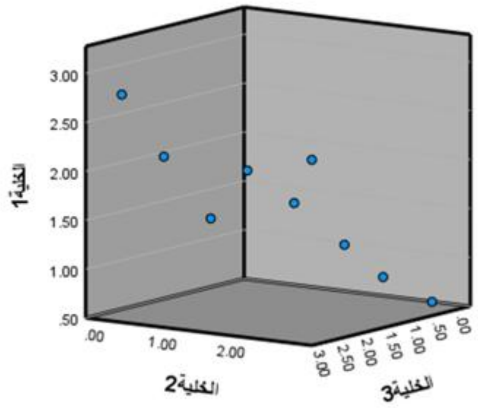
	الخلية1	الخلية 2	الخلية 3	الخلية 4
الجين 1	3	0.25	2.8	0.1
الجين 2	2.9	0.8	2.2	1.8
الجين 3	2.2	1	1.5	3.2
الجين 4	2	1.4	2	0.3
الجين 5	1.3	1.6	1.6	0
الجين 6	1.5	2	2.1	3
الجين 7	1.1	2.2	1.2	2.8
الجين 8	1	2.7	0.9	0.3
الجين 9	0.4	3	0.6	0.1

سنقوم بتعيين موقع قياس كل جين على رسم بياني ثنائي البعد بين خليتين فقط.

	الخلية 1	الخلية 2
الجين 1	3	0.25
الجين 2	2.9	0.8
الجين 3	2.2	1
الجين 4	2	1.4
الجين 5	1.3	1.6
الجين 6	1.5	2
الجين 7	1.1	2.2
الجين 8	1	2.7
الجين 9	0.4	3



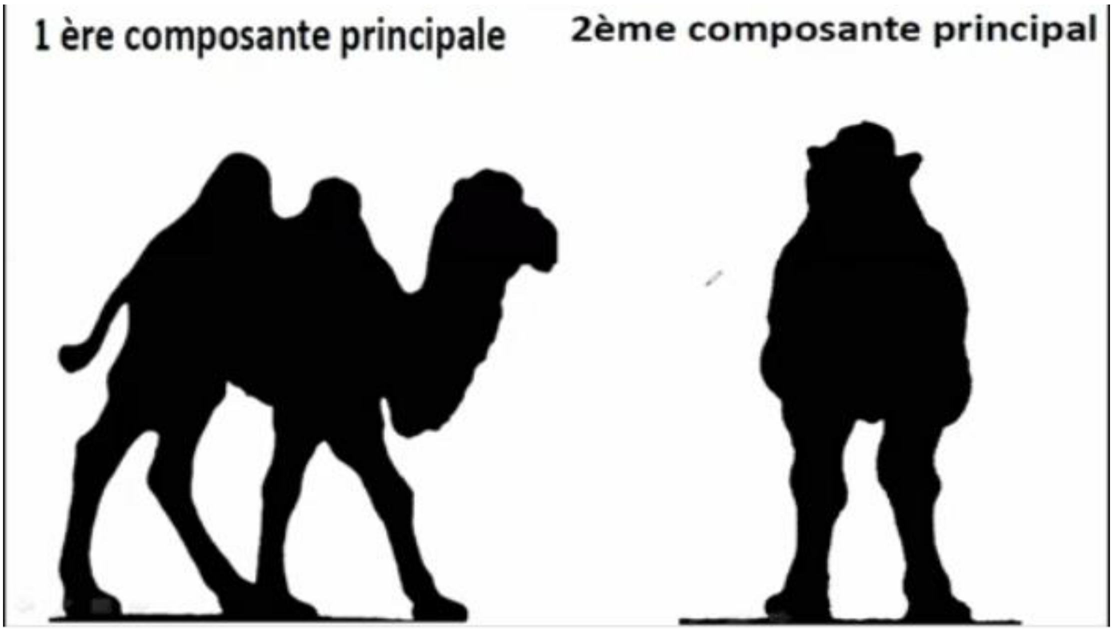
يظهر من الشكل أعلاه أنّ الجين الأول منسوخ في الخلية الأولى بشكل جيد، ومنسوخ في الخلية الثانية بشكل ضعيف، وكذلك الحال بالنسبة للجين الثاني، وهكذا. وبشكل عام يظهر أنّ هناك علاقة عكسية بين الخلية الأولى والثانية، وهذا يعني أنهما نوعان مختلفان من الخلايا بما أنهما يستخدمان نوعا مختلفا من الجينات. لنفترض الآن أننا نريد تحديد العلاقة بين ثلاثة خلايا، سنحتاج إلى ثلاث رسومات بيانية لتمثيل العلاقة بين كل خليتين، أو بدلاً من ذلك نرسم رسم بياني واحد ثلاثي الأبعاد يوضع العلاقة بين الخلايا الثلاثة.



يمكننا ملاحظة أن الجين الأول منسوخ في الخليتين الأولى والثالثة بشكل جيد، ومنسوخ في الخلية الثانية بشكل ضعيف، أمّا الجين التاسع فمنسوخ في الخلية 2 بشكل جيد، ومنسوخ في الخليتين 1 و3 بشكل ضعيف.

إلى هذا الحد الأمر يسير نوعا ما، لكن لو كان لدينا 4 خلايا أو أكثر مثلا، فلدينا حلين، إمّا مقارنة كل خليتين على حدة، وسنحتاج إلى رسم العديد من الرسومات ومحاولة فهم كل رسم لوحده، أو رسم بياني واحد بأربعة أبعاد أو أكثر ونجعل الأمر صعبًا جدًا.

إضافة إلى ذلك لنلاحظ الآن الصورة التي أمامنا من بعد واحد (1ére composante principal) ومن بعدين (2éme composante principal).



قبل ذلك يقول الخبراء إن الرؤية ثلاثية الأبعاد تنتج عنها مشاكل بصرية أكبر من مجرد صعوبة الرؤية مثل مشكلة كسل العين. لو نظرنا فقط للصورة الموجودة على اليمين (2éme composante principal) فهل سنكون متأكدين بنسبة مائة في المائة أنها صورة جمل، بالطبع لا، ربما تكون صورة

لحصان أو إنسان متخفي...الخ. سنكون متأكدين بنسبة 20 في المائة أنّه جمل. الآن لو ننظر إلى الصورة على اليسار يمكننا التأكد بنسبة 70 في المائة من أنه جمل. وإذا نظرنا للصورتين معا سنكون متأكدين بنسبة 90 في المائة أن الصورة لجمل.

إنّ النتيجة المهمة من هذه التجربة أنه عند انتقالنا من رؤية ثلاثية الأبعاد إلى رؤية ثنائية الأبعاد سنحتفظ بتسعين في المائة من المعلومات، والعشرة في المائة الضائعة لم تأثر على قرارنا في أن الصورة لجمل. بمعنى آخر أنّ تقليل عدد الأبعاد مع خسارة طفيفة في المعلومات يكون أفضل لنا لاتخاذ القرارات الصائبة.

ملاحظة: يمكن أن تكون العشرة في المائة الضائعة تمثل لون الشعر والعينين وتفاصيل الفم....الخ.

من أجل ذلك فإنّ التحليل بالمركبات الرئيسية يوفر لنا حلاً يسيرًا، إذ يحول الترابط بين كل خليتين من عدمه إلى رسم ثنائي الأبعاد، الخلايا ذات الترابط الكبير (المتشابهة) تُجمع مع بعضها، والأقل ارتباطا تجمع مع بعضها، وهكذا... ثم يتم ترتيب المحاور بحسب أهميتها، فالاختلافات على طول المحور الرئيسي الأول تكون أكثر أهمية من الاختلافات على طول المحور الرئيسي الثاني...إلخ.

وعلى ذلك، فإنّ الهدف من التحليل بواسطة المركبات الرئيسية هو البحث عن فضاء شعاعي جزئي أقل بعد يسمح لنا بأحسن تمثيل وأكبر كمية من المعلومات، بشرط أن تكون كل المعطيات ذات طبيعة كمية، أو بمعنى آخر الهدف من التحليل بالمركبات الرئيسية هو إيجاد نظام للمحاور يسمح لنا بإعادة بناء وضعية كل نقطة بالنسبة للأخرى من خلال إسقاط سحابات النقاط على هذه المحاور، مما يسمح بتقليص الفضاء المتعدد الأبعاد إلى فضاء ذو بعدين، أي الانتقال من p بعد إلى 2 أو 3 أبعاد ومن 100 في المائة من المعلومات إلى 90 أو 80 في المائة من المعلومات. كما تسمح لنا الطريقة بتحديد الأفراد المتشابهين والأفراد المختلفين، بالإضافة إلى اكتشاف المتغيرات المسؤولة عن التشابه أو الاختلاف بين الأفراد.

أولا، معارف أساسية لفهم كيفية عمل الطريقة:

البيانات: إنّ نقطة الانطلاق في التحليل بالمركبات الرئيسية، هو جدول X يحتوي على عدد p من المتغيرات المستمرة تم قياسها على عدد n من الأفراد. على سبيل المثال، نقاط 5 طلبة في ثلاثة مواد، لدينا ثلاثة متغيرات وخمسة أفراد، يتم قياس كل متغير على جميع الأفراد، ووصف كل فرد بعدد من المتغيرات. وفي بعض الحالات قد نستخدم المتغيرات النوعية.

مثال: يحتوي الجدول الموالي على بيانات عن عشرة (10) منتجات، تشمل تقييم سعر المنتج وتقييم جودته من خلال بعض المواصفات، كالذوق، حجم المنتج، مدى توفّره في السوق، جودة التغليف، وطريقة عرضه (صورة المنتج) (على سلّم من 1 إلى 10). ويريد مدير التسويق توزيع هذه المنتجات في السوق على حسب السعر والمواصفات.

	السعر	الذوق	الحجم	التوفر	التغليف	الصورة
المنتج 1	4,4	8,18	6,32	9,37	8,32	9,3
المنتج 2	4,32	7,12	5,9	8,68	7,18	8,2
المنتج 3	6,11	5,33	5,12	6,79	5,84	5,63
المنتج 4	4,62	7,07	5,93	6,7	7,78	7,97
المنتج 5	5	6,72	6,38	6,14	6,73	7,38
المنتج 6	7,28	5,08	5,18	8,67	5,02	5,8
المنتج 7	8,32	2,83	3,13	7,88	2,83	3,65
المنتج 8	7,93	3,42	4	7,63	3,64	4,15
المنتج 9	4,82	4,34	3,73	5,29	7,3	6,39
المنتج 10	7,59	2,38	2,81	4,43	2,38	2,76

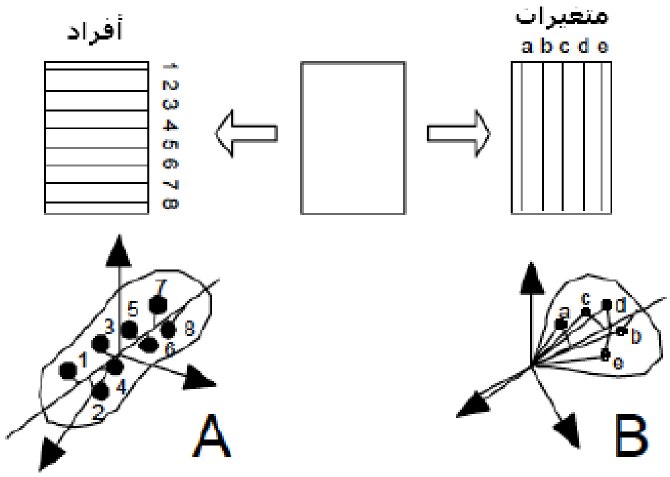
ملاحظة هامة: قبل إجراء التحليل، علينا دائمًا إجراء تحليل استكشافي للبيانات، مثل حساب القيم القصوى والدنيا، المدى، مقاييس النزعة المركزية، مقاييس التشتت، والنظر في توزيع المتغيرات (مثل مخططات الصندوق والمدرج التكراري). يساعدنا ذلك في تحديد القيم المتطرفة أو الأخطاء أو غيرها من الحالات الشاذة في البيانات، التي من الممكن أن تأثر في التحليل وتجعل النتائج عديمة القيمة.

2. البيانات النشطة والتكميلية: إنّ الهدف من وراء تحليل معطيات الجدول أعلاه هو مقارنة المنتجات العشرة اعتمادا على السعر وجودة المنتج، أو على السعر والجودة ومدى توفر المنتج في السوق. يمكننا بالتأكيد تضمين جميع المتغيرات وإجراء التحليل، لكن من الأفضل تحديد مجموعة من المتغيرات التي تكون أكثر تجانساً وفقًا لموضوع معين، وأكثر توافقًا مع أهداف التحليل، كأن نستخدم السعر والجودة (ممثلة في الذوق والحجم والتغليف) لمقارنة المنتجات مع بعضها البعض، ويكون تفسير التشابه أو الاختلاف بين المنتجات أسهل. نسمي المتغيرات المختارة، **بالمتغيرات النشطة** (active variables)، وتشمل العناصر التي سيتم استخدامها لمقارنة المنتجات فيما بينها. ونسمي بقية المتغيرات غير النشطة **بالمتغيرات التكميلية** (supplementary variables). وذلك لا يعني أنه لن يتم استخدام المتغيرات التكميلية، بل يتم استخدامها كمعلومات إضافية قد تساعدنا في شرح التشابه من عدمه بين المنتجات العشرة.

3. **المسافات (Distances):** يمكننا من خلال التحليل بالمكونات الأساسية تصور الاختلافات بين المنتجات العشرة وفقا للسعر، بالإضافة إلى تصور الارتباط بين المتغيرات، وللحصول على هذه التصورات المرئية، نستخدم تقريبًا هندسيا بسيطا فيما يتعلق بمصفوفة البيانات، فننظر لها من جانبين، جانب يخص الصفوف وجانب خاص بالأعمدة. ويتضمن كل تصور النظر في سحابة من النقاط، واحدة للأفراد وواحدة للمتغيرات.

عرض الطريقة:

إنّ نقطة الانطلاق في التحليل بالمركبات الرئيسية، هو جدول X يحتوي على عدد p من المتغيرات تم قياسها على عدد n من الأفراد. على سبيل المثال، نقاط 5 طلبة في ثلاثة مواد، لدينا ثلاثة متغيرات وخمسة أفراد، يتم قياس كل متغير على عدد الأفراد، ووصف كل فرد بعدد من المتغيرات.

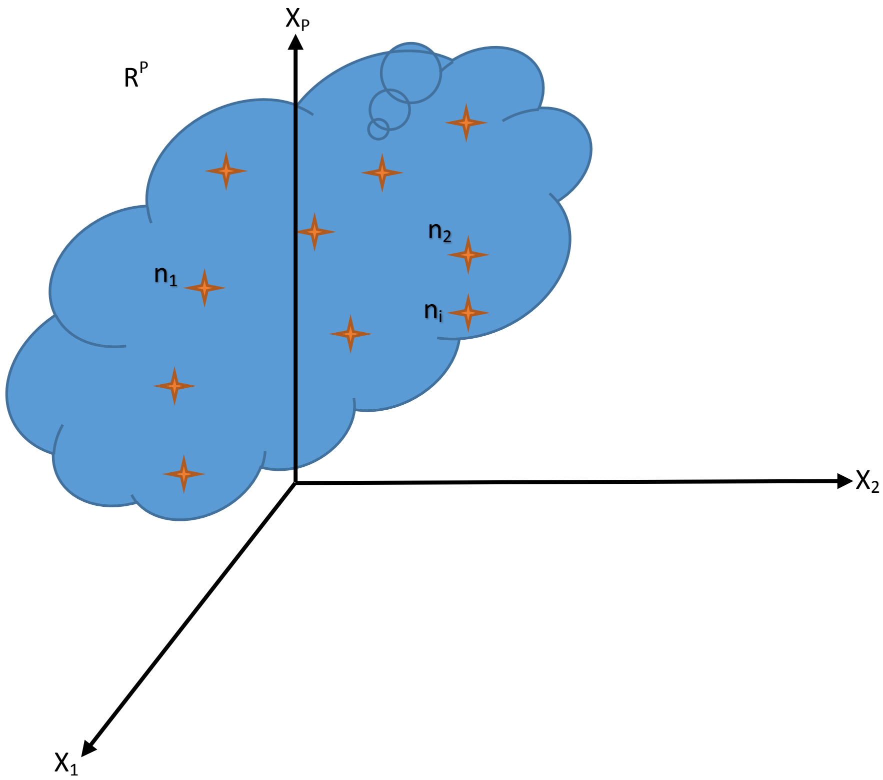


الوضعية A: عدد n من الأفراد في فضاء المتغيرات. مجموع نقط الأفراد في هذا الفضاء تشكل ما يسمى بسحابة النقط NI (point cloud)(nuage de points).

في الصورة أعلاه سحابة من النقاط (الطيور) $(e_1,e_2,\ldots\ldots,e_i,\ldots.e_p)$ في فضاء ذو بعد p ‏(p=2)، كل نقطة تقابل الطائر الذي يتم وصفه بواسطة p من المتغيرات الكمية $(X_1, X_2,\ldots\ldots X_j,\ldots.X_p)$.

الوضعية B: عدد p من المتغيرات في فضاء الأفراد. مجموع نقط المتغيرات في هذا الفضاء تشكل ما يسمى بسحابة النقط NP.

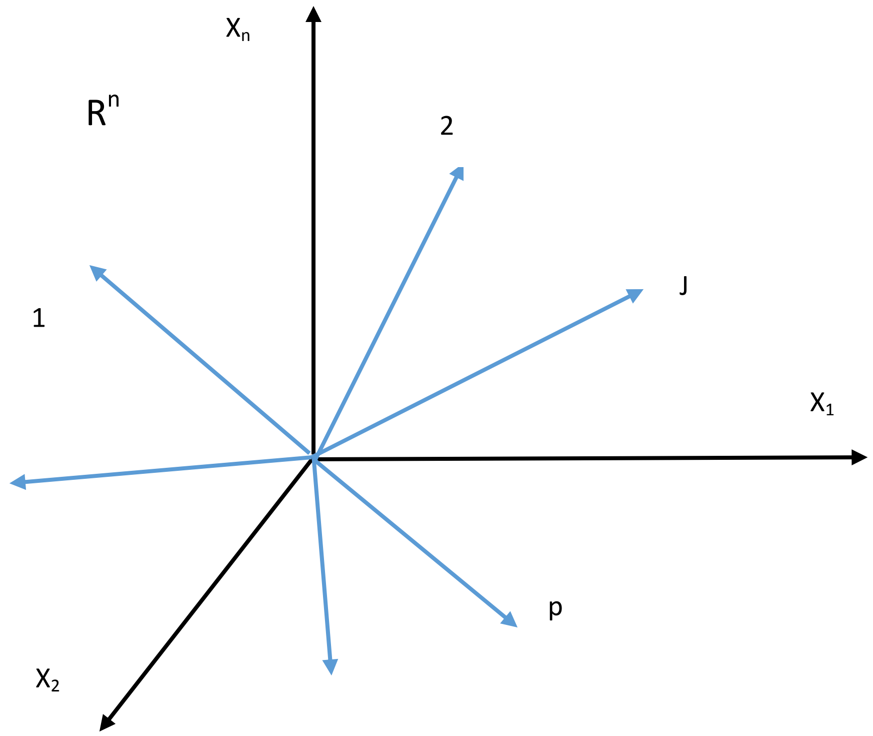
والهدف من ذلك، إسقاط سحابة النقط على المحاور التي تعظّم القصور الذاتي (العطالة) (نعبر عنه بالتباين الكلي) (inertia) (inertie) لهذه السحابة.



في هذه السحابة تشير كل نقطتين قريبتين من بعضهما إلى منتجين بقيم متشابهة في السعر والمواصفات، في المقابل تشير كل نقطتين متباعدتين عن بعضهما إلى منتجين بأسعار مختلفة للغاية. ولقياس مفهوم التقارب بين نقاط الصف (المنتجات)، نحتاج إلى تحديد مقياس للمسافة، والمقياس الأكثر بديهية لقياس المسافة في الفضاء، هي المسافة الإقليدية بين نقطتين، والتي تُعطى بالعلاقة:

$$d^2(e,e')=\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2$$

2.3. سحابة المتغيرات: تمثل هذه السحابة الارتباطات بين المتغيرات، فكل متغير يقيس خاصية ملحوظة على المنتجات. وبالتالي، نحتاج إلى تحديد المسافة بين نقاط المتغيرات التي تلتقط كثافة وطبيعة الارتباط بين المتغيرات. حيث تشير نقطتي متغيرين قريبين من بعضهما البعض إلى أنهما يأخذان قيمًا مرتبطة في مجموعة المنتجات العشرة، فإذا عرفنا قيم أحد المتغيرات، يمكننا معرفة قيم المتغير الآخر. ويتم قياس الارتباط بين المتغيرات باستخدام معامل الارتباط الخطي، وعند استخدامنا لهذا المعامل كمسافة، فإنّ تصور المتغيرات يصبح عرضًا مرئيًا لمصفوفة الارتباطات بين المتغيرات.



4. مصفوفة الأوزان: يعبر الوزن المخصص لكل فرد عن الأهمية التي نريد أن نعطيها له في الدراسة (تمثيل العينة المدروسة في المجتمع). ويمكن أن يكون لكل فرد وزن i ، خاصة عندما لا يكون للأفراد نفس الأهمية في الدراسة (عينات معدلة، بيانات مجمعة،...الخ).

$$\sum_{i=1}^{n} p_i = 1$$

ونكتب المصفوفة القطرية للأوزان على الشكل التالي:

$$D_P = \begin{bmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_i & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & p_n \end{bmatrix}$$

وهي مصفوفة متماثلة:

$$D_P^T = D_P$$

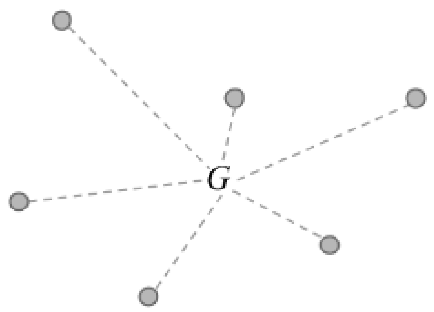
فإذا كان لجميع الأفراد نفس الوزن، تكون:

$$p_i = \frac{1}{n}$$

$$D_P = \frac{1}{n} I_n$$

حيث: I_n مصفوفة الوحدة.

5. مركز ثقل سحابة النقط: لكل سحابة من النقط نقطة متوسطة تسمى بمركز ثقل سحابة النقط.



وتعطى احداثيات مركز الثقل بالعلاقة:

$$G_j = \sum_{i=1}^{n} p_i X_i$$

وبما أنّ النقطة المتوسطة تتوافق مع المفهوم الإحصائي المتوسط، نكتب:

$$\bar{X} = \sum_{i=1}^{n} p_i x_i = p_1 x_1 + p_2 x_2 + \dots + p_n x_n$$

وإذا كان لكل الأفراد نفس الوزن، نكتب:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

ونكتب شعاع مركز الثقل G_X كما يلي:

$$G_X = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ \dots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{bmatrix} = \begin{bmatrix} \overline{X_1} \\ \overline{X_2} \\ \dots \\ \overline{X_p} \end{bmatrix} \in M(p, 1).$$

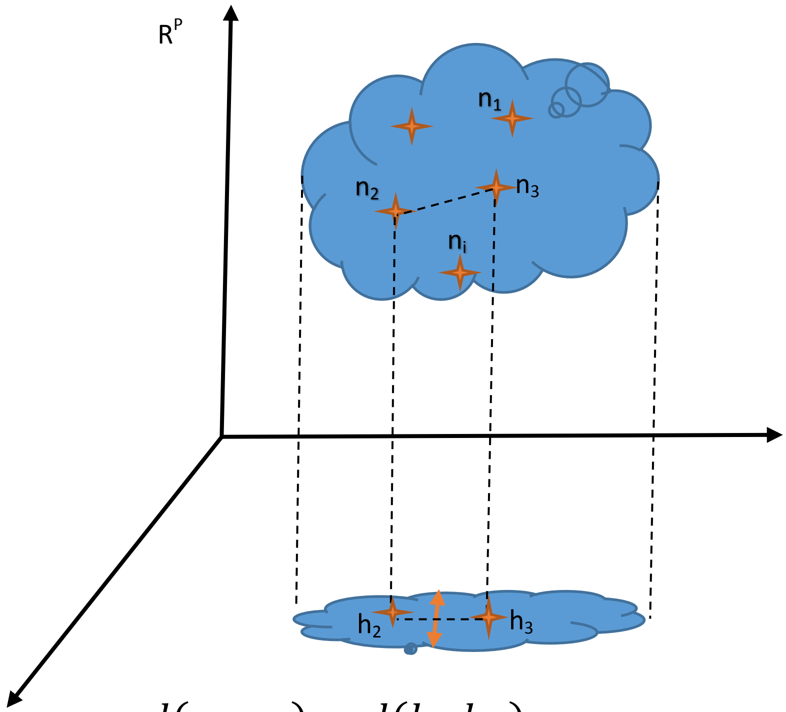
كما يمكن حسابه من الصيغة المصفوفية التالية:

$$G_X = X^T D_P I_n$$

حيث: X^T منقول مصفوفة البيانات، D_P مصفوفة الأوزان، I_n متجه الوحدة.

6. المسافات بين النقاط: نظرًا لأن سحابتي الأفراد والمتغيرات تقعان في مساحة متعددة الأبعاد، فلا يمكننا ملاحظتها مباشرة، لذلك علينا البحث عن مستوى نعرض عليه سحابة النقاط بطريقة تجعل التكوين الذي تم الحصول عليه أقرب ما يمكن من التكوين الأصلي للسحابة في الفضاء متعدد الأبعاد. نسمي هذا المستوى بالمستوى العاملي، ولكي نحصل عليه، علينا أن نجعل المسافات الإجمالية بين النقاط المسقطة أقرب ما يمكن إلى المسافات الحقيقية بين النقاط في الفضاء الأصلي.

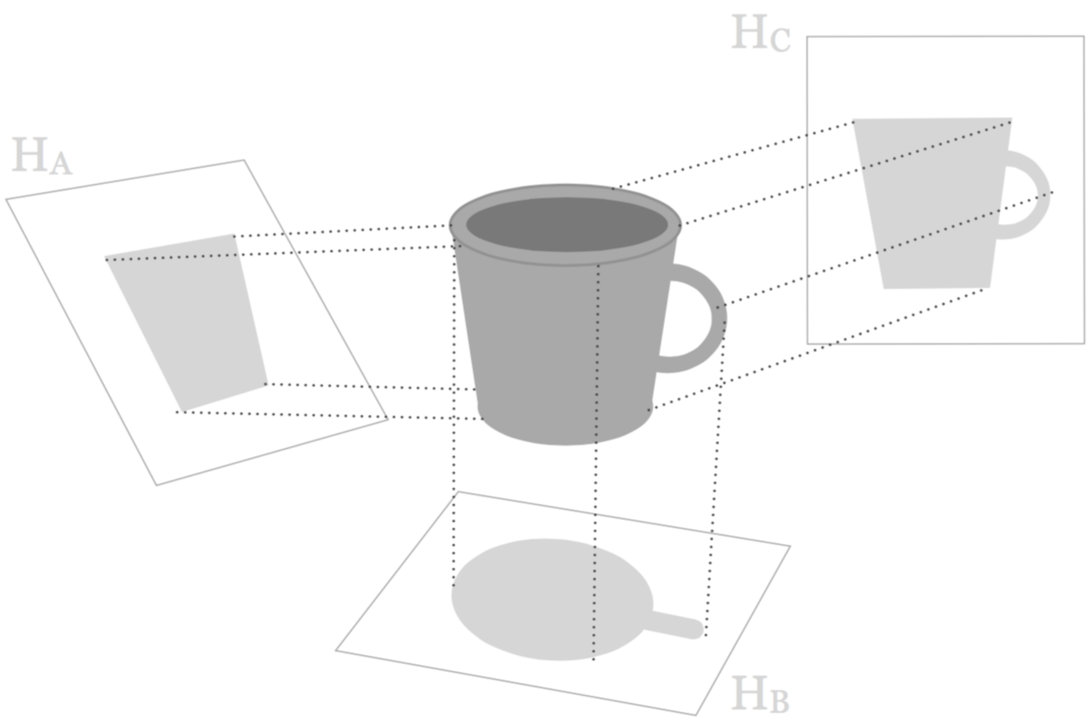
إذا فالمبدأ الأساسي للتحليل بالمركبات الأساسية هو اسقاط سحابة نقط الأفراد والمتغيرات من فضاء متعدد الأبعاد إلى فضاء جزئي ببعدين أو ثلاثة فقط، مع المحافظة قدر المستطاع على المسافة بين الأفراد والمتغيرات، أي بتعبير آخر التقليل أقصى ما يمكن من التشوه الذي قد يطرأ على الساحبتين نتيجة الاسقاط.



$$d(n_i, n_{i'}) = d(h_i, h_{i'})$$

حيث: x_i هي نقطة في الفضاء، و h_i هي اسقاطها العمودي.

لكن السؤال المطروح كيف نحدد هذا المستوى وعلى أي أساس نختاره؟ للاجابة على هذا السؤال ننظر في الصورة الموالية:



بالنظر إلى الصورة يمكن القول أنّ الإسقاط على المستوى H_A أكثر فائدة من الإسقاط على المستوى H_B، فعلى الأقل يمكننا أن نرى أن الشكل في H_A له علاقة بجسم طويل، وأن أحد نهاياته أوسع من الأخرى، وعلى النقيض من ذلك، فإن جميع نقاط السحابة المسقطة على المستوى H_B مشوشة، ولا تنقل فكرة واضحة عن السحابة الأصلية، باستثناء ظل المقبض. وأفضل إسقاط بين الاسقاطات الثلاث هو الاسقاط H_C.

ولكي نحصل على الاسقاط H_C يجب علينا البحث عن المستوى الذي يجعل تشتت نقاط السحابة أكبر ما يمكن، ونكتب:

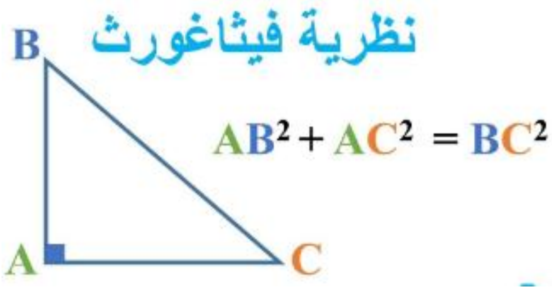
$$\text{Max}_H \sum_i \sum_{i'} d_H^2\,(i,i')$$

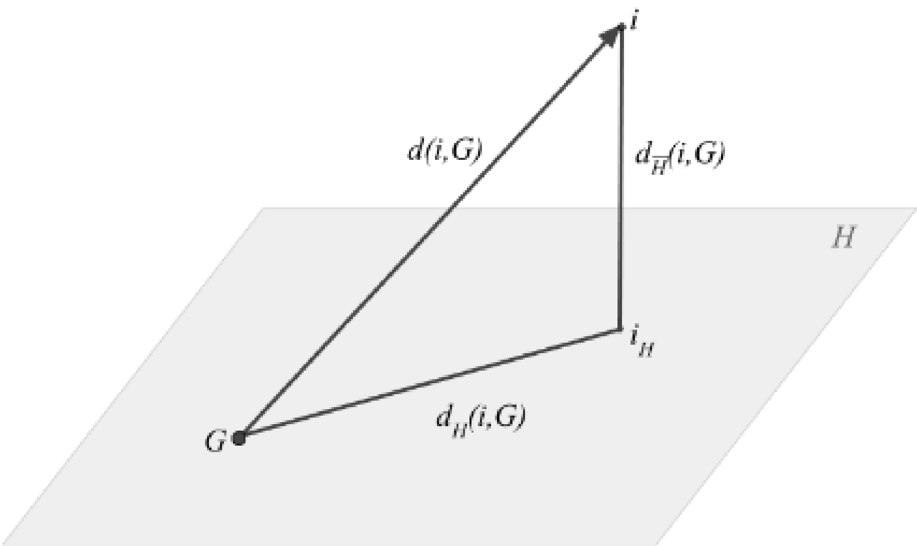
فإذا كان H يمثل فضاء جزئي للإسقاط، يمكن كتابة البحث عن الحد الأقصى على النحو التالي:

$$\text{Max}_H \sum_i \sum_{i'} d_H^2\,(i,i') = \text{Max}_H\{2n \sum_i d_H^2\,(i,G)\}$$

حيث: G هو مركز ثقل سحابة النقط. فتصبح مشكلة الحفاظ على المسافات المسقطة بين جميع أزواج النقاط مشكلة في الحفاظ على المسافات بين كل نقطة ومركز الثقل.

ولحل هذه المشكلة، نستخدم نظرية فيثاغورس (Pythagoras)، التي تنص على أنّ مربّع الوتر في مثلث قائم الزاوية يساوي مجموع مربّع الضلعين.





نلاحظ من الشكل أعلاه أنّ الاسقاط العمودي للنقطة i هو النقطة i_H، وحسب نظرية فيتاغورس، فمربّع المسافة بين النقطة i ومركز ثقل سحابة النقط G تساوي مربّع المسافة بين النقطة i واسقاطها i_H زائد مربّع المسافة بين النقطة i_H ومركز الثقل، ونكتب:

$$\sum_i d^2 (i,G)=\sum_i d_H^2 (i,G)+\sum_i d_{\bar{H}}^2 (i,G)$$

بهذه الطريقة، فإن مستوى الإسقاط الذي يضمن أقصى قدر من التشتت بين النقاط، هو أيضًا المستوى الذي يقترب قدر الإمكان من السحابة الأصلية.

$$\text{Max} \sum_i d_H^2 (i,G) \Leftrightarrow \text{Min} \sum_i d_{\bar{H}}^2 (i,G)$$

وفي بعض الأحيان قد نكون مهتمين بتخصيص أوزان للأفراد بناءً على أهميتهم النسبية أو ملائمتهم. فعندما يكون لجميع الأفراد نفس الأهمية، يمكننا إعطاء وزن يساوي $\frac{1}{n}$ لكل منهم، فيصبح المعيار الملائم على النحو التالي:

$$\text{Max} \frac{1}{n} \sum_i d^2 (i,G)= \text{Max} \frac{\sum_i (x_{iH}-\bar{x}_H)^2}{n}$$

في الحالة العامة حيث يكون لكل فرد وزنه الخاص p_i حيث:

$$\text{Max} \sum_i p_i {d_H}^2 (i,G)$$

حاصل ضرب وزن نقطة i في مربع المسافة التي تفصلها عن مركز الثقل $(d_H^2(i,G))$. يُعرف باسم القصور الذاتي للنقطة. في هذه الحالة، تتضمن المشكلة البحث عن مستوى الإسقاط الذي يزيد من القصور الذاتي المتوقع.

7. تحليل القصور الذاتي (التباين الكلي):

$$I_T= \sum_i p_i d^2 (i,G)$$

يمكن تقسيم القصور الذاتي الكلي إلى جزئين: القصور الذاتي المتوقع على مساحة جزئية H والقصور الذاتي متعامد الإسقاط على مساحة جزئية $\bar{H}$.

$$I_T=I_H + I_{\bar{H}}$$

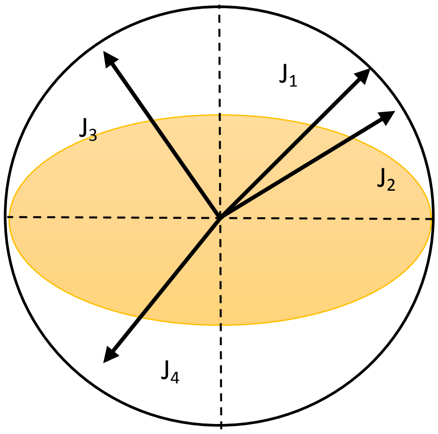
من أجل العثور على الفضاء الجزئي الأمثل، نبدأ بالبحث عن فضاء أحادي البعد يحقق أقصى قدر من القصور الذاتي المتوقع. فإذا كان لجميع الأفراد نفس الوزن، فإن هذا الفضاء الجزئي الأول يتوافق مع اتجاه أقصى امتداد للسحابة. وبعد العثور على أول فضاء فرعي أحادي البعد، فإن الخطوة التالية تتضمن إيجاد فضاء فرعي ثنائي الأبعاد (أي مستوى) بأقصى قدر من القصور الذاتي المتوقع. ثم نبحث عن فضاء ثلاثي الأبعاد، وهكذا دواليك. وفي كل خطوة، نبحث عن مساحة ذات أبعاد أعلى بحيث يكون القصور الذاتي المتوقع أكبر ما يمكن.

$$I_T = I_1 + I_2 + \cdots + I_p$$

8. الارتباط بين المتغيرات: لو افترضنا أن المتغيرات قد تمت مركزتها وترجيحها بواسطة الانحراف المعياري، فإنّ أول ما يجب فعله هو قياس المسافة بين المتغيرات. والعلاقة التي تزودنا بفكرة عن التقارب بين المتغيرات معطاة بالصيغة التالية:

$$d^2(j,j')= (2(1-cor(j,j'))$$

فإذا كان المتغيرين j و j' يقيسان الشيء نفسه (بمعنى وجود ارتباط خطي= 1)، فستكون نقاطهما متطابقة. وفي الحالة التي يكون فيها للمتغيرين ارتباطا خطيا يساوي -1 (عندما يزيد أحدهما، ينقص الآخر)، سيكون لنقاط المتغيرين مسافة قصوى في اتجاهين متعاكسين. وعندما لا يوفر أحد المتغيرات أي معلومات عن الآخر، يكون معامل الارتباط يساوي الصفر، ويتوافق ذلك مع مسافة وسيطة تشكل فيها المتغيرات زاوية متعامدة.



تسمى الدائرة بدائرة الارتباط، نصف قطر هذه الدائرة يساوي الواحد (1)، والذي يعبر عن الارتباط التام. وكل متغير يمثل بمتجه ينطلق من مركز الدائرة نحو محيطها. حيث يشكل كل متغيرين قريبين من بعضهما زاوية صغيرة، ويعني ذلك أنهما يرتبطان بقوة، ويشكل كل متغيرين مستقلين (معامل الارتباط =0) زاوية قائمة. فيما يكون معامل الارتباط لمتغيرين متعاكسين في الإتجاه (متقابلين بالنسبة للدائرة) مساويا لـ -1.

وبشكل مشابه لسحابة نقاط الأفراد، نسعى أيضًا إلى العثور على مستوى إسقاط يوفر أكبر قدر من المعلومات حول الارتباطات بين المتغيرات. أو بتعبير أدق، نبحث عن مستوى يوفر معلومات حول الزوايا بين المتغيرات، أي حول ارتباطاتها.

9. طريقة المركبات الرئيسية البسيطة (non-normalized) والمرجّحة (Normalized):

1.9. طريقة المركبات الرئيسية البسيطة (non-normalized): تستخدم هذه الطريقة في حالة كانت جميع المتغيرات متجانسة، أي لها وحدة القياس نفسها، مثل نقاط طلبة في ثلاث مواد، لدينا ثلاثة متغيرات كلها تقاس بالدرجة. ويتم العمل في هذه الحالة على جدول البيانات الممركز، حيث يتم مركزة أو توسيط البيانات حول متوسطها الحسابي.

ملاحظة: نقوم بحساب المتوسطات الحسابية من أجل مركزة البيانات وتسهيل العمليات الحسابية، لأنّ شكل سحابة النقط لا يتغير عند مركزة البيانات.

ملاحظة: النقاط التي تكون قريبة من مركز ثقل سحابة النقط، تمثل المعلومات التي يتم فقدها عند الاسقاط، لأنها ستكون ممثلة أفضل في المحور الرابع أو الخامس...الخ.

2.9. **طريقة المركبات الرئيسية المرجّحة (Normalized):** تستخدم هذه الطريقة في حالة كانت المتغيرات غير متجانسة وليس لها وحدة القياس نفسها. ويتم العمل على جدول الكميات الممركز والمرجّح. وغالبا ما يكون العمل مع المتغيرات المخفضّة أكثر ملائمة، أي التي يكون انحرافها المعياري 1، من أجل جعلها قابلة للمقارنة مع بعضها البعض، واختزال وحدات القياس لتصبح كل المتغيرات بدون وحدة قياس، بحيث تكون المسافات مستقلة عن وحدات القياس.

ومن أجل فهم السبب وراء اختزال البيانات نأخذ المثال الموالي: في الجدولين أدناه لدينا عينتين تمثلان نفس المجتمع (تفاح).

	العينة 1		العينة 2	
	الوزن (غ)	القطر (ملمتر)	الوزن (غ)	القطر (سنتمتر)
	100	70	100	7
	95	65	95	6.5
التباين	6.25	6.25	6.25	0.00625

في الحالة الأولى عند البحث عن المركبة الأساسية الأولى، فإنّ الوزن والقطر سيأثران بالتساوي في تشكيل المحور (لهما نفس التباين)، بينما في الحالة الثانية الوزن يفوق بكثير القطر.

ثانيا، خطوات التحليل بالمركبات الرئيسة:

1. جدول الأفراد مقابل المتغيرات (مصفوفة البيانات): تستخدم طريقة المركبات الأساسية مع الجداول التي تحتوي على بيانات ذات طبيعة كمية، حيث نشير بالرمز X_{ij} إلى قيمة المتغير X_j التي لوحظت على الفرد e_i. وتتم الإشارة إلى جدول البيانات الأولية الذي سنقوم من خلاله بعملية التحليل بالرمز X ويأخذ الشكل التالي:

	X_1	X_2	X_j	X_p
1	X_{11}	X_{12}	X_{1j}	X_{1p}
2	X_{21}	X_{22}	X_{2j}	X_{2p}
i	X_{i1}	X_{i2}	X_{ij}	X_{ip}
.....				
n	X_{n1}	X_{n2}	X_{nj}	X_{np}

X_j يمثل المتغير j، (j=1,…j=p). X_{ij} تمثل قيمة المتغير X_j عند الفرد i. ونكتب مصفوفة البيانات الأولية X كما يلي:

$$X=\begin{bmatrix} x_{11} & x_{12} & x_{1j} & x_{1p} \\ x_{21} & x_{22} & x_{2j} & x_{2p} \\ x_{i1} & x_{i2} & x_{ij} & x_{ip} \\ .. & .. & .. & .. \\ x_{n1} & x_{n2} & x_{nj} & x_{np} \end{bmatrix} \in M(n$$

2. مصفوفة الأوزان واحداثيات مركز ثقل سحابة النقط:

1.2. مصفوفة الأوزان: في حالة أعطينا لكل فرد في الدراسة نفس الوزن (الأهمية)، نحسب مصفوفة الأوزان، من العلاقة:

$$D_P=\frac{1}{n}I_n$$

2.2. مركز ثقل سحابة النقط:

$$G_j=\sum_{i=1}^{n} p_i X_i=\frac{1}{n}\sum_{i=1}^{n} X_i$$

$$G_X=\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ ..... \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{bmatrix}=\begin{bmatrix} \overline{X_1} \\ \overline{X_2} \\ \overline{...} \\ \overline{X_p} \end{bmatrix}$$

كما يمكن حسابه من الصيغة المصفوفية:

$$G_X=X^T D_P I_n$$

حيث: X^T منقول مصفوفة البيانات، D_P مصفوفة الأوزان، I_n متجه الوحدة.

3. المصفوفة المركزة: : نحسب المصفوفة المركزة X_C من الصيغة المصفوفية:

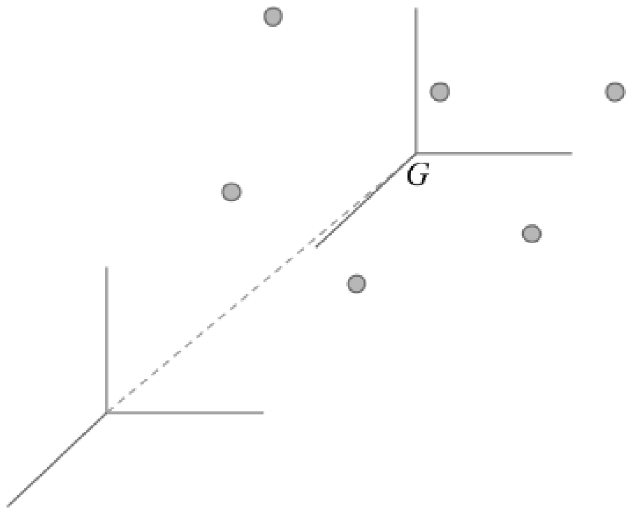
$$X_C = X - 1_n G_X^T$$

حيث: X مصفوفة البيانات الأولية، 1_n متجه الوحدة، G_X^T منقول مركز ثقل سحابة النقط.

$$1_n G_X^T = \begin{bmatrix} 1 \\ .. \\ 1 \\ .. \\ 1 \end{bmatrix} * \begin{bmatrix} g_1 & g_2 & \cdots & g_p \end{bmatrix} = \begin{bmatrix} g_1 & \cdots & g_p \\ \vdots & \ddots & \vdots \\ g_1 & \cdots & g_p \end{bmatrix} \in M(n, p)$$

وتكتب المصفوفة المركزة كما يلي:

$$X_C = \begin{bmatrix} x_{11} - g_1 & \cdots & x_{1p} - g_p \\ \cdots & \cdots & \cdots\cdot \\ x_{n1} - g_1 & \cdots & x_{np} - g_p \end{bmatrix} \in M(n, p)$$



ويمكننا كذلك حساب الجدول المركز مباشرة بطرح الوسط الحسابي لكل متغير من قيم العمود المقابل له.

$$\widetilde{\boldsymbol{X}}_{\boldsymbol{ij}} = \boldsymbol{X}_{\boldsymbol{ij}} - \bar{X}_j$$

ملاحظة: عند مركزة البيانات يصبح مركز ثقل سحابة النقط الجديد يساوي 0.

$G_X = (0,0,0)$.

4. المصفوفة المعيارية: لاختزال البيانات نقسم احداثيات كل عمود في المصفوفة المركزة على الانحراف المعياري المقابل له. وعليه، نحسب أوّلا الانحراف المعياري.

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{ij} - g_j)^2} \ldots\ldots\ldots j=1,\ldots\ldots,p$$

ولدينا المصفوفة القطرية لمقلوب الإنحرافات المعيارية:

$$D_{\frac{1}{s}}=\begin{bmatrix} 1/s_1 & 0 & \dots. & 0 \\ 0 & \dots. & 0 & \dots. \\ \dots.. & 0 & 1/s_{p-1} & 0 \\ 0 & \dots. & 0 & 1/s_p \end{bmatrix}$$

ونحسب المصفوفة المعيارية من العلاقة التالية:

$$X_{CR}=X_C * D_{\frac{1}{s}}$$

وتكون المصفوفة الممركزة والمعيارية أو المرجحة على النحو التالي:

$$X_{CR}=\begin{bmatrix} \frac{x_{11}-g_1}{s_1} & \dots & \frac{x_{1p}-g_p}{s_p} \\ \dots & \dots & \dots. \\ \frac{x_{n1}-g_1}{s_1} & \dots & \frac{x_{np}-g_p}{s_p} \end{bmatrix} \in M(n, p)$$

ويمكن كذلك حساب الجدول الممركز والمرجّح مباشرة بقسمة احداثيات كل عمود في الجدول الممركز على الانحراف المعياري المقابل له. وفقا للعلاقة:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$$

5. حساب مصفوفة التباين والتباين المشترك V: في حالة كانت المتغيرات متجانسة نستخدم هذه المصفوفة، ويتم حساب التباين والتباين المشترك من الصيغة المصفوفية:

$$V=X_C^T D_P X_C$$

حيث: X_C المصفوفة الممركزة، X_C^T منقول المصفوفة X_C. D_P المصفوفة القطرية لمصفوفة الأوزان.

ملاحظة: عندما يكون لجميع الأفراد نفس الوزن:

$$D_P = \frac{1}{n}$$

أو بحساب التباين من العلاقة الموالية:

$$V(X,X) = \frac{1}{n}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)^2$$

والتباين المشترك من العلاقة التالية:

$$Cov(X,Y) = \frac{1}{n}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j) * (y_{ij} - \bar{y}_j)$$

ملاحظة: مصفوفة التباين والتباين المشترك، مصفوفة متماثلة (تساوي منقولها):

$$V^T=\frac{1}{n}X_C X_C^T=V$$

ملاحظة: عند استخدام مصفوفة التباين والتباين المشترك، يحسب القصور الذاتي الكلي من العلاقة:

$$I_T = \text{Trace }(V) = \sum_{j=1}^{p} Var(j)$$

6. **مصفوفة الارتباطات:** في حالة كانت المتغيرات غير متجانسة نستخدم مصفوفة الارتباطات، ويتم حساب الارتباط بين المتغيرات باستخدام الصيغة المصفوفية الآتية:

$$R = X_{CR}^{T} D_P X_{CR}$$

ملاحظة: عندما يكون لجميع الأفراد نفس الوزن:

$$D_P = \frac{1}{n}$$

أو من الصيغة المصفوفية التالية:

$$R = D_{\frac{1}{S}} V D_{\frac{1}{S}}$$

أو باستخدام معامل الارتباط المعطى بالعلاقة:

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

$$r_{XY} = \frac{Cov\ (X,Y)}{\delta_X \delta_Y}$$

ملاحظة: مصفوفة الارتباط مصفوفة متماثلة (تساوي منقولها):

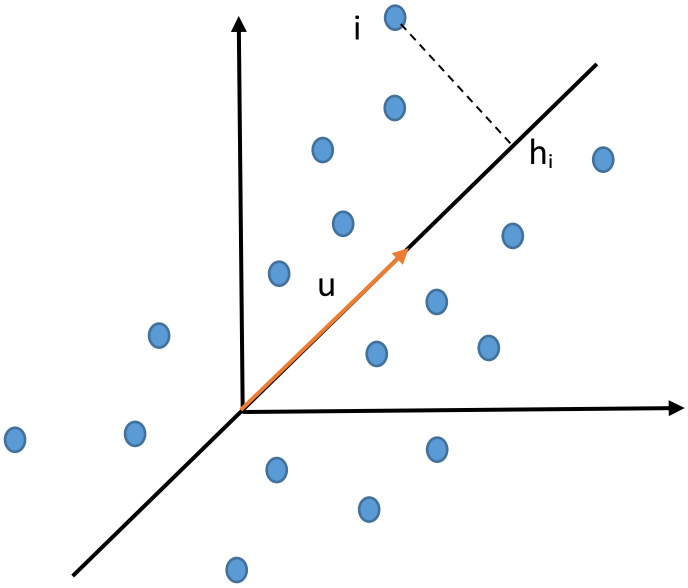
$$R^T = \frac{1}{n} X_{CR} X_{CR}^{T} = \frac{1}{n} X_{CR}^{T} X_{CR} = R$$

ملاحظة: عند استخدام مصفوفة الارتباطات، يحسب القصور الذاتي الكلي من العلاقة:

$$I_T = \text{Trace }(R) = P$$

P : عدد المتغيرات.

7. **القيم والمتجهات الذاتية:** تتمثل الخطوة الأولى من طريقة التحليل بالمركبات الرئيسية في البحث عن أفضل محور من الفضاء الجزئي، المعرّف بواسطة متجه وحدوي u، تكون النقاط المسقطة عليه تتمتع بأقصى قدر من التباين. بمعنى آخر، سيتم ايجاد المتجه u بحيث تكون المسافات بين أزواج النقاط المسقطة قريبة قدر الإمكان من المسافات الأصلية.



اسقاط (احداثيات) نقطة الصف (الفرد) i على الخط الموجّه بواسطة الشعاع u تعطى بالعلاقة:

$$h_i = \sum_{j=1}^{p} (x_{ij} - \bar{x}_j)\, u_j$$

القصور الذاتي لجميع النقاط المسقطة على u تعطى بالعلاقة:

$$\sum_{i=1}^{n} p_i h_i^2 = \lambda$$

إذا فالهدف هو البحث عن المتجه u الذي يعظّم القيمة الذاتية λ.

تبين الأشعة الذاتية اتجاهات القصور الذاتي الأقصى، ونطلق عليها اسم المحاور العاملية. في هذه الاتجاهات نقوم باسقاط الأفراد ونسميها المركبات الأساسية (يتم الحصول على كل مركبة أساسية كمجموعة خطية من المتغيرات الأصلية).

$$F_\alpha = u_1 x_1 + \ldots\ldots + \ldots\ldots + u_p x_p$$

وتحتوي كل مركبة على تباين يساوي القيمة الذاتية المرتبطة بها.

$$V(F_\alpha) = \lambda\alpha$$

باختصار، يمكن النظر إلى تحليل المركبات الأساسية باعتباره أسلوبًا ننتقل فيه من المتغيرات الأصلية، ولكل منها أهمية معينة من خلال تباينها، إلى متغيرات جديدة، هذه المتغيرات الجديدة هي مزيج خطي من المتغيرات الأصلية، ولها أهمية تُعطى من خلال تباينها الذي يتضح أنه قيمها الذاتية.

1.7. **القيم الذاتية:** لحساب القيم الذاتية للمصفوفة V (مصفوفة التباين والتباين المشترك) نعتمد على العلاقة:

$$VU = \lambda U$$

حيث $\vec{\ }$ هو الشعاع الذاتي للمصفوفة V.

ويمكن كتابة العلاقة السابقة كما يلي:

$$(V-\lambda I)U=0$$

وحتى تقبل المعادلة حلا غير صفري، لا بد أن يكون:

$$|V-\lambda I|=0$$

وهي العلاقة التي تسمح بحساب القيم الذاتية للمصفوفة V.

ولحساب القيم الذاتية للمصفوفة R (مصفوفة الارتباط) نعتمد على العلاقة:

$$RU=\lambda U$$

حيث $\vec{\ }$ هو الشعاع الذاتي للمصفوفة R.

ويمكن كتابة العلاقة كما يلي:

$$(R-\lambda I)U=0$$

وحتى تقبل المعادلة حلا غير صفري، لا بد أن يكون:

$$|R-\lambda I|=0$$

وهي العلاقة التي تسمح بحساب القيم الذاتية للمصفوفة R.

2.7. **المتجهات الذاتية:** كل قيمة ذاتية يقابلها شعاع ذاتي، ولتحديد الأشعة الذاتية نعتمد على ما يلي:

– الشعاع الذاتي الأول U_1 الذي يرافق القيمة الذاتية الأولى، يحدد من العلاقة:

$$(V-\lambda_1 I)U_1=0$$

أو

$$(R-\lambda_1 I)U_1=0$$

حيث يكون:

$$U_1^T * U_1=1$$

– الشعاع الذاتي الثاني U_2 الذي يرافق القيمة الذاتية الثانية، يحدد من العلاقة:

$$(V-\lambda_2 I)U_2=0$$

أو

$$(R-\lambda_2 I)U_2=0$$

حيث يكون:

$$U_2^T * U_2=1$$

$$U_1^T * U_2=0$$

بمعنى أنّ الأشعة الذاتية تكون متعامدة مثنى مثنى.

8. **نسبة القصور الذاتي (التباين) المفسّر بواسطة المركبات الأساسية:** من الممكن تحديد القصور الذاتي الكلي المفسر بواسطة كل مركبة أساسية من القيم الذاتية (كل قيمة ذاتية تعتبر كمركبة أساسية)، وكذلك كمية المعلومات التي يحفظها كل محور عاملي. فبعد حساب القيم الذاتية لمصفوفة الارتباط نقوم بترتيبها تنازليا من الأكبر إلى الأصغر، ثم نشكل جدول القيم الذاتية كما يلي:

الرقم	القيمة الذاتية	النسبة المئوية	النسبة المئوية الصاعدة
1	λ_1	$\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}$
2	λ_2	$\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i}$
3	λ_3	$\frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_1+\lambda_2+\lambda_3}{\sum_{i=1}^{p}\lambda_i}$
.....	....	....	....
i	λ_i	$\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_1+\lambda_2+\cdots\lambda_i}{\sum_{i=1}^{p}\lambda_i}$
....	.....	....	....
p	λ_p	$\frac{\lambda_p}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{p}\lambda_i}=1$
	$\sum_{i=1}^{p}\lambda_i=p$	1	

– كل قيمة ذاتية λ_i تمثل المساهمة المطلقة للمحور α في التباين الكلي (Inertie). فاعتمادا على نظرية heygens يمكننا تفكيك التباين الكلي للأفراد كما يلي:

$$I_0= I_{\Delta *_1}+I_{\Delta *_2}+\ldots\ldots+I_{\Delta *_p}=\lambda_1+\lambda_2+\cdots+\lambda_p$$

– تقاس المساهمة المطلقة للمحور في التباين الكلي بالعلاقة:

$$C_a(\Delta_k/I_0)=\lambda_k$$

معنى ذلك أنّ كل محور تقاس مساهمته المطلقة بالقيمة الذاتية المقابلة له.

– تقاس المساهمة النسبية للمحور α من العلاقة:

$$C_r(\Delta_k/I_0)=\frac{\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_p}=\frac{\lambda_k}{Trac(V)}$$

– تمثل القيمة $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\ _i}$ نسبة القصور الذاتي الكلي المفسر بالمستوى العاملي الأول (المحور الأول والثاني)، وتقاس نسبة تمثيل المستوى من العلاقة:

$$C_r(\Delta_1\oplus\Delta_2/I_0)=\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\cdots+\lambda_p}=\frac{\lambda_1+\lambda_2}{Trac(V)}$$

وكقاعدة عامة تكون المركبات الرئيسية المتحصل عليها جيدة وكافية لإعطاء صورة واضحة عن سحابة النقاط في المخطط العاملي إذا كانت نسبة القصور الذاتي الكلي المفسر بالمركبات الرئيسية أكبر من 75 في المائة.

ملاحظة: القيمة الذاتية الأولى تعطينا الشعاع الذاتي الأول، والذي يعطينا بدوره المحور العاملي الأول (المستقيم: $\Delta 1$)، وهذا الأخير يعطينا المركبة الرئيسية الأولى (F_1) وبنفس الطريقة القيمة الذاتية الثانية تعطينا الشعاع الذاتي الثاني، والذي يعطينا بدوره المحور العاملي الثاني، وهذا الأخير يعطينا المركبة الرئيسية الثانية. ثم المحورين 1 و 2 يشكلان لنا المستوي العاملي الأول، حيث يكون الشعاع الذاتي الأول متعامد مع الشعاع الذاتي الثاني.

9. تحديد عدد المحاور التي تأخذ في التحليل: توجد عدة معايير تسمح بتحديد عدد المحاور التي تؤخذ في التحليل، من بينها:

أ. محك كايزر (Kaiser): حسب هذا المعيار نأخذ فقط العوامل التي تزيد قيمها الذاتية عن 1.

ب. معيار التمثيل البياني للقيم الذاتية: بعد إنشاء التمثيل البياني بالأعمدة للقيم الذاتية نقوم برسم مستقيم يربط بين أكبر عدد من القيم الذاتية، والقيم الكبيرة التي لا تنتمي للمستقيم تمثل المحاور التي تؤخذ في التحليل .أو نربط بين رؤوس الأعمدة بخطوط مستقيمة، وأين تشكل لنا شكل مرفق (Code) نتوقف عند تلك القيمة الذاتية (المرفق يؤخذ).

وعند استخدام برنامج SPSS يتم استخدام محك كاتل، حيث يتم القيام بفحص الرسم Scree plot واختيار العوامل التي تقع قبل تحول المنحنى إلى مسار أفقي.

ج. معيار نسبة المستوي الأول: إذا كانت قيمة النسبة $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i}$ أكبر من 80%، فذلك دليل على أنّ فقدان المعلومات قليل، ولا داعي لإنشاء المستوى العاملي الثاني.

د. معيار نسبة المحور الثالث: إذا كانت قيمة النسبة $\frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i}$ أكبر من 15% فلا بد من إنشاء المستوى العاملي الثاني.

9. التحليل في فضاء المتغيرات وفي فضاء الأفراد: من وجهة نظر مصفوفية، يتكون PCA من دراسة مصفوفة البيانات X_{CR} المحددة بمصفوفة مترية I_p في الفضاء R^p والمصفوفة المترية N في الفضاء R^n تأتي المصفوفة X_{CR} محددة بالطريقة التالية:

– في الفضاء R^p:

$$X_{CR}^{T} N X_{CR} U = \lambda U \text{..........} U^T U = 1$$

– في الفضاء R^n:

$N^{\frac{1}{2}}X_{CR}X_{CR}^{T}NV=\lambda V.........V^{T}V=1$

وتكتب علاقة العبور كالتالي:

$$U=\frac{1}{\sqrt{\lambda}}X_{CR}^{T}P^{\frac{1}{2}}V$$

$$V=\frac{1}{\sqrt{\lambda}}P^{\frac{1}{2}}X_{CR}U$$

المصفوفة المتماثلة التي نريد تقطيرها $(X_{CR}^{T}NX_{CR})$، تتزامن هذه المصفوفة مع مصفوفة الارتباطات في حالة استخدام الطريقة المرجحة؛ أو مع مصفوفة التباين المشترك في حالة الطريقة البسيطة.

1.9. التحليل في فضاء المتغيرات (الفضاء الشعاعي R^p): تحليل الأفراد في فضاء المتغيرات.

1.1.9. المركبات الأساسية للأفراد على المحاور: المركبات الرئيسية هي المتغيرات التي تحددها العوامل الرئيسية، وتحتوي على إحداثيات الأفراد، التي يتم حسابها عن طريق الإسقاط المتعامد لصفوف المصفوفة المركزة أو المعيارية على الاتجاهات المحددة بواسطة المتجهات الذاتية U_α. ويتم حساب المركبات الرئيسية من خلال تقطير المصفوفة V أو R.

.1.1.1.9 في حالة استخدام مصفوفة التباين والتباين المشترك: إذا كان U_α هو المتجه الذاتي للعدد n من الأفراد، يتم الحصول على إحداثيات إسقاطات n من النقاط ببساطة عن طريق:

$$F_\alpha=\begin{bmatrix}F_{1\propto}\\F_{2\propto}\\ ..\\F_{n\propto}\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^{p}(x_{1j}-\bar{X}_j)U_{j\propto}\\\sum_{j=1}^{p}(x_{2j}-\bar{X}_j)U_{j\propto}\\ ..\\\sum_{j=1}^{p}(x_{nj}-\bar{X}_j)U_{j\propto}\end{bmatrix}........................\boldsymbol{\widetilde{X}_{ij}}=\boldsymbol{X_{ij}}-\bar{X}_j$$

ومنه:

$$F_\alpha=\sum_{j=1}^{p}\boldsymbol{\widetilde{X}_{ij}}\,U_{j\propto}$$

وتكتب بالصيغة المصفوفية الآتية:

$$F_\alpha=X_CU_\alpha$$

حيث: F هي المركبة الرئيسية، X_C المصفوفة المركزة، U_α الشعاع الذاتي المرافق للقيمة الذاتية λ_α.

ملاحظات:

متوسطات المركبات الرئيسية معدومة (لأنّ المركبات ممركزة).

$$\overline{F_\alpha}=0$$

كما أنّ تباين كل مركبة أساسية يساوي القيمة الذاتية المرتبطة بها.

$$V(F_\alpha)=\lambda_\alpha$$

.2.1.1.9 حالة استخدام مصفوفة الارتباطات: يتم حساب إحداثيات الأفراد على المحور α كما يلي:

$$F_\alpha=\begin{bmatrix}F_{1\propto}\\F_{2\propto}\\..\\F_{n\propto}\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^{p}(\frac{x_{1j}-\bar{X}_j}{S_j})U_{j\propto}\\\sum_{j=1}^{p}(\frac{x_{2j}-\bar{X}_j}{S_j})U_{j\propto}\\..\\\sum_{j=1}^{p}(\frac{x_{nj}-\bar{X}_j}{S_j})U_{j\propto}\end{bmatrix}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\boldsymbol{Z}=\frac{X_{ij}-\bar{X}_j}{S_j}$$

ومنه:

$$F_\alpha=\sum_{j=1}^{p}\frac{X_{ij}-\bar{X}_j}{S_j}u_{j\propto}$$

وتكتب بالصيغة المصفوفية التالية:

$$F_\alpha= X_{CR}U_\alpha$$

حيث: F هي المركبة الرئيسية، X_{CR} المصفوفة الممركزة، U_α الشعاع الذاتي المرافق للقيمة الذاتية. وتكون متوسطات المركبات الرئيسية معدومة (لأنّ المركبات ممركزة).

$$\overline{F_\alpha}=0$$

كما أنّ تباين كل مركبة أساسية يساوي القيمة الذاتية.

$$V(F_\alpha) = \lambda_\alpha$$

2.1.9. **جودة تمثيل الأفراد على المحاور**: في الواقع العملي قد يكون عدد المتغيرات كبير جدا، مما يجعل من الصعب ملاحظة موقع كل نقطة على المحاور العاملية، لذلك فنحن بحاجة إلى حساب جودة تمثيل كل نقطة على المحاور، حتى يسهل تفسير المخطط. ويكون تمثيل الفرد جيدا على المحور العاملي، إذا كانت قيمة تجب الزاوية التي يشكلها الفرد مع المستوى مساوية للواحد.

$$Cos^2(\theta)=1$$

ويتم قياس جودة تمثيل الأفراد على المحاور، باستخدام تجب الزاوية التي يشكلها مع المستوى، وفقا للعلاقة:

$$Cos^2\theta_{i\alpha}=\frac{F_\alpha^2(i)}{\sum_{\alpha=1}^{p}F_\alpha^2(i)}$$

كلما اقتربت هذه القيمة من الواحد، كان الفرد ممثلا أحسن تمثيل.

3.1.9. **نسبة مساهمة الأفراد في تشكيل المحاور**: نحسب نسبة مساهمة الأفراد في تشكيل المحاور، لكي نتمكن من تفسير المركبات الأساسية. وهي كمية التباين لكل فرد التي يتم تفسيرها بواسطة المحور العاملي. ويحسب بتربيع المركبات الأساسية للأفراد ثم القسمة على ثقل المصفوفة (عدد الأفراد) مضروبة في القيمة الذاتية المقابلة لها:

$$\boldsymbol{C_i^\alpha}=\frac{F_\alpha^2(i)}{\boldsymbol{n\lambda_\alpha}}$$

2.9. التحليل في فضاء الأفراد (الفضاء الشعاعي R^n) (تحليل المتغيرات في فضاء الأفراد): على نفس الخطوات التي استخدمناها في تمثيل الأفراد في فضاء المتغيرات، سنقوم بتمثيل المتغيرات في فضاء الأفراد، غير أننا سنقوم بتغيير الأساس في هذا الفضاء.

1.2.9. احداثيات المتغيرات النشطة: يتم الحصول على إحداثيات المتغيرات النشطة عن طريق الإسقاط المتعامد لأعمدة المصفوفة الممركزة أو المعيارية على الاتجاهات المحددة بواسطة المتجهات الذاتية V_α، والموزونة.

$$G_\alpha = X_{CR}^T P^{\frac{1}{2}} V_\alpha = \frac{1}{\sqrt{\lambda}} X_{CR}^T P X_{CR} U_\alpha = D_{\frac{1}{\sqrt{\lambda}}} R U = \sqrt{\lambda_\alpha} U_\alpha$$

$$G_\alpha = \sqrt{\lambda_\alpha} U_\alpha$$

2.2.9. الارتباط الخطي بين المتغيرات الأصلية والمركبات الأساسية: من أجل دراسة الارتباط بين المكونات الأساسية المتحصل عليها والمتغيرات الأولية، نحن مهتمون بدراسة العلاقات المتبادلة بين المتغيرات الجديدة والأصلية. أولاً نحدد التباين المشترك بين المتغير الأولي X_j والمكون الرئيسي F_α.

$$Cov(X_j, F_\alpha) = \frac{1}{n} F_\alpha^T X_j = \frac{1}{n} (XU_\alpha)^T X_j = \frac{1}{n} U_\alpha^T X^T X_j$$

$$= \frac{1}{n} U_\alpha^T X^T X_j \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = U_\alpha^T V \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \lambda_\alpha U_\alpha^T \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \lambda_\alpha U_{\alpha j}$$

نستنتج الارتباط بين X_j و F_α:

$$Cor(X_j, F_\alpha) = \frac{Cov(X_j, F_\alpha)}{\sqrt{Var(X_j) Var(F_\alpha)}} = \frac{\lambda_\alpha U_{\alpha j}}{\sqrt{S_j^2 \lambda_\alpha}} = \sqrt{\lambda_\alpha} \frac{U_{\alpha j}}{S_j}$$

$$Cor(X_j, F_\alpha) = \sqrt{\lambda_\alpha} \frac{U_{\alpha j}}{S_j}$$

في حالة استخدام طريقة المركبات الرئيسية البسيطة تكتب العلاقة السابقة بالصيغة المصفوفية الآتية:

$$Cor = D_{\sqrt{\lambda}} * U^T * D_{\frac{1}{S}}$$

حيث: $D_{\sqrt{\lambda_\alpha}}$: المصفوفة القطرية لجذور القيم الذاتية λ. U^T: منقول مصفوفة المتجهات الذاتية، $D_{\frac{1}{S}}$: المصفوفة القطرية لمقلوبات الانحرافات المعيارية.

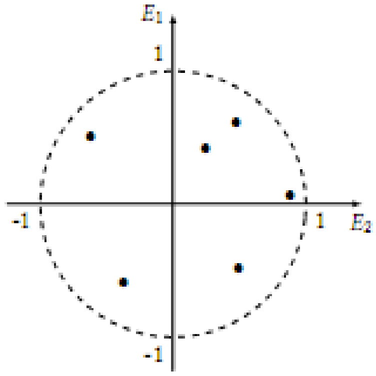
وفي حالة استخدام طريقة المركبات الرئيسية المرجحة تكتب العلاقة السابقة على الشكل:

$$Cor(X_j, F_\alpha) = \sqrt{\lambda_\alpha} U\alpha j$$

وتكتب مصفوفيا على النحو التالي:

$$Cor = U D_{\sqrt{\lambda}}$$

وتعطي هذه الأخيرة جودة تمثيل المتغير j على المحور E_α، وكلما اقترنا من 1 في القيمة المطلقة، كان تمثيل المتغير أفضل. وتسمح لنا علامة الارتباط بمعرفة ما إذا كان المتغير يساهم بشكل إيجابي أو سلبي في تشكيل المحور. وغالبًا ما نمثل هذه البيانات بيانياً بترابطات ثنائية الأبعاد باستخدام ما يسمى بدائرة الارتباط. بتعبير أدق، يتم اختيار محورين E_j و E_k؛ ونرسم النقاط التي تمثل إحداثياتها ارتباطات كل من المتغيرات مع المكونين الأساسيين F_1 و F_2 .



على المستوى العاملي الأول، يتوافق الإحداثي السيني مع الارتباط بين كل متغير أولي والمكون الرئيسي F_1؛ وينسق العلاقة مع المركبة الثانية F_2. وكلما كان المتغير أقرب إلى دائرة الارتباط، كان تمثيله أفضل بالمستوى قيد الدراسة. وإذا كان هناك متغيرين قريبين من دائرة الارتباط (وبالتالي يتم تمثيلهما جيدًا في المستوى المدروس) وقريبان من بعضهما البعض، فعندئذٍ يكونان مرتبطان ارتباطًا إيجابيًا قويًا. وعلى العكس من ذلك، فإذا كان متغيرين قريبين من دائرة الارتباط ولكنهما متعاكسان بشكل متماثل فيما يتعلق بالأصل سيكونان مرتبطين سلبًا بقوة.

3.2.9. المركبات الرئيسية للمتغيرات على المحاور:

$$G_\alpha = X_{CR}^T V_\alpha$$

$$V = \frac{1}{\sqrt{\lambda}} D_P^{\frac{1}{2}} X_{CR} U$$

4.3.9. جودة تمثيل المتغيرات على المحاور:

$$Cos^2(j,\propto) = \frac{G_{j\propto}^2}{d^2(j,0)}$$

تتطابق مسافة المتغير إلى الأصل مع الانحراف المعياري للمتغير تحت PCA البسيط، وفي المقابل، عند إجراء PCA مرجّح، تكون المسافة مساوية لـ 1.

$$Cos^2(j,\alpha) = Cor^2(j,\alpha)$$

5.3.9. نسبة مساهمة المتغيرات في تشكيل المحاور:

القصور الذاتي المتوقع على المحور في الفضاء R^n:

$$\lambda_{\propto} = \sum_{j}^{p} G_{j\propto}^{2}$$

مساهمة المتغيرات في القصور الذاتي لهذا المحور:

$$Ctr_{(j,\propto)} = \frac{G_{j\propto}^{2}}{\lambda_{\propto}}$$

مع مراعاة صيغة حساب إحداثيات المتغيرات:

$$Ctr_{(j,\propto)} = u_{j\propto}^{2} * 100$$

10. تمثيل الأفراد والمتغيرات على المستوى: في هذا المخطط يتم تمثيل احداثيات كل من الأفراد والمتغيرات على نفس المخطط. ولتفسير هذا المخطط، نتبع القواعد التالية:

- تقارب نقطتي فردين تعني التشابه.
- تقارب نقطتي متغيرين تعني الترابط.
- تقارب نقطة فرد مع نقطة متغير تعني أنّ المتغير يأثر في سلوك ذلك الفرد.

11. إضافة أفراد أو متغيرات:

1.11. إضافة أفراد:

$$X = \frac{\begin{bmatrix} X_{11} & X_{1j} & X_{1p} \\ X_{i1} & X_{ij} & X \\ X_{n1} & X_{nj} & X_{np} \end{bmatrix}}{X_{ij}^{+}} \Longrightarrow X_C = \begin{bmatrix} \dfrac{X_{ij} = \dfrac{x_{ij} - \overline{X_J}}{s_j}}{X_C^{+} \dashrightarrow X_{ij}^{+} = \dfrac{X_{ij}^{+} - \overline{X_J}}{s_j}} \end{bmatrix}$$

المركبات الأساسية للفرد المضاف في حالة استخدام طريقة المركبات الأساسية البسيطة:

$$F_{\alpha}^{+} = X_C^{+} U_{\alpha}$$

وفي حالة استخدام طريقة المركبات الأساسية المرجحة:

$$F_{\alpha}^{+} = X_{CR}^{+} U_{\alpha}$$

2.11. المتغيرات التكميلية:

$$R = \begin{bmatrix} r_{11} & r_{1j} & r_{1p} \\ r_{i1} & r_{ij} & r_{ip} \\ r_{n1} & r_{nj} & r_{np} \end{bmatrix} (r_{ij}^{+}) \Longrightarrow X = \left[X \to X_{ij} = \frac{r_{ij} - \bar{r_J}}{s_j} \right] \left[X^{+} \to X_{ij}^{+} = \frac{r_{ij}^{+} - \bar{r_J}^{+}}{s_j^{+}} \right]$$

احداثيات المتغيرات التكميلية: يتم تحديد المتغيرات التكميلية باستخدام القاعدة السابقة حول حساب الإحداثيات، دع X_{CR}^{+T} مصفوفة البيانات التي تحتوي على المتغيرات التكميلية، مع الأخذ في الاعتبار العلاقات الانتقالية لدينا ما يلي:

في حالة استخدام طريقة المركبات البسيطة:

$$G_{\alpha}^{+}=\frac{1}{n}X_{C}^{+T}\left(\frac{X_{C}U_{\propto}}{\sqrt{\lambda_{\propto}}}\right)$$

وفي حالة استخدام طريقة المركبات الأساسية المرجحة:

$$G_{\alpha}^{+}=\frac{1}{n}X_{CR}^{+T}\left(\frac{X_{CR}U_{\propto}}{\sqrt{\lambda_{\propto}}}\right)$$

يتم حساب إسقاط المتغيرات التكميلية من هذه العلاقة بين إحداثيات المتغير وإسقاط الأفراد. في PCA المرجّح، يكون هذا الإسقاط مساويًا للعلاقة بين المتغيرات والمكون الرئيسي.

تمارين محلولة

التمرين الأول: لديك المصفوفة التالية:

$$X=\begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

المطلوب:

1. أحسب العلاقة المصفوفية X^TX، هل المصفوفة الناتجة مربعة، وهل هي متناظرة.
2. أحسب القيم الذاتية والمتجهات الذاتية المرافقة لها.
3. أثبت أنّ:

$$\text{trac}(X^TX)=\text{trac}(D)=\sum_i \lambda_i$$

الحل:

.1

$$X^T=\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}$$

$$X^TX= =\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

المصفوفة الناتجة مصفوفة مربعة $(3*3)$، وهي متناظرة.

2. حساب القيم والمتجهات الذاتية:

$$\det(X^TX-\lambda I)=0 \rightarrow \begin{vmatrix} 1-\lambda & 0 & -1 \\ 0 & 1-\lambda & -1 \\ -1 & -1 & 2-\lambda \end{vmatrix}=0$$

$$(1-\lambda)((1-\lambda)(2-\lambda)-(-1)(-1))+(-1)((-1)(0)-(1-\lambda)(-1))=$$

$$(1-\lambda)((1-\lambda)(2-\lambda)-1)-(1-\lambda)= (1-\lambda)(1-\lambda)(2-\lambda)-(1-\lambda)-(1-\lambda)=$$

$$(1-\lambda)[(1-\lambda)(2-\lambda)-1-1]=(1-\lambda)(2-\lambda-2\lambda+\lambda^2-1-1)=(1-\lambda)(\lambda^2-3\lambda)=$$

$$\lambda(1-\lambda)(\lambda-3)=0 \rightarrow \lambda_1=3,\ \lambda_2=1,\ \lambda_3=0$$

$$D=\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$\lambda_1=3$

$X^TXu = \lambda u$

$$(X^TX-\lambda)\, u= \rightarrow \begin{bmatrix} -2 & 0 & -1 \\ 0 & -2 & -1 \\ -1 & -1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

إذاً لدينا نظام متجانس من المعادلات الخطية، نقوم بحلها بواسطة حذف كاوس:

$$\begin{bmatrix}-2 & 0 & -1\\ 0 & -2 & -1\\ -1 & -1 & -1\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix} \boxed{\begin{matrix}*\text{-}0.5\\ +\end{matrix}} \begin{bmatrix}1 & 0 & \frac{1}{2}\\ 0 & -2 & -1\\ 0 & -1 & -\frac{1}{2}\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix} \boxed{\begin{matrix}*\text{-}0.5\\ +\end{matrix}} \begin{bmatrix}1 & 0 & \frac{1}{2}\\ 0 & 1 & \frac{1}{2}\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix}$$

$X=-\frac{1}{2}z$

$Y=-\frac{1}{2}z$

$Z=Z$

$$U=\begin{pmatrix}-\frac{1}{2}z\\ -\frac{1}{2}z\\ z\end{pmatrix} \rightarrow U=z\begin{pmatrix}-\frac{1}{2}\\ -\frac{1}{2}\\ 1\end{pmatrix} \ldots\ldots z=1 \rightarrow U_1=\begin{pmatrix}-\frac{1}{2}\\ -\frac{1}{2}\\ 1\end{pmatrix}$$

$$\|U1\|=\sqrt{(-\tfrac{1}{2})^2+(-\tfrac{1}{2})^2+(1)^2}=\sqrt{\frac{3}{2}}$$

$$v_1=\begin{pmatrix}-\frac{\sqrt{6}}{6}\\ -\frac{\sqrt{6}}{6}\\ \frac{\sqrt{6}}{3}\end{pmatrix}$$

وبالطريقة نفسها نحسب بقية المتجهات الذاتية.

$$v_2=\begin{pmatrix}-\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\\ 0\end{pmatrix},\ v_3=\begin{pmatrix}\frac{\sqrt{3}}{3}\\ \frac{\sqrt{3}}{3}\\ \frac{\sqrt{3}}{3}\end{pmatrix}$$

.3

$$D=\begin{bmatrix}3 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix},\ U=\begin{bmatrix}\frac{-1}{2} & -1 & 1\\ \frac{-1}{2} & 1 & 1\\ 1 & 0 & 1\end{bmatrix}$$

$D=U^{-1}X^TXU \dashrightarrow tr(D)=tr(U^{-1}X^TXU)=tr(U^{-1}U\,X^TX)=tr(I\,X^TX)$.

$tr(X^TX)=1+1+2=4=tr(D)=3+1+0$.

التمرين الثاني: لديك المصفوفة التالية.

$$X=\begin{bmatrix}4 & 5\\ 6 & 7\\ 8 & 0\end{bmatrix}$$

1. أحسب المصفوفة المركزة والمعيارية.
2. احسب مصفوفة التباين والتباين المشترك (بطريقتين).

3. أحسب القيم الذاتية لمصفوفة التباين والتباين المشترك.

الحل:

1. حساب المصفوفة المركزة والمعيارية:

1.1. المصفوفة المركزة:

$$X_C = X - 1_n G_X^T$$

$G_X = \begin{bmatrix} (4+6+8)/3 \\ (5+7+0)/3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, $G_X^T = [6 \quad 4]$, $1_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$X_C = \begin{bmatrix} 4 & 5 \\ 6 & 7 \\ 8 & 0 \end{bmatrix} - \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix} * [6 \quad 4] \right] = \begin{bmatrix} -2 & 1 \\ 0 & 3 \\ 2 & -4 \end{bmatrix}$

2.1. المصفوفة المعيارية:

$$X_{CR} = X_C * D_{\frac{1}{S}}$$

$S_1 = \sqrt{\frac{1}{3}((-2)^2 + (2)^2)} = 1.63$, $S_2 = 2.94$, $D_{\frac{1}{S}} = \begin{bmatrix} 0.61 & 0 \\ 0 & 0.34 \end{bmatrix}$

$X_{CR} = \begin{bmatrix} -2 & 1 \\ 0 & 3 \\ 2 & -4 \end{bmatrix} * \begin{bmatrix} 0.61 & 0 \\ 0 & 0.34 \end{bmatrix} = \begin{bmatrix} -1.22 & 0.34 \\ 0 & 1.02 \\ 1.22 & -1.36 \end{bmatrix}$

2. حساب مصفوفة التباين والتباين المشترك.

ط1. باستخدام علاقة التباين وعلاقة التباين المشترك لايجاد العناصر: V_{11}، V_{22}، V_{12}.

$$V(X,X) = \frac{1}{n}\sum_{j=1}^{p} (x_{ij} - \bar{x}_j)^2$$

$V_{11} = \frac{1}{3}(4+0+4) = 2.666$, $V_{22} = \frac{1}{3}(1+9+16) = 8.666$.

والتباين المشترك من العلاقة التالية:

$$Cov(X,Y) = \frac{1}{n}\sum_{j=1}^{p} (x_{ij} - \bar{x}_j) * (y_{ij} - \bar{y}_j)$$

$V_{12} = V_{21} = \frac{1}{3}((-2)(1)+(0)(3)+(2)(-4)) = -3.333$

$V = \begin{bmatrix} 2.666 & -3.333 \\ -3.333 & 8.666 \end{bmatrix}$

ط2. نستخدم الصيغة المصفوفية:

$$V = \frac{1}{n} X_C^T X_C$$

$V = \frac{1}{3} \begin{bmatrix} -2 & 0 & 2 \\ 1 & 3 & -4 \end{bmatrix} * \begin{bmatrix} -2 & 1 \\ 0 & 3 \\ 2 & -4 \end{bmatrix} = \begin{bmatrix} 2.666 & -3.333 \\ -3.333 & 8.666 \end{bmatrix}$

3. حساب القيم الذاتية للمصفوفة V.

$\det(V-\lambda I)=0 \leftarrow> \begin{bmatrix} 2.666-\lambda & -3.333 \\ -3.333 & 8.666-\lambda \end{bmatrix}=0 (2.666-\lambda)(8.666-\lambda)-$

$(-3.333)(-3.333)= 23.10-2.666\lambda-8.666\lambda+\lambda^2-11.10=0$

$\lambda^2-11.33\lambda+12=0$

$\Delta= b^2-4ac= (-11.33)^2-4(1)(12)=80.36$

$\lambda_1=\frac{-(-11.33)+\sqrt{80.36}}{2}= 10.14,\ \lambda_2=\frac{-(-11.33)-\sqrt{80.36}}{2}=1.18.$

التمرين الثالث: يحتوي الجدول الموالي على نقاط 6 طلبة في ثلاثة مواد (تحليل المعطيات، جزئي، رياضيات).

	تحليل المعطيات	جزئي	رياضيات
صابر	8	1	0
فريد	4	6	5
أحلام	6	8	7
محمد	10	4	7
سهام	8	2	5
يوسف	0	3	6

المطلوب: تحليل الجدول بواسطة المركبات الأساسية باتباع الخطوات الموالية.

1. تشكيل مصفوفة البيانات وحساب مركز ثقلها ومصفوفة الأوزان، وحساب الانحرافات المعيارية، ومن ثم إيجاد كل من المصفوفة الممركزة والمصفوفة المعيارية (المخفضّة).
2. حدد طريقة المركبات الأساسية التي تستخدم في التحليل (البسيطة- المرجحة).
3. حساب مصفوفة التباين والتباين المشترك (V) ومصفوفة الارتباطات.
4. حساب القيم الذاتية والأشعة الذاتية.
5. حساب التباين الكلي بطريقتين.
6. تشكيل جدول القيم الذاتية، وتحديد موثوقية المركبات الرئيسية المتحصل عليها.
7. حساب المركبات الأساسية للأفراد على المحاور.
8. حساب جودة تمثيل الأفراد على المحاور.
9. حساب نسبة مساهمة الأفراد في تشكيل المحاور.
10. حساب احداثيات المتغيرات على المحاور ومعاملات الارتباط بين المتغيرات والمركبات الرئيسية، ثم رسم دائرة الارتباط.

11. حساب جودة تمثيل المتغيرات على المحاور.

12. حساب نسبة مساهمة المتغيرات في تشكيل المحاور.

13. تمثيل الأفراد والمتغيرات على المخطط العاملي والتعليق عليه.

14. مثل الأفراد والمتغيرات على المخطط العاملي وفسّره.

14. لنفرض أننا أضفنا طالب جديد (صالح) كانت نقاطه على الترتيب (8، 7، 11)، أحسب المركبات الأساسية للفرد المضاف.

15. لنفرض أننا أضفنا مقياس جديد (تاريخ الفكر) (4، 3،6، 10،4،9).، أحسب احداثياته.

16. مثل الفرد المضاف والمتغير المضاف على المخطط العاملي.

الحل:

1. تشكيل مصفوفة البيانات وحساب مركز ثقلها مصفوفة الأوزان، وحساب الانحرافات المعيارية، ومن ثم إيجاد كل من المصفوفة المركزة والمصفوفة المعيارية (المخفضّة).

1.1. مصفوفة البيانات:

$$X=\begin{bmatrix} 8 & 1 & 0 \\ 4 & 6 & 5 \\ 6 & 8 & 7 \\ 10 & 4 & 7 \\ 8 & 2 & 5 \\ 0 & 3 & 6 \end{bmatrix}$$

2.1. مركز ثقل سحابة النقط للمصفوفة X.

$$G_X=\begin{matrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \dots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ \dots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{matrix}$$

$$G_X=\frac{1}{6}\left(\begin{bmatrix} 8+4+6+10+8+0 \\ 1+6+8+4+2+3 \\ 0+5+7+7+5+6 \end{bmatrix}\right)=\begin{bmatrix} 6 \\ 4 \\ 5 \end{bmatrix}$$

3.1. مصفوفة الأوزان:

$$D_p=\frac{1}{n}I_n$$

$$D_p=\frac{1}{6}I_6=\frac{1}{6}*\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}=\frac{1}{6}$$

4.1. حساب الانحرافات المعيارية:

$$S_1 = \sqrt{\frac{\sum_{j=1}^{6}(8-6)^2+(4-6)^2+(6-6)^2+(10-6)^2+(8-6)^2+(0-6)^2}{6}} = \sqrt{10} = 3.266$$

$$S_2 = \sqrt{\frac{\sum_{j=1}^{6}(1-4)^2+(6-4)^2+(8-4)^2+(4-4)^2+(2-4)^2+(3-4)^2}{6}} = \sqrt{5.667} = 2.3805$$

$$S_3 = \sqrt{\frac{\sum_{j=1}^{6}(0-5)^2+(5-5)^2+(7-5)^2+(7-5)^2+(5-5)^2+(6-5)^2}{6}} = \sqrt{5.667} = 2.3805$$

$$D_{\frac{1}{S}} = \begin{bmatrix} 0.306 & 0 & 0 \\ 0 & 0.420 & 0 \\ 0 & 0 & 0.420 \end{bmatrix}$$

5.1. المصفوفة المركزة:

$$X_C = X - 1_n G_X^T$$

$$1_6 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad G_X^T = [6 \quad 4 \quad 5]$$

$$X_C = \begin{bmatrix} 8 & 1 & 0 \\ 4 & 6 & 5 \\ 6 & 8 & 7 \\ 10 & 4 & 7 \\ 8 & 2 & 5 \\ 0 & 3 & 6 \end{bmatrix} - \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * [6 \quad 4 \quad 5] \right) = \begin{bmatrix} 8 & 1 & 0 \\ 4 & 6 & 5 \\ 6 & 8 & 7 \\ 10 & 4 & 7 \\ 8 & 2 & 5 \\ 0 & 3 & 6 \end{bmatrix} - \begin{bmatrix} 6 & 4 & 5 \\ 6 & 4 & 5 \\ 6 & 4 & 5 \\ 6 & 4 & 5 \\ 6 & 4 & 5 \\ 6 & 4 & 5 \end{bmatrix} \begin{bmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{bmatrix} =$$

6.1. المصفوفة المركزة المعيارية:

$$X_{CR} = X_C * D_{\frac{1}{S}}$$

$$X_{CR} = \begin{bmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{bmatrix} * \begin{bmatrix} 0.306 & 0 & 0 \\ 0 & 0.420 & 0 \\ 0 & 0 & 0.420 \end{bmatrix} = \begin{bmatrix} 0.612 & -1.260 & -2.1 \\ -0.612 & 0.84 & 0 \\ 0 & 1.68 & 0.84 \\ 1.224 & 0 & 0.84 \\ 0.612 & -0.840 & 0 \\ -1.837 & -0.420 & 0.42 \end{bmatrix}$$

2. طريقة المركبات الأساسية التي تستخدم في التحليل (البسيطة- المرجحة): بما أنّ المتغيرات متجانسة ولها نفس وحدات القياس، يمكننا استخدام طريقة المركبات الأساسية البسيطة.

3. حساب مصفوفة التباين والتباين المشترك (V): يمكننا حسابها بطريقتين:

1.3. باستخدام علاقتي التباين والتباين المشترك لايجاد العناصر: V_{11}، V_{22}، V_{33}، V_{12}، V_{13}، V_{23}.

$$V(X,X) = \frac{1}{n}\sum_{j=1}^{p}\left(x_{ij} - \bar{x}_j\right)^2$$

$V_{11}=\frac{1}{6}(4+4+16+4+36)=10.666$, $V_{22}=\frac{1}{6}(9+4+16+4+1)=5.666$, $V_{33}=\frac{1}{6}(25+4+4+1)=5.666$

التباين المشترك من العلاقة التالية:

$$Cov(X,Y)=\frac{1}{n}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)*(y_{ij}-\bar{y}_j)$$

$V_{12}=V_{21}=\frac{1}{6}((2)(-3)+(-2)(2)+(0)(4)+(4)(0)+(2)(-2)+(-6)(-1))=-1.333$

$V_{13}=V_{31}=\frac{1}{6}((2)(-5)+(-2)(0)+(0)(2)+(4)(2)+(2)(0)+(-6)(\ 1))=-1.333$

$V_{23}=V_{32}=\frac{1}{6}((-3)(-5)+(\ 2)(0)+(4)(2)+(0)(2)+(-2)(0)+(-1)(\ 1))=3.666$

$$V=\begin{bmatrix}10.666 & -1.333 & -1.333\\ -1.333 & 5.666 & 3.666\\ -1.333 & 3.666 & 5.666\end{bmatrix}$$

2.3. نقوم الآن بحسابها مستخدمين الصيغة المصفوفية:

$$V=\frac{1}{n}X_C^TX_C$$

$$V=\frac{1}{6}\begin{bmatrix}2 & -2 & 0 & 4 & 2 & -6\\ -3 & 2 & 4 & 0 & -2 & -1\\ -5 & 0 & 2 & 2 & 0 & 1\end{bmatrix}*\begin{bmatrix}2 & -3 & -5\\ -2 & 2 & 0\\ 0 & 4 & 2\\ 4 & 0 & 2\\ 2 & -2 & 0\\ -6 & -1 & 1\end{bmatrix}=\begin{bmatrix}10.666 & -1.333 & -1.333\\ -1.333 & 5.666 & 3.666\\ -1.333 & 3.666 & 5.666\end{bmatrix}$$

3.3. حساب مصفوفة الارتباطات:

$$R=D_{\frac{1}{S}}VD_{\frac{1}{S}}$$

$$R=\begin{bmatrix}0.306 & 0 & 0\\ 0 & 0.420 & 0\\ 0 & 0 & 0.420\end{bmatrix}*\begin{bmatrix}10.666 & -1.333 & -1.333\\ -1.333 & 5.666 & 3.666\\ -1.333 & 3.666 & 5.666\end{bmatrix}*\begin{bmatrix}0.306 & 0 & 0\\ 0 & 0.420 & 0\\ 0 & 0 & 0.420\end{bmatrix}=$$

$$\begin{matrix}1 & -0.171 & -0.171\\ -0.171 & 1 & 0.647\\ -0.171 & 0.647 & 1\end{matrix}$$

4. حساب القيم الذاتية والأشعة الذاتية لمصفوفة التباين والتباين المشترك.

1.4. القيم الذاتية: نبحث عن جذور كثير الحدود المميز، المعرف بمحدد $(V-\lambda I_3)$.

$$V-\lambda I_3=\begin{bmatrix}10.666 & -1.333 & -1.333\\ -1.333 & 5.666 & 3.666\\ -1.333 & 3.666 & 5.666\end{bmatrix}-\begin{bmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{bmatrix}$$

$$\det(V-\lambda I_3)=\begin{bmatrix}10.666-\lambda & -1.333 & -1.333\\ -1.333 & 5.666-\lambda & 3.666\\ -1.333 & 3.666 & 5.666-\lambda\end{bmatrix}=-\lambda^3+22\lambda^2-136\lambda+192$$

$$-\lambda^3+22\lambda^2-136\lambda+192=0$$

$$\lambda_1= 12,\ \lambda_2=8\ ,\ \lambda_3=2$$

$$D=\begin{matrix} 12 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 2 \end{matrix}$$

2.4. المتجهات الذاتية: لكل قيمة ذاتية نحسب متجهها الخاص:

$\lambda_1= 12$

VU= λU

$$\begin{bmatrix} 10.666 & -1.333 & -1.333 \\ -1.333 & 5.666 & 3.666 \\ -1.333 & 3.666 & 5.666 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}=12\begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

نستخدم حذف كاوس في حل النظام المتجانس من المعادلات الخطية.

$$\begin{bmatrix} -1.333 & -1.333 & -1.333 \\ -1.333 & -6.333 & 3.666 \\ -1.333 & 3.666 & -6.333 \end{bmatrix} \boxed{\begin{matrix} *-1 \\ + \end{matrix}} \begin{bmatrix} -1.333 & -1.333 & -1.333 \\ -1.333 & -6.333 & 3.666 \\ 0 & 5 & -5 \end{bmatrix} \boxed{\begin{matrix} *-1 \\ + \end{matrix}}$$

$$\begin{bmatrix} -1.333 & -1.333 & -1.333 \\ 0 & -5 & 5 \\ 0 & 5 & -5 \end{bmatrix} \boxed{\begin{matrix} *1 \\ + \end{matrix}} \begin{bmatrix} -1.333 & -1.333 & -1.333 \\ 0 & -5 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

$-1.333a-1.333b-1.333c=0......1$

$-5b+5c=0$----------→ b=c...2

c=c

بتعويض 2 في 1 نجد:

$-1.333a-2.666b=0$----→ a=−2b

$$U_1=\begin{bmatrix} -2b \\ b \\ b \end{bmatrix}=b\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

ولكي يكون متجه موجه للمحور العاملي الأول (الرئيسي)، يجب أن تكون طويلته تساوي 1.

$$\|U_1\|=\sqrt{(-2)^2 + (1)^2 + (1)^2}= 2.499$$

نقوم بتعديله، بقسمة مركباته على طويلته:

$$U_1=\begin{bmatrix} {}^{-2}/_{2.499} \\ {}^{1}/_{2.499} \\ {}^{1}/_{2.499} \end{bmatrix}=\begin{bmatrix} -0.8 \\ 0.4 \\ 0.4 \end{bmatrix}$$

$\lambda_1= 8$

$U_2=\begin{bmatrix}0.577\\0.577\\0.577\end{bmatrix}$, $U_3=\begin{bmatrix}0\\-0.707\\0.707\end{bmatrix}$........... $U=\begin{bmatrix}-0.8 & 0.577 & 0\\0.4 & 0.577 & -0.707\\0.4 & 0.577 & 0.707\end{bmatrix}$

5. حساب التباين الكلي بطريقتين.

1.5. الطريقة الأولى: التباين الكلي يساوي أثر مصفوفة التباين والتباين المشترك.

I_T= trac(V)= 10.666+5.666+5.666= 22

2.5. الطريقة الثانية: التباين يساوي مجموع القيم الذاتية للمصفوفة V.

$I_T=\sum_{j=1}^{p}\lambda_i= 12+8+2=22$

6. تشكيل جدول القيم الذاتية، وتحديد موثوقية المركبات الرئيسية المتحصل عليها.

F_i	القيمة الذاتية	النسبة المئوية	النسبة المئوية المتصاعدة
F_1	12	54.545	54.545
F_2	8	36.364	90.909
F_3	2	9.091	100

تكون المركبات الرئيسية المتحصل عليها جيدة وكافية لإعطاء صورة واضحة عن سحابة النقط في المخطط العاملي إذا كانت نسبة التباين الكلي المفسر بالمركبات الرئيسية أكبر من 75 في المائة. نلاحظ أنّ نسبة التباين المفسّر بواسطة المركبة الأساسية الأولى يساوي 54.54 في المائة، ونسبة التباين المفسّر بالمركبة الرئيسية الثانية يساوي 36.36 في المائة. فتكون نسبة التباين المفسّر بواسطة المركبتين معا 90.90 في المائة، وهي نسبة كافية لإعطاء صورة واضحة جدا لسحابة النقاط على المخطط العاملي، أي حفظ أكبر قدر من المعلومات.

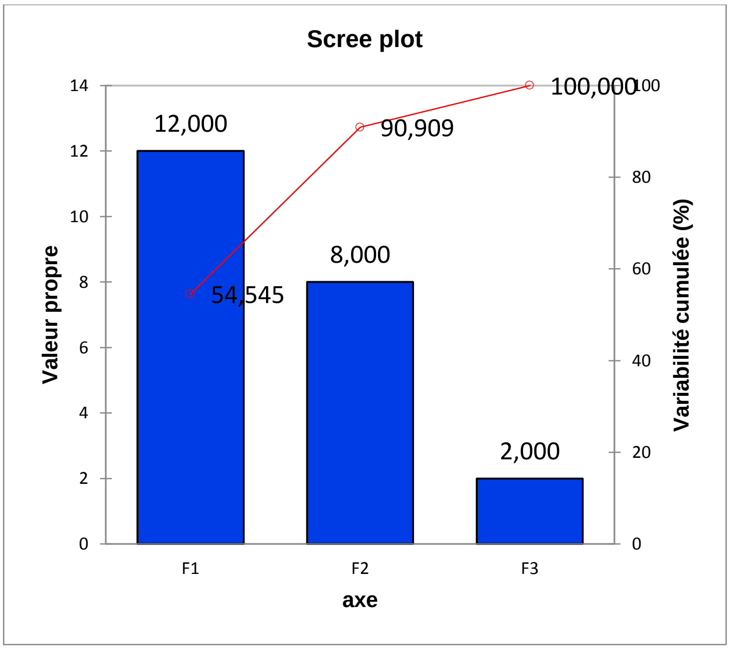
* معايير اختيار عدد المحاور:

- حسب محك كايزر كل قيمة ذاتية تفوق متوسط القيم الذاتية تأخذ كمركبة رئيسية.

$\bar{\lambda}=\frac{1}{p}\sum_{i=1}^{p}\lambda_i=\frac{1}{3}(12+8+2)= 7.333$

وعليه، نحتفظ بالمحورين الأول والثاني فقط.

- حسب معيار التمثيل البياني:



بحسب معيار التمثيل البياني نحتفظ فقط بالمحورين الأول والثاني.

7. حساب احداثيات الأفراد على المحاور.

$$F_\alpha = X_C U_\alpha$$

$$F_\alpha = \begin{bmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{bmatrix} * \begin{bmatrix} \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -4.9 & -3.46 & -1.41 \\ 2.45 & 0 & -1.41 \\ 2.45 & 3.46 & -1.41 \\ -2.45 & 3.46 & 1.41 \\ -2.45 & 0 & 1.41 \\ 4.9 & -3.46 & 1.41 \end{bmatrix}$$

$$\overline{F_\alpha} = 0$$

	F1	F2	F3
صابر	-4,9	-3,464	-1,414
فريد	2,45	0,000	-1,414
أحلام	2,45	3,464	-1,414
محمد	-2,45	3,464	1,414
سهام	-2,45	0,000	1,414
يوسف	4, 9	-3,464	1,414

8. حساب جودة تمثيل الأفراد على المحاور.

$$\text{Cos}^2\theta_{i\alpha} = \frac{F_\alpha^2(i)}{\sum_{\alpha=1}^{p} F_\alpha^2(i)}$$

كلما اقتربت هذه القيمة من الواحد، كان الفرد ممثلا أحسن تمثيل.

$$\text{Cos}^2\,\theta_{11} = \frac{-4.9^2}{(-4.9)^2+(-3.446)^2+(-1.414)^2} = 0.632$$

	F1	F2	F3
صابر	**0,632**	0,316	0,053
فريد	**0,750**	0,000	0,250
أحلام	0,300	**0,600**	0,100
محمد	0,300	**0,600**	0,100
سهام	**0,750**	0,000	0,250
يوسف	**0,632**	0,316	0,053

يظهر من الجدول أنّ صابر وفريد وسهام ويوسف ممثلين جيدا على المحور الأول. أحلام ومحمد ممثلين جيدا على المحور الثاني. وكل الأفراد غير ممثلين بشكل جيد على المحور الثالث.

9. حساب نسبة مساهمة الأفراد في تشكيل المحاور.

$$C_i^{\alpha}=\frac{F_{\alpha}^2(i)}{n\lambda_{\alpha}}$$

$C^1_1=\frac{F_1^2(1)}{6\lambda_1}=\frac{(-4.9)^2}{6*12}=0.333,\ C^2_1=\frac{F_2^2(1)}{6\lambda_2}=\frac{(-3.464)^2}{6*8}=0.25$

وبنفس الطريقة حتى نحسب جميع القيم.

	F1	F2	F3
صابر	33,333	25,000	16,667
فريد	8,333	0,000	16,667
أحلام	8,333	25,000	16,667
محمد	8,333	25,000	16,667
سهام	8,333	0,000	16,667
يوسف	33,333	25,000	16,667

نسبة مساهمة صابر ويوسف في تشكيل المحور الأول متساوية ، ونسبة مساهمة صابر وأحلام ومحمد ويوسف في تشكيل المحور الثاني متساوية.

10. حساب احداثيات المتغيرات على المحاور، ورسم دائرة الارتباط.

1.10. حساب احداثيات المتغيرات على المحاور.

$$G_{\alpha}=\sqrt{\lambda_{\alpha}}U_{\alpha}$$

$G_1=\sqrt{\lambda_1}U_1=\sqrt{12}\begin{bmatrix}-0.8\\0.4\\0.4\end{bmatrix}=\begin{bmatrix}-2.771\\1.414\\1.414\end{bmatrix},\quad G_2=\sqrt{8}\begin{bmatrix}0.577\\0.577\\0.577\end{bmatrix}=\begin{bmatrix}1.632\\1.632\\1.632\end{bmatrix},\quad G_3=\sqrt{2}\begin{bmatrix}0\\-0.707\\0.707\end{bmatrix}=\begin{bmatrix}0\\-1\\1\end{bmatrix}$

	F1	F2	F3
تحليل المعطيات	-2,771	1,632	0,000
جزئي	1,385	1,632	-1,000
رياضيات	1,385	1,632	1,000

2.10. الارتباط بين المتغيرات والمركبات الأساسية.

$$Cor= D_{\sqrt{\lambda}} * U^T * D_{\frac{1}{s}}$$

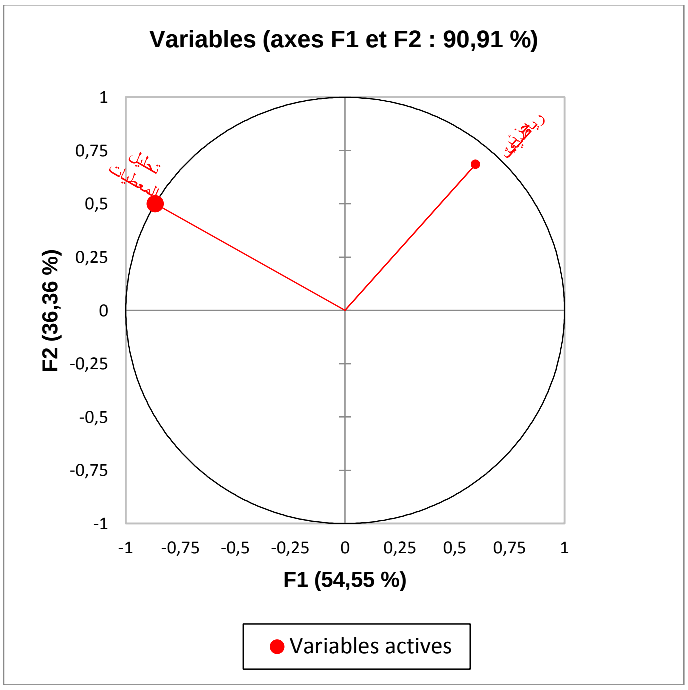
$$D_{\sqrt{\lambda}} \begin{bmatrix} 3.464 & 0 & 0 \\ 0 & 2.828 & 0 \\ 0 & 0 & 1.414 \end{bmatrix}$$

$$Cor= \begin{bmatrix} 3.464 & 0 & 0 \\ 0 & 2.828 & 0 \\ 0 & 0 & 1.414 \end{bmatrix} * \begin{bmatrix} -0.8 & 0.4 & 0.4 \\ 0.577 & 0.577 & 0.577 \\ 0 & -0.707 & 0.707 \end{bmatrix} * \begin{bmatrix} 0.306 & 0 & 0 \\ 0 & 0.42 & 0 \\ 0 & 0 & 0.42 \end{bmatrix} =$$

$$\begin{bmatrix} -0.848 & 0.582 & 0.582 \\ 0.499 & 0.685 & 0.685 \\ 0 & -0.42 & 0.42 \end{bmatrix}$$

	F1	F2	F3
تحليل المعطيات	**-0.848**	0.582	0.582
جزئي	0.499	**0.685**	0.685
رياضيات	0	-0.420	0.420

2.10. رسم دائرة الارتباط: نرسم دائرة الارتباط بتحديد احداثية كل متغير على المخطط العاملي، ثم نرسم دائرة انطلاقا من المركز نصف قطرها يساوي 1.



يظهر أنّ الارتباط بين متغير تحليل المعطيات والمركبة الأساسية الأولى قوي وعكسي (سالب)، كما يرتبط ايجابيا بشكل متوسط مع المركبة الأساسية الثانية. في حين أنّ الارتباط بين متغير الجزئي والمركبة الأساسية الأولى متوسط، وهو متوسط كذلك مع المركبة الثانية. ولا يوجد ارتباط بين متغير الرياضيات مع المركبة الأولى، ويرتبط بالمركبة الثانية ارتباطا سلبيا متوسطا.

11. جودة تمثيل المتغيرات على المحاور:

$$Cos^2(j,\propto) = \frac{G_{j\propto}^2}{d^2(j,0)}$$

$Cos^2(1,1) = \frac{G_{11}^2}{d^2(1,0)}$.................$d^2(1{,}0)=(3.262)^2 = 10.666$

في التحليل بالمركبات الرئيسية البسيطة تتطابق المسافة إلى الأصل مع الانحراف المعياري للمتغير.

$Cos^2(1,1) = \frac{G_{11}^2}{d^2(1,0)} = \frac{(-2.828)^2}{10.666} = 0.749$

$Cos^2(1,2) = \frac{G_{12}^2}{d^2(1,0)} = \frac{(1.632)^2}{10.666} = 0.49$, $Cos^2(2,1) = \frac{G_{21}^2}{d^2(2,0)} = \frac{(1.385)^2}{5.666} = 0.338$

وبعد اتمام جميع العمليات نجد:

	F1	F2	F3
تحليل المعطيات	**0.749**	2.50	0
جزئي	0.338	**0.470**	0.176
رياضيات	0.338	**0.470**	0.176

12. نسبة مساهمة المتغيرات في تشكيل المحاور:

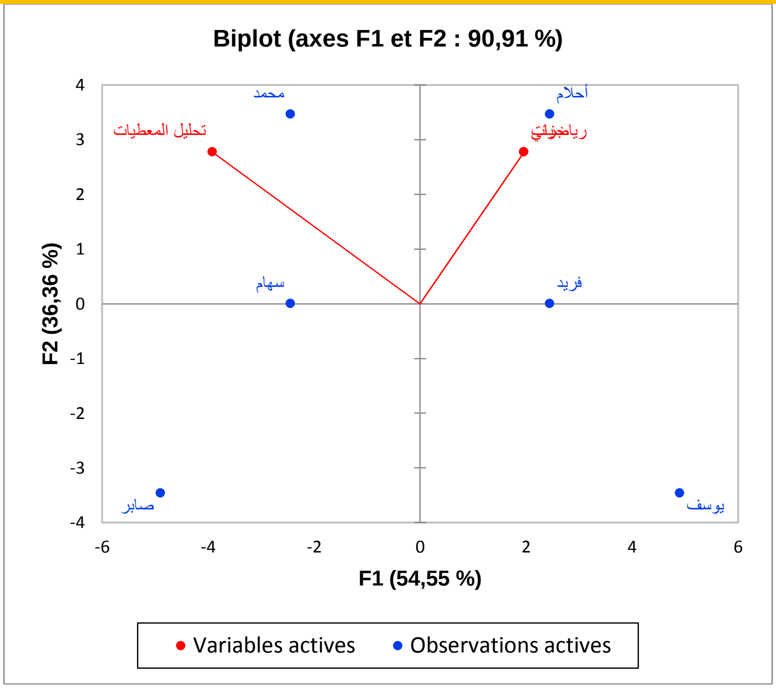
$$Ctr_{(j,\propto)} = \frac{G_{j\propto}^2}{\lambda_\propto}$$

$Ctr_{(1,1)} = \frac{G_{11}^2}{\lambda_1} = \frac{(-2.828)^2}{12} = 0.6664$

وبعد اتمام جميع العمليات نجد:

	F1	F2	F3
تحليل المعطيات	66.64	33.33	0,000
جزئي	15.98	33.33	50
رياضيات	15.98	33.33	50

13. تمثيل الأفراد والمتغيرات على المخطط العاملي والتعليق عليه: في هذا المخطط يتم تمثيل احداثيات كل من الأفراد والمتغيرات على نفس المخطط.



لتفسير هذا المخطط، نتبع القواعد التالية:

- تقارب نقطتي فردين تعني التشابه.
- تقارب نقطتي متغيرين تعني الترابط.
- تقارب نقطة فرد مع نقطة متغير تعني أنّ المتغير يأثر في سلوك ذلك الفرد.

ومن خلال المخطط أعلاه نستنتج ما يلي:

- يرتبط متغير تحليل المعطيات مع المحور الأول بقوة وفي اتجاه عكسي، معنى ذلك أنّ الأفراد المتواجدين على يمين المخطط نقاطهم ضعيفة في تحليل المعطيات (أحلام، فريد، يوسف)، نقطة يوسف ضعيفة جدا، والأفراد المتواجدين يسار المخطط نقاطهم أفضل (صابر، محمد، سهام).
- نقاط أحلام في الجزئي والرياضيات متقاربة وأفضل من بقية الأفراد، نقطة محمد في تحليل المعطيات أفضل من البقية.
- نقطتي فريد وسهام متعاكسة في تحليل المعطيات، سهام أفضل، ونقطة فريد في الجزئي أفضل من سهام، بينما تتقارب نقاطهما في الرياضيات (يقعان على المحور الأول ولا ترتبط الرياضيات مع هذا المحور).
- نقاط صابر ضعيفة جدا في الجزئي والرياضيات، وجيدة في تحليل المعطيات (تمثيل صابر جيد على المحور الأول، ونسبة مساهمته في تشكيل المحور أفضل من البقية، كما أنّ الجزئي والرياضيات لا يرتبطان بالحمور الأول، في حين يرتبط تحليل المعطيات بقوة مع المحور الأول).

14. حساب المركبات الأساسية للفرد المضاف.

المركبات الأساسية للفرد المضاف في حالة استخدام طريقة المركبات الأساسية البسيطة:

$$F_\alpha^{+} = X_C^{+} U_\alpha$$

$X = [8 \quad 7 \quad 11] \Rightarrow X_C^+ = [2 \quad 3 \quad 6]$

$$F_\alpha^+ = [2 \quad 3 \quad 2] * \begin{bmatrix} -0.8 & 0.577 & 0 \\ 0.4 & 0.577 & -0.707 \\ 0.4 & 0.577 & 0.707 \end{bmatrix} = [0.4 \quad 4.039 \quad -0.707]$$

15. احداثية المتغير التكميلي، (تاريخ الفكر) (4، 3،6، 10،4،9).

$$G_\alpha^+ = \frac{1}{n} X_C^{+T} \left(\frac{X_C U_\propto}{\sqrt{\lambda_\propto}}\right)$$

$$X_C^+ = \begin{bmatrix} 2 & -3 & -5 & -2 \\ -2 & 2 & 0 & 0 \\ 0 & 4 & 2 & -3 \\ 4 & 0 & 2 & 4 \\ 2 & -2 & 0 & -2 \\ -6 & -1 & 1 & 3 \end{bmatrix}$$

$$G_1^+ = \frac{1}{6}\frac{1}{\sqrt{12}} \begin{bmatrix} 2 & -2 & 0 & 4 & 2 & -6 \\ -3 & 2 & 4 & 0 & -2 & -1 \\ -5 & 0 & 2 & 2 & 0 & 1 \\ -2 & 0 & -3 & 4 & -2 & 3 \end{bmatrix} * \begin{bmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{bmatrix} * \begin{bmatrix} \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{matrix} -2.822 \\ 1.411 \\ 1.411 \\ 0.588 \end{matrix}$$

$$G_2^+ = \frac{1}{6}\frac{1}{\sqrt{8}} \begin{bmatrix} 2 & -2 & 0 & 4 & 2 & -6 \\ -3 & 2 & 4 & 0 & -2 & -1 \\ -5 & 0 & 2 & 2 & 0 & 1 \\ -2 & 0 & -3 & 4 & -2 & 3 \end{bmatrix} * \begin{bmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{matrix} 1.630 \\ 1.630 \\ 1.630 \\ 0 \end{matrix}$$

التمرين الرابع: يحتوي جدول المعطيات الكمية الموالي على معدل تساقط الأمطار (precipitation) ودرجات الحرارة القصوى (extreme temperatures) والدنيا (lower temperatures) بستة مدن جزائرية.

	P	T-extreme	T-lower
جيجل	12.04	23.7	5.9
سطيف	17.18	15.5	-1.8
خنشلة	11.83	13.1	2.8
تبسة	6.23	13.5	-2.4
عنابة	16.99	21.1	7.2
البيض	3.87	20.3	-0.9

لدينا ستة ولايات (أفراد) مقابل ثلاثة متغيرات (معدل التساقط، درجات الحرارة القصوى، درجات الحرارة الدنيا)، يمكن ملاحظة أنّ متوسط معدل التساقط في ولاية جيجل 12.04، ومتوسط درجات الحرارة القصوى 23.7، والدنيا 5.9. ونكتب مصفوفة البيانات الأولية X كما يلي:

$$X=\begin{bmatrix} 12.04 & 23.7 & 5.9 \\ 17.18 & 15.5 & -1.8 \\ 11.83 & 13.1 & 2.8 \\ 6.23 & 13.5 & -2.4 \\ 16.99 & 21.1 & 7.2 \\ 3.87 & 20.3 & -0.9 \end{bmatrix}$$

1. مصفوفة الأوزان المرتبطة بالأفراد:

بما أنّ جميع الأفراد في مثالنا لهم نفس الأهمية في الدراسة، يكون لهم نفس الوزن، ونكتب:

$$D_P= \frac{1}{n} I_n$$

$$I_6=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \ldots\ldots D_P=\frac{1}{6}$$

2. مركز ثقل سحابة النقط.

$$G_j= \sum_{i=1}^{n} p_i X_i= \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$G_1= \frac{1}{6}(12.04 + 17.18 + 11.83 + 6.23 + 16.99 + 3.87)=11.36$$

$$G_2= \frac{1}{6}(23.7+15.5+13.1+13.5+21.1+20.3)= 17.87$$

$$G_3= \frac{1}{6}(5.9 - 1.8 + 2.8 - 2.4 + 7.2 - 0.9)=1.8$$

$$G_X=\begin{bmatrix}11.36\\17.87\\1.8\end{bmatrix}$$

كما يمكن حسابه من الصيغة المصفوفية:

$$G_X= X^T D_P I_n$$

$$G_X=\begin{bmatrix}12.04 & 17.18 & 11.83 & 6.23 & 16.99 & 3.87\\23.7 & 15.5 & 13.1 & 13.5 & 21.1 & 20.3\\5.9 & -1.8 & 2.8 & -2.4 & 7.2 & -0.9\end{bmatrix}*\begin{bmatrix}\frac{1}{6} & 0 & 0 & 0 & 0 & 0\\0 & \frac{1}{6} & 0 & 0 & 0 & 0\\0 & 0 & \frac{1}{6} & 0 & 0 & 0\\0 & 0 & 0 & \frac{1}{6} & 0 & 0\\0 & 0 & 0 & 0 & \frac{1}{6} & 0\\0 & 0 & 0 & 0 & 0 & \frac{1}{6}\end{bmatrix}*\begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}=\begin{bmatrix}11.36\\17.87\\1.8\end{bmatrix}$$

3. المصفوفة الممركزة: نحسب المصفوفة الممركزة X_C من الصيغة المصفوفية:

$$X_C= X-1_n G_X^T$$

حيث: X مصفوفة البيانات الأولية، 1_n متجه الوحدة، G_X^T منقول مركز ثقل سحابة النقط.

$$G_X=\begin{bmatrix}11.36\\17.87\\1.8\end{bmatrix}\in M(3,1),\ 1_6=\begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}\in M(6,1)$$

$$X_C=\begin{bmatrix}12.04 & 23.7 & 5.9\\17.18 & 15.5 & -1.8\\11.83 & 13.1 & 2.8\\6.23 & 13.5 & -2.4\\16.99 & 21.1 & 7.2\\3.87 & 20.3 & -0.9\end{bmatrix}-\left(\begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}*\begin{bmatrix}11.36 & 17.87 & 1.8\end{bmatrix}\right)\begin{bmatrix}0.68 & 5.83 & 4.1\\5.82 & -2.37 & -3.6\\0.47 & -4.77 & 1\\-5.13 & -4.37 & -4.2\\5.63 & 3.23 & 5.4\\-7.49 & 2.43 & -2.7\end{bmatrix}$$

ويمكننا كذلك حساب الجدول الممركز من العلاقة:

$$\boldsymbol{\tilde{X}_{ij}}=\boldsymbol{X_{ij}}-\bar{X}_j$$

	P	T-extreme	T-lower
جيجل	12.04-11.36	23.7-17.87	5.9-1.8
سطيف	17.18-1136	15.5-17.87	-1.8-1.8
خنشلة	11.83-11.36	13.1-17.87	2.8-1.8
تبسة	6.23-11.36	13.5-17.87	-2.4-1.8
عنابة	16.99-11.36	21.1-17.87	7.2-1.8

البيض	3.87-11.36	20.3-17.87	-0.9-1.8
G_X	11.36-11.36	17.87-17.87	1.8-1.8

ملاحظة: عند مركزة البيانات يصبح مركز ثقل سحابة النقط الجديد يساوي 0.

$G_X=(0,0,0)$.

4. المصفوفة المركزة المعيارية (تخفيض البيانات):

نحسب الانحراف المعياري.

$$s_j=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{ij}-g_j)^2}\text{.......}j=1,\ldots\ldots,p$$

ونحسب المصفوفة المعيارية من العلاقة التالية:

$$X_{CR}=X_C{}^{*}\ D_{\frac{1}{S}}$$

الانحرافات المعيارية للمتغيرات:

$$\delta_P=\sqrt{\frac{\sum_{j=1}^{6}(12.04-11.36)^2+(17.18-11.36)^2+(11.83-11.36)^2+(6.23-11.36)^2+(16.99-11.36)^2+(3.87-11.36)^2}{6}}=4.98$$

$$\delta_{\text{T-extreme}}=4.04\ ,\ \delta_{\text{T-lower}}=3.76$$

$$Z_{11}=\frac{X_{11}-\bar{X}_1}{\sigma_1}=\frac{0.68}{4.98}=0.136$$

حساب المصفوفة المركزة والمعيارية:

$$D_{\frac{1}{S}}=\begin{bmatrix}\frac{1}{4.98} & 0 & 0\\ 0 & \frac{1}{4.04} & 0\\ 0 & 0 & \frac{1}{3.76}\end{bmatrix}=\begin{bmatrix}0.201 & 0 & 0\\ 0 & 0.248 & 0\\ 0 & 0 & 0.266\end{bmatrix}$$

$$X_{CR}=\begin{bmatrix}0.68 & 5.83 & 4.1\\ 5.82 & -2.37 & -3.6\\ 0.47 & -4.77 & 1\\ -5.13 & -4.37 & -4.2\\ 5.63 & 3.23 & 5.4\\ -7.49 & 2.43 & -2.7\end{bmatrix}*\begin{bmatrix}0.201 & 0 & 0\\ 0 & 0.248 & 0\\ 0 & 0 & 0.266\end{bmatrix}=$$

$$\begin{bmatrix}0.137 & 1.443 & 1.090\\ 1.169 & -0.587 & -0.957\\ 0.094 & -1.181 & 0.266\\ -1.030 & -1.082 & -1.117\\ 1.131 & 0.800 & 1.436\\ -1.504 & 0.601 & -0.718\end{bmatrix}$$

ويمكن كذلك حساب الجدول المركز المعياري مباشرة بقسمة احداثيات كل عمود في الجدول المركز على الانحراف المعياري المقابل له. وفقا للعلاقة:

$$Z_{ij}=\frac{X_{ij}-\bar{X}_j}{S_j}$$

$$Z_{11} = \frac{X_{11}-\bar{X}_1}{S_1} = \frac{12.04-11.36}{4.98}$$

	P	T-extreme	T-lower
جيجل	0.137	1.443	1.090
سطيف	1.169	-0.587	-0.957
خنشلة	0.094	-1.181	0.266
تبسة	-1.030	-1.082	-1.117
عنابة	1.131	0.800	1.436
البيض	-1.504	0.601	-0.718
G_X	0	0	0
S	1	1	1

5. تحديد طريقة المركبات الأساسية المستخدمة:

في حالتنا هذه لا يمكننا استخدام طريقة المركبات الأساسية البسيطة، لأنّ وحدات المتغيرات غير متجانسة (درجة الحرارة تقاس بالدرجة، والتساقط بالمللتر)، وعلينا اختزال الوحدات، وبالتالي استخدام طريقة المركبات الأساسية المرجّحة.

6. مصفوفة الارتباطات: في حالة كانت المتغيرات غير متجانسة نستخدم مصفوفة الارتباطات، ويتم حساب الارتباط بين المتغيرات باستخدام الصيغة المصفوفية الآتية:

$$R = X_{CR}^T D_P X_{CR}$$

$$R=\frac{1}{6}\begin{bmatrix} 0.137 & 1.169 & 0.094 & -1.030 & 1.131 & -1.504 \\ 1.443 & -0.587 & -1.181 & -1.082 & 0.800 & 0.601 \\ 1.090 & -0.957 & 0.266 & -1.117 & 1.436 & -0.718 \end{bmatrix} * \begin{bmatrix} 0.137 & 1.443 & 1.090 \\ 1.169 & -0.587 & -0.957 \\ 0.094 & -1.181 & 0.266 \\ -1.030 & -1.082 & -1.117 \\ 1.131 & 0.800 & 1.436 \\ -1.504 & 0.601 & -0.718 \end{bmatrix}$$

$$R=\begin{bmatrix} 1 & 0.09 & 0.49 \\ 0.09 & 1 & 0.62 \\ 0.49 & 0.62 & 1 \end{bmatrix}$$

باستخدام معامل الارتباط المعطى بالعلاقة:

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

$$r_{XY} = \frac{Cov\,(X,Y)}{\delta_X \delta_Y}$$

$$r_{12}=r_{21}=\frac{1.722}{4.98*4.04}=0.09$$

وبعد اتمام جميع العمليات تكون مصفوفة الارتباطات كما يلي:

R= $\begin{bmatrix} 1 & 0.09 & 0.49 \\ 0.09 & 1 & 0.62 \\ 0.49 & 0.62 & 1 \end{bmatrix}$

يظهر من مصفوفة الارتباطات أن متغير معدل تساقط الأمطار مستقل عن متغير درجات الحرارة القصوى، معامل الارتباط=0.09، ويرتبط بشكل متوسط مع متغير درجات الحرارة الدنيا (r=0.49). في حين يرتبط متغيري درجات الحرارة الدنيا والقصوى بقوة (r=0.62).

7. القيم والمتجهات الذاتية:

نحسب القيم الذاتية لمصفوفة الارتباط باستخدام المحدد.

det($\begin{bmatrix} 1 & 0.09 & 0.49 \\ 0.09 & 1 & 0.62 \\ 0.49 & 0.62 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$)=0→det($\begin{bmatrix} 1-\lambda & 0.09 & 0.49 \\ 0.09 & 1-\lambda & 0.62 \\ 0.49 & 0.62 & 1-\lambda \end{bmatrix}$) $= 0$

$(1-\lambda)[(1-\lambda)(1-\lambda) - 0.62 * 0.62] - 0.09[(1-\lambda)(0.09) - 0.62 * 0.49] + 0.49[(0.62)(0.09) - 0.49 * (1-\lambda)] = -\lambda^3 + 3\lambda^2 - 2.367\lambda + 0.422 = 0$

$\lambda_1 = 1.83,\ \lambda_2 = 0.92,\ \lambda_3 = 0.25.$

ونكتب المصفوفة القطرية للقيم الذاتية على الشكل التالي:

D= $\begin{bmatrix} 1.83 & 0 & 0 \\ 0 & 0.92 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$

لكل قيمة ذاتية نحسب المتجه الذاتي المرافق لها.

$\lambda_1 = 1.83$

$(R - \lambda_1 I)U1 = 0 \longrightarrow \begin{bmatrix} 1-1.83 & 0.09 & 0.49 \\ 0.09 & 1-183 & 0.62 \\ 0.49 & 0.62 & 1-1.83 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

إذاً لدينا نظام متجانس من المعادلات الخطية، نقوم بحلها بواسطة حذف كاوس:

$\begin{bmatrix} -0.835 & 0.09 & 0.49 \\ 0.09 & -0.835 & 0.62 \\ 0.49 & 0.62 & -0.835 \end{bmatrix}$ *-1.197 $\begin{bmatrix} 1 & -0.108 & -0.586 \\ 0.09 & -0.835 & 0.62 \\ 0.49 & 0.62 & -0.835 \end{bmatrix}$ *-0.09 +

$\begin{bmatrix} 1 & -0.108 & -0.586 \\ 0 & -0.826 & 0.673 \\ 0.49 & 0.62 & -0.835 \end{bmatrix}$ *-0.49 + $\begin{bmatrix} 1 & -0.108 & -0.586 \\ 0 & -0.826 & 0.673 \\ 0 & 0.673 & -0.548 \end{bmatrix}$ *-1.211

$\begin{bmatrix} 1 & -0.108 & -0.586 \\ 0 & 1 & -0.815 \\ 0 & 0.673 & -0.548 \end{bmatrix}$ *-0.673 + $\begin{bmatrix} 1 & -0.108 & -0.586 \\ 0 & 1 & -0.815 \\ 0 & 0 & 0 \end{bmatrix}$

$U_1 - 0.108U_2 - 0.586U_3 = 0$-----→$U_1 = 0.108U_2 + 0.586U_3$

$U_2 - 0.815U_3 = 0$--------→ $U_2 = 0.815U_3$

$U_3 = U_3$

$U_1= 0.674\ U_3$

$$U=\begin{bmatrix}0.674U_3\\0.815U_3\\U_3=U_3\end{bmatrix} \longleftrightarrow U=U_3\begin{bmatrix}0.674\\0.815\\1\end{bmatrix} \ldots\ldots U_3=1 \longrightarrow U=\begin{bmatrix}0.674\\0.815\\1\end{bmatrix}$$

علينا تحويله إلى شعاع الوحدة، نحسب طويلته ثم نقسم مركباته على الطويلة.

$$\|U\|=\sqrt{0.674^2+0.815^2+1^2}=1.455$$

$$U=\begin{bmatrix}0.674/1.455\\0.815/1.455\\1/1.455\end{bmatrix}=\begin{bmatrix}0.46\\0.56\\0.69\end{bmatrix}$$

$$\|U\|=\sqrt{0.46^2+0.56^2+0.69^2}=1$$

وبعد حساب بقية المتجهات الذاتية، نكتب مصفوفة المتجهات الذاتية الوحدوية على النحو التالي:

$$U=\begin{bmatrix}0.46&0.79&0.41\\0.56&-0.61&0.56\\0.69&-0.03&-0.72\end{bmatrix}$$

ويمكننا إثبات أنّ:

$U_1^T*U_1=1$

$$\begin{bmatrix}0.46&0.56&0.69\end{bmatrix}*\begin{bmatrix}0.46\\0.56\\0.69\end{bmatrix}=1$$

$U_1^T*U_2=0$

$$\begin{bmatrix}0.46&0.56&0.69\end{bmatrix}*\begin{bmatrix}0.79\\-0.61\\-0.03\end{bmatrix}=0$$

كما أنّ مصفوفة المتجهات الذاتية U مصفوفة قابلة للعكسية ومعكوسها يساوي منقولها ($U^{-1}=U^T$). وعلى ذلك يمكننا إثبات أنّ مصفوفة الارتباط قابلة للاستقطار، حيث:

$R=UDU^{-1}$

$$R=\begin{bmatrix}0.46&0.79&0.41\\0.56&-0.61&0.56\\0.69&-0.03&-0.72\end{bmatrix}*\begin{bmatrix}1.83&0&0\\0&0.92&0\\0&0&0.25\end{bmatrix}*\begin{bmatrix}0.46&0.56&0.69\\0.79&-0.61&-0.03\\0.41&0.56&-0.72\end{bmatrix}=R$$

8. نسبة التباين المفسّر بواسطة المركبات الأساسية:

F_i	القيمة الذاتية	نسبة التباين المفسّر	النسبة المتصاعدة
F_1	1.835	$61=100*\frac{1.83}{1.83+0.92+0.25}$	61
F_2	0.92	30.67	90.67
F_3	0.25	8.33	100

يحتفظ المحور الأول بـ 61 % من المعلومات، ويحفظ المحور الثاني 30.67 % من المعلومات، في حين يحتفظ المحور الثالث بنسبة 8.33 % من المعلومات.

وبما أنّ كل قيمة ذاتية تمثل المساهمة المطلقة للمحور في التباين الكلي، نجد:

– المساهمة المطلقة للمحور العاملي الأول في التباين الكلي تساوي 0.61.

$$C_r(\Delta_1/I_0)=\frac{\lambda_1}{\lambda_1+\lambda_2+\cdots+\lambda_p}=\frac{\lambda_1}{Trac(R)}=\frac{1.83}{3}=0.6100$$

– المساهمة المطلقة للمحور العاملي الثاني في التباين الكلي تساوي 0.3051.

$$C_r(\Delta_2/I_0)=\frac{\lambda_2}{\lambda_1+\lambda_2+\cdots+\lambda_p}=\frac{\lambda_2}{Trac(R)}=\frac{0.92}{3}=0.3067$$

– المساهمة المطلقة للمحور العاملي الثالث في التباين الكلي تساوي 0.0832.

– المساهمة النسبية للمحور العاملي الأول في التباين الكلي تساوي 61.17 في المائة.

– المساهمة النسبية للمحور العاملي الثاني في التباين الكلي تساوي 30.51 في المائة.

– المساهمة النسبية للمحور العاملي الثالث في التباين الكلي تساوي 8.32 في المائة.

– نسبة التمثيل للمستوى الأول:

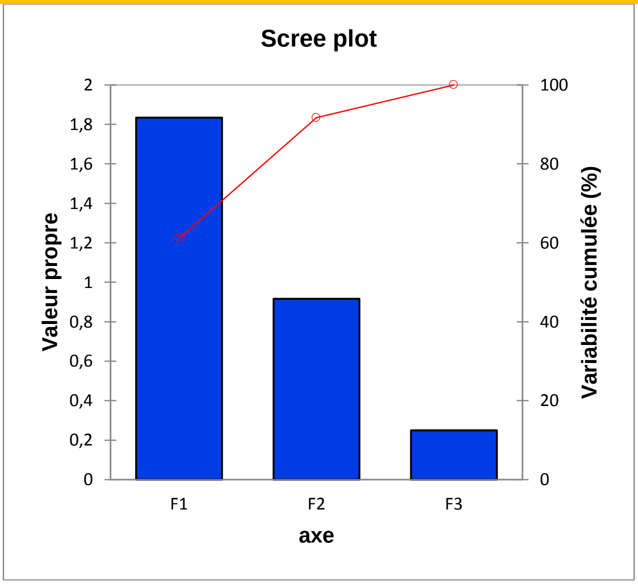
$$C_r(\Delta_1 \oplus \Delta_2/I_0)=\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\cdots+\lambda_p}=\frac{\lambda_1+\lambda_2}{Trac(V)}=\frac{1.83+0.92}{3}=91.66\%$$

وبناء عليه، يمكن القول: إنّ نسبة التمثيل الخاصة بالمركبة الرئيسية الأولى F_1 تساوي 61 في المائة من التباين الكلي، ونسبة التمثيل الخاصة بالمركبة الرئيسية الثانية تساوي 30.67 في المائة. فتكون نسبة التمثيل على المخطط العاملي في الفضاء R^2 المتكون من المحورين الأول والثاني، والممثلة بنسبة تباين كلي 91.66 في المائة، مما يعني إعطاء صورة واضحة لسحابة النقاط على المخطط العاملي.

9. تحديد عدد المحاور التي تأخذ في التحليل:

1. حسب معيار كايزر، نأخذ فقط العوامل التي تزيد قيمها الذاتية عن 1، وعليه نأخذ مركبة واحدة.

2. معيار التمثيل البياني:



حسب هذا المعيار نأخذ القيمتين الأولى والثانية.

3. حسب هذا المعيار نتوقف.

$$C_r(\Delta_1 \oplus \Delta_2/I_0)=\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\cdots+\lambda_p}=\frac{\lambda_1+\lambda_2}{Trac(V)}=\frac{1.83+0.92}{3}=91.66\%$$

4. حسب هذا المعيار لا داعي لإنشاء المستوى العاملي الثاني.

$$\frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i}=\frac{0.25}{3}*100=8.33$$

10. احداثيات الأفراد على المحاور:

$$F_\alpha= X_{CR}U_\alpha$$

حيث: F هي المركبة الرئيسية، X_{CR} المصفوفة المركزة، U_α الشعاع الذاتي المرافق للقيمة الذاتية. أو باستخدام العلاقة التالية:

$$F_\alpha=\sum_{j=1}^{p}\frac{x_{ij}}{s_j}u_{j\propto}$$

$$F=\begin{bmatrix}0.137 & 1.443 & 1.090\\ 1.169 & -0.587 & -0.957\\ 0.094 & -1.181 & 0.266\\ -1.030 & -1.082 & -1.117\\ 1.131 & 0.800 & 1.436\\ -1.504 & 0.601 & -0.718\end{bmatrix}*\begin{bmatrix}0.46 & 0.79 & 0.41\\ 0.56 & -0.61 & 0.56\\ 0.69 & -0.03 & -0.72\end{bmatrix}=\begin{bmatrix}1.623 & -0.805 & 0.079\\ -0.451 & 1.310 & 0.840\\ -0.435 & 0.787 & -0.814\\ -1.850 & -0.120 & -0.224\\ 1.959 & 0.362 & -0.122\\ -0.851 & -1.533 & 0.237\end{bmatrix}$$

$$\overline{F_1}=0,\ \overline{F_2}=0,\ \overline{F_3}=0$$

$$V(F_1)=\frac{1}{6}\sum_{i=1}^{6}F_1^2(i)=\frac{1}{6}((1.623)^2+(-0.451)^2+(-0.435)^2+(-1.850)^2+(1.959)^2+(-0.851)^2)=1.835=\lambda_1$$

لو نستخدم العلاقة السابقة نحصل على النتيجة نفسها:

$$F_{i\alpha}=\sum_{j=1}^{p}\frac{x_{ij}}{s_j}u_{j\propto}\ \cdots$$

$F_{11}=(\frac{x_{11}}{s_1}u_{11}+\frac{x_{12}}{s_2}u_{21}+\frac{x_{13}}{s_3}u_{31})=((0.137*0.46)+(1.443*0.56)+(1.090*0.69))=1.623$

λ	1.835	0.92	0.25
	F_1	F_2	F_3
جيجل	1.623	-0.805	0.079
سطيف	-0.451	1.310	0.840
خنشلة	-0.435	0.787	-0.812
تبسة	-1.850	-0.120	-0.224
عنابة	1.959	0.362	-0.122
البيض	0.851	-1.533	0.237

11. جودة تمثيل الأفراد على المحاور:

$$\cos^2\theta_{i\alpha}=\frac{F_\alpha^2(i)}{\sum_{\alpha=1}^{p}F_\alpha^2(i)}$$

$$\cos^2\theta_{41}=\frac{F_1^2(4)}{\sum_{\alpha=1}^{3}F_1^2(4)}=\frac{(-1.844)^2}{(-1.844)^2+(-0.119)^2+0.224^2}=0.982$$

وبعد اتمام جميع العمليات نجد:

	$\cos^2\theta_{i1}$	$\cos^2\theta_{i2}$	$\cos^2\theta_{i3}$
جيجل	0.802	0.197	0.999
سطيف	0.078	0.653	0.731
خنشلة	0.128	0.423	0.551
تبسة	0.982	0.004	0.986
عنابة	0.962	0.034	0.996
البيض	0.228	0.753	0.981

من النتائج أعلاه يمكن القول أنّ جيجل وتبسة وعنابة ممثلين بشكل جيد على المحور الأول، وغير ممثلين في المحور الثاني، بينما بقية الولايات تمثيلهم أفضل على المحور الثاني.

12. نسبة مساهمة الأفراد في تشكيل المحاور:

$$C_i^\alpha=\frac{F_\alpha^2(i)}{n\lambda_\alpha}$$

$C_1^1=\frac{F_1^2(1)}{6\lambda_1}=\frac{(1.629)^2}{6*1.835}=0.241$, $C_1^2=\frac{F_2^2(1)}{6\lambda_2}=\frac{(-0.804)^2}{6*0.92}=0.25$

وبنفس الطريقة حتى نحسب جميع القيم.

	C^1_i	C^2_i	C^3_i
جيجل	0.241	0.117	0.198
سطيف	0.019	0.312	0.116
خنشلة	0.017	0.113	0.049
تبسة	0.310	0.003	0.208
عنابة	0.347	0.024	0.240
البيض	0.065	0.430	0.187

يظهر من الجدول أن نسب مساهمة ولايات: جيجل تبسة وعنابة في تشكيل المحور الأول متقاربة، بينما لم يكن لسطيف وخنشلة والبيض مساهمة في تشكيل المحور الأول، وكان للبيض نسبة لا بأس بها في تشكيل المحور الثاني مع سطيف.

13. احدثيات المتغيرات النشطة: يتم الحصول على إحداثيات المتغيرات النشطة عن طريق الإسقاط المتعامد لأعمدة المصفوفة المركزة و/أو المعيارية على الاتجاهات المحددة بواسطة المتجهات الذاتية V_α،

$$G_\alpha = \sqrt{\lambda_\alpha}\,U_\alpha$$

$G_1=\sqrt{\lambda_1}\,U_1=\sqrt{1.835}\begin{bmatrix}0.46\\0.56\\0.69\end{bmatrix}=\begin{bmatrix}0.62\\0.76\\0.93\end{bmatrix}$, $G_2=\sqrt{\lambda_2}\,U_2=\sqrt{0.92}\begin{bmatrix}0.79\\-0.61\\-0.03\end{bmatrix}=\begin{bmatrix}0.76\\-0.59\\-0.03\end{bmatrix}$

$G_3=\sqrt{\lambda_3}\,U_3=\sqrt{0.25}\begin{bmatrix}0.41\\0.56\\-0.72\end{bmatrix}=\begin{bmatrix}0.21\\0.28\\-0.36\end{bmatrix}$

وتكون احداثيات المتغيرات النشطة على المحاور كما يلي:

	G_1	G_2	G_3
P	0.62	0.76	0.21
T-max	0.76	-0.59	0.28
T-min	0.93	-0.03	-0.36

14. حساب معاملات الارتباط الخطي بين المتغيرات الأصلية والمركبات الأساسية:

$$Cor = UD_{\sqrt{\lambda}}$$

$$Cor=\begin{bmatrix}0.46 & 0.79 & 0.41\\0.56 & -0.61 & 0.56\\0.69 & -0.03 & -0.72\end{bmatrix}*\begin{bmatrix}\sqrt{1.835} & 0 & 0\\0 & \sqrt{0.92} & 0\\0 & 0 & \sqrt{0.25}\end{bmatrix}=$$

$$\begin{bmatrix}0.623 & 0.758 & 0.205\\0.759 & -0.585 & 0.28\\0.935 & -0.029 & -0.360\end{bmatrix}$$

يمكننا حساب V_α:

$$V_1=\frac{1}{\lambda_1}P^{\frac{1}{2}}X_{CR}U_1=\frac{1}{\sqrt{1.835}}*\left(\frac{1}{6}\right)^{\frac{1}{2}}\begin{bmatrix}0.137 & 1.443 & 1.090\\ 1.169 & -0.587 & -0.957\\ 0.094 & -1.181 & 0.266\\ -1.030 & -1.082 & -1.117\\ 1.131 & 0.800 & 1.436\\ -1.504 & 0.601 & -0.718\end{bmatrix}*\begin{bmatrix}0.62\\0.76\\0.93\end{bmatrix}=\begin{bmatrix}0.49\\-0.14\\-0.13\\-0.56\\0.59\\-0.25\end{bmatrix}$$

$$V_2=\begin{bmatrix}-0.34\\0.56\\0.34\\-0.05\\0.16\\-0.66\end{bmatrix},\ V_3=\begin{bmatrix}0.06\\0.69\\-0.66\\-0.17\\-0.11\\0.20\end{bmatrix}\text{---------------------------}V=\begin{bmatrix}0.49 & -0.34 & 0.06\\ -0.14 & 0.56 & 0.69\\ -0.13 & 0.34 & -0.66\\ -0.56 & -0.05 & -0.17\\ 0.59 & 0.16 & -0.11\\ -0.25 & -0.66 & 0.20\end{bmatrix}$$

$$Cor= X_{CR}^{T}N^{\frac{1}{2}}V_\alpha$$

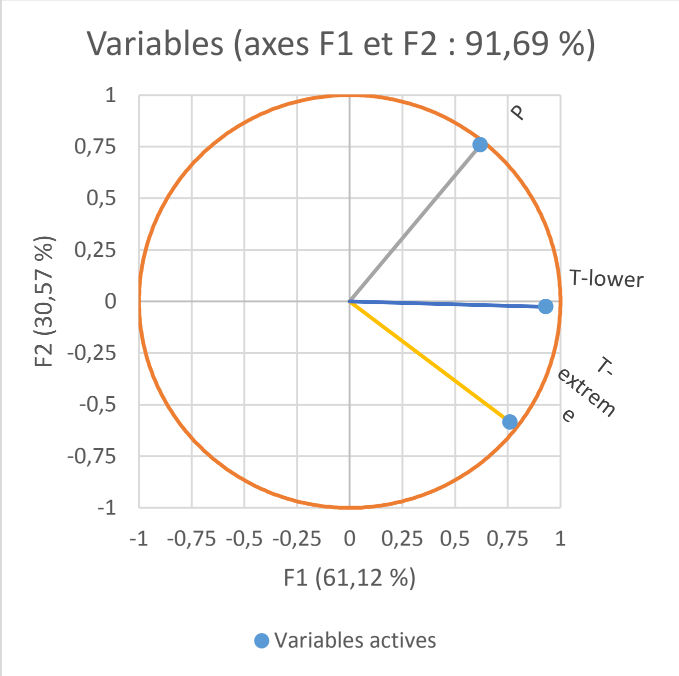
$$=\frac{1}{\sqrt{6}}\begin{bmatrix}0.137 & 1.169 & 0.094 & -1.030 & 1.131 & -1.504\\ 1.443 & -0.587 & -1.181 & -1.082 & 0.800 & 0.601\\ 1.090 & -0.957 & 0.266 & -1.117 & 1.436 & -0.718\end{bmatrix}*\begin{bmatrix}0.49 & -0.34 & 0.06\\ -0.14 & 0.56 & 0.69\\ -0.13 & 0.34 & -0.66\\ -0.56 & -0.05 & -0.17\\ 0.59 & 0.16 & -0.11\\ -0.25 & -0.66 & 0.20\end{bmatrix}=$$

$$\begin{bmatrix}0.617 & 0.761 & 0.205\\ 0.763 & -0.586 & 0.276\\ 0.933 & -0.023 & -0.360\end{bmatrix}$$

وتكون المركبات الأساسية للمتغيرات على المحاور كما يلي:

	G_1	G_2	G_3
P	0.62	0.76	0.21
T-max	0.76	-0.59	0.28
T-min	0.93	-0.03	-0.36

ونرسم دائرة الارتباط بتحديد احداثية كل متغير على المخطط العاملي، ثم نرسم دائرة انطلاقا من مركز الثقل لسحابة النقط نصف قطرها يساوي 1.



يمكن ملاحظة أنّ هناك علاقة قوية بين درجة الحرارة الدنيا والمركبة الأساسية الأولى، كما أنّ المتغيرات الثلاثة لهم علاقة موجبة مع المحور الأول. وعلاقة سالبة بين كل من درجة الحرارة القصوى والدنيا مع المحور الثاني.

15. المركبات الرئيسية للمتغيرات على المحاور:

$$G_\alpha = X_{CR}^T V_\alpha = \frac{1}{\lambda_\alpha} X_{CR}^T P^{\frac{1}{2}} X_{CR} U_\alpha$$

$$G_\alpha = X_{CR}^T V_\alpha =$$

$$\left[\begin{bmatrix} 0.137 & 1.169 & 0.094 & -1.030 & 1.131 & -1.504 \\ 1.443 & -0.587 & -1.181 & -1.082 & 0.800 & 0.601 \\ 1.090 & -0.957 & 0.266 & -1.117 & 1.436 & -0.718 \end{bmatrix}\right] * \begin{bmatrix} 0.49 & -0.34 & 0.06 \\ -0.14 & 0.56 & 0.69 \\ -0.13 & 0.34 & -0.66 \\ -0.56 & -0.05 & -0.17 \\ 0.59 & 0.16 & -0.11 \\ -0.25 & -0.66 & 0.20 \end{bmatrix} =$$

$$\begin{bmatrix} 1.515 & 1.865 & 0.503 \\ 1.870 & -1.435 & 0.677 \\ 2.286 & -0.057 & -0.882 \end{bmatrix}$$

16. جودة تمثيل المتغيرات على لمحاور:

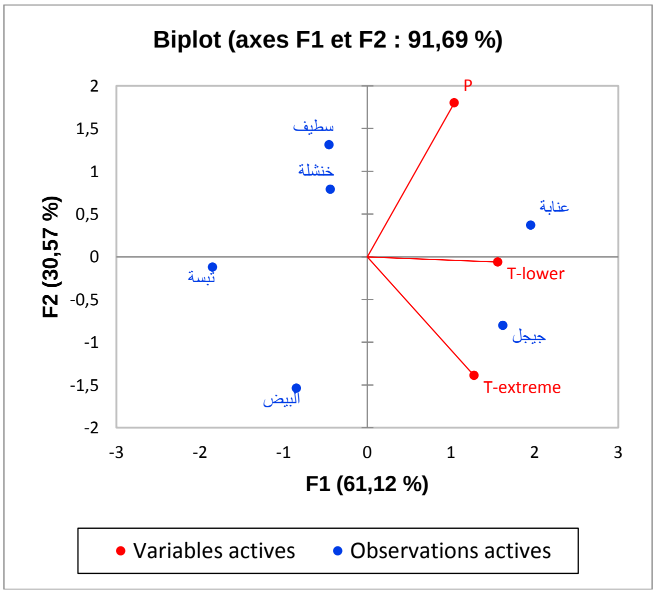
$$Cos^2(j,\propto) = \frac{G_{j\propto}^2}{d^2(j,0)}$$

	F1	F2	F3
P	0,385	**0,574**	0,041
T-extreme	**0,580**	0,342	0,077
T-lower	**0,868**	0,001	0,131

17. نسبة مساهمة المتغيرات في تشكيل المحاور:

	F1	F2	F3
P	20,992	62,616	16,392
T-extreme	31,658	37,309	31,033
T-lower	47,350	0,075	52,574

18. تمثيل الأفراد والمتغيرات على المخطط العاملي: في هذا المخطط يتم تمثيل احداثيات كل من الأفراد والمتغيرات على نفس المخطط.



لتفسير هذا المخطط، نتبع القواعد التالية:

- تقارب نقطتي فردين تعني التشابه.
- تقارب نقطتي متغيرين تعني الترابط.
- تقارب نقطة فرد مع نقطة متغير تعني أنّ المتغير يأثر في سلوك ذلك الفرد.

ومن خلال المخطط أعلاه نستنتج ما يلي:

- يرتبط متغير التساقط مع المحور الثاني، ويرتبط متغير درجة الحرارة الدنيا مع المحور الأول ايجابيا، فيما يرتبط متغير درجة الحرارة القصوى سلبيا مع المحور الثاني.
- المدن (عنابة، جيجل) التي تتواجد على يمين المخطط العاملي لها احداثيات عالية مع المحور الأول، ستكون لهم قيم مرتفعة بالنسبة لدرجات الحرارة الدنيا، لأنّ متغير درجات الحرارة الدنيا يرتبط بقوة مع المحور الأول.
- المدن التي تقع يسار المخطط، لها احداثيات ضعيفة مع المحور الأول، وبالتالي ستكون لها قيم منخفضة بالنسبة لدرجات الحرارة الدنيا.
- المدن التي تقع أعلى المخطط نسبة تساقط الأمطار فيها مرتفعة (سطيف، خنشلة، عنابة، جيجل)، والمدن التي تقع أسفل المخطط نسبة التساقط فيها ضعيفة (تبسة، البيض).

– سطيف وخنشلة تتشابهان في نسب التساقط ودرجات الحرارة الدنيا والقصوى، بينما مدينة البيض أبرد من مدينة تبسة، ونسبة تساقط الأمطار في تبسة أعلى من البيض.

التمرين الخامس: يحتوي الجدول الموالي على قياسات ضغط الدم الانقباضي وضغط الدم الانبساطي لستة أفراد.

	SBP	DBP
عادل	126	78
أمين	128	80
جود	128	82
أفنان	130	82
رشيد	130	84
وسيم	132	86

المطلوب: إجراء التحليل بالمركبات الرئيسية متبعا جميع الخطوات.

الحل:

1. حساب المتوسطات الحسابية والانحرافات المعيارية لجدول البيانات الكمية (المصفوفة X):

$$X=\begin{bmatrix}126 & 78\\128 & 80\\128 & 82\\130 & 82\\130 & 84\\132 & 86\end{bmatrix}$$

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_{ij}$$

$$\overline{X_{SBP}}=\frac{1}{6}(126+128+128+130+130+132)=129$$

$$\overline{X_{\text{DBP}}}=\frac{1}{6}(78+80+82+82+84+86)=82$$

$$\delta_j=\sqrt{\frac{\sum_{j=1}^{p}(X_J-\overline{X_J})^2}{n}}$$

$$\delta_1=\sqrt{\frac{\sum_{j=1}^{6}(126-129)^2+(128-129)^2+(128-129)^2+(130-129)^2+(130-129)^2+(132-129)^2}{6}}=2.098$$

$$\delta_2=\sqrt{\frac{\sum_{j=1}^{6}(78-82)^2+(80-82)^2+(82-82)^2+(82-82)^2+(84-82)^2+(86-82)^2}{6}}=2.828$$

2. يمكننا استنتاج مركز ثقل سحابة النقط لجدول البيانات الكمية.

$$G_X=\begin{bmatrix}\overline{X_1}\\\overline{X_2}\end{bmatrix}=\begin{bmatrix}129\\82\end{bmatrix}$$

ويكون الجدول على النحو الآتي:

	SBP	DBP
عادل	126	78
أمين	128	80
جود	128	82
أفنان	130	82
رشيد	130	84
وسيم	132	86
G_X	129	82
δ_j	2.098	2.828

3. حساب جدول البيانات الكمية المركز والمعياري:

1.3. جدول البيانات الكمية المركز:

$$\widetilde{X_{ij}} = X_{ij} - \overline{X_j}$$

	SBP	DBP
عادل	-3	-4
أمين	-1	-2
جود	-1	0
أفنان	1	0
رشيد	1	2
وسيم	3	4

ويمكن كذلك حساب المصفوفة المركزة من الصيغة المصفوفية التالية:

$$X_C = X - 1_n G_X^T$$

$$1_n = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad G_X^T = [129 \quad 82]$$

$$X_C = \begin{bmatrix} 126 & 78 \\ 128 & 80 \\ 128 & 82 \\ 130 & 82 \\ 130 & 84 \\ 132 & 86 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * [129 \quad 82] = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

2.3. جدول البيانات المركز والمعياري:

$$z = \frac{x_{ij} - \overline{x_j}}{\sigma_j}$$

	SBP	DBP
عادل	-1.428	-1.412
أمين	-0.476	-0.706
جود	-0.476	0
أفنان	0.476	0
رشيد	0.476	0.706
وسيم	1.428	1.412

ويمكننا حساب المصفوفة المركز المعيارية من الصيغة الآتية:

$$X_{CR}=X_C * D_{\frac{1}{S}}$$

$$D_{\frac{1}{S}}=\begin{bmatrix}\frac{1}{2.098} & 0\\ 0 & \frac{1}{2.828}\end{bmatrix}=\begin{bmatrix}0.476 & 0\\ 0 & 0.353\end{bmatrix}$$

$$X_{CR}=\begin{bmatrix}-3 & -4\\ -1 & -2\\ -1 & 0\\ 1 & 0\\ 1 & 2\\ 3 & 4\end{bmatrix} * \begin{bmatrix}0.476 & 0\\ 0 & 0.353\end{bmatrix}=\begin{bmatrix}-1.428 & -1.412\\ -0.476 & -0.706\\ -0.476 & 0\\ 0.476 & 0\\ 0.476 & 0.706\\ 1.428 & 1.412\end{bmatrix}$$

4. تحديد طريقة المركبات الرئيسية المستخدمة في التحليل: المتغيرين متجانسين ولهما نفس وحدة القياس، لذلك يمكننا استخدام طريقة المركبات الرئيسية البسيطة.

5. حساب مصفوفة التباين والتباين المشترك:

نحسب تباين المتغير من المعادلة الموالية:

$$V(X,X)=\frac{1}{N}\sum_{k=1}^{N}\left(x_{ij}-\bar{x}_j\right)^2$$

وتباين المتغيرين X1 و X2 من المعادلة التالية:

$$cov(X,Y)=\frac{1}{N}\sum_{k=1}^{N}(Xij-\bar{X}_j)*(Yij-\bar{Y}_j)$$

$V_{11}=\frac{1}{6}((126-129)^2+(128-129)^2+(128-129)^2+(130-129)^2+(130-129)^2+(132-129)^2)=3.667$

$V_{22}=\frac{1}{6}((78-82)^2+(80-82)^2+(82-82)^2+(82-82)^2+(84-82)^2+(86-82)^2)=6.667$

$cov(SBP,DBP)=cov(DBP,SBP)=\frac{1}{6}(12+2+0+0+2+12)=4.667$

$$V=\begin{bmatrix}3.667 & 4.667\\ 4.667 & 6.667\end{bmatrix}$$

ويمكننا كذلك حساب مصفوفة التباين والتباين المشترك من الصيغة المصفوفية التالية:

$$V=\frac{1}{n}X_c^TX_c$$

$$V=\frac{1}{6}\begin{bmatrix}-3 & -1 & -1 & 1 & 1 & 3\\ -4 & -2 & 0 & 0 & 2 & 4\end{bmatrix}*\begin{bmatrix}-3 & -4\\ -1 & -2\\ -1 & 0\\ 1 & 0\\ 1 & 2\\ 3 & 4\end{bmatrix}=\begin{bmatrix}3.667 & 4.667\\ 4.667 & 6.667\end{bmatrix}$$

6. حساب القيم والمتجهات الذاتية لمصفوفة التباين والتباين المشترك:

1.6. حساب القيم الذاتية باستخدام المحدد:

$$\det(V-\lambda I)=0$$

$$\det\left(\begin{bmatrix}3.667 & 4.667\\ 4.667 & 6.667\end{bmatrix}-\lambda\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)=0 \rightarrow \det\left(\begin{bmatrix}3.667-\lambda & 4.667\\ 4.667 & 6.667-\lambda\end{bmatrix}\right)=0$$

$$((3.667-\lambda)(6.667-\lambda))-((4.667)(4.667))=24.447-3.667\lambda-6.667\lambda+\lambda^2-21.78=0$$

$$\lambda^2-10.334\lambda+2.667=0$$

$$\Delta=b^2-4ac=(-10.334)^2-4*1*2.667=96.123$$

وبما أنّ قيمة المميز أكبر من الصفر، فهي تقبل حلان.

$$\lambda=\frac{-b\pm\sqrt{\Delta}}{2a}$$

$$\lambda=\frac{-(-10.334)+\sqrt{96.123}}{2*1}=10.068,\ \lambda=\frac{-(-10.334)-\sqrt{96.123}}{2*1}=0.265$$

ومنه:

$$\lambda_1=10.068,\ \lambda_2=0.265$$

ونكتب مصفوفة القيم الذاتية D على النحو التالي:

$$D=\begin{bmatrix}10.068 & 0\\ 0 & 0.265\end{bmatrix}$$

F2	F1	
0,265	10.068	القيمة الذاتية
2,563	97,437	النسبة(%)
100,000	97,437	الصاعد%

يظهر من الجدول أعلاه أنّ نسبة التباين المفسر بواسطة المركبتين الرئيسيتين الأولى والثانية يساوي 100.

2.6. حساب المتجهات الذاتية المرافقة:

$$(V-\lambda I)U=0$$

$$\lambda=10.068$$

$$\left(\begin{bmatrix}3.667 & 4.667\\4.667 & 6.667\end{bmatrix}-\lambda\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)U=0 \rightarrow \begin{bmatrix}3.667-10.068 & 4.667\\4.667 & 6.667-10.068\end{bmatrix}U=0$$

$$\begin{bmatrix}-6.402 & 4.667\\4.667 & -3.402\end{bmatrix}*\begin{bmatrix}U1\\U2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

إذاً لدينا نظام متجانس من المعادلات الخطية، نقوم بحلها بواسطة حذف كاوس:

$$\begin{bmatrix}-6.402 & 4.667\\4.667 & -3.402\end{bmatrix}\begin{vmatrix}0\\0\end{vmatrix}\ \boxed{*-0.156}\ \begin{bmatrix}1 & -0.729\\4.667 & -3.402\end{bmatrix}\begin{vmatrix}0\\0\end{vmatrix}\ \boxed{\begin{matrix}*-4.667\\+\end{matrix}}$$

$$\begin{bmatrix}1 & -0.729\\0 & 0\end{bmatrix}\begin{vmatrix}0\\0\end{vmatrix}$$

$U1-0.729U2=0 \dashrightarrow U1=0.729U2$

$U2=U2$

$$U=\begin{bmatrix}0.729U2\\U2=U2\end{bmatrix} \longleftrightarrow U=U_2\begin{vmatrix}0.729\\1\end{vmatrix}\ldots\ldots\ U_2=1\rightarrow U=\begin{matrix}0.729\\1\end{matrix}$$

شعاع الوحدة المرافق للقيمة الذاتية الأولى:

$$\|U_2\|=\sqrt{(0.729)^2*(1)^2}=1.237$$

$$U=\begin{vmatrix}\frac{0.729}{1.237}\\\frac{1}{1.237}\end{vmatrix}=\begin{vmatrix}0.589\\0.808\end{vmatrix}$$

$\lambda=0.265$

$$\begin{bmatrix}3.402 & 4.667\\4.667 & 6.402\end{bmatrix}*\begin{bmatrix}U1\\U2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

إذاً لدينا نظام متجانس من المعادلات الخطية، نقوم بحلها بواسطة حذف كاوس:

$$\begin{bmatrix}3.402 & 4.667\\4.667 & 6.402\end{bmatrix}\begin{vmatrix}0\\0\end{vmatrix}\ \boxed{*0.294}\ \begin{bmatrix}1 & 1.372\\4.667 & 6.402\end{bmatrix}\begin{vmatrix}0\\0\end{vmatrix}\ \boxed{\begin{matrix}*-4.667\\+\end{matrix}}$$

$$\begin{bmatrix}1 & 1.372\\0 & 0\end{bmatrix}\begin{vmatrix}0\\0\end{vmatrix}$$

$U1+1.372U2=0 \dashrightarrow U1=-1.372U2$

$U2=U2$

$$U=\begin{bmatrix}-1.372U2\\U2=U2\end{bmatrix} \longleftrightarrow U=U_2\begin{vmatrix}-1.372\\1\end{vmatrix}\ldots\ldots\ U_2=1\rightarrow U=\begin{vmatrix}-1.372\\1\end{vmatrix}$$

شعاع الوحدة المرافق للقيمة الذاتية الثانية:

$$\|U\|=\sqrt{(-1.372)^2*(1)^2}=1.697$$

$$U=\begin{vmatrix}\frac{-1.372}{1.697}\\\frac{1}{1.697}\end{vmatrix}=\begin{vmatrix}-0.808\\0.589\end{vmatrix}$$

ونكتب مصفوفة الأشعة الذاتية:

$$U=\begin{bmatrix}0.589 & -0.808\\0.808 & 0.589\end{bmatrix}$$

7. حساب احداثيات الأفراد على المحاور:

$$F_\alpha = X_C U_\alpha$$

$$F_\alpha = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} * \begin{bmatrix} 0.589 & -0.808 \\ 0.808 & 0.589 \end{bmatrix} = \begin{bmatrix} -5 & -068 \\ -2.205 & 0.370 \\ -0.589 & -0.808 \\ 0.589 & 0.808 \\ 2.205 & -0.370 \\ 5 & 0.068 \end{bmatrix}$$

	F1	F2
عادل	-5,000	-0,068
أمين	-2,205	0,370
جود	-0,589	-0,808
أفنان	0,589	0,808
رشيد	2,205	-0,370
وسيم	5,000	0,068

8. نسبة مساهمة الأفراد في تشكيل المحاور:

$$C_i^\alpha = \frac{F_\alpha^2(i)}{n\lambda_\alpha}$$

$$C^1_1 = \frac{F_1^2(1)}{6\lambda_1} = \frac{(-5)^2}{6*10.068} = 0.41376$$

وبنفس الطريقة حتى نحسب جميع القيم.

	F1	F2
عادل	41,376	0,291
أمين	8,050	8,617
جود	0,574	41,092
أفنان	0,574	41,092
رشيد	8,050	8,617
وسيم	41,376	0,291

عادل ووسيم يساهمان بالنسبة نفسها في تشكيل المحور الأول، ويساهمان معا بنسبة 82.752%. في حين يساهم كل من جود وأفنان في تشكيل المحور الثاني بنفس النسبة، كما أنهما معا يساهمان بنسبة 82.184% في تشكيل المحور الثاني.

9. جودة تمثيل الأفراد على المحاور:

$$\cos^2 \theta_{i\alpha} = \frac{F_\alpha^2(i)}{\sum_{\alpha=1}^{p} F_\alpha^2(i)}$$

$$\text{Cos}^2\ \theta_{11} = \frac{(-5)^2}{(-5)^2 + (-0.068)^2} = 1$$

	F1	F2
عادل	**1,000**	0,000
أمين	**0,973**	0,027
جود	0,347	**0,653**
أفنان	0,347	**0,653**
رشيد	**0,973**	0,027
وسيم	**1,000**	0,000

عادل، أمين، رشيد، وسيم، ممثلين على المحور الأول أحسن تمثيل. بينما تمثيل جود وأفنان كان على المحور الثاني بشكل جيد.

10. احداثيات المتغيرات على المحاور:

$$G_\alpha=\sqrt{\lambda_\alpha}U_\alpha$$

$G_1=\sqrt{\lambda_1}U_1=\sqrt{10.068}\begin{vmatrix}0.589\\0.808\end{vmatrix}=\begin{vmatrix}1.869\\2.564\end{vmatrix}$, $G_2=\sqrt{\lambda_2}U_2=\sqrt{0.265}\begin{vmatrix}-0.808\\0.589\end{vmatrix}=\begin{vmatrix}-0.416\\0.303\end{vmatrix}$

	F1	F2
SBP	1,869	-0,416
DBP	2,564	0,303

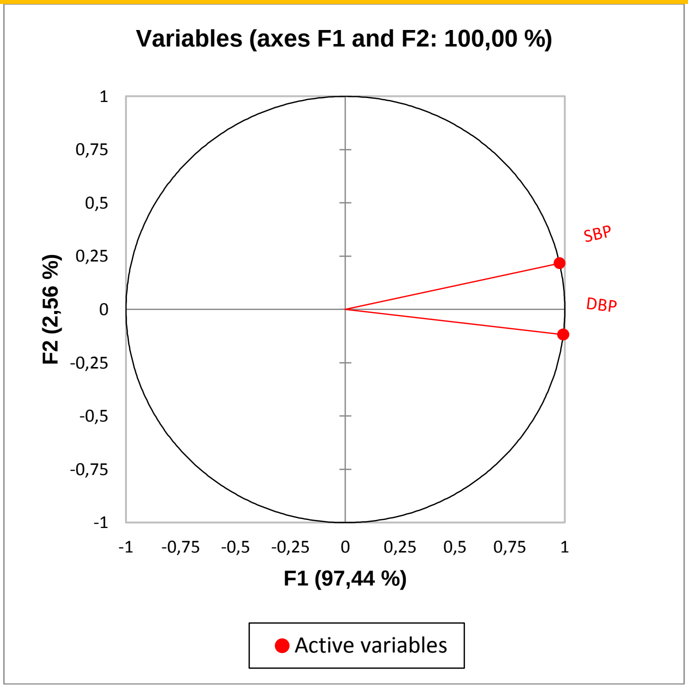
11. الارتباط بين المتغيرات الأصلية والمركبات الرئيسية:

$$Cor= D_{\sqrt{\lambda}} * U^T * D_{\frac{1}{S}}$$

$$D_{\sqrt{\lambda}}=\begin{bmatrix}\sqrt{10.068} & 0\\ 0 & \sqrt{0.265}\end{bmatrix}=\begin{bmatrix}3.173 & 0\\ 0 & 0.514\end{bmatrix}$$

$$Cor=\begin{bmatrix}3.173 & 0\\ 0 & 0.514\end{bmatrix}*\begin{bmatrix}0.589 & 0.808\\ -0.808 & 0.589\end{bmatrix}*\begin{bmatrix}0.476 & 0\\ 0 & 0.353\end{bmatrix}=\begin{matrix}0.890 & 0.905\\ -0.198 & 0.107\end{matrix}$$

	F1	F2
SBP	0.890	0.905
DBP	-0.198	0.107


Variables (axes F1 and F2: 100,00 %)
F2 (2,56 %)
F1 (97,44 %)
1
0,75
0,5
0,25
0
-0,25
-0,5
-0,75
-1
-1
-0,75
-0,5
-0,25
0
0,25
0,5
0,75
1
SBP
DBP
Active variables

التطبيق باستخدام XLSTAT

تقديم البرنامج:

ظهر برنامج XLSTAT سنة 1995، ويعد من أفضل البرامج التي تتيح تحليل البيانات عبر العديد من الحلول الإحصائية، وهو برنامج متكامل وموثوق وسهل الاستخدام لأي شخص. كما أنّه متوافق مع جميع إصدارات Microsoft Excel، بدءًا من Excel 97 حتى Excel 2016. ويأتي مع واجهة بعدة لغات (للأسف لا يتوفر على اللغة العربية، ربما بسبب عدم استخدامه في البيئة العربية)، مع توفّر نسخة تجريبية بكامل الصلاحيات لمدة شهر، يمكن تنزيلها من موقع البرنامج.

www.xlstat.com

طبعا البرنامج غير مجاني، ومكلف نوعا ما بالنسبة للطلبة والباحثين بصفة عامة، وقد اجتهدت في توفير نسخة مجانية تعمل بكل الصلاحيات، يمكن تنزيلها من قناتي على اليوتوب، مع شرح كيفية التثبيت.

ولقد تطورت بنية البرنامج بشكل كبير على مدى السنوات الخمس الماضية من أجل مراعاة التطورات في Microsoft Excel وقضايا التوافق بين الأنظمة الأساسية. يعتمد البرنامج على تطبيق Visual Basic للواجهة، وعلى C ++ للحسابات الرياضية والإحصائية.

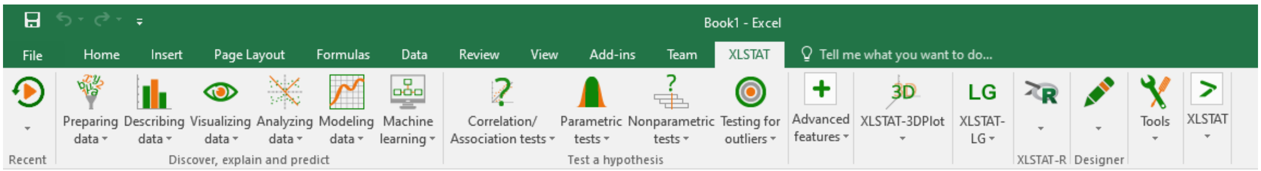
والملفت للانتاه في البرنامج أنّه يحتوي على الغالبية العظمى من الاختبارات الاحصائية، كما أنّه يتفوق على كل البرامج الأخرى من الناحية الرسومية والاختيارات المتاحة أثناء تنفيذ بعض الاختبارات، كما سنرى ذلك في هذا التطبيق.

مثال: يحتوي الجدول الموالي على بيانات عن عشرة (10) منتجات، تشمل سعر المنتج وتقييم جودته من خلال بعض المواصفات، كالذوق، حجم المنتج، مدى توفّره في السوق، جودة التغليف، وطريقة عرضه (صورة المنتج). ويريد مدير التسويق توزيع هذه المنتجات في السوق على حسب السعر والمواصفات.

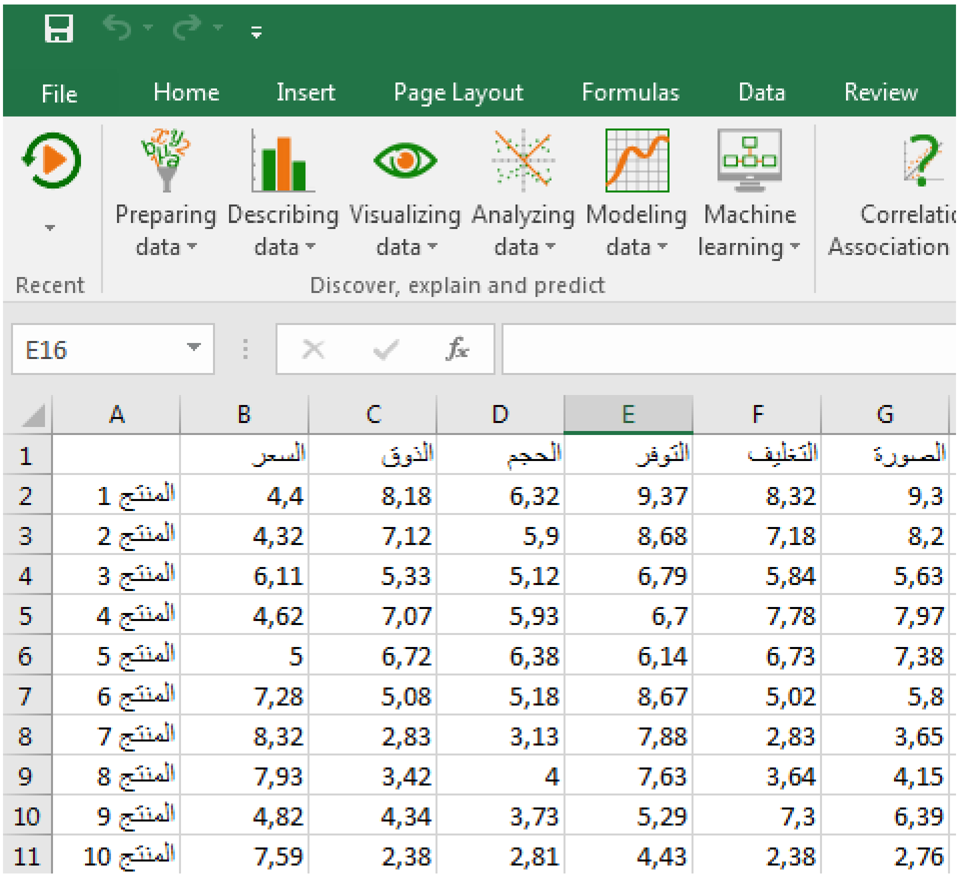
	السعر	الذوق	الحجم	التوفر	التغليف	الصورة
المنتج 1	4,4	8,18	6,32	9,37	8,32	9,3
المنتج 2	4,32	7,12	5,9	8,68	7,18	8,2
المنتج 3	6,11	5,33	5,12	6,79	5,84	5,63
المنتج 4	4,62	7,07	5,93	6,7	7,78	7,97
المنتج 5	5	6,72	6,38	6,14	6,73	7,38
المنتج 6	7,28	5,08	5,18	8,67	5,02	5,8
المنتج 7	8,32	2,83	3,13	7,88	2,83	3,65
المنتج 8	7,93	3,42	4	7,63	3,64	4,15
المنتج 9	4,82	4,34	3,73	5,29	7,3	6,39
المنتج 10	7,59	2,38	2,81	4,43	2,38	2,76

ولتنفيذ التحليل في برنامج XLSTAT نتبع الخطوات الآتية:

في الخطوة الأولى وبعد تنصيب XLSTAT على برنامج Excel يمكنك فتحه مباشرة من الأيقونة التي ستظهر على سطح المكتب لديك، أو من خلال برنامج Excel.

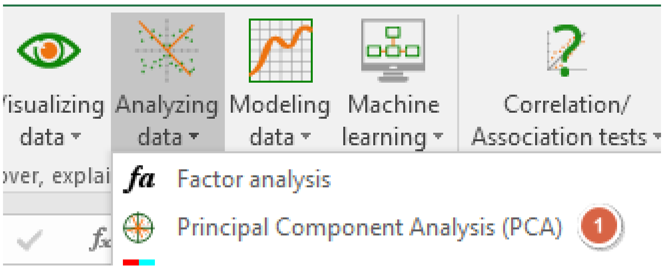


في الخطوة الثانية ننقل الجدول إلى برنامج Excel.

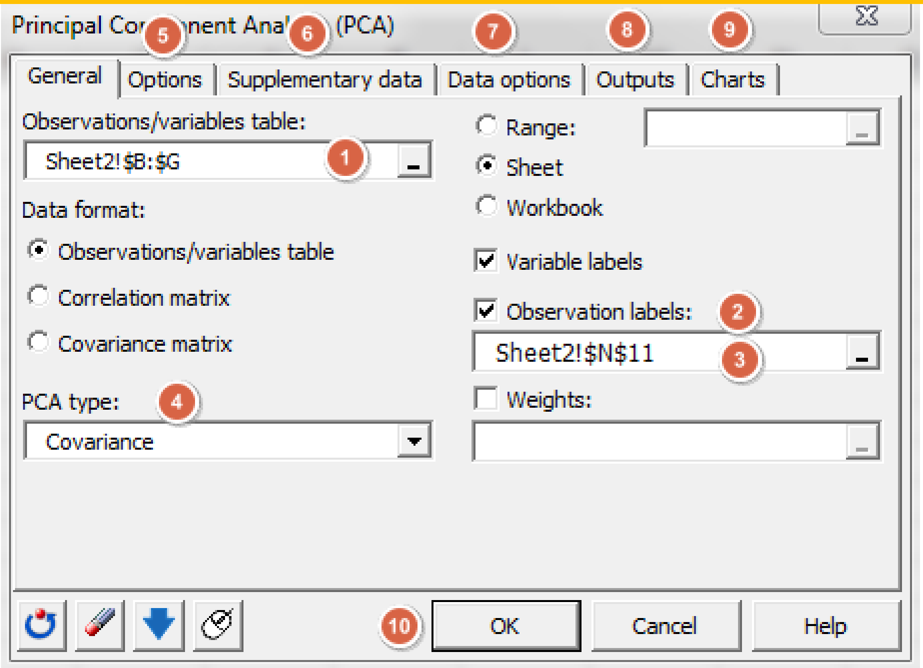


	A	B	C	D	E	F	G
1		السعر	الذوق	الحجم	التوفر	التغليف	الصورة
2	المنتج 1	4,4	8,18	6,32	9,37	8,32	9,3
3	المنتج 2	4,32	7,12	5,9	8,68	7,18	8,2
4	المنتج 3	6,11	5,33	5,12	6,79	5,84	5,63
5	المنتج 4	4,62	7,07	5,93	6,7	7,78	7,97
6	المنتج 5	5	6,72	6,38	6,14	6,73	7,38
7	المنتج 6	7,28	5,08	5,18	8,67	5,02	5,8
8	المنتج 7	8,32	2,83	3,13	7,88	2,83	3,65
9	المنتج 8	7,93	3,42	4	7,63	3,64	4,15
10	المنتج 9	4,82	4,34	3,73	5,29	7,3	6,39
11	المنتج 10	7,59	2,38	2,81	4,43	2,38	2,76

نضغط على Analyzing data.

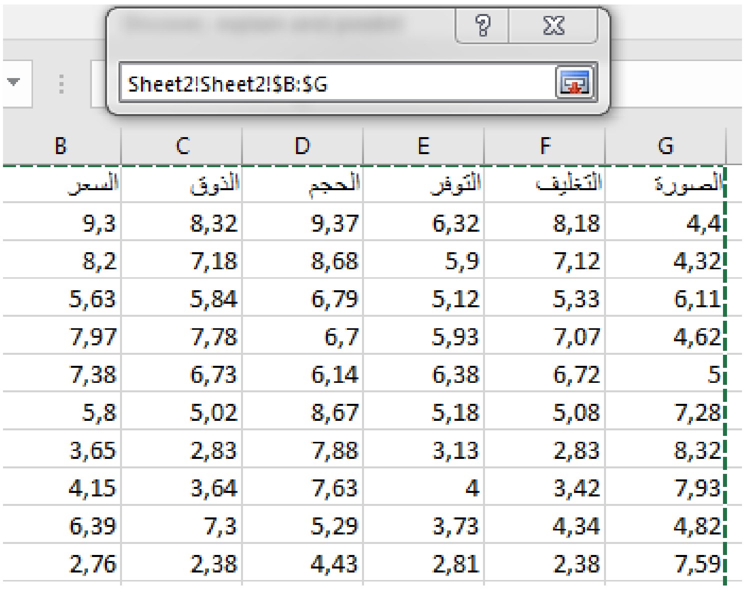


من القائمة المنسدلة نختار PCA، فيظهر صندوق الحوار التالي:



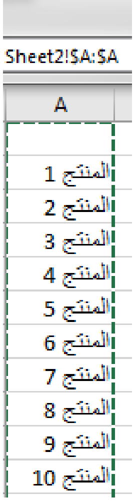
1. نحدد أعمدة المتغيرات بالضغط على العلامة – ثم تحديد كل المتغيرات.

تحت الخيار Data formay (شكل البيانات)، يمكنك تحديد إن كانت البيانات عبارة عند جدول بيانات، أو مصفوفة الارتباط او مصفوفة التباين المشترك.

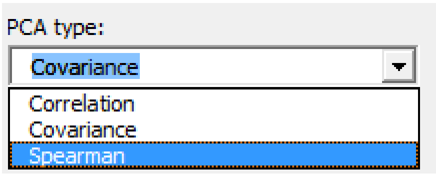


B	C	D	E	F	G
السعر	الذوق	الحجم	التوفر	التغليف	الصورة
9,3	8,32	9,37	6,32	8,18	4,4
8,2	7,18	8,68	5,9	7,12	4,32
5,63	5,84	6,79	5,12	5,33	6,11
7,97	7,78	6,7	5,93	7,07	4,62
7,38	6,73	6,14	6,38	6,72	5
5,8	5,02	8,67	5,18	5,08	7,28
3,65	2,83	7,88	3,13	2,83	8,32
4,15	3,64	7,63	4	3,42	7,93
6,39	7,3	5,29	3,73	4,34	4,82
2,76	2,38	4,43	2,81	2,38	7,59

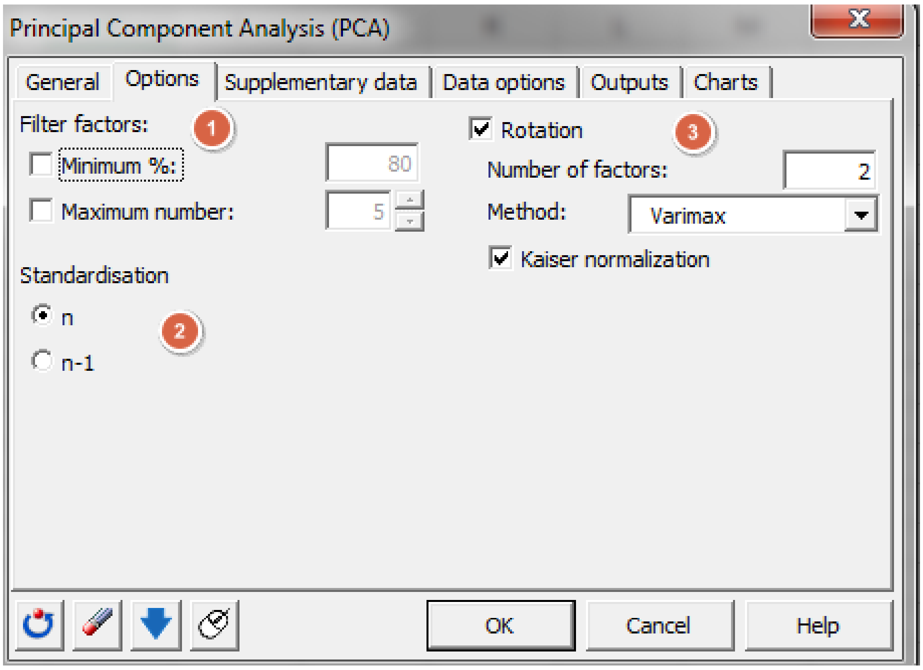
2. نعود لمربع الحوار الرئيسي ونضغط على علامة – لتحديد عمود تسميات الأفراد، في مثالنا النتجات.



4. من PCA type نختار طريقة المركبت الرئيسية المستخدمة (البسيطة، المرجحة)، يستخدم البرنامج بشكل آلي مصفوفة الإرتباط، يمكنك اختيار استخدام مصفوفة التباين والتباين المشترك أو مصفوفة الإرتباط، أو سبيرمان.



5. نضغط Options:

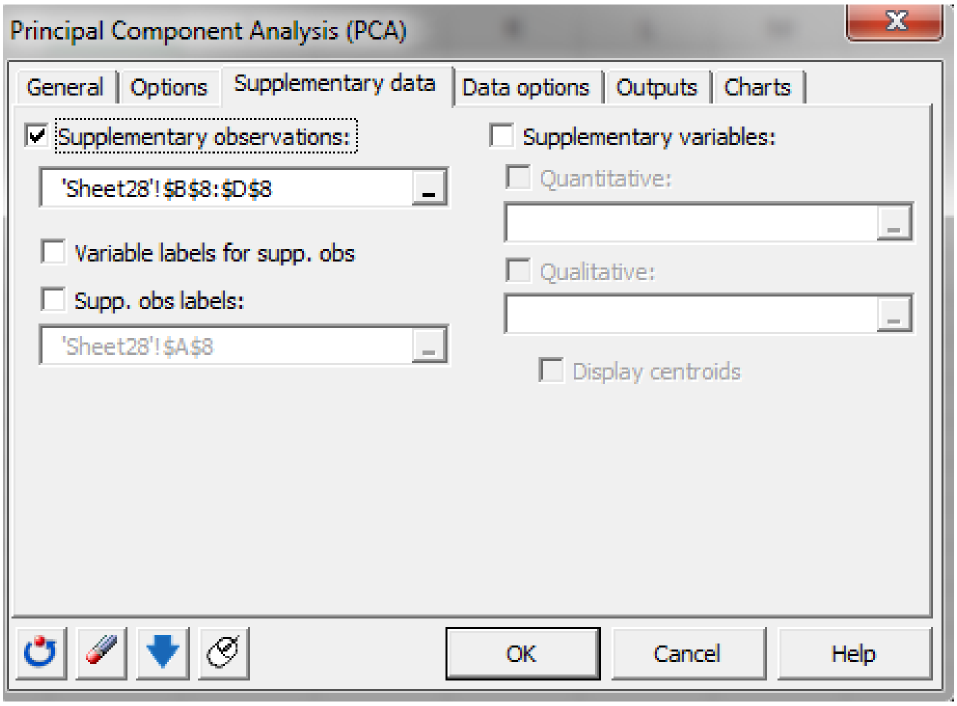


1.5. تصفية العوامل: يمكنك تنشيط أحد الخيارين التاليين لتقليل عدد العوامل التي يتم عرض نتائجها:
– الحد الأدنى: قم بتنشيط هذا الخيار ثم أدخل الحد الأدنى للتباين الكلي الذي يجب أن تفسّره العوامل المختارة .– العدد الأقصى: قم بتنشيط هذا الخيار لتعيين عدد العوامل التي يجب مراعاتها.

2.5. التوحيد القياسي: إذا كان تنسيق بياناتك هو "الملاحظات/المتغيرات"، يمكنك اختيار كيفية حساب الارتباط (أو التباين): بالمقام (n) أو $(n-1)$.

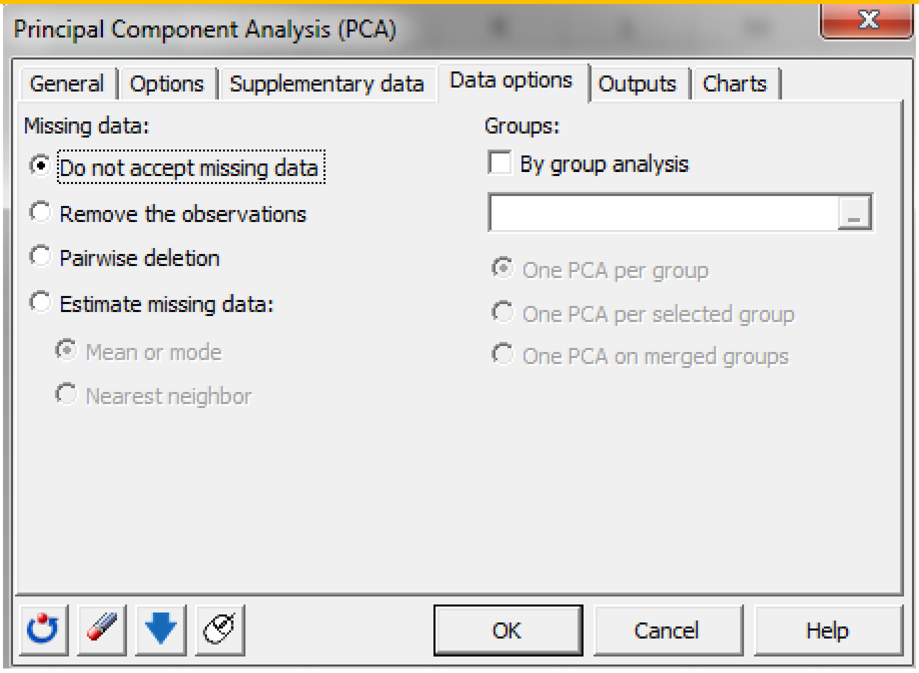
3.5. التدوير: أدخل عدد العوامل التي سيتم تطبيق التدوير عليها، ثم اختر طريقة التدوير المراد تطبيقها، وحدد Kaiser Normalization لتطبيق تسوية كايزر أثناء حساب الدوران.

6. الأفراد والمتغيرات المضافة (Supplementary observations):



قم بتنشيط هذا الخيار إذا كنت تريد حساب إحداثيات الأفراد والمتغيرات المضافة. لا تؤخذ المتغيرات والأفراد المضافة في الاعتبار عند حساب مصفوفة الارتباط وللحسابات اللاحقة.

7. خيارات البيانات (Data options).



البيانات مفقودة (Missing data): عدم قبول البيانات المفقودة: قم بتنشيط هذا الخيار بحيث لا يتابع XLSTAT العمليات الحسابية إذا تم الكشف عن القيم المفقودة.

إزالة المشاهدات (Remove the observations): قم بتنشيط هذا الخيار لإزالة المشاهدات التي تحتوي على بيانات مفقودة.

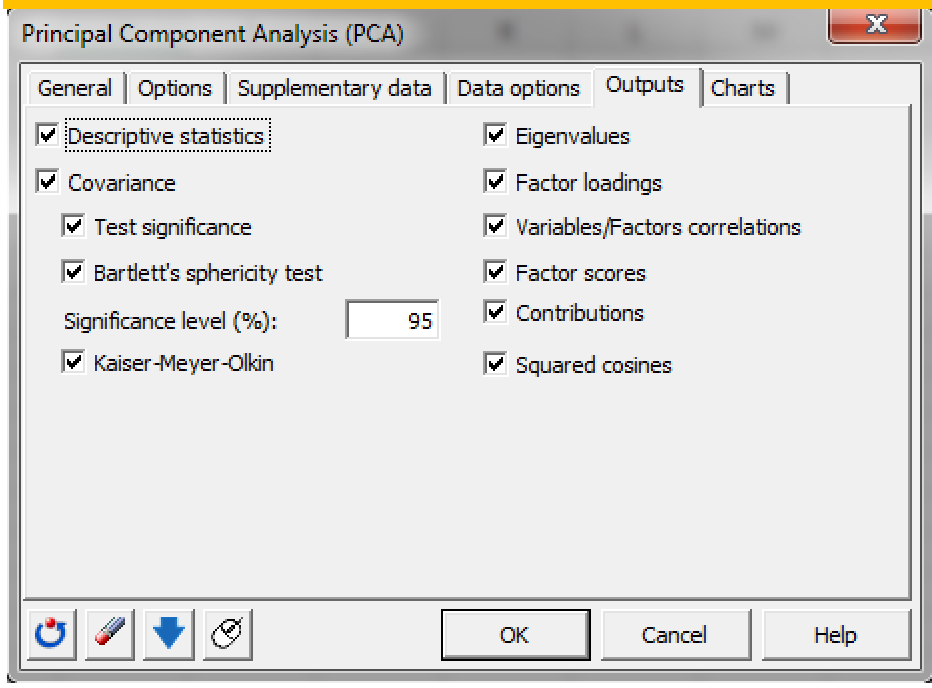
الحذف الزوجي (Pairwise deletion): قم بتنشيط هذا الخيار لإزالة الملاحظات التي تحتوي على بيانات مفقودة فقط عندما تكون المتغيرات المضمنة في العمليات الحسابية بها بيانات مفقودة .على سبيل المثال، عند حساب الارتباط بين متغيرين، لن يتم تجاهل الملاحظة إلا إذا كانت البيانات المقابلة لأحد المتغيرين مفقودة.

تقدير البيانات المفقودة (Estimate missing data): قم بتنشيط هذا الخيار لتقدير البيانات المفقودة قبل بدء الحساب.

- المتوسط أو الوسيط (Mean or mode): قم بتنشيط هذا الخيار لتقدير البيانات المفقودة باستخدام المتوسط (المتغيرات الكمية) أو الوسيط (المتغيرات النوعية).
- أقرب جار: قم بتنشيط هذا الخيار لتقدير البيانات المفقودة لملاحظة من خلال البحث عن أقرب جار للملاحظة.

استبدل البيانات المفقودة بـ 0: إذا كان تنسيق بياناتك هو "مصفوفة الارتباط" أو "مصفوفة التغاير"، يمكنك تنشيط هذا الخيار لاستبدال البيانات المفقودة بـ 0.

8. المخرجات (Outputs):



– الإحصاء الوصفي (Descriptive statistics): قم بتنشيط هذا الخيار لعرض الإحصائات الوصفية للمتغيرات المحددة.

– الارتباطات (Correlations): قم بتنشيط هذا الخيار لعرض مصفوفة الارتباط أو التغاير وفقًا لنوع الخيارات المختارة في علامة التبويب عام (General).

* مستوى دلالة الاختبار (Test significance): عند اختيار ارتباط في علامة التبويب "عام" في مربع الحوار، قم بتنشيط هذا الخيار لاختبار مستوى دلالة الارتباطات.

* اختبار بارتليت للكروية (Bartlett's sphericity): قم بتنشيط هذا الخيار لإجراء اختبار بارتليت للكروية (مؤشر للعلاقة بين المتغيرات، إذ يجب أن يكون مستوى الدلالة لهذه العلاقة أقل من 0.05 وذلك حتى نستطيع التأكيد على أنّ هذه العلاقة دالة إحصائيا، وأنّ مصفوفة الارتباط الأصلية ليست بمصفوفة الوحدة). وحدد مستوى الدلالة للاختبار (الذي تقبل أو ترفض عنده فرضية العدم).

– مقياس كايز–ماير–أولكين (KMO): قم بتنشيط هذا الخيار لحساب مقياس Kaiser–Meyer–Olkin لكفاية أخذ العينات.

– القيم الذاتية (Eigenvalues): حدد هذا الخيار لعرض الجدول والرسم البياني للقيم الذاتية (scree plot)، لتحديد عدد المركبات الرئيسية.

– تحميل العوامل (Factor loadings): حدد هذا الخيار لعرض المركبات الرئيسية للمتغيرات على المحاور العاملية.

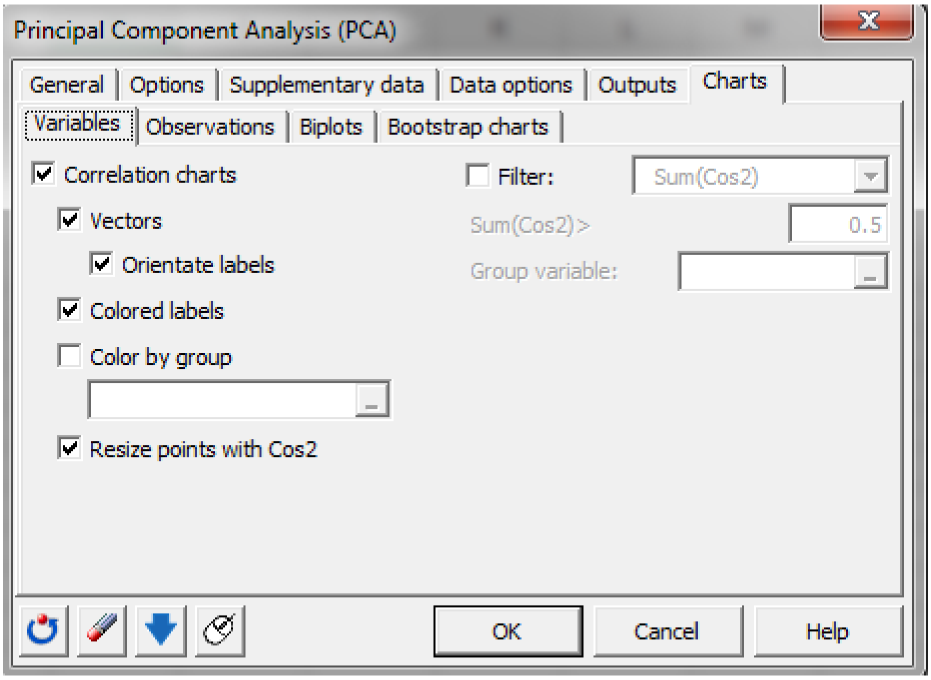
– الارتباط بين المتغيرات والمركبات الرئيسية (Variables/Factors correlations): حدد هذا الخيار لعرض جدول الارتباطات بين المركبات الرئيسية والمتغيرات الأصلية.

– إحداثيات المتغيرات على المحاور (Factor scores): حدد هذا الخيار لعرض جدول احداثيات المتغيرات على المحاور.

– المساهمات (Contributions): قم بتنشيط هذا الخيار لعرض جداول المساهمة للمتغيرات والأفراد في تشكيل المحاور.

– جيب التمام التربيعي (Squared cosines): قم بتنشيط هذا الخيار لعرض جودة تمثيل الأفراد والمتغيرات على المحاور.

9. الرسوم البيانية (Charts):



تحتوي هذه النافذة على أربعة نوافذ فرعية:

– المتغيرات:

* مخططات الارتباط (Correlations charts): قم بتنشيط هذا الخيار لعرض دائرة الارتباطات بين المكونات والمتغيرات الأولية.

* متجهات (Vectors): حدد هذا الخيار لعرض المتغيرات الأصلية في شكل متجهات داخل الدائرة.

* توجيه التسميات (Orientate labels): يسمح هذا الخيار (متوفر فقط مع Excel 2010 والإصدارات الأحدث) بعرض تسميات المتغيرات مع المتجه.

* تسميات ملونة (Colored labels): قم بتنشيط هذا الخيار لإظهار التسميات بنفس لون النقاط.

* لون حسب المجموعة (Color by group): قم بتنشيط هذا الخيار، إذا كنت تريد تلوين نقاط متغيرة وفقًا لمستويات متغير نوعي، ثم حدد سلسلة عمودية من البيانات تحتوي على عدد من الصفوف يساوي عدد المتغيرات النشطة. إذا تم تحديد رؤوس للجدول الرئيسي، فيجب تضمين رأس في هذا التحديد.

* تغيير حجم النقاط باستخدام جيب التمام التربيعي (Cos^2): قم بتنشيط هذا الخيار بحيث تتناسب أحجام النقاط المتغيرة مع المجموع المقابل لجيب التمام المربّع داخل الفضاء الفرعي المحدد.

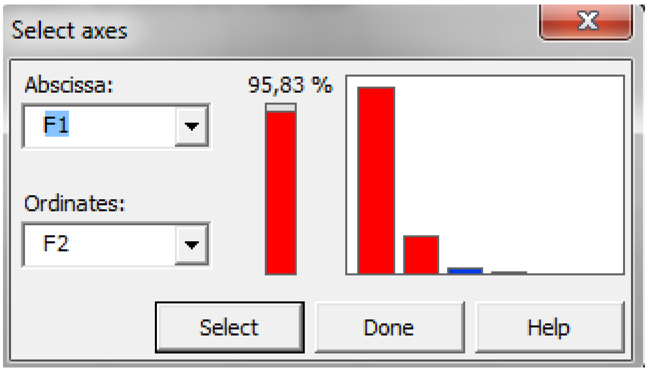
* عامل التصفية: قم بتنشيط هذا الخيار لتعديل عدد المتغيرات المعروضة.

– المشاهدات (observations): قم بتنشيط هذا الخيار لعرض المخططات التي تمثل المشاهدات في المساحة الجديدة.

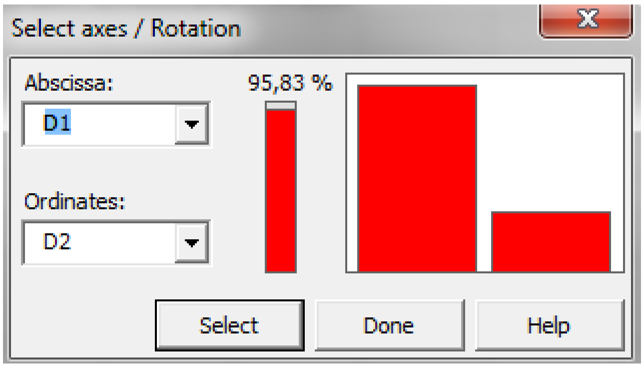
– الرسومات (Biplots): قم بتنشيط هذا الخيار لعرض المخططات التي تمثل الملاحظات والمتغيرات في نفس الوقت في المساحة الجديدة.

ملاحظة: لم يتم التطرق لبعض الخيارات التي تحتاج لفهم بعض متقدم في الغحصاء وتحليل البيانات.

10. نضغط ok فيقوم البرنامج بالعمليات الحسابية، ثم يظهر لك مربع حوار صغير لاختيار عدد المحاور.



تم استخلاص 6 مركبات رئيسية، المركبتين الأولي والثانية تفسران 95.83% من التباين، لذلك سنكتفي بهما، ونضغط Done، للمواصلة.



يمكنك تحديد المحاور التي سيتم تدويرها، في مثالنا نريد تدوير المحورين الأول والثاني. نضغط Done.

1. الجدول الأول: ملخص إحصائي.

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
السعر	10	0	10	2,760	9,300	6,123	2,138
الذوق	10	0	10	2,380	8,320	5,702	2,134
الحجم	10	0	10	4,430	9,370	7,158	1,583

التوفر	10	0	10	2,810	6,380	4,850	1,335
التغليف	10	0	10	2,380	8,180	5,247	1,996
الصورة	10	0	10	4,320	8,320	6,039	1,597

يعرض الجدول أعلاه المتغيرات، وعدد المشاهدات لكل متغير، والمشاهدات التي تحتوي على قيم مفقودة، وأقل وأكبر قيمة، إضافة إلى المتوسط والانحراف المعياري لكل متغير. هذا الجدول مفيد في فحص اولي للبيانات لاكتشاف القيم المفقودة والشاذة، وشكل انتشار البيانات...الخ. نلاحظ مثلا أن عدد المشاهدات لكل متغير 10، ولا توجد لدينا قيم مفقودة، كما يظهر أنّ القيم الدنيا لمتغيري الذوق والتغليف متساويتين والأمر نفسه تقريبا بالنسبة للقيم القصوى، مما يدل على تشابه تفضيلات الأفراد بالنسبة للذوق والتغليف....وهكذا يمكننا من خلال هذا الجدول استنتاج العديد من الأمور التي تساعدنا فيما بعد في عملية التحليل.

2. الجدول الثاني: مصفوفة الارتباطات.

Variables	السعر	الذوق	الحجم	التوفر	التغليف	الصورة
السعر	**1**	**-0,859**	**-0,742**	**-0,092**	**-0,953**	**-0,913**
الذوق	**-0,859**	**1**	**0,960**	**0,479**	**0,904**	**0,974**
الحجم	**-0,742**	**0,960**	**1**	**0,484**	**0,807**	**0,896**
التوفر	**-0,092**	**0,479**	**0,484**	**1**	**0,263**	**0,450**
التغليف	**-0,953**	**0,904**	**0,807**	**0,263**	**1**	**0,960**
الصورة	**-0,913**	**0,974**	**0,896**	**0,450**	**0,960**	**1**

القيم المكتوبة بخط ثخين داخل الجدول تعني أنّ الارتباط بين المتغيرين دال احصائيا عند مستوى دلالة 0.05 فأقل. يمكننا ملاحظة أنّ جميع الارتباطات عالية بين المتغيرات، ايجابا أو سلبا. في الحقيقة هذا الأمر يمثل مشكلة التعدد الخطي في التحليل العاملي، إذ إنّ أبرز شروطه أن تكون الارتباطات بين المتغيرات غير مبالغ فيها (أكبر من 0.9). وفي الغالب يمكننا معالجة هذا الأمر بالنظر في الارتباط بين متغيرين، يزيد ارتباطهما عن 0.9 ونحذف احداهما ثم نعيد التحليل.

3. الجدول الثالث: اختبار بارتليت للدائرية (الكروية):

Chi-square (Observed value)	83,235
Chi-square (Critical value)	7,261
DF	15
p-value (Two-tailed)	< 0,0001
alpha	0,95

أختبار Bartilett للدائرية sphericity مؤشر للعلاقة بين المتغيرات، إذ يجب أن يكون مستوى الدلالة لهذه العلاقة أقل من 0.05 وذلك حتى نستطيع التأكيد على أنّ هذه العلاقة دالة إحصائيا، وأنّ مصفوفة الارتباط الأصلية ليست بمصفوفة الوحدة (بمعنى أن المتغيرات مستقلة عن بعضها البعض). الفرضية الصفرية لاختبار بارتليت للكروية، هي أن مصفوفة الإرتباط هي مصفوفة الوحدة، أو أن المصفوفة لها واحد على القطر الرئيسي وصفرا فيما عدا ذلك. ويظهر من النتائج أنّ مستوى دلالة الاختبار أقل من 0.05 (p-value < 0,0001)، مما يدل على أنّ مصفوفة الارتباط الأصلية ليست بمصفوفة الوحدة. ولذلك نرفض فرضية العدم.

الجدول الرابع: مقياس كايزر–ماير–أولكين لمدى كفاية أخذ العينات (Kaiser–Meyer–Olkin):

السعر	0,712
الذوق	0,745
الحجم	0,735
التوفر	0,358
التغليف	0,816
الصورة	0,728
KMO	0,713

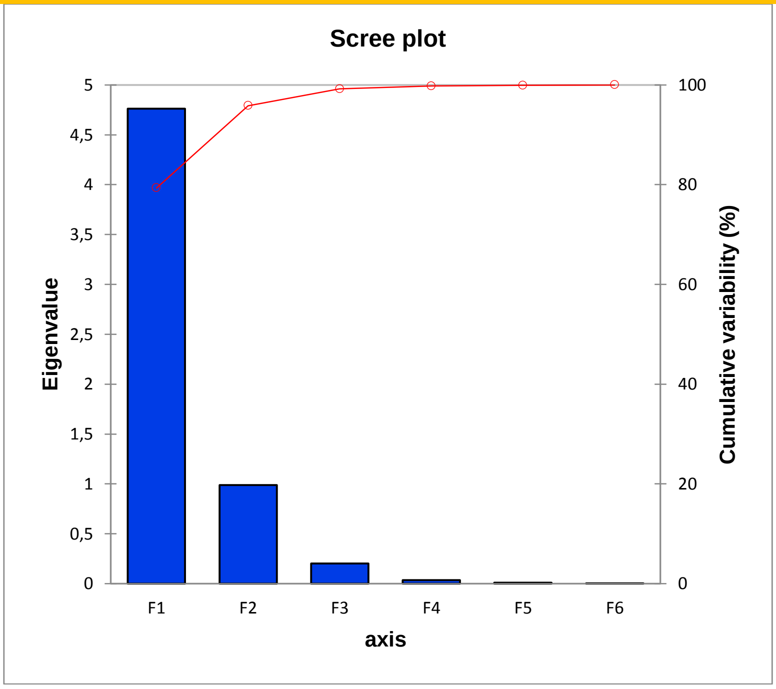
اختبار (KMO) هو مقياس لمدى ملاءمة بياناتك لتحليل العوامل، يقيس الاختبار مدى كفاية أخذ العينات لكل متغير في النموذج وللنموذج الكامل. ويأخذ KMO قيمًا بين 0 و 1. وكقاعدة أساسية تشير قيم KMO بين 0.8 و 1 إلى أن أخذ العينات مناسب، وتشير قيم KMO الأقل من 0.6 إلى أن أخذ العينات غير كاف وأنه ينبغي اتخاذ إجراء علاجي. وقد وضع بعض المؤلفين هذه القيمة عند 0.5، لذلك استخدم حكمك الخاص للقيم بين 0.5 و 0.6. وتعني قيم KMO القريبة من الصفر أن هناك ارتباطات جزئية كبيرة مقارنة بمجموع الارتباطات. بمعنى آخر، هناك ارتباط واسع الانتشار يمثل مشكلة كبيرة لتحليل العوامل.

الجدول الخامس: القيم الذاتية لمصفوفة الارتباط:

	F1	F2	F3	F4	F5	F6
Eigenvalue	4,761	0,989	0,202	0,035	0,009	0,005
Variability (%)	79,343	16,487	3,365	0,576	0,151	0,077
Cumulative %	79,343	95,831	99,196	99,772	99,923	100,000

كل قيمة ذاتية تعتبر مركبة أساسية. يعرض الجدول أعلاه عدد القيم الذاتية لمصفوفة الارتباط، والتباين المفسر بواسطة كل مركبة أساسية. المركبتين الأولى والثانية تفسران معا 95.831 في المائة من التباين الكلي. لذلك سنكتفي فقط بالمركبتين الأولى والثانية.

الشكل رقم 1: التمثيل البياني للقيم الذاتية:



يمثل قيم الجذور الكامنة لكل عامل على المحور الصادي ورقم المكون على المحور السيني، ويعتبر الرسم البياني معيارا آخر يمكن استخدامه بالإضافة إلى معيار الإبقاء على العوامل التي يزيد جذرها الكامن عن الواحد الصحيح لتحديد العوامل في التحليل العاملي والإبقاء فقط على تلك التي تكون في المنطقة شديدة الانحدار. لاحظ أنّ هناك عامل واحد فقط جذره الكامن أكبر من واحد، ويوجد عامل ثاني يقع في المرفق، ويساوي الواحد تقريبا، يمكننا أخذه إذا اعتمدنا على محكات أخرى بخلاف محك كايزر.

الجدول السادس: المتجهات الذاتية المرتبطة بالقيم الذاتية لمصفوفة الارتباط.

	F1	F2	F3	F4	F5	F6
السعر	-0,414	0,392	0,301	0,572	0,504	0,056
الذوق	0,453	0,056	0,249	-0,288	0,570	-0,567
الحجم	0,426	0,133	0,751	0,174	-0,419	0,173
التوفر	0,208	0,881	-0,355	-0,076	-0,208	-0,072
التغليف	0,436	-0,221	-0,356	0,743	-0,070	-0,277
الصورة	0,456	-0,014	-0,174	-0,033	0,444	0,751

لكل قيمة ذاتية متجهها الخاص.

الجدول السابع: (تحميل العوامل) احداثيات المتغيرات على المحاور.

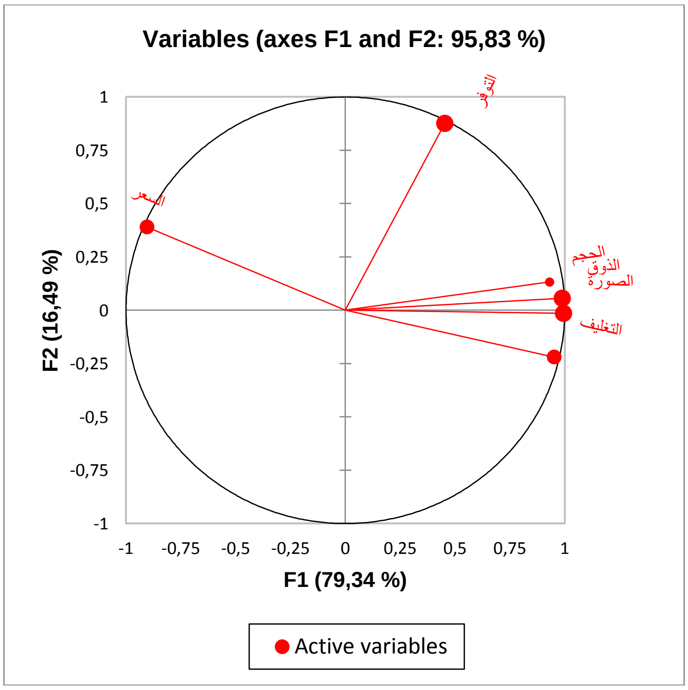
	F1	F2	F3	F4	F5	F6
السعر	-0,903	0,390	0,135	0,106	0,048	0,004
الذوق	0,988	0,056	0,112	-0,054	0,054	-0,039
الحجم	0,930	0,132	0,338	0,032	-0,040	0,012
التوفر	0,454	0,876	-0,160	-0,014	-0,020	-0,005
التغليف	0,952	-0,220	-0,160	0,138	-0,007	-0,019
الصورة	0,995	-0,014	-0,078	-0,006	0,042	0,051

الجدول الثامن: مصفوفة الارتباطات بين المتغيرات والمركبات الرئيسية.

	F1	F2	F3	F4	F5	F6
السعر	-0,903	0,390	0,135	0,106	0,048	0,004
الذوق	0,988	0,056	0,112	-0,054	0,054	-0,039
الحجم	0,930	0,132	0,338	0,032	-0,040	0,012
التوفر	0,454	0,876	-0,160	-0,014	-0,020	-0,005
التغليف	0,952	-0,220	-0,160	0,138	-0,007	-0,019
الصورة	0,995	-0,014	-0,078	-0,006	0,042	0,051

يظهر أنّ المتغيرات (الذوق، الحجم، التغليف، الصورة) ترتبط ايجابيا وبقوة مع المحور الأول، أمّا متغير السعر، فيرتبط سلبا وبقوة مع المحور الأول، ويرتبط ايجابيا مع المحور الثاني، لكن الارتباط ضعيف (0.39). فيما يرتبط متغير التوفر بقوة مع المحور الثاني.

الشكل الثاني: دائرة الارتباط.



المتغيرات التي تقع يمين الرسم على وبالقرب من المحور الأول، ارتباطها قوي وموجب معه، ومتغير سعر المنتج يقع يسار الرسم تقريبا على محيط الدائرة وبالقرب من المحور الأول، مما يعني أنّه مرتبط سلبا وبقوة مع المحور الأول، ويرتبط بالمحور الثاني ارتباطا ضعيفا. امّا متغير التوفر فهو قريب من المحور الثاني وفي يمين المخطط، مما يعني ارتباطه بالمحور الثاني ايجابيا.

الجدول التاسع: نسبة مساهمة المتغيرات في تشكيل المحاور.

	F1	F2	F3	F4	F5	F6
السعر	17,144	15,355	9,083	32,723	25,384	0,312
الذوق	20,525	0,314	6,181	8,320	32,526	32,135
الحجم	18,187	1,763	56,454	3,031	17,564	3,002
التوفر	4,321	77,657	12,608	0,582	4,307	0,525

التغليف	19,043	4,891	12,653	55,234	0,485	7,693
الصورة	20,780	0,020	3,021	0,111	19,734	56,334

متغيرات (السعر، الذوق، التوفر، التغليف، الصورة) تساهم بنسب متقاربة في تشكيل المحور الأول، فيما يساهم متغير التوفر بنسبة كبيرة في تشكيل المحور الثاني، ولا يساهم في تشكيل المحور الأول. ومتغير السعر فقط من يساهم بنسبة مقاربة في تشكيل المحور الثاني لنسبته في تشكيل المحور الأول.

الجدول العاشر: جودة تمثيل المتغيرات على المحاور.

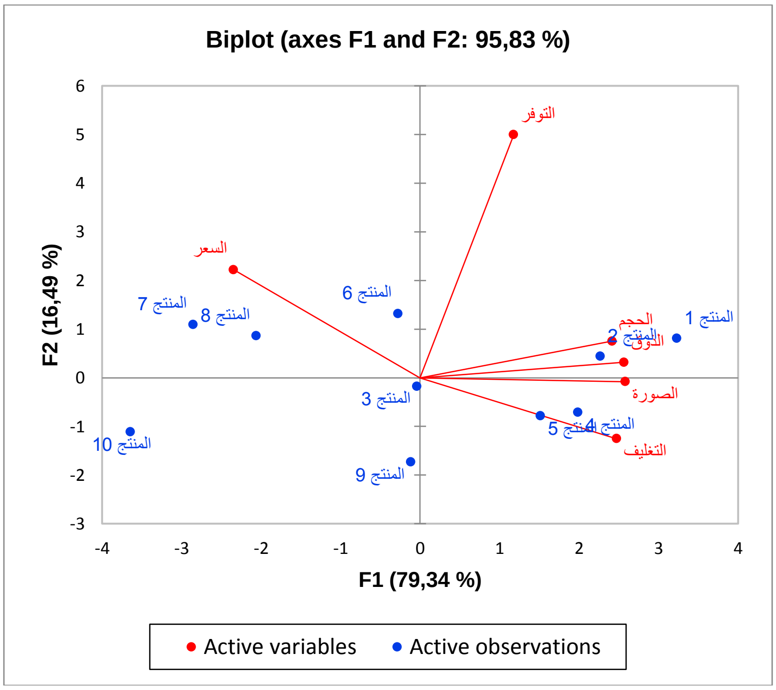
	F1	F2	F3	F4	F5	F6
السعر	**0,816**	0,152	0,018	0,011	0,002	0,000
الذوق	**0,977**	0,003	0,012	0,003	0,003	0,001
الحجم	**0,866**	0,017	0,114	0,001	0,002	0,000
التوفر	0,206	**0,768**	0,025	0,000	0,000	0,000
التغليف	**0,907**	0,048	0,026	0,019	0,000	0,000
الصورة	**0,989**	0,000	0,006	0,000	0,002	0,003

القيم بالخط الثخين (قيم كبيرة)، وتعني أنّ المتغير ممثل بشكل جيد على المحور. متغيرات (السعر، الذوق، الحجم، التغليف، الصورة) ممثلة بشكل جيد على المحور الأول، ومتغير التوفر ممثل بشكل جيد على المحور الثاني.

الجدول الحادي عشر: إحداثيات الأفراد على المحاور.

	F1	F2	F3	F4	F5	F6
المنتج 1	3,229	0,807	-0,324	-0,067	0,152	-0,027
المنتج 2	2,267	0,438	-0,271	-0,359	-0,161	0,012
المنتج 3	-0,041	-0,178	0,290	0,129	-0,103	-0,169
المنتج 4	1,987	-0,709	0,182	0,091	0,116	-0,029
المنتج 5	1,515	-0,783	0,846	0,002	-0,027	0,103
المنتج 6	-0,278	1,315	0,211	0,218	-0,003	0,042
المنتج 7	-2,856	1,093	-0,338	-0,057	0,057	0,016
المنتج 8	-2,063	0,862	0,052	0,127	-0,067	0,030
المنتج 9	-0,115	-1,732	-0,888	0,201	-0,046	0,043
المنتج 10	-3,646	-1,112	0,239	-0,284	0,082	-0,021

الشكل الثالث: تمثيل الأفراد والمتغيرات على المخطط العاملي.



ترتبط المتغيرات (الحجم، الذوق، الصورة، التغليف) مع المحور الأول طرديا وبقوة ويقابلها في جهة اليسار سعر المنتج (يرتبط عكسيا وبقوة مع المحور الأول) (المحور الأول يعرض جودة المنتجات مقارنة بالسعر)، لذلك فالمنتجات التي تقع يمين الرسم (تحمل المواصفات السابقة)، والممثلة جيدا على المحور الأول، منتجات بمواصفات أو جودة عالية (المنتجات: 1،2،4،5)، وسعرها مرتفع، وعلى المؤسسة توزيع تلك المنتجات أو بيعها في مناطق الأغنياء. بينما المنتجات التي تتموضع على يسار المحور الأول (7،8،10)، فهي منتجات أقل جودة، وأسعارها منخفضة.

أمّا المحور الثاني فيعرض لنا التوفّر، فالمنتجات التي تكون أعلى هذا المحور (قريب من مواصفة التوفر) هي منتجات متوفرة في السوق ويمكن الحصول عليها بسهولة، والمنتجات التي تتموضع أسفله هي منتجات غير متوفرة في السوق ولا يمكن الحصول عليها بسهولة. فالمنتج رقم 6 هو المنتج الأكثر توفّرًا في السوق (لاحظ أنّه قريب من مكان وجود خاصية التوفّر). والمنتج التاسع غير متوفّر في السوق، أمّا المنتج الثالث فهو غير ممثل جيدا عى المحورين (الأول والثاني)، بمعنى أنه منتج غير مرغوب فيه.

الجدول الثاني عشر: نسبة مساهمة الأفراد في تشكيل المحاور.

	F1	F2	F3	F4	F5	F6
المنتج 1	21,902	6,580	5,212	1,301	25,546	1,585
المنتج 2	10,800	1,939	3,634	37,257	28,748	0,326
المنتج 3	0,003	0,322	4,170	4,791	11,636	61,726
المنتج 4	8,295	5,075	1,649	2,378	14,913	1,868
المنتج 5	4,819	6,201	35,484	0,001	0,834	23,163
المنتج 6	0,163	17,471	2,198	13,687	0,010	3,779
المنتج 7	17,129	12,074	5,652	0,948	3,606	0,541
المنتج 8	8,937	7,511	0,134	4,664	4,895	1,938
المنتج 9	0,028	30,328	39,035	11,690	2,384	4,084
المنتج 10	27,925	12,498	2,832	23,281	7,429	0,988

الجدول الثالث عشر: جودة تمثيل الأفراد على المحاور.

	F1	F2	F3	F4	F5	F6
المنتج 1	**0,930**	0,058	0,009	0,000	0,002	0,000
المنتج 2	**0,924**	0,034	0,013	0,023	0,005	0,000
المنتج 3	0,010	0,184	**0,486**	0,096	0,061	0,165
المنتج 4	**0,876**	0,111	0,007	0,002	0,003	0,000
المنتج 5	**0,631**	0,169	0,197	0,000	0,000	0,003
المنتج 6	0,041	**0,910**	0,023	0,025	0,000	0,001
المنتج 7	**0,861**	0,126	0,012	0,000	0,000	0,000
المنتج 8	**0,847**	0,148	0,001	0,003	0,001	0,000
المنتج 9	0,003	**0,780**	0,205	0,011	0,001	0,000
المنتج 10	**0,906**	0,084	0,004	0,005	0,000	0,000

المنتجات (10، 8، 7، 5، 4، 2، 1) ممثلة بشكل جيد على المحور الأول، والمنتجين 6، 9 ممثلين على المحور الثاني. فيما يظهر أنّ المنتج الثالث غير ممثل على المحورين الأول والثاني، وممثل بشكل أفضل على المحور الثالث، وهذا سبب اقترابه من مركز الثقل، ودليل على أنّ هذا المنتج غير مرغوب فيه، وعلى المؤسسة سحبه من السوق.

التطبيق على برنامج SPSS

نستخدم نفس المثال السابق. علينا نقل معطيات الجدول إلى البرنامج.

	المنتجات	السعر	الذوق	الحجم	التوفر	التغليف	الصورة
1	منتج1	4,40	8,18	6,32	9,37	8,32	9,30
2	منتج2	4,32	7,12	5,90	8,68	7,18	8,20
3	منتج3	6,11	5,33	5,12	6,79	5,84	5,63
4	منتج4	4,62	7,07	5,93	6,70	7,78	7,97
5	منتج5	5,00	6,72	6,38	6,14	6,73	7,38
6	منتج6	7,28	5,08	5,18	8,67	5,02	5,80
7	منتج7	8,32	2,83	3,13	7,88	2,83	3,65
8	منتج8	7,93	3,42	4,00	7,63	3,64	4,15
9	منتج9	4,82	4,34	3,73	5,29	7,30	6,39
10	منتج10	7,59	2,38	2,81	4,43	2,38	2,76

سنقوم الآن بعملية التحليل، لكن عندما نختار تحليل البيانات باستخدام المركبات الرئيسية، فإن جزءًا من العملية يتضمن التحقق من إمكانية تحليل بياناتنا باستخدام التحليل بالمركبات الرئيسية. وللقيام بذلك على بيانتنا أن تجتاز أربعة افتراضات مطلوبة حتى تكون النتائج صحيحة. ولا يجب عليك أن تتفاجأ إذا تم انتهاك واحد أو أكثر من هذه الافتراضات (أي لم يتم الوفاء بها). لن يكون ذلك موجودا عند العمل مع بيانات العالم الحقيقي بدلاً من أمثلة الكتب التعليمية. ومع ذلك، حتى عندما تفشل بياناتك في افتراضات معينة، غالبًا ما يكون هناك حل لمحاولة التغلب على ذلك. أولاً، دعنا نلقي نظرة على هذه الافتراضات الأربعة:

الافتراض رقم 1: لديك متغيرات متعددة يجب قياسها على المستوى المستمر (على الرغم من استخدام المتغيرات الترتيبية كثيرًا). تتضمن أمثلة المتغيرات المستمرة (أي متغيرات النسبة أو الفاصل الزمني) وقت المراجعة (يقاس بالساعات)، والذكاء (المقاس باستخدام درجة الذكاء)، وأداء الاختبار (يقاس من 0 إلى 100)، والوزن (يقاس بالكيلوغرام) ، وما إلى ذلك. تتضمن أمثلة المتغيرات الترتيبية المستخدمة بشكل شائع في PCA نطاقًا واسعًا من مقاييس ليكرت (على سبيل المثال، مقياس مكون من 7 نقاط من أوافق بشدة إلى لا أوافق بشدة).

الافتراض رقم 2: يجب أن تكون هناك علاقة خطية بين جميع المتغيرات. سبب هذا الافتراض هو أن التحليل بالمركبات الرئيسية يعتمد على معاملات ارتباط بيرسون. وعلى هذا النحو يجب أن تكون هناك علاقة خطية بين المتغيرات.

الافتراض رقم 3: يجب أن يكون لديك كفاية لأخذ العينات، وهو ما يعني ببساطة أنه لكي تحصل على نتائج موثوقة من التحليل بالمركبات الرئيسية، يلزم وجود أحجام عينات كبيرة بما يكفي. بشكل عام تمت التوصية بحد أدنى من 150 مشاهدة، أو من 5 إلى 10 مشاهدات لكل متغير كحد أدنى لحجم العينة. وهناك عدة

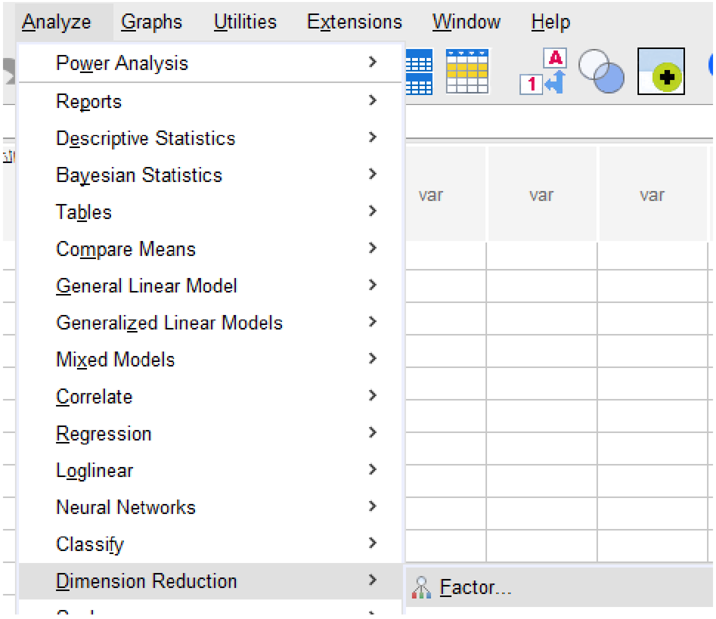
طرق لاكتشاف كفاية أخذ العينات: مقياس كايزر-ماير-أولكين (KMO) لكفاية أخذ العينات لمجموعة البيانات الإجمالية؛ ومقياس KMO لكل متغير فردي.

الافتراض رقم 4: يجب أن تكون بياناتك مناسبة للتحليل العاملي. إذ يجب أن يكون لديك ارتباطات كافية بين المتغيرات من أجل تقليل المتغيرات إلى عدد أقل من المكونات. الطريقة التي يستخدمها SPSS لاكتشاف ذلك هي اختبار Bartlett للكروية.

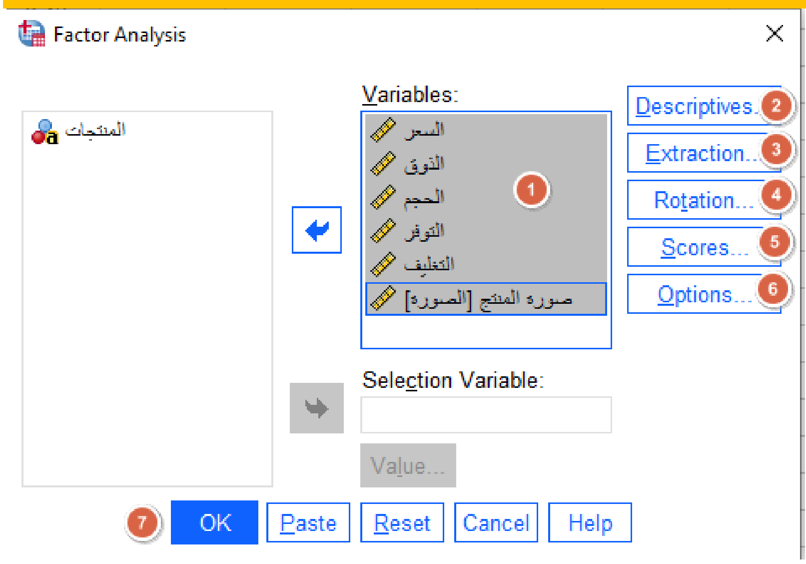
الافتراض رقم 5: يجب ألاّ يكون هناك قيم متطرفة كبيرة. القيم المتطرفة مهمة لأنها يمكن أن يكون لها تأثير غير متناسب على نتائجك.

ملاحظة: يمكنك التحقق من الافتراضات 2 و 3 و 4 و 5 باستخدام SPSS فقط. تذكر أنه إذا لم تقم بإجراء الاختبارات الإحصائية على هذه الافتراضات بشكل صحيح، فقد لا تكون النتائج التي تحصل عليها صالحة.

من القائمة الرئيسية Analyze اختر Dimension Reduction ثمFactor .

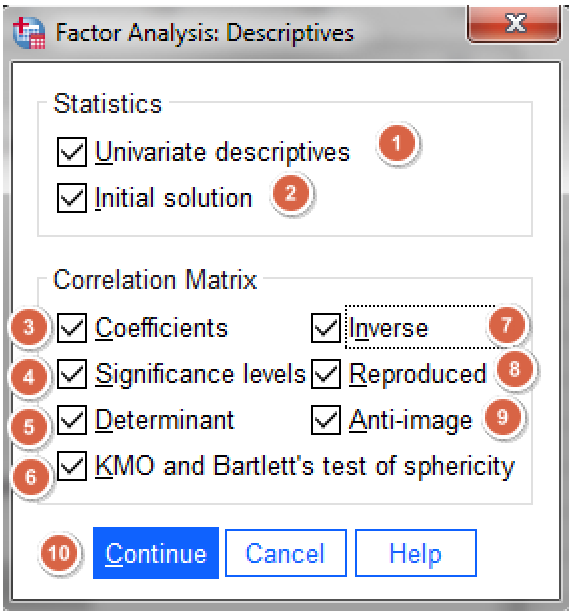


Analyze> Dimension Reduction>Factor...



1- ننقل المتغيرات إلى مربع Variable. (المتغيرات الكمية فقط).

2- نضغط Descriptives:



1. Univariate descriptives (الوصف وحيد المتغير): حدد هذا الخيار لعرض الاحصاءات الوصفية للمتغيرات، مثل المتوسط الحسابي والانحراف المعياري وعدد المشاهدات الصالحة لكل متغير.

2. Initial solution (الحل المبدئي): حدد هذا الخيار لعرض الاشتراكيات (الشيوع) (Communalities) والجذور الكامنة (القيم الذاتية) (Eigenvalues)، النسب المئوية والمتراكمة للتباين المفسّر (Percentage and Cumulative percentage of Variance).

الشيوع (Communalities): مجموع إسهامات المتغير في العوامل المختلفة التي أمكن استخلاصها في المصفوفة العاملية – وحيث أن المتغير الواحد يسهم بمقادير مختلفة في كل عامل، سواءًا كانت إسهاماته

جوهرية أو كانت غير ذات دلالة، فإن مجموع مربعات هذه الإسهامات أو التشبعات على عوامل المصفوفة هي قيمة شيوع المتغير أو الاشتراكيات.

ثم من Correlation Matrix نختار:

3. Coefficients لحساب معاملات مصفوفة الارتباط.

4. من خلال التأشير على مستويات الأهمية (الدلالة) (significance levels) نحصل على مصفوفة العلاقات من أجل الاطلاع عليها والتأكد من شرط عدم وجود ارتباط عالي بين المتغيرات، أي أعلى من 90 % بين أي متغيرين، حيث يتم استبعاد تلك المتغيرات التي بينها هذه النسبة العالية من الارتباط.

5. من خلال التأشير على المحدد (Determinant) نحصل على محدد المصفوفة، وذلك لقياس مشكلة الارتباط الذاتي، إذ يجب ألاّ تقل قيمة المحدد عن 0.0001، فإذا كانت قيمته أقل من ذلك ننظر إلى المتغيرات المرتبطة عاليا، أكثر من 0.80 ونحذف أحدهما.

6. ونؤشر على KMO AND Bartletts test of sphericty لنحصل من خلال قياس KMO على مدى كفاية عدد أفراد العينة ويجب أن تكون قيمته أكبر من 0.5 حتى تكون العينة كافية، وهذا شرط أساسي يجب تحقيقه، أما فيما يتعلق بأختبار Bartilett للدائرية sphericity فهو مؤشر للعلاقة بين المتغيرات، إذ يجب أن يكون مستوى الدلالة لهذه العلاقة أقل من 0.05 وذلك حتى نستطيع التأكيد على أنّ هذه العلاقة دالة إحصائيا، وأنّ مصفوفة الارتباط الأصلية ليست بمصفوفة الوحدة.

الفرضية الصفرية لاختبار بارتليت للكروية، هي أن مصفوفة الإرتباط هي مصفوفة الوحدة، أو أن المصفوفة لها واحد على القطر الرئيسي وصفرا فيما عدا ذلك.

ملاحظة: ما هو اختبار(KMO) Kaiser-Meyer-Olkin ؟

اختبار (KMO) هو مقياس لمدى ملاءمة بياناتك لتحليل العوامل، يقيس الاختبار مدى كفاية أخذ العينات لكل متغير في النموذج وللنموذج الكامل.

ويأخذ KMO قيمًا بين 0 و 1. وكقاعدة أساسية تشير قيم KMO بين 0.8 و 1 إلى أن أخذ العينات مناسب، وتشير قيم KMO الأقل من 0.6 إلى أن أخذ العينات غير كاف وأنه ينبغي اتخاذ إجراء علاجي. وقد وضع بعض المؤلفين هذه القيمة عند 0.5، لذا استخدم حكمك الخاص للقيم بين 0.5 و 0.6. وتعني قيم KMO القريبة من الصفر أن هناك ارتباطات جزئية كبيرة مقارنة بمجموع الارتباطات. بمعنى آخر، هناك ارتباط واسع الانتشار يمثل مشكلة كبيرة لتحليل العوامل.

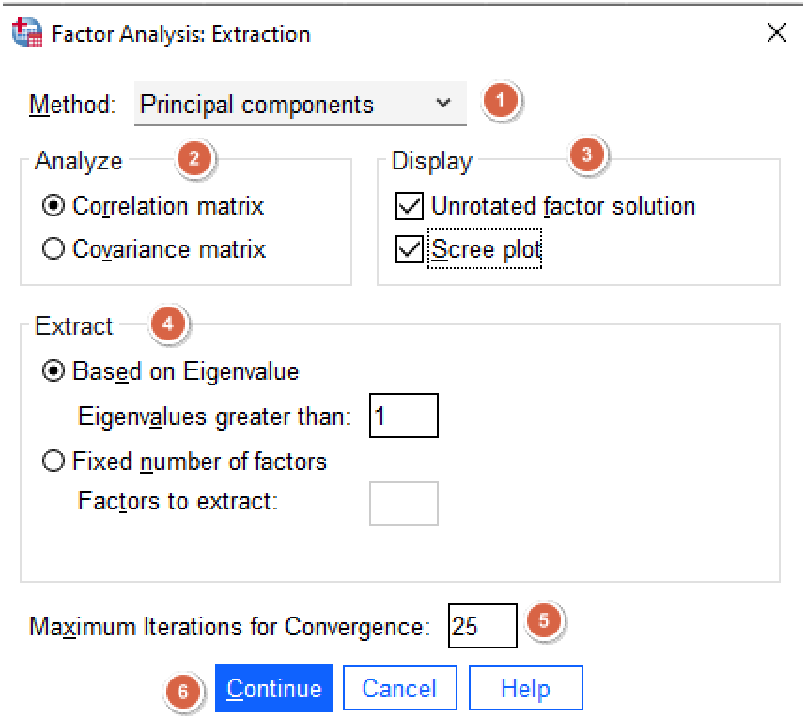
7. بالتأشير على Inverse (المعكوسات)،

8. بالتأشير على إعادة تقدير المصفوفة (Reproduced)،

9. نؤشر على Anti-Image (مصفوفة الارتباطات الجزئية الصورية) (العكسية) لاختبار كفاية أخذ العينات لكل متغير، ويتم الحصول على مصفوفة الارتباطات الجزئية، التي يمكننا من خلالها معرفة مدى كفاية أخذ

العينات لكل متغير على حدة. ويجب أن تكون قيم معاملات الارتباط في قطر المصفوفة أكبر من أو تساوي 0.5، وكل متغير معامل ارتباطه أقل من 0.5 يحذف ويعاد التحليل.

3- نضغط Extraction :



1.3. نختار طريقة الاستخراج، وتوجد سبعة طرق، ما يهمنا منها طريقة المكونات الأساسية Principle Components Analysis.

2.3. Analyze: تعني المصفوفة المراد تحليلها وتتضمن:

- من خلال التأشير على الحقل Correlation Matrix نحصل على مصفوفة الارتباط.
- من خلال التأشير على الحقل Covariance Matrix نحصل على مصفوفة التباين المشترك.

3.3. تحت الخيار Display، من خلال التأشير على Unrotated factor solution نحصل على حل العوامل قبل التدوير، وذلك من أجل مقارنة نتائجه مع نتائج rotated factor solution فإذا كانت متفقة مع بعضها من حيث عدد العوامل تكون النتائج دقيقة، أما إذا اختلفت النتيجتان فإننا نقوم بفحص الاشتراكيات لنقرر عدد العوامل، وعادة يستخدم هذا الخيار عندما يكون عدد المتغيرات كبيرا، أي أكثر من 200 متغير.

ومن خلال التأشير على Scree plot والذي يعني الرسم البياني (سكري) (يطلق عليه كذلك اسم منحنى الصخرة) ويستخدم لتحديد العدد الأقصى من العوامل التي يمكن استخلاصها قبل أن يبدأ التباين الخاص في السيطرة على التباين العام. يقوم البرنامج بالتمثيل البياني لرقم الجذر المميز على المحور الأفقي وقيمة الجذر على المحور الرأسي، فيكون عدد العوامل مساويا لرقم الجذر المميز المناظر إلى الإنحناء المرفق

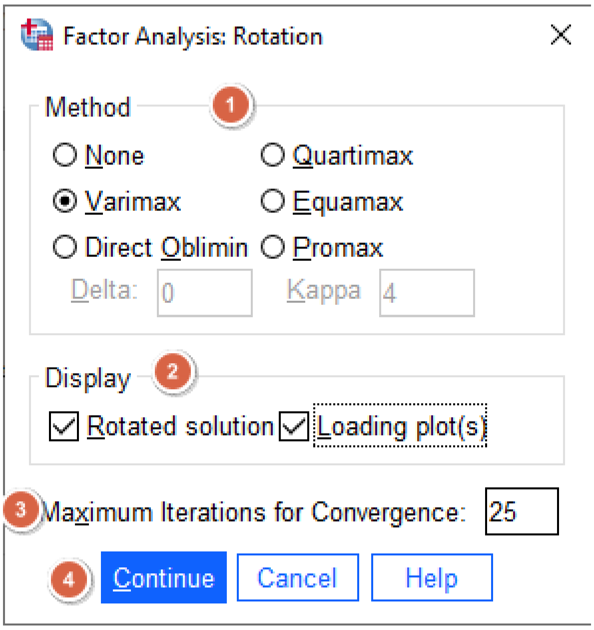
للمنحنى (Corresponding to an elbow in curve)، فالنقطة التي ينتهي عندها هذا الإنحناء تتوقف عندها الجذور المميزة الكبيرة وتبدأ الجذور المميزة الصغيرة، والتي يتم استبعادها من التحليل.

3.4. Extract: بالتأشير على Eigenvalues نحصل على قيمة الجذر الكامن، والبرنامج يحدد قيمة الجذر الكامن لتكون أكبر من 1 (حسب محك كايزر). ومن خلال التأشير على الحقل Number of Factorsيتم استخراج عدد من العوامل بعد أن يتم تحديدها من قبل الباحث. استخدام هذا الخيار يلغي الخيار الأول، والمتعلق بقيمة الجذر الكامن.

3.5. يوجد في أسفل صندوق الحوار خيارا لتحديد الحد الأعلى لعدد تكرارات الخوارزمية الضرورية للوصول للحل المناسب (Maximum iterations for Convergence 25) وبإمكان الباحث أن يغير هذا الرقم المحدد مسبقا من قبل البرنامج بما يتناسب مع أهداف وطبيعة البحث.

3.6. نضغط Continue.

4- نضغط Rotation:



4.1. نختار طريقة التدوير (Method)، ليتم تدوير العوامل. وتوجد خمسة طرق لعملية التدوير:

- **None** وتعني عدم إجراء عملية التدوير.
- طريقة التدوير المتعامدة التي تؤدي إلى زيادة تباين مربع تشبعات العوامل على كافة المتغيرات (Varimax)، وهي من أفضل طرق التدوير المتعامد، لافتراضها وجود استقلالية بين العوامل.
- قائمة على طريقة التدوير المائل (Direct Oblimin)، وتؤدي إلى قيم أعلى للجذور الكامنة.
- طريقة (Quartimax): وهي طريقة أخرى للتدوير المتعامد، تؤدي إلى تخفيض عدد العوامل التي تحتاجها لتفسير كل متغير.
- طريقة (Equamax): تقع في الوسط بين طريقتي Varimax و Quartimax.

– Promax : طريقة أخرى للتدوير المائل وهي أسرع في العمليات الحسابية من طريقة Direct Oblimin، لذلك فهي تستخدم في بعض الأحيان في العينات الكبيرة العدد.

2.4. كما يتضمن صندوق الحوار خياران للعرض (Display):

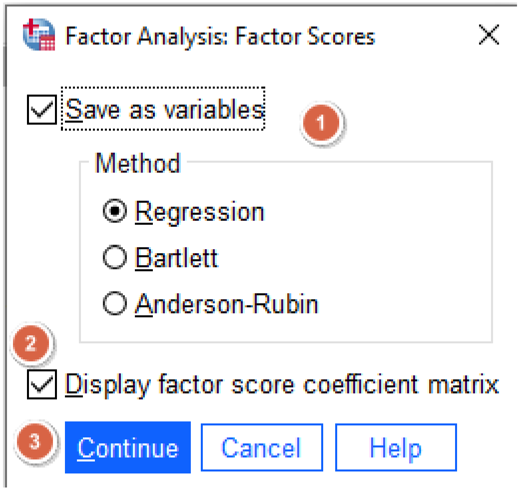
– العوامل بعد التدوير Rotated Solution وهذا الحقل محدد سلفا من قبل البرنامج.

– الرسوم البيانية للتشبعات Loading Plot(s) بالإمكان تحديد هذا الحقل واضافته للحقل الأول.

3.4. Maximum Iterations for Convergence: عدد مرات تدوير العوامل، نتركها كما هي 25 دورة، في غالب الأحيان ربما لا يصل البرنامج لهذا العدد من الدورات.

4.4. نضغط Continue.

5– نضغط Scores:



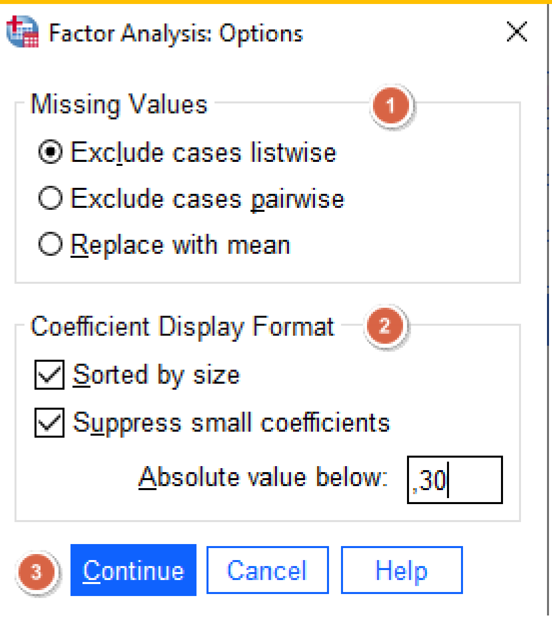
1.5. نختار Save as variables ليتم حفظ العوامل المستخرجة كمتغيرات. وعند تحديد هذا الحقل فإنه سيتم تفعيل طرق حساب الدرجات، التي يمكن استخدامها في إجراء عمليات إحصائية إضافية وفقا لاحتياجات البحث، ويمكن اختيار طريقة الحفظ: Regression، أو طريقة Bartlett، أو بطريقة Anderson–Rubin.

ملاحظة: في مثالنا سنحدد طريقة الانحدار، لاستخدام العوامل المستخرجة في تحليل الانحدار فيما بعد.

2.5. نؤشّر على Display factor score coefficient matrix لعرض مصفوفة معاملات الدرجات.

3.5. نضغط Continue.

6– نضغط Options:



1.6. نختار طريقة معالجة البيانات المفقودة.

2.6. نختار طريقة عرض المعاملات، بطريقة الإخراج حسب الحجم (Sorted by size)، يتم ترتيب التشبعات على العوامل وفقا لمقدارها. كما يمكننا إخفاء عرض القيم المطلقة للتشبعات التي تقل عن قيمة معينة (Suppress absolute values less than)، وبالتأشير على المربع الصغير أمام هذا الخيار يتم تفعيل القيم التي يرغب الباحث بوضعها لإخفاء المعلومات المتعلقة بالقيم الأقل (التشبعات)، علما بأن هذه القيمة محددة سلفا في البرنامج 0.10.

ملاحظة: في مثالنا جعلناه 0.30، كل متغير تشبعه على عامل معين أقل من 0.30 يتم إخفائه.

3.6. نضغط Continue.

7- نضغط أخيرا على ok فتظهر النتائج التالية:

Warnings

Warnings
Only one component was extracted. Component plots cannot be produced.

يخبرنا هذا التحذير أنه قد تم فقط استخراج مركبة أساسية واحدة، لذلك لن يتم عرض التمثيل البياني للمحاور العاملية. وقد حصلنا على هذه النتيجة لأننا استخدمنا محك كايزر في تحديد عدد العوامل المستخرجة، مما يعني وجود عامل واحد فقط جذره الكامن أكبر من الواحد الصحيح. كما سنرى فيما بعد عند تحليل جدول العوامل الكامنة.

الجدول الأول: جدول الإحصاءات الوصفية.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
السعر	6,0390	1,59687	10
الذوق	5,2470	1,99612	10
الحجم	4,8500	1,33539	10
التوفر	7,1580	1,58298	10
التغليف	5,7020	2,13404	10
صورة المنتج	6,1230	2,13790	10

يعرض هذا الجدول المتوسط الحسابي والانحراف المعياري للمتغيرات التي أدخلناها للتحليل وعدد المفردات (N) (الأفراد مقابل المتغيرات).

الجدول الثاني: مصفوفة الارتباطات.

Correlation Matrix[a]

		السعر	الذوق	الحجم	التوفر	التغليف	صورة المنتج
Correlation	السعر	1,000	-,859	-,742	-,092	-,953	-,913
	الذوق	-,859	1,000	,960	,479	,904	,974
	الحجم	-,742	,960	1,000	,484	,807	,896
	التوفر	-,092	,479	,484	1,000	,263	,450
	التغليف	-,953	,904	,807	,263	1,000	,960
	صورة المنتج	-,913	,974	,896	,450	,960	1,000
Sig. (1-tailed)	السعر		<,001	,007	,400	<,001	<,001
	الذوق	,001		,000	,081	,000	,000
	الحجم	,007	,000		,078	,002	,000
	التوفر	,400	,081	,078		,232	,096
	التغليف	,000	,000	,002	,232		,000
	صورة المنتج	,000	,000	,000	,096	,000	

a. Determinant = 1,374E-6

يتضح من الجدول بأننا حصلنا على مصفوفة معاملات الارتباطات البينية، التي تعد الحل الأولي للعلاقات بين المتغيرات الداخلة في التحليل العاملي. ونلاحظ أن قيمة المحدد تساوي 0.000001374 (Determinant = 1,374E−6) أقل بكثير من 0.0001 وبالتالي يتوجب علينا أن ننظر في المتغيرات التي قيمة معامل الارتباط بينها أعلى من 0.80 ونحذف إحداها، ثم نعيد التحليل. يمكننا من مصفوفة الارتباط ملاحظة أنّ معامل الارتباط بين الحجم والذوق مرتفع جدا (0.96)، ودال احصائيا، وكذلك الحال بين التغليف والسعر وصورة المنتج والسعر (ارتباط عكسي ذو دلالة احصائية). وبين صورة المنتج والذوق...الخ.

الجدول الثالث: قياس جودة الاختبار، ومدى صلاحية البيانات للتحليل العاملي.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,713
Bartlett's Test of Sphericity	Approx. Chi-Square	83,235
	df	15
	Sig.	,000

يتضح من الجدول بأننا قد حصلنا على قيمة قياس KMO وهي أكبر من 0.5 وهذا يدل على زيادة الاعتمادية للعوامل التي نحصل عليها من التحليل العاملي، وكذلك نحكم بكفاية حجم العينة، كما نجد أن قيمة مستوى الدلالة لاختبار بارتلت للدائرية تساوي 0.0001 وهي أقل من 0.05 وهذا يؤكد على وجود علاقة دالة إحصائيا بين المتغيرات ومصفوفة الارتباط ليست بمصفوفة الوحدة، وبذلك يمكن إجراء التحليل العاملي.

الجدول الرابع: مصفوفة الارتباطات الجزئية.

Anti-image Matrices

		السعر	الذوق	الحجم	التوفر	التغليف	صورة المنتج
Anti-image Covariance	السعر	,026	,005	-,015	-,048	,003	,006
	الذوق	,005	,009	-,015	-,005	,006	-,004
	الحجم	-,015	-,015	,034	,024	-,005	,002
	التوفر	-,048	-,005	,024	,134	,019	-,020
	التغليف	,003	,006	-,005	,019	,030	-,010
	صورة المنتج	,006	-,004	,002	-,020	-,010	,007
Anti-image Correlation	السعر	,712[a]	,316	-,508	-,815	,122	,441
	الذوق	,316	,745[a]	-,850	-,138	,380	-,512
	الحجم	-,508	-,850	,735[a]	,358	-,150	,104
	التوفر	-,815	-,138	,358	,358[a]	,299	-,657
	التغليف	,122	,380	-,150	,299	,816[a]	-,699
	صورة المنتج	,441	-,512	,104	-,657	-,699	,728[a]

a. Measures of Sampling Adequacy(MSA)

تبيّن لنا هذه المصفوفة مدى كفاية كل متغير للتحليل العاملي، لذلك يجب أن تكون الارتباطات الجزئية لكل متغير في القطر أكبر من 0.5، وكل متغير أقل من هذه القيمة علينا حذفه. في حالتنا علينا حذف متغير التوفر (0.358)، واعادة التحليل.

الجدول الرابع: الاشتراكيات (الشيوع):

Communalities

	Initial	Extraction
السعر	1,000	,816
الذوق	1,000	,977
الحجم	1,000	,866
التوفر	1,000	,206
التغليف	1,000	,907
صورة المنتج	1,000	,989

Extraction Method: Principal Component Analysis.

القاعدة أنّه كلما كانت قيمة الاستخلاص (Extraction) أكبر من 75 بالمائة، فذلك دليل على أنّها تمثل البيانات تمثيلا جيدا. مثلا الذوق لديه أكبر تشبع (0.977).

الجدول الخامس: جدول التباين المفسّر، وهو من أهم الجداول لدينا.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,761	79,343	79,343	4,761	79,343	79,343
2	,989	16,487	95,831			
3	,202	3,365	99,196			
4	,035	,576	99,772			
5	,009	,151	99,923			
6	,005	,077	100,000			

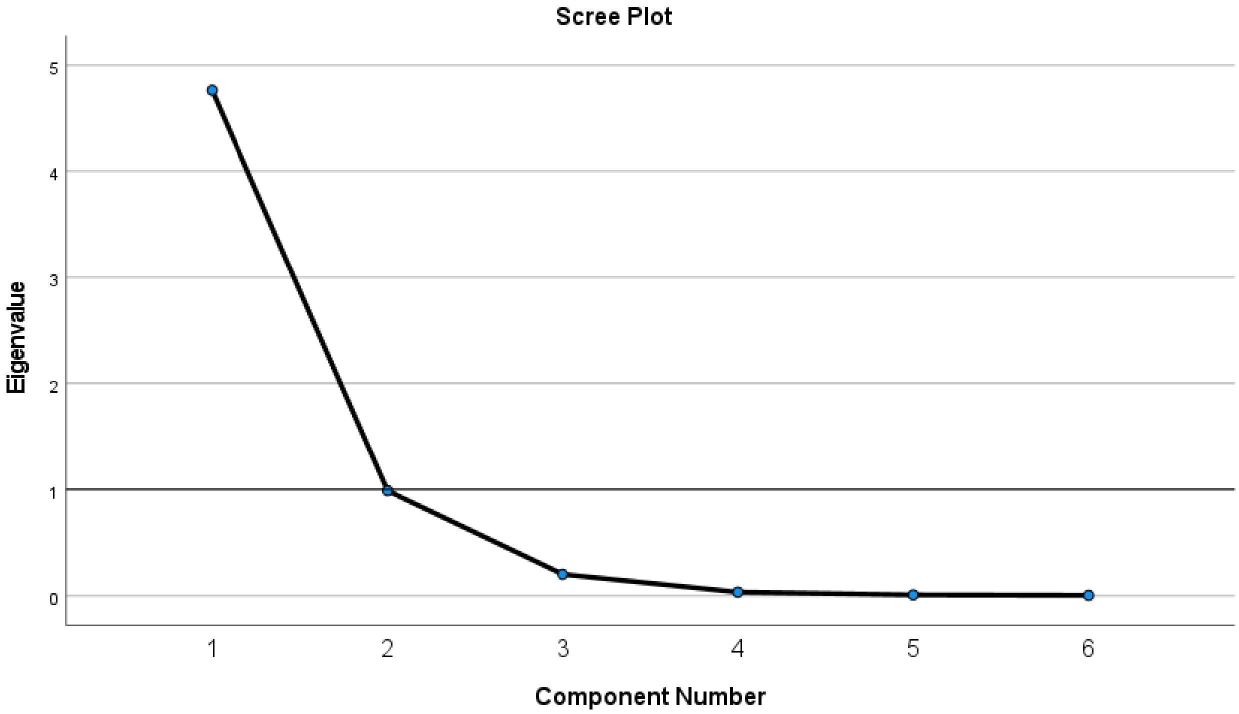
Extraction Method: Principal Component Analysis.

يفسّر هذا الجدول العوامل، وأيّها أكثر تأثيرا، من خلال قيمة Eigenvalues (الجذر الكامن)، فكلما كانت قيمة هذا الأخير أكبر من واحد (1) كان المكوّن المتحصل عليه يساهم في تفسير التباين بين المتغيرات، ونلاحظ أنّ Eigenvalues للعامل الأول يساوي 4.761 وهو أكبر من واحد، ونسبة تفسيره للتباين 79.34 بالمائة. وعلى ذلك، واستنادا لمحك كايزر فقد تم استخراج العامل الأول فقط، واستبعدت بقية العوامل.

يمثل العمود الأول من الجدول عدد العوامل المستخلصة، 6 عوامل، ويحتوي العمود الثاني على قيمة الجذر الكامن لكل عامل، فيما يعرض العمود الثالث نسبة تفسير التباين لكل عامل (النسبة 79.34 تحسب بقسمة 4.761 على 6). ويعرض العمود الرابع على التكرار النسبي التجميعي.

ملاحظة مهمة: تم عرض العوامل قبل التدوير فقط، ولم يعرض العوامل بعد التدوير، بسبب أنّ هناك عامل واحد فقط قد تم استخلاصه.

ملاحظة: عدد الجذور الكامنة يساوي عدد المتغيرات.



إن الرسم البياني يمثل قيم الجذور الكامنة لكل عامل على المحور الصادي ورقم المكون على المحور السيني، ويعتبر الرسم البياني معيارا آخر يمكن استخدامه بالإضافة إلى معيار الإبقاء على العوامل التي يزيد جذرها الكامن عن الواحد الصحيح لتحديد العوامل في التحليل العاملي والإبقاء فقط على تلك التي تكون في المنطقة شديدة الانحدار. لاحظ أنّ هناك عامل واحد فقط جذره الكامن أكبر من واحد، ويوجد عامل ثاني يقع في المرفق، ويساوي الواحد تقريبا، يمكننا أخذه إذا اعتمدنا على محكات أخرى بخلاف محك كايزر.

ولو عدنا إلى ملف الداتا لدينا سنجد أن البرنامج قد استخرج العامل الأول وأنشأ له متغير جديد (FAC1_1,).

FAC1_1
1,40399
,98590
-,01768
,86402
,65855
-,12103
-1,24162
-,89683
-,04998
-1,58531

الجدول السادس: مصفوفة العوامل قبل التدوير.

Component Matrix[a]

	Component
	1
صورة المنتج	,995
الذوق	,988
التغليف	,952
الحجم	,930
السعر	-,903
التوفر	,454

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

يمكننا من خلال هذا الجدول إلقاء نظرة أولية على تمثيل المتغيرات في العامل المستخرج، القاعدة أنّ كل متغير تقترب قيمة تشبّعه على العامل من 1 يكون أكبر تأثيرا في ذلك العامل. مثلا متغير صورة المنتج قيمة تشبّعه على العامل 0.995، وكذلك متغيرات: الذوق والتغليف والحجم والسعر، فكلها قيم تشبّعها على العامل الأول مرتفعة، ماعدة متغير التوفر فقيمة تشبعه ضعيفة نوعا ما. كما تشير القيم السالبة إلى ارتباط سالب للمتغير مع العامل.

الجدول السابع: مصفوفة الارتباط المقدرة.

Reproduced Correlations

		السعر	الذوق	الحجم	التوفر	التغليف	صورة المنتج
Reproduced Correlation	السعر	,816[a]	-,893	-,841	-,410	-,860	-,899
	الذوق	-,893	,977[a]	,920	,448	,941	,983
	الحجم	-,841	,920	,866[a]	,422	,886	,925
	التوفر	-,410	,448	,422	,206[a]	,432	,451
	التغليف	-,860	,941	,886	,432	,907[a]	,947
	صورة المنتج	-,899	,983	,925	,451	,947	,989[a]
Residual[b]	السعر		,034	,099	,318	-,093	-,014
	الذوق	,034		,041	,031	-,037	-,009
	الحجم	,099	,041		,062	-,078	-,030
	التوفر	,318	,031	,062		-,169	-,001
	التغليف	-,093	-,037	-,078	-,169		,013
	صورة المنتج	-,014	-,009	-,030	-,001	,013	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 6 (40,0%) nonredundant residuals with absolute values greater than 0.05.

الجدول الثامن: مصفوفة العوامل بعد التدوير.

Rotated Component Matrix[a]

a. Only one component was extracted. The solution cannot be rotated.

بسب استخلاص عامل واحد فقط لن تتم عملية التدوير.

الجدول التاسع:

Component Score Coefficient Matrix

	Component
	1
السعر	-,190
الاوق	,208
الحجم	,195
التوفر	,095
التغليف	,200
صوره المنتج	,209

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

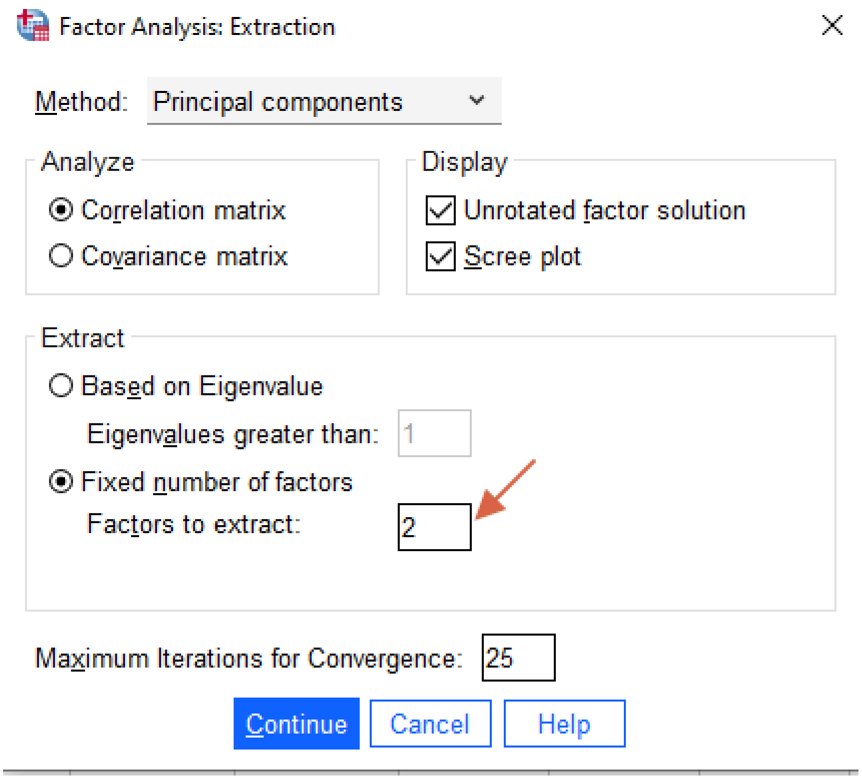
الجدول العاشر:

Component Score Covariance Matrix

Component	1
1	1,000

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

وكما أسلفنا، وبالنظر إلى معيار التمثيل البياني للقيم الذاتية، يمكننا أخذ عامل ثاني. نعيد التحليل بنفس الطريقة، من مربع الحوار Extraction فقط سنحدد عدد العوامل المستخلصة ونجعلها 2.



ونعود لمربع الحوار الرئيسي ونضغط موافق. لن يتغير أي شيء في أغلب الجداول فقط سيكون التغيير في بعضها، كما يلي:

جدول التباين المفسّر.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,761	79,343	79,343	4,761	79,343	79,343	4,338	72,305	72,305
2	,989	16,487	95,831	,989	16,487	95,831	1,412	23,525	95,831
3	,202	3,365	99,196						
4	,035	,576	99,772						
5	,009	,151	99,923						
6	,005	,077	100,000						

Extraction Method: Principal Component Analysis.

نلاحظ أنّه عند تحديد عدد العوامل المستخلصة بـ 2 للعامل، العاملين معًا يفسّران 95.83 من التباين. وقد عرض لنا البرنامج العوامل بعد التدوير، أي أنّ عملية التدوير قد تمت.

ولو عدنا إلى ملف الداتا لدينا سنجد أن البرنامج قد استخرج العاملين الأول والثاني وأنشأ لهما متغيرين جديدين (FAC1_1, FAC2_1).

FAC1_1	FAC2_1
1,06555	1,19499
,78926	,72361
,04030	-,16633
1,04036	-,34775
,87057	-,48365
-,53365	1,14117
-1,51886	,56685
-1,12025	,47467
,50574	-1,57362
-1,13902	-1,52993

مصفوفة العوامل قبل التدوير:

Component Matrix[a]

	Component	
	1	2
صورة المنتج	,995	
الذوق	,988	
التغليف	,952	
الحجم	,930	
السعر	-,903	,390
التوفر	,454	,876

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

كما عرفنا سابقا، القاعدة أنّ كل متغير تقترب قيمة تشبّعه على العامل من 1 يكون أكبر تأثيرا في ذلك العامل، فمثلا متغير صورة المنتج قيمة تشبّعه على العامل الأول 0.995 وغير متشبع على العامل الثاني، فيكون في العامل الأول، وكذلك العوامل: الذوق والتغليف والحجم، كلها قيم تشبّعها على العامل الأول أفضل من الثاني. بينما نلاحظ أنّ متغير السعر متشبع على المحور الأول، لكن بقيمة سالبة، القيمة السالبة هنا تعني أنّ هذا المتغير له ارتباط عكسي مع العامل. أمّا متغير التوفّر فإنّه متشبّع على العامل الثاني أفضل من الأول.

مصفوفة العوامل المدورة:

Rotated Component Matrix[a]

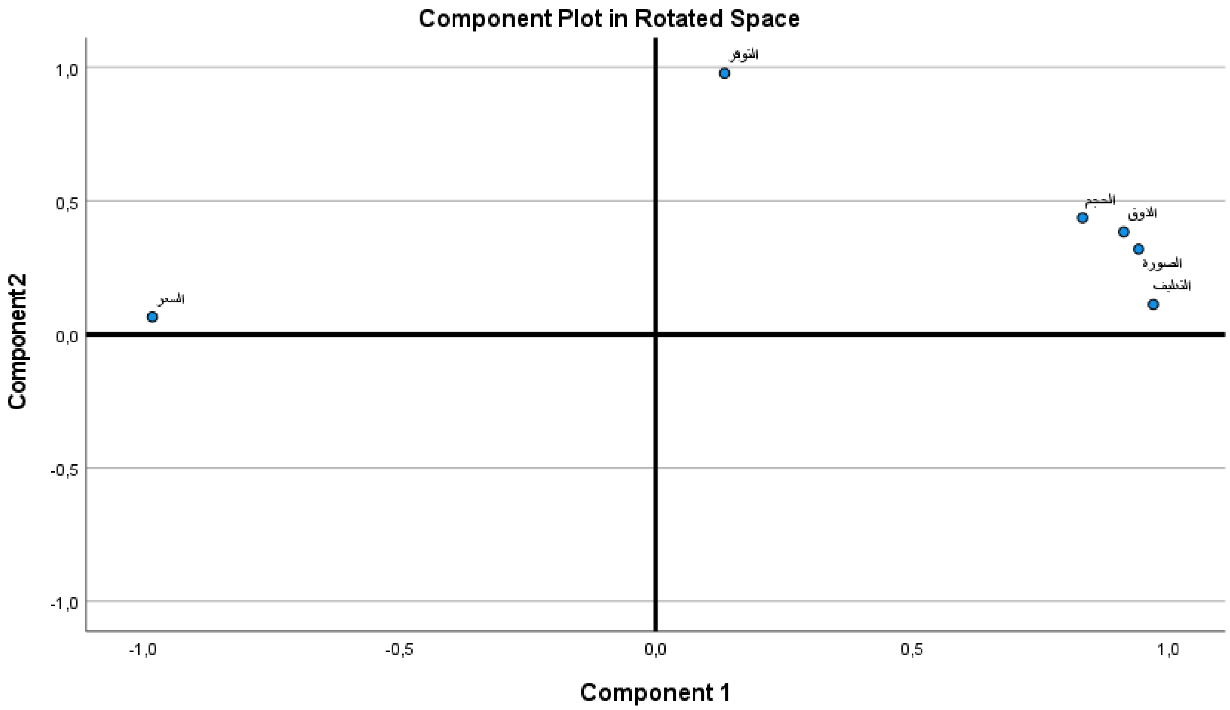
	Component	
	1	2
السعر	-,982	
التغليف	,971	
صورة المنتج	,942	,320
الذوق	,913	,383
الحجم	,833	,436
التوفر		,978

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

يمكننا من خلال هذه المصفوفة ملاحظة أنه فيما عدى متغير التوفر الذي يتشبع على العامل الثاني أفضل من العامل الأول، بقية المتغيرات تتشبع على العامل الأول.

المحاور العاملية:



لم يتم انشاء هذا الرسم البياني عند اختيار عامل واحد (مركبة أساسية واحدة)، لكن عند اختيار عاملين تم انشاءه. ويمكننا من خلاله اتخاذ القرار بشأن توزيع منتجاتنا في السوق.

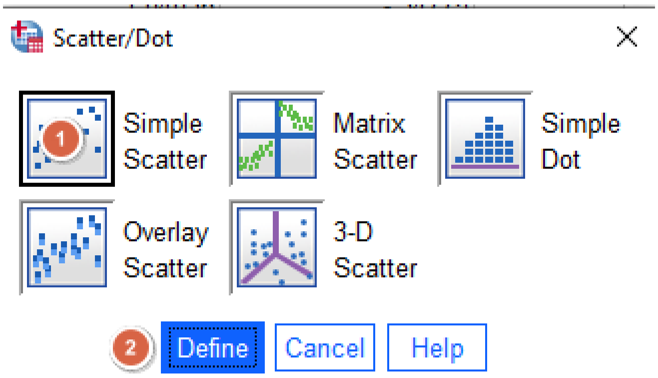
يمكننا ملاحظة أنّ جميع المتغيرات، ما عدى متغير التوفر، قريبة من الخط (المحور الأول) الذي يمثل العامل الأول (المركبة الرئيسية الأولى)، ومتغير التوفر قريب من الخط (المحور الثاني) الذي يمثل العامل الثاني. لكن ماذا تعني لنا هذه النتائج؟ ماذا يعرض لنا المحور الأول وماذا يعرض المحور الثاني؟

لاحظ تموضع مواصفات المنتجات الأربعة (الحجم، الذوق، الصورة، التغليف) حول المحور الأول في جهة اليمين، ويقابله في جهة اليسار السعر، فالمحور الأول يعرض جودة المنتجات مقارنة بالسعر، هذا يعني أنّ المنتجات التي تحمل المواصفات الأربعة السابقة، منتجات بمواصفات أو جودة عالية، وسعرها مرتفع، وعلى المؤسسة توزيع تلك المنتجات أو بيعها في مناطق الأغنياء. بينما المنتجات التي تتموضع في يسار المحور الأول، فهي منتجات متوسطة أو قليلة الجودة، وسعرها منخفض.

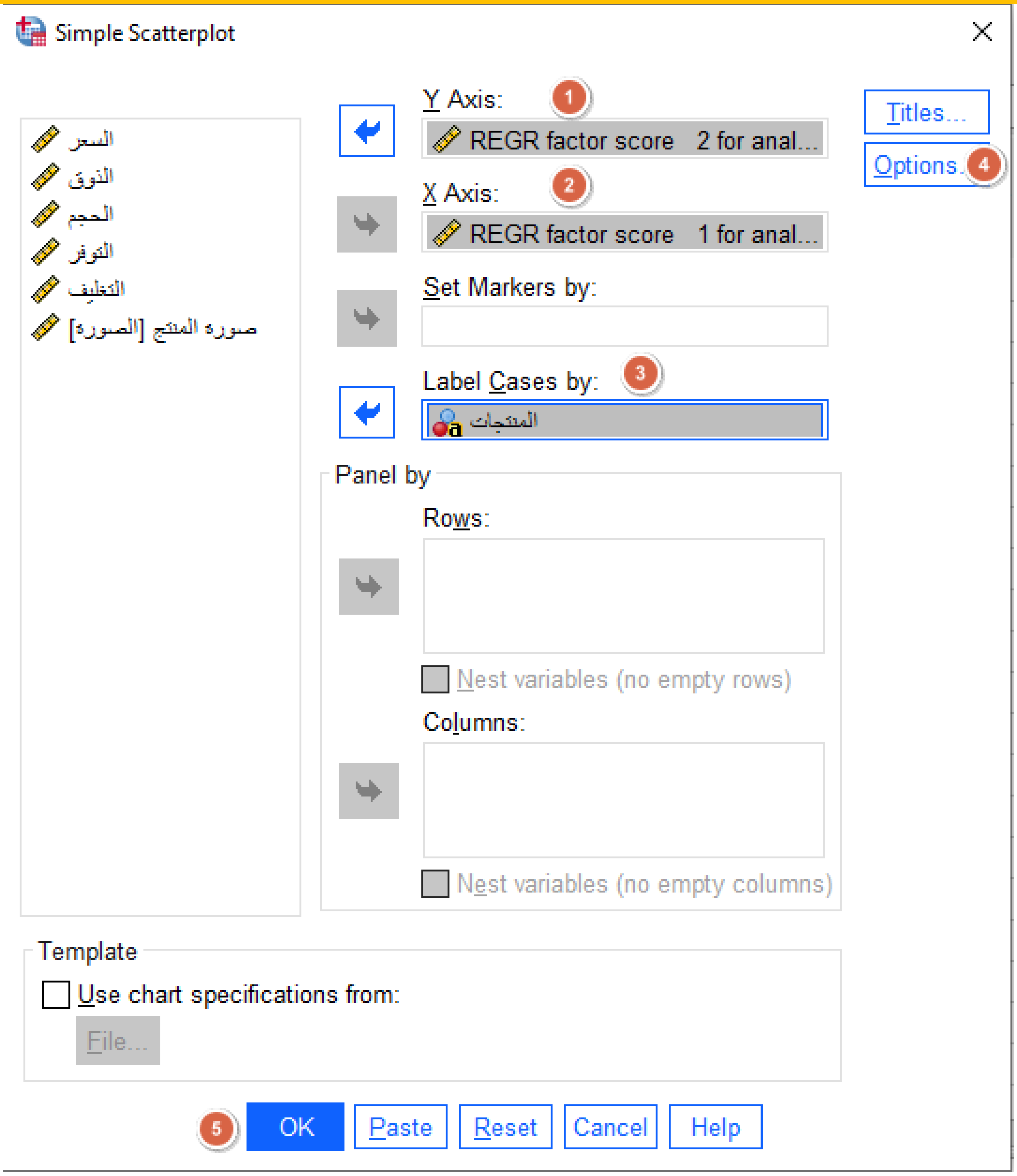
أمّا المحور الثاني فيعرض لنا التوفّر، فالمنتجات التي تكون أعلى هذا المحور (قريب من مواصفة التوفر) هي منتجات متوفرة في السوق ويمكن الحصول عليها بسهولة، والمنتجات التي تتموضع أسفله هي منتجات غير متوفرة في السوق ولا يمكن الحصول عليها بسهولة.

ملاحظة: برنامج SPSS ينشأ فقط رسما للمتغيرات في فضاء الأفراد ولتحديد الأفراد في فضاء المتغيرات، في مثالنا المنتجات (تموضع المنتجات وكيفية توزيعها في السوق على حسب مواصفات كل منتج وسعره)، نقوم برسم المنحنى البياني الخاص بالمنتجات، من خلال عمل الآتي:

Graphs>Legacy Dialogs>Scatter/Dot



نختار Simple Scatter ثم Define.

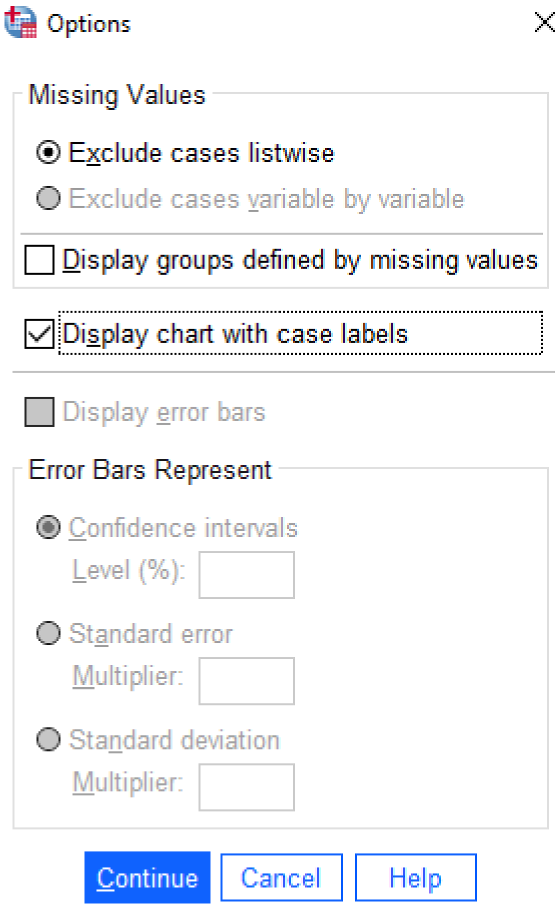


1- ننقل العامل Fac2_2 إلى محور العينات (Y).

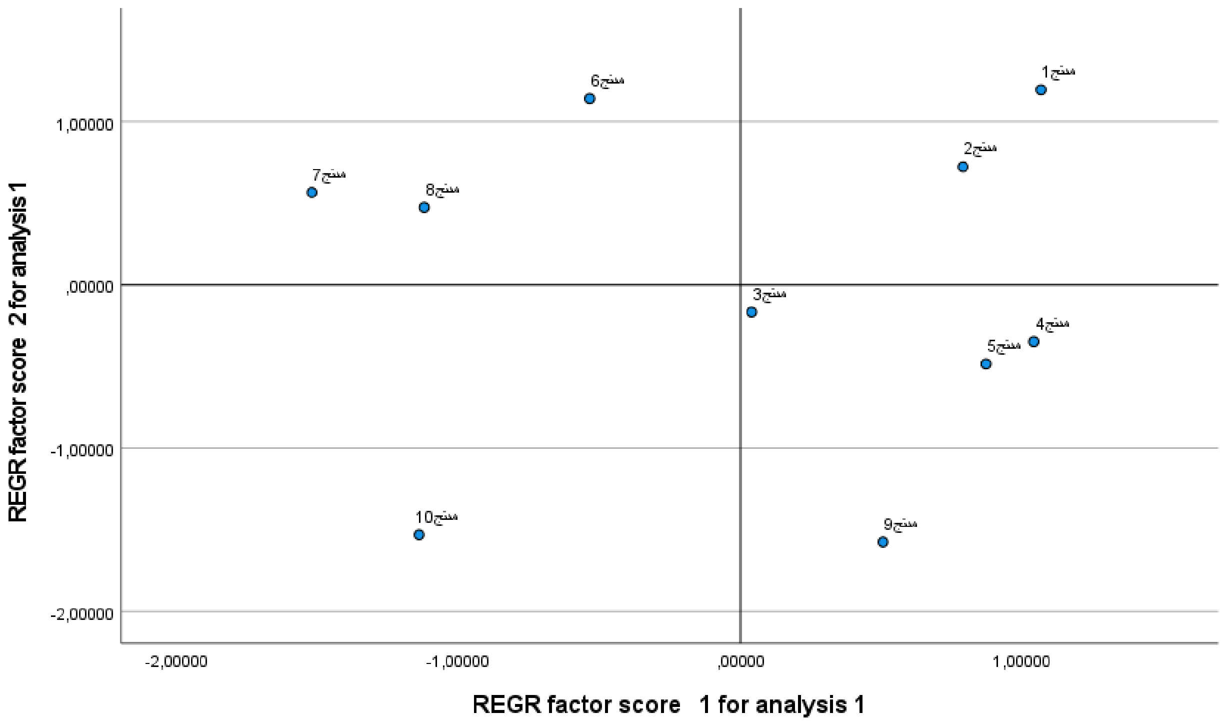
2- ننقل العامل Fac2_2 إلى محور السينات (X).

3- ننقل متغير المنتجات إلى خانة Label Cases by (تسمية الحالات بواسطة).

4- نضغط Options ونختار Display chart with case labels ثم Continue.


Options
Missing Values
Exclude cases listwise
Exclude cases variable by variable
Display groups defined by missing values
Display chart with case labels
Display error bars
Error Bars Represent
Confidence intervals
Level (%):
Standard error
Multiplier:
Standard deviation
Multiplier:
Continue
Cancel
Help

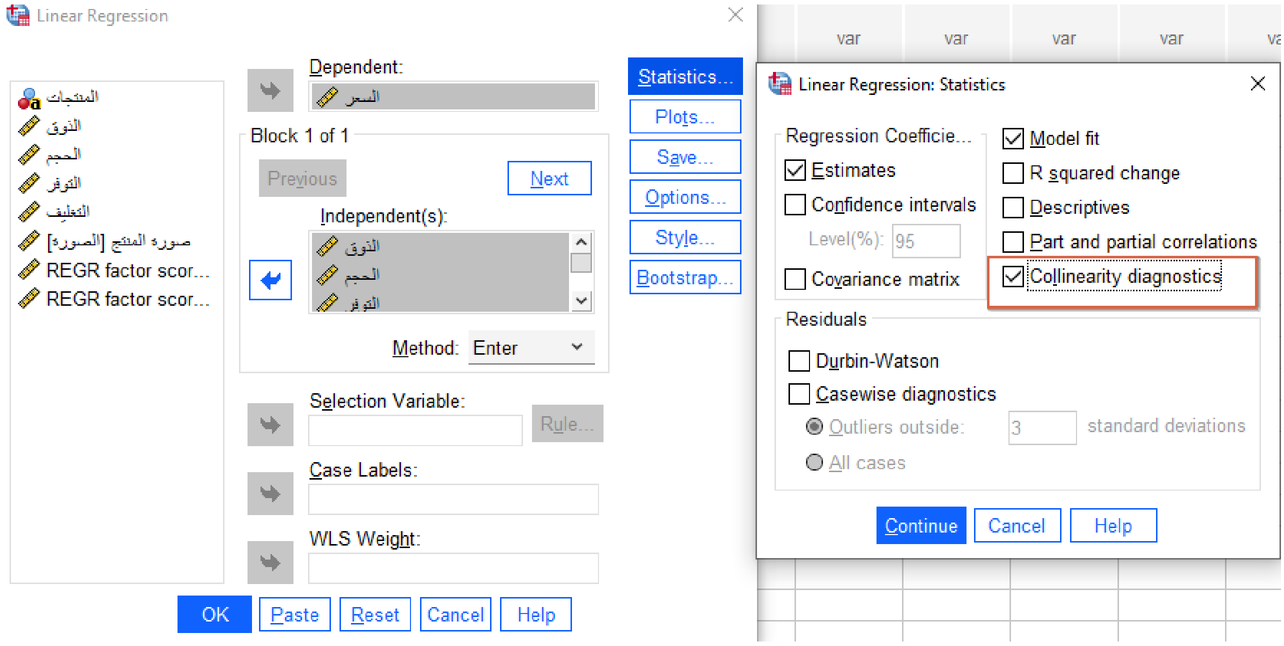
5- نضغط ok.


منتج1
منتج2
منتج3
منتج4
منتج5
منتج6
منتج7
منتج8
منتج9
منتج10
1,00000
,00000
-1,00000
-2,00000
-2,00000
-1,00000
,00000
1,00000
REGR factor score 2 for analysis 1
REGR factor score 1 for analysis 1

لو قمنا بمطابقة الرسمين البيانيين، يمكننا ملاحظة أنّ المنتجات: الأول والثاني والرابع والخامس تتموضع على يمين المحور الأول، مما يدل على أنها منتجات بجودة عالية، وسعرها مرتفع. بينما المنتجات 7،8،10 تتموضع يسار المحور الأول، مما يعني أنها منتجات أقل جودة، وأسعارها منخفضة. والمنتج رقم 6 هو المنتج الأكثر توفّرًا في السوق (لاحظ أنّه قريب من مكان وجود خاصية التوفّر). والمنتج التاسع غير متوفّر في السوق، أمّا المنتج الثالث فهو غير ممثل تمثيلا جيد في التحليل، بمعنى أنه منتج غير مرغوب فيه.

ملاحظة: تساعدنا طريقة المكونات الأساسية في التخلص من مشكلة التعددية الخطية بين المتغيرات، أثناء استخدام تحلي الانحدار المتعدد، فلو اختبرنا تحليل الانحدار بين السعر ومواصفات المنتجات العشرة (الذوق، الحجم، التوفر، التغليف، الصورة).

Analyze>Regression>Linear…



النتيجة:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,987[a]	,974	,942	,38431

a. Predictors: (Constant), الذوق, التغليف, الحجم, التوفر, صورة المنتج

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22,359	5	4,472	30,277	,003[b]
	Residual	,591	4	,148		
	Total	22,950	9			

a. Dependent Variable: السعر

b. Predictors: (Constant), الذوق, التغليف, الحجم, التوفر, صورة المنتج

نلاحظ أن هناك ارتباط قوي جدا بين السعر ومواصفات المنتج، بلغ معامل الارتباط 0.987، كما أنّ المواصفات الخمسة للمنتج تفسّر 97 بالمائة من تغير السعر ارتفاعا أو انخفاضا، بمعنى أنّه إذا كانت المواصفات جيدة فالسعر يكون مرتفع كثيرا والعكس صحيح. ويظهر من جدول تحليل التباين ANOVA أنّ الاختبار دال إحصائيا عند مستوى دلالة 0.01 فأقل (P=0.003). لنرى الآن إن كان هناك تعدد خطي بين المتغيرات المستقلة (مواصفات المنتج)، والجدول الموالي يوضّح النتائج.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	7,462	1,033		7,223	,002		
	الذوق	-,422	,633	-,528	-,667	,541	,010	97,432
	الحجم	,531	,450	,444	1,179	,304	,045	21,995
	التوفر	,361	,128	,358	2,813	,048	,398	2,513
	التغليف	-,085	,347	-,113	-,245	,819	,030	33,373
	صورة المنتج	-,634	,645	-,848	-,982	,382	,009	115,983

a. Dependent Variable: السعر

لاحظ قيم معامل تضخم التباين VIF (variance inflation factor)، إذا كانت قيم هذا المعامل بين 5 و10، فذلك يشير إلى أنّ الارتباط عالي وقد يكون مشكلة، وإذا تجاوزت قيمته 10، فيمكنك افتراض أن معاملات الانحدار يتم تقديرها بشكل سيء بسبب التعدد الخطي. وفي مثالنا هذا أغلبية قيم معامل تضخم التباين عالية، ما عدى القيمة المقابلة لصفة التوفر، مما يدل على وجود تعددية خطية بين المتغيرات.

نجري الآن تحليل الانحدار المتعدد بين السعر والعوامل التي استخلصناها بطريقة المكونات الأساسية، بنفس الطريقة، فقط في مكان المتغيرات المستقلة (المواصفات) نختار العوامل المستخلصة، وهما العاملين FAC1_1 و FAC2_1.

النتيجة:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,984[a]	,968	,959	,32373

a. Predictors: (Constant), REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22,216	2	11,108	105,993	,000[b]
	Residual	,734	7	,105		
	Total	22,950	9			

a. Dependent Variable: السعر

b. Predictors: (Constant), REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

لاحظ أنّ معاملي الارتباط والتفسير تقريبا نفسهما، كما أنّ الاختبار معنوي عند مستوى دلالة 0.0001.

لنرى الآن قيم معامل تضخم التباين في الجدول الموالي.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	6,039	,102		58,991	,000		
	REGR factor score 1 for analysis 1	-1,443	,108	-,903	-13,369	,000	1,000	1,000
	REGR factor score 2 for analysis 1	,622	,108	,390	5,767	,001	1,000	1,000

a. Dependent Variable: السعر

لاحظ أنّ قيم معامل تضخم التباين VIF تساوي واحد، وهي أقل من 5، مما يعني عدم وجود تعددية خطية.

تمارين مقترحة:

1. يحتوي الجدول الموالي على معطيات تقييم ستة أفراد لثلاثة علامات من الشكلاطة (جيد=2، متوسط=1، رديئ=0).

	A	B	C
فريد	2	1	0
علي	1	1	0
حنان	0	0	1
وسيم	2	0	1
أمين	1	0	1
نرجس	0	1	0

المطلوب: إجراء التحليل بالمركبات الرئيسية على الجدول مع التمثيل البياني.

2. إليك مجموعة من صور شخصيات كرتونية (أبطال خارقين).

السؤال المطروح عليك، أي واحد من هؤلاء الأبطال الخارقين تشعر أنّه يشبهك؟ ضع اجاباتك في الجدول الموالي مستعينا بالاقتراحات التالية:

- لديه قوى خارقة (0 أو 1)
- يرتدي لباس متميز (تصنيف بين 1 و 3)
- يعمل في فريق (التصنيف بين 1 و 10)
- لديه معدات خاصة (تصنيف بين 1 و 10)
- ذكر / أنثى (1 أو 0)

	لديه قوى خارقة	يرتدي لباس متميز	يعمل في فريق	لديه معدات خاصة	الجنس
Superman					
Batman					
Spiderman					
Hulk					
Ironman					
Catwoman					
x-or					
Daredevil					
Wonderwoman					
Bioman					
x-men					
Tortues Ninja					

بعد ملأ الجدول قم بتحليله بواسطة المركبات الأساسية.

3. في دراسة للجودة في أربعة مطاعم تتضمن جودة الخدمة والطعام والسعر، طلب من خبير في التذوق تقييم المطاعم الأربعة على مقياس من −3 إلى 3. وقد تم جمع البيانات في الجدول الموالي.

	الخدمة	الطعام	السعر
R1	−2	3	−1
R2	−1	1	0
R3	2	−1	−1
R4	1	−3	2

المطلوب: إجراء التحليل بالمركبات الرئيسية.

الفصل الرابع: التحليل العاملي للتوفيقات

التحليل العاملي للتوفيقات

تعد طريقة التحليل بالمعاملات للتوفيقات (Correspondence Analysis) ضمن أهم الطرائق العاملية لاستكشاف البيانات ومن أبرزها في التحليل المتعدد الأبعاد، وأكثرها انتشارا وشيوعا في الأبحاث والدراسات العلمية، خاصة في مجال العلوم الانسانية، بسبب ارتباطها الوثيق بتحليل البيانات النوعية. وتشبه هذه الطريقة في تطبيقها طريقة التحليل بالمركبات الأساسية التي تستخدم لتحليل البيانات الكمية، غير أنها تدرس العلاقات بين متغيرين قياسهما نوعي فئوي أو ترتيبي.

وسيتم في هذا الفصل شرح طريقة عملها بشيء من التفصيل الذي يزيل اللبس والغموض الذي يكتنفها، عبر التطرق إلى خطواتها واحدة تلوى الأخرى، مستعينين في ذلك بأمثلة توضيحية وتمارين محلولة، إضافة إلى توضيح كيفية استخدام بعض البرامج الإحصائية المشهورة في إجراء التحليل.

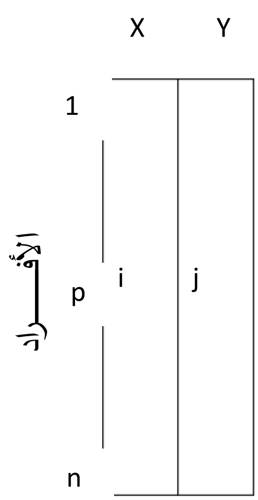
أولا، مفهوم التحليل العاملي للتوفيقات:

هو أسلوب تحليل عاملي مخصص لمعالجة جداول البيانات حيث تكون القيم موجبة ومتجانسة، مثل جداول التقاطع (التي تشكل الجزء الأكبر من الجداول التي تتم معالجتها بهذه الطريقة). وقد تم تقديم هذا النوع من التحليل في الستينيات من قبل الفرنسي بن زكري (J .P.Benzecri) خلال الفترة 1970–1990. ويهدف التحليل العاملي التقابلي إلى تحديد الروابط الممكنة بين المتغيرات النوعية، ويستخدم المبادئ المشتركة نفسها بين جميع طرق التحليل العاملي لحساب الروابط بين المتغيرات، والسماح بتقليل أبعاد البيانات لإعادتها مرة أخرى إلى مساحة تحفظ أو تشرح أقصى ما يمكن من المعلومات.

ويطبق التحليل العاملي للتوفيقات بشكل أساسي على جداول التوافيق (Contingency table). هذه الجداول تحتوي على أفراد عند تقاطع السطر i والعمود j. ويتعلق الأمر بتقسيم مجتمع أو عينة الدراسة وفقًا لأي حرفين، X في السطر وY في العمود مثلا، ولذلك فهي خصائص نوعية اسمية و/أو ترتيبية. وغالبًا ما تركز الدراسات التقليدية لمثل هذه الجداول على الاعتمادية (التوافق) أو الاستقلالية بين المتغيرين، ويتم ذلك عادةً باستخدام اختبار كاي تربيع.

وترتكـز الدارسـة في التحليل العاملي للتوفيقات حـول معطيـات كيفيـة مسـتمدة مـن دراسـة ظـاهرة معينـة، مشـكلّة مـن عدد p من مستويات التصنيف التي تمـثّل متغيرين نوعيين (X, Y) والحكم علـى مـدى التوافـق الموجـود بينهمـا.

لدينا العدد n من الأفراد ومتغيرين نوعيين (qualitative)، المتغير النوعي يحتوي على مستوى أو مجموعة من المستويات أو الخصائص، مثلا: (اللون: بني) هذا المتغير بخاصية واحدة؛ (اللون: بني، أحمر) هذا المتغير بخاصيتين، وهكذا. متغيرين (Y, X)، مع خاصيتين، (الخاصية p للمتغير X) (الخاصية q للمتغير Y).



المتغيرات

الأفراد

	X	Y
1	n_{11}	n_{12}
2	n_{21}	n_{22}
3	n_{31}	n_{32}
.....		
p	n_{p1}	n_{p2}
.....		
n	n_{n1}	n_{n2}

خصائص أو مستويات المتغير النوعي: هي مختلف القيم التي يمكن أن يأخذها المتغير، إذا عبرنا عنها بأرقام. فعلى سبيل المثال، خصائص متغير نوع السيارة رونو ومارسيدس، فإذا رمزنا لرونو بالرمز 1 ولمرسيدس بالرمز 2، فمتغير نوع السيارة يأخذ القيمة 1 أو 2.

يمكننا من خلال المعطيات بناء جدول يحتوي على متغيرين ومجموعة من الخصائص، هذا الجدول يساعدنا في دراسة سحابتين نقطيتين تصفان التقارب بين نقاط الأسطر (خصائص المتغير الأول) مع نقاط الأعمدة (خصائص المتغير الثاني).

ثانيا، جداول التوافيق أو التقاطع (Contingency table):

هو جدول يسجل تكرار كل تركيب ممكن من متغيرين أو أكثر. يستخدم هذا النوع من الجداول عادة في تحليل البيانات لاستكشاف العلاقة بين متغيرين فئويين. وعادةً ما يتم عرض المتغيرات كأعمدة وصفوف في الجدول، ويتم عرض التكرارات في خلايا الجدول.

فعلى سبيل المثال، يمكن استخدام جدول التوافيق لدراسة العلاقة بين الجنس وتفضيلات الأكل لدى البالغين، فيكون المتغيران هما الجنس وتفضيلات الأكل، مع الفئات الممكنة لكل متغير (على سبيل المثال، ذكر وأنثى للجنس، ولحم، نباتي، ونباتي صرف لتفضيلات الأكل). وتسجل التكرارات لكل تركيب ممكن من الجنس وتفضيلات الأكل في خلايا جدول التوافيق.

ويتم استخدام جدول التوافيق لحساب مقاييس الارتباط بين المتغيرات، مثل معامل ارتباط بيرسون أو معامل كاي تربيع للاستقلالية. كما يمكن استخدامه أيضًا لتحديد العلاقات بين المتغيرات ولتحديد الاختلافات بين المجموعات. ويتم الحصول على جدول التوافيق N من تقاطع المتغيرين $_1$ وX_2. ويكون على الشكل التالي:

جدول التوافيق

المتغير X بـ p من الخصائص (المستويات)

المتغير Y بعدد q من الخصائص (المستويات)

	1		j		q	$n_{i.}$
1	n_{11}		n_{1j}		n_{1q}	$n_{1.}$
i	n_{i1}		n_{ij}		n_{iq}	$n_{i.}$
p	n_{p1}		n_{pj}		n_{pq}	$n_{p.}$
$n_{.j}$	$n_{.1}$		$n_{.j}$		$n_{.q}$	$n_{..}$

n_{ij} تعبر عن عدد الأفراد الذين يحققون النموذج بالنسبة للخاصية i وفي نفس الوقت الخاصية j (تحدد عدد المفردات التي تشكل في نفس الوقت خصائص المتغيرين X Y). أو بتعبير آخر عدد التكرارات التي من أجلها يأخذ المتغير X القيمة $_i$ ويأخذ المتغير Y القيمة Y_j.

$n_{i.}$ تعبر عن مجموع أفراد السطر i، النقطة بعد الحرف i تعني مجموع الأفراد في السطر i.

$$n_{i.}=\sum_{j=1}^{q} n_{ij}$$

$n_{.j}$ تعبّر عن مجموع أفراد العمود j، النقطة قبل الحرف j تعني المجموع.

$$n_{.j}=\sum_{i=1}^{p} n_{ij}$$

$n_{..}$ تعبر عن مجموع الأفراد في الأسطر والأعمدة، أو بعبارة أخرى حجم العينة.

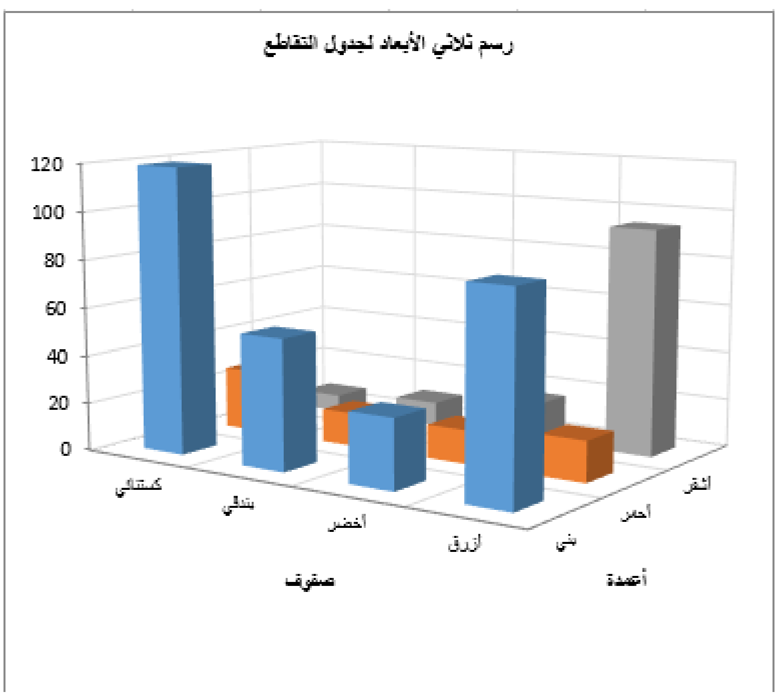
$$n_{..}=\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}$$

مثال: نحن مهتمون بالعلاقة بين لون العينين ولون الشعر لمجموعة إناث تتكون من 484 فردا. وجدول الاقتران أو التقاطع الموالي يبين تقاطع المتغيرين الوصفيين، لون العينين في الأسطر ولون الشعر في الأعمدة. حيث يشمل متغير لون العينين أربعة خصائص، هي: كستنائي، عسلي، أخضر، أزرق؛ في حين يشمل متغير لون الشعر ثلاثة خصائص، هي: بني، أحمر، أشقر.

الجدول رقم 1.

	بني	أحمر	أشقر
كستنائي	119	26	7
عسلي	54	14	10
أخضر	29	14	16
أزرق	84	17	94

سيكون من المثير للاهتمام معرفة ما إذا كانت العلاقات موجودة بين خصائص المتغيرين، وكذلك لعمل تمثيل مرئي للبيانات يسلط الضوء على أقوى الارتباطات.



يمكننا ملاحظة أنّ الأشخاص ذوي الأعين الزرقاء يملكون في أغلبهم شعرًا أشقرا، والأشخاص ذوي الشعر البني هم الأكثر بين من يملكون أعينًا كستنائية. نحن قادرون على رؤية هذا لأنّ الجدول صغير جدًا، وسيكون شرح أو وصف الرابط بين متغيرين نوعيين أكثر صعوبة عندما يكون عدد الخصائص أكبر أو عندما يكون المشروع استكشافيًا (أي ليس لدينا فكرة مسبقة عن الارتباط بين المتغيرات).

في التحليل العاملي للتوفيقات، سنستخدم فكرة المسافة بين الصفوف والمسافة بين الأعمدة لتوضيح جميع الخصائص على نفس الرسم البياني من أجل تصور العلاقات بين تلك الخصائص. لذلك سنسعى إلى تمثيل خصائص المتغيرين بشكل مرئي في نفس المستوى بحيث تكون خاصيتين مترابطتين إيجابيا، (خاصيتين يكون الرقم فيهما أكبر مما كان متوقعًا في ظل الاستقلال) قريبتين من بعضهما البعض.

ثالثا، عرض الطريقة:

1. بناء جدول التقاطع: لتشكيل جدول التقاطع (**Contingency table**) نظيف إلى الجدول السابق (الجدول رقم 1) عمودا على يمينه، تحتوي كل خلية منه على مجموع قيم السطر المقابل لها. ونضيف أسفله سطر تحتوي كل خلية منه على مجموع قيم العمود المقابل لها. ويمكن كتابة الجدول السابق على الشكل التالي:

جدول التقاطع

	بني	أحمر	أشقر	$n_{i.}$
كستنائي	119	26	7	**152**
عسلي	54	14	10	**78**
أخضر	29	14	16	**59**
أزرق	84	17	94	**195**
$n_{.j}$	**286**	**71**	**127**	**484**

العمود باللون الأصفر يطلق عليه هامش السطر (سمي كذلك لأنه عمود لكنه يحتوي على مجموع أرقام كل سطر. فعلى سبيل المثال، هامش السطر الأول= 152=7+26+119، ويحتوي على مجموع الأفراد بلون عينين كستنائي مهما كان لون شعرهم. والسطر باللون الأخضر يطلق عليه هامش العمود (سمي كذلك لأنه سطر لكنه يحتوي على مجموع أرقام كل عمود، فعلى سبيل المثال، هامش العمود الأول= 286=84+29+54+119، ويحتوي على مجموع الأفراد بلون الشعر بني مهما كان لون أعينهم.

ويمكن من جدول التقاطع ملاحظة أنّه لدينا متغير لون العينين X بأربعة خصائص (كستنائي، عسلي، أخضر، أزرق)، هذا المتغير فئوي. ولدينا متغير لون الشعر Y، هذا المتغير كذلك فئوي بثلاثة خصائص (بني، أحمر، أشقر). على سبيل المثال، لون العينين كستنائي خاصية من خصائص المتغير X ولون الشعر بني خاصية من خصائص المتغير Y، عدد الأفراد في العينة الذين يملكون الخاصيتين معًا 119 فردا، بمعنى آخر لدينا في العينة 119 فردًا لون أعينهم كستنائي ولون شعرهم بني. ويمكن التعبير عن الكلام السابق رياضيا على الشكل التالي:

$n_{11} = 119$

ويمكن كذلك اختصار الجدول السابق على شكل مصفوفة، كالآتي:

$$N^{*}=\begin{bmatrix} 119 & 26 & 7 \\ 54 & 14 & 10 \\ 29 & 14 & 16 \\ 84 & 17 & 94 \end{bmatrix}$$

نحسب المجموع الكلي:

$$n_{..}=\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}$$

$n_{..}= 484$

2. تحويل البيانات:

لأنّ طريقة التحليل العاملي للتوفيقات تعتمد اعتمادا كليا في تحليل البيانات على التكرارات النسبية، فإنّ استخدام جدول التقاطع لن يكون مفيدا. لذلك تعتبر المجاميع الهامشية على السطر والعمود جد مهمة، وعلى افتراض وجود ارتباط (علاقة) بين المتغيرين، فوصف هذه العلاقة يقتضي تحويل الجدول الابتدائي من التكرارات المطلقة إلى التكرارات النسبية. وعلى ذلك، فإنّ أول خطوة في التحليل هي تحويل البيانات، والعمل على جدول التكرارات النسبية، وجدولي الأسطر والأعمدة (Profils Lignes et Profils Colonnes).

1.2. جدول التكرارات النسبية: لبناء جدول التكرارات النسبية نحتاج إلى حساب التكرارا النسبي في كل خلية من خلايا جدول التقاطع، بقسمة قيمتها على المجموع الكلي، وفقا للصيغة الموالية:

$$f_{ij}=\frac{n_{ij}}{n_{..}}= P(X=i, Y=j)$$

حيث تعبر $_{ij}$ عن التكرار النسبي في كل خلية. أي نسبة الأفراد الذين يحققون الخاصية i للمتغير X والخاصية j للمتغير Y معا. كما يمكن كذلك اعتبارها احتمالا لتحقق الخاصيتين معا. ويكون جدول التكرارات النسبية على الشكل التالي:

	1		j		q	$f_{i.}$
1	f_{11}		f_j		f_{1p}	$f_{1.}$
i	f_{i1}		f_{ij}		f_{iq}	$f_{i.}$
p	f_{p1}		f_{pj}		f_{pq}	$f_{p.}$
$f_{.j}$	$f_{.1}$		$f_{.j}$		$f_{.q}$	$f_{..}$

تمثل $f_{i.}$ نسبة الأفراد الذين يحققون الخاصية i للمتغير X مهما كانت الخاصية j للمتغير Y. وتمثل $f_{.j}$ نسبة الأفراد الذين يحققون الخاصية j للمتغير Y مهما كانت الخاصية i للمتغير X. $f_{..}$ تمثل نسبة كل الأفراد في العينة المدروسة، وتساوي الواحد. كما أنّ مجموع التكرارات النسبية في هامش السطر يساوي الواحد، ومجموع التكرارات النسبية في هامش العمود يساوي الواحد، ونكتب:

$$\sum_{i=1}^{p} f_{ij}=f_{i.},\ \sum_{j=1}^{q} f_{ij}=f_{.j},\ f_{..}=\sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij}=1$$

وتكتب مصفوفة التكرارات النسبية على الشكل التالي:

$$N=\begin{pmatrix} f_{11} & \cdots & f_{1q} \\ \vdots & \ddots & \vdots \\ f_{p1} & \cdots & f_{pq} \end{pmatrix}=\frac{1}{n_{..}}*N^{*}\ldots\ldots\ N=\frac{1}{484}\begin{bmatrix} 119 & 26 & 7 \\ 54 & 14 & 10 \\ 29 & 14 & 16 \\ 84 & 17 & 94 \end{bmatrix}$$

$$N=\begin{bmatrix} 0.246 & 0.054 & 0.014 \\ 0.112 & 0.029 & 0.021 \\ 0.060 & 0.029 & 0.033 \\ 0.174 & 0.035 & 0.194 \end{bmatrix}$$

ويكون جدول التكرارات النسبية كالآتي:

جدول التكرارات النسبية

	بني	أحمر	أشقر	$f_{i.}$
كستنائي	0.246	0.054	0.014	0.314
عسلي	0.112	0.029	0.021	0.161
أخضر	0.060	0.029	0.033	0.122
أزرق	0.174	0.035	0.194	0.403
$f_{.j}$	0.591	0.147	0.262	1

وتحسب التكرارات الهامشية للأسطر والأعمدة كما يلي:

$$f_{i.}=\sum_{j=1}^{q} f_{ij}=P(X=i),\ f_{.j}=\sum_{i=1}^{p} f_{ij}=P(X=j)$$

$$\sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij}=\sum_{i=1}^{p} f_{i.}=\sum_{j=1}^{q} f_{.j}=1$$

$$f_{1.}=\sum_{j=1}^{3} f_{1j}=\left(\frac{119}{484}+\frac{26}{484}+\frac{7}{484}\right)=\frac{152}{484}=0.314$$

$$f_{.1}=\sum_{i=1}^{4} f_{i1}=\left(\frac{119}{484}+\frac{54}{484}+\frac{29}{484}+\frac{84}{484}\right)=\frac{286}{484}=0.591$$

يظهر من الجدول أعلاه مثلا، أنّ نسبة 24.6 في المائة من الأفراد في العينة المدروسة لون أعينهم كستنائي ولون شعرهم بني، كما أنّ ما نسبته 31.4 في المائة من الأفراد ذوي أعين كستنائية اللون مهما كان لون شعرهم، ونسبة 59.1 في المائة من الأفراد لون شعرهم بني مهما كان لون أعينهم، وهذا ما يطلق عليه الاحتمال الشرطي.

2.2. جدول التكرارات النسبية للأسطر: ويسمى كذلك بملف تعريف الأسطر، ونحصل عليه بقسمة كل قيمة في جدول التقاطع على إجمالي العمود المقابل، أو بقسمة كل قيمة في جدول التكرارات النسبية على إجمالي العمود المقابل، بتعبير آخر قسمة كل سطر i في المصفوفة N على ثقله $(f_{i.})$. مما يعني أننا سنتعامل مع الجدول كأسطر فقط، والقيمة $\frac{f_{ij}}{f_{i.}}$ تمثل احتمال الحصول على الخاصية j من المتغير Y مع العلم أن الخاصية i من المتغير X محققة. ونكتب:

$$f_{j/i}=P(Y=j,\ X=i)=\frac{P(Y=j,X=i)}{P(X=i)}=\frac{f_{ij}}{f_{i.}}\ /\ \sum_{j=1}^{q} f_{i/j}=1$$

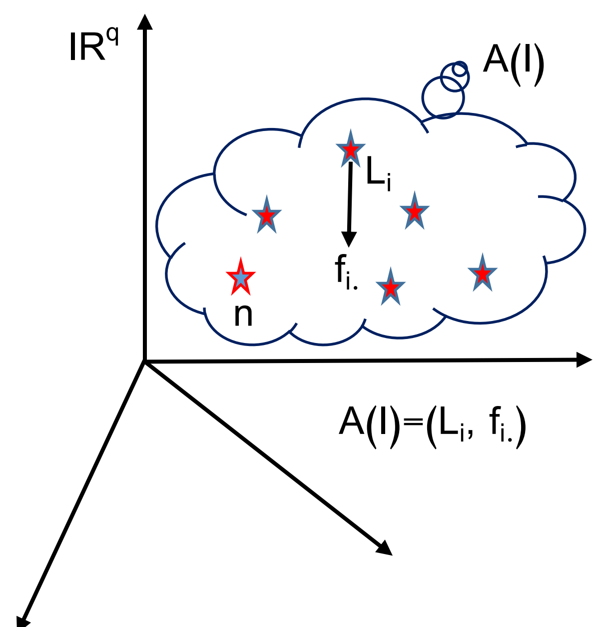
هذا الجدول يمكن كذلك تمثيله بـ p من النقط، ونكتب:

$$L_i=\left(\frac{f_{i1}}{f_{i.}},\frac{f_{i2}}{f_{i.}},\dots\frac{f_{ij}}{f_{i.}},\dots\frac{f_{iq}}{f_{i.}}\right)\dots\dots\dots\dots/(i=1\dots\dots,\dots,p).$$

$$L_1==\left(\frac{f_{11}}{f_{1.}},\frac{f_{12}}{f_{1.}},\frac{f_{13}}{f_{1.}}\right)=\left(\frac{0.246}{0.314},\frac{0.053}{0.314},\frac{0.014}{0.314}\right)=(0.783,0.169,0.046\ ).$$

في فضاء ذو بعد q (R^q) كل نقطة يتم وزنها بالاحتمال $_{i.}$.

$$L_i=\left(\frac{f_{i1}}{f_{i.}},\frac{f_{i2}}{f_{i.}},\dots\frac{f_{ij}}{f_{i.}},\dots\frac{f_{iq}}{f_{i.}}\right)$$



كل سطر يتم تمثيله في الفضاء R^q، كل نقطة L_i مرتبطة بالوزن $f_{i.}$، هذا الوزن يتناسب وعدد المفردات في السطر $L_{i.}$. وبناء على المعطيات السابقة يكون جدول التكرارات النسبية للأسطر في مثالنا على النحو الآتي:

جدول التكرارات النسبية للأسطر

	بني	أحمر	أشقر	المجموع
كستنائي	0,783	0,169	0,046	1
عسلي	0,692	0,179	0,128	1
أخضر	0,492	0,237	0,271	1
أزرق	0,431	0,087	0,482	1
المتوسط	0.591	0.147	0.262	1

يظهر من السطر الثاني، العمود الأول، أنّ 69٪ من الأشخاص ذوي الأعين العسلية لديهم شعر بني. كما أنّ 59.1٪ من الأفراد لديهم شعر بني. كما يعبر السطر باللون الأخضر عن المتوسطات الهامشية، التي تمثل مركز سحابة النقاط في الفضاء R^q، ويسمى هذا السطر كذلك بالسطر المتوسط L_m.

$$\sum_{i=1}^{p} \frac{n_{i.}}{n} * \frac{n_{ij}}{n_{i.}} = \frac{n_{.j}}{n}$$

$L_m = [f_{.1} \quad f_{.j} \quad f_{.q}] = [0.591 \quad 0.147 \quad 0.262]$

ونحدد المصفوفة القطرية للتكرارات النسبية الهامشية للأسطر على الشكل التالي:

$$D_r = \begin{pmatrix} f_{1.} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{p.} \end{pmatrix} = \begin{bmatrix} 0.314 & 0 & 0 & 0 \\ 0 & 0.161 & 0 & 0 \\ 0 & 0 & 0.122 & 0 \\ 0 & 0 & 0 & 0.403 \end{bmatrix}$$

وتكتب مصفوفة التكرارات النسبية للأسطر على النحو التالي:

$$X_r = \begin{pmatrix} \frac{f_{11}}{f_{1.}} & \cdots & \frac{f_{1q}}{f_{1.}} \\ \vdots & \ddots & \vdots \\ \frac{f_{p1}}{f_{p.}} & \cdots & \frac{f_{pq}}{f_{p.}} \end{pmatrix} = \begin{bmatrix} \frac{0.246}{0.314} & \frac{0.054}{0.314} & \frac{0.014}{0.314} \\ \frac{0.112}{0.161} & \frac{0.029}{0.161} & \frac{0.021}{0.161} \\ \frac{0.060}{0.122} & \frac{0.029}{0.122} & \frac{0.033}{0.122} \\ \frac{0.174}{0.403} & \frac{0.035}{0.403} & \frac{0.194}{0.403} \end{bmatrix} = \begin{bmatrix} 0.783 & 0.172 & 0.045 \\ 0.696 & 0.180 & 0.130 \\ 0.492 & 0.238 & 0.270 \\ 0.432 & 0.087 & 0.481 \end{bmatrix}$$

أو تحسب من الصيغة المصفوفية الآتية:

$$X_r = D_r^{-1}.N$$

$$X_r = D_r^{-1}.N = \begin{bmatrix} \frac{1}{0.314} & 0 & 0 & 0 \\ 0 & \frac{1}{0.161} & 0 & 0 \\ 0 & 0 & \frac{1}{0.122} & 0 \\ 0 & 0 & 0 & \frac{1}{0.403} \end{bmatrix} * \begin{bmatrix} 0.246 & 0.054 & 0.014 \\ 0.112 & 0.029 & 0.021 \\ 0.060 & 0.029 & 0.033 \\ 0.174 & 0.035 & 0.194 \end{bmatrix} = \begin{bmatrix} 0.783 & 0.172 & 0.045 \\ 0.696 & 0.180 & 0.130 \\ 0.492 & 0.238 & 0.270 \\ 0.432 & 0.087 & 0.481 \end{bmatrix}$$

$$D_r^{-1} = \begin{pmatrix} \frac{1}{f_{1.}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{f_{p.}} \end{pmatrix}, \quad D_r^{-1/2} = \begin{pmatrix} \frac{1}{\sqrt{f_{1.}}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{f_{p.}}} \end{pmatrix}$$

ملاحظة: إذا كان لون الشعر ولون العينين متغيرين مستقلين، فإن الأسطر في جدول الأسطر ستكون جميعها مماثلة للسطر المتوسط. وبالتالي، فإنّ الاعتمادية بين المتغيرات ستكون دالة على التشابه (أو المسافة) بين الأسطر.

3.2. **جدول التكرارات النسبية للأعمدة:** ويسمى كذلك بملف تعريف الأعمدة، ونحصل عليه بقسمة كل قيمة على إجمالي الصف المقابل. بتعبير آخر قسمة كل عمود j في المصفوفة N على ثقله $(f_{.j})$. مما يعني أننا سنتعامل مع الجدول كأعمدة فقط، والقيمة $\frac{f_{ij}}{f_{.j}}$ تمثل احتمال الحصول على الخاصية i من المتغير X مع العلم أن الخاصية j من المتغير Y محققة.

$$f_{i/j}=P(X=i,\ Y=j)=\frac{P(X=i,Y=j)}{P(Y=j)}=\frac{f_{ij}}{f_{.j}}\ /\sum_{i=1}^{p} f_{i/j}=1$$

جدول التكرارات النسبية للأعمدة

	بني	أحمر	أشقر	المتوسط
كستنائي	0,416	0,366	0,055	0.314
عسلي	0,189	0,197	0,079	0.161
أخضر	0,101	0,197	0,126	0.123
أزرق	0,294	0,239	0,740	0.403
المجموع	1	1	1	1

يظهر من العمود المتوسط، مثلا، أنّ 40.3٪ من الأفراد لديهم أعين زرقاء اللون.

كما يعبر العمود باللون الأخضر عن المتوسطات الهامشية، التي تمثل مركز سحابة النقاط في الفضاء R^p، ويسمى هذا العمود كذلك بالعمود المتوسط C_m.

$$\sum_{j=1}^{q} \frac{n_{.j}}{n} * \frac{n_{ij}}{n_{.j}}=\frac{n_{i.}}{n}$$

$C_m=[0.314 \quad 0.161 \quad 0.123 \quad 0.403]$

وعند دراسة وتحميل خصائص المتغير Y في فضاء خصائص المتغير X، تكون احداثيات النقطة j في الفضاء R^P هي:

$$F_j^X=\left(\frac{f_{1j}}{f_{.j}},\ \dots,\ \dots \frac{f_{pj}}{f_{.j}}\right)^T,\ j=1\dots,q$$

ونحدد المصفوفة القطرية للتكرارات النسبية الهامشية للأعمدة على الشكل التالي:

$$D_c=\begin{pmatrix} f_{.1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{.q} \end{pmatrix}=\begin{bmatrix} 0.591 & 0 & 0 \\ 0 & 0.147 & 0 \\ 0 & 0 & 0.262 \end{bmatrix}$$

$$D_C^{-1}=\begin{pmatrix} \frac{1}{f_{.1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{f_{.q}} \end{pmatrix}=\begin{bmatrix} 1.692 & 0 & 0 \\ 0 & 6.803 & 0 \\ 0 & 0 & 3.817 \end{bmatrix},\ D_C^{-1/2}=\begin{pmatrix} \frac{1}{\sqrt{f_{.1}}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{f_{.q}}} \end{pmatrix}$$

وتكتب مصفوفة التكرارات النسبية للأعمدة على النحو التالي:

$$X_c=\begin{pmatrix}\frac{f_{11}}{f_{.1}} & \cdots & \frac{f_{p1}}{f_{.1}}\\ \vdots & \ddots & \vdots\\ \frac{f_{1q}}{f_{.q}} & \cdots & \frac{f_{pq}}{f_{.q}}\end{pmatrix}=\begin{bmatrix}0.416 & 0.189 & 0.101 & 0.294\\ 0.366 & 0.197 & 0.197 & 0.239\\ 0.055 & 0.079 & 0.126 & 0.740\end{bmatrix}$$

أو يمكن حسابها من الصيغة المصفوفية الآتية:

$$X_c=D_c^{-1}.N^T$$

$$\begin{bmatrix}1.692 & 0 & 0\\ 0 & 6.803 & 0\\ 0 & 0 & 3.817\end{bmatrix}*\begin{bmatrix}0.246 & 0.112 & 0.060 & 0.174\\ 0.054 & 0.029 & 0.029 & 0.035\\ 0.014 & 0.021 & 0.033 & 0.194\end{bmatrix}$$

ملاحظة: كما كان الحال مع جدول الأسطر، إذا كان لون الشعر والعينين متغيرين مستقلين، فإن الأعمدة في جدول الأعمدة ستكون جميعها مماثلة للعمود المتوسط، وبالتالي، فإنّ الاعتمادية بين المتغيرات ستكون دالة على التشابه (أو المسافة) بين الأعمدة.

4.2. **مركز الثقل لجدولي الأسطر والأعمدة:** يحدد مركز ثقل الأسطر والأعمدة ، كما يلي:

$$G_r=(f_{.1},\ldots,f_{.q})^T,\ G_c(f_{1.},\ldots f_{p.})^T.$$

حيث:

G_r متجه ذاتي لـ $1*q$ و G_C متجه ذاتي لـ $1*p$.

$G_r=[0.314\quad 0.161\quad 0.123\quad 0.403]^T$

$G_c=[0.591\quad 0.147\quad 0.262]^T$

3. المسافة إلى الأصل في جدول الأسطر والأعمدة:

يقصد بالمسافة إلى الأصل، المسافة بين كل سطر والسطر المتوسط في جدول الأسطر، والمسافة بين كل عمود والعمود المتوسط في جدول الأعمدة. ولحساب المسافة بين نقطتين في فضاء ثنائي البعد نستخدم المسافة الأقليدية أو مسافة كاي تربيع، وسواء استخدمنا أيّا منهما سنصل إلى نفس الاستنتاجات، لكن ما يميز مسافة كاي تربيع، هو ضمان التكافؤ التوزيعي، والعلاقة شبه المركزية بين السحب النقطية.

ملاحظات:

– تظل المسافة المربعة بين صفين كما هي سواء قبل أو بعد تجميع عمودين بنفس العمود.

– تظل المسافة المربعة بين عمودين كما هي سواء قبل أو بعد تجميع صفين بنفس السطر.

1.3. المسافة إلى الأصل في جدول الأسطر: من أجل حساب المسافة إلى الأصل في جدول الأسطر، نحسب مسافة كاي تربيع بين كل سطر والسطر المتوسط، اعتمادا على العلاقة الآتية:

$$d^2_{X^2}(i,\ L_m)=\sum_{j=1}^{q}\frac{1}{f.j}\left(\frac{fij}{fi.}-f.j\right)^2$$

نعتمد في الحساب على جدول التكرارات النسبية للأسطر.

	بني	أحمر	أشقر	المجموع
كستنائي	0,783	0,171	0,046	1
عسلي	0,692	0,179	0,128	1
أخضر	0,492	0,237	0,271	1
أزرق	0,431	0,087	0,482	1
المتوسط	0.591	0.147	0.262	1

d^2(كستنائي, L_m) $= \frac{1}{0.591}(0.783-0.591)^2 + \frac{1}{0.147}(0.171-0.147)^2 + \frac{1}{0.262}(0.046-0.262)^2 = 0.245$

وعليه، يكون جدول المسافات بين الأسطر والسطر المتوسط كما يلي:

	$DISTO^2$
كستنائي	0.245
عسلي	0.093
أخضر	0.073
أزرق	0.251

يمكننا ملاحظة أنّ اللون الكستنائي للعينين يختلف اختلافا كبيرا عن اللونين العسلي والأخضر ويتشابه مع اللون الأزرق.

.2.3 المسافة إلى الأصل في جدول الأعمدة: نحسب المسافة بين كل خاصية في جدول الأعمدة والعمود المتوسط (الموقع بالنسبة للمتوسط).

$$d^2_{x^2}(j,\ c_m)=\sum_{j=1}^{p}\frac{1}{f_{i.}}\left(\frac{f_{ij}}{f_{.j}}-f_{i.}\right)^2$$

نعتمد في الحساب على جدول التكرارات النسبية للأعمدة.

	بني	أحمر	أشقر	المتوسط
كستنائي	0,416	0,366	0,055	0.314
عسلي	0,189	0,197	0,079	0.161
أخضر	0,101	0,197	0,126	0.123
أزرق	0,294	0,239	0,740	0.403
المجموع	1	1	1	

$d^2_{x^2}$ (بني, C_m) $= \frac{1}{0.314}(0.416-0.314)^2 + \frac{1}{0.161}(0.189-0.161)^2 + \frac{1}{0.123}(0.101-0.123)^2 + \frac{1}{0.403}(0.294-0.403)^2 = 0.071$

	بني	أحمر	أشقر
$DISTO^2$	0.071	0.130	0.538

يمكننا ملاحظة أن أصحاب الشعر الأشقر مختلفين في لون العينين عن بيقة الأفراد.

4. المسافة بين الأسطر في جدول الأسطر، والمسافة بين الأعمدة في جدول الأعمدة:

1.4. المسافة بين الأسطر: قد نرغب في إجراء تحليل أكثر عمقا، ولمعرفة ذلك علينا حساب المسافة بين كل سطرين (نعني خاصيتين) في جدول الأسطر، ولعمل ذلك نستخدم المسافة الأقليدية في فضاء ثنائي البعد:

$$d^2(i, i')= \sum_{j=1}^{p}\left(\frac{f_{ij}}{f_{i.}}-\frac{f_{i'j}}{f_{i'.}}\right)^2$$

لكن، هذه المسافة لا تأخذ في الاعتبار أهمية كل عمود. الخيار الأكثر حكمة هو أخذ مسافة مربع كاي، والتي تأخذ في الاعتبار أهمية كل عمود.

$$d^2_{\chi^2}(i, i')= \sum_{j=1}^{p}\frac{1}{f_{.j}}\left(\frac{f_{ij}}{f_{i.}}-\frac{f_{i'j}}{f_{i'.}}\right)^2$$

نضع:

$$\frac{f_{ij}}{f_{i.}}=x_{ij},\ \frac{f_{i'j}}{f_{i'.}}=x_{i'j}$$

وعليه، يمكن كتابة العلاقة بالصيغة المصفوفية التالية:

$$d^2_{\chi^2}(i, i')=(x_{ij}-x_{i'j})'\begin{bmatrix}\frac{1}{f_{.1}} & & \\ & \frac{1}{f_{.j}} & \\ & & \frac{1}{f_{.q}}\end{bmatrix}(x_{ij}-x_{i'j})$$

$$d^2_{\chi^2}(i, i')=(x_{ij}-x_{i'j})D_C^{-1}(x_{ij}-x_{i'j})$$

فكل سطر في جدول التكرارات النسبية للأسطر (L_i) يعتبر كنقطة M_i في الفضاء R^q، وهذا يعني أنّ جدول التكرارات النسبية للأسطر يمثل سحابة من النقط في الفضاء R^q. ونكتب:

$$M_i=(B_{i1}, B_{i2}, \ldots. B_{ij}, \ldots. B_{iq})\ldots\ldots\ldots/\ B_{ij}=\frac{1}{\sqrt{f_{.j}}}x_{ij}=\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}}$$

نعود إلى جدول التكرارات النسبية والهامشية ونقوم بحساب المسافات بين خصائص المتغير الأول. سنحسب المسافة بين اللون الكستنائي للعينين واللون العسلي، وبين اللونين الكستنائي والأخضر، والكستنائي والأزرق، ثم العسلي والأخضر، العسلي والأزرق، الأخضر والأزرق.

المجموع	أشقر	أحمر	بني	
1	0,046	0,171	0,783	كستنائي
1	0,128	0,179	0,692	عسلي
1	0,271	0,237	0,492	أخضر
1	0,482	0,087	0,431	أزرق
1	0.262	0.147	0.591	

$d2(\text{كستنائي}, \text{عسلي}) = \frac{1}{0.591}(0.783 - 0.692)^2 + \frac{1}{0.147}(0.171 - 0.179)^2 + \frac{1}{0.262}(0.046 - 0.128)^2 = 0.04$

وبنفس الطريقة نحسب بقية المسافات، ونلخص النتائج في الجدول الموالي:

	كستنائي	عسلي	أخضر	أزرق
كستنائي	0	0.04	0.365	0.982
عسلي	0.04	0	0.167	0.650
أخضر	0.365	0.167	0	0.328
أزرق	0.982	0.650	0.328	0

يمكننا ملاحظة، مثلا، أنّ الأشخاص ذوو الأعين الكستنائية لديهم شعر يشبه الأشخاص ذوي الأعين العسلية أكثر من الأشخاص ذوي العيون الزرقاء، وذوي الأعين العسلية لديهم شعر يشبه شعر ذوي الأعين الخضراء أكثر من ذوي الأعين الزرقاء، وذوي الأعين الكستنائية لديهم شعر يشبه شعر ذوي الأعين الخضراء أكثر من ذوي الأعين الزرقاء.

2.4. حساب المسافة بين كل عمودين (تحليل خاص) (بين كل نقطتين): حساب المسافة بين كل عمودين، وفقا للعلاقة الموالية:

$$d^2_{\chi^2}(j, j') = \sum_{i=1}^{n} \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}}\right)^2$$

سنحسب المسافة بين لون الشعر بني وبني داكن، وبين بني وأحمر، بني وأشقر، ثم بني داكن وأحمر، بني داكن وأشقر، وأخيرا أحمر مع أشقر. بالاعتماد على جدول التكرارات النسبية للأعمدة.

	بني	أحمر	أشقر	
كستنائي	0,416	0,366	0,055	0.314
عسلي	0,189	0,197	0,079	0.161
أخضر	0,101	0,197	0,126	0.123
أزرق	0,294	0,239	0,740	0.403
المجموع	1	1	1	1

$d^2_{\chi^2}(\text{بني}, \text{أحمر}) = \frac{1}{0.314}(0.416-0.3.66)^2 + \frac{1}{161}(0.189-0.197)^2 + \frac{1}{0.123}(0.101-0.197)^2 + \frac{1}{0.403}(0.294-0.239)^2 = 0.08$

	بني	أحمر	أشقر
بني	0	0.08	0.983
أحمر	0.08	0	1.026
أشقر	0.983	1.026	0

أصحاب الشعر الأحمر أقرب إلى أصحاب الشعر البني من البقية.

5. مؤشر التجاذب والتنافر: يستخدم هذا المؤشر للتحليل المعمّق للارتباطات بين شروط المتغيرين، ويستخرج من العلاقة التالية:

$$t_{ij}=\frac{f_{ij}}{f_{i.}f_{.j}}$$

$t_{ij>1}$ نقول أنّ i و j تتجاذب، والعكس صحيح.

$t_{ij} = 1$ نقول أنّه لا يوجد ارتباط بين السطر i والعمود j.

بالعودة إلى مثالنا، نحسب مؤشر التجاذب والتنافر للعنصر في الخلية الأولى من السطر الأول أين يتقاطع السطر الأول مع العمود الأول (كستنائي، بني)، t_{11}.

$$t_{11}= \frac{0.246}{0.314*0.591}=1.325$$

وبعد إتمام العمليات نجد:

الجدول رقم 5: مؤشر التجاذب والتنافر بين شروط المتغيرين.

	بني	أحمر	أشقر
كستنائي	3251.	1.166	0.175
عسلي	1.172	1.224	0.489
أخضر	0.838	1.629	1.041
أزرق	0.730	0.595	1.841

يظهر من النتائج ما يلي:

- توجد علاقة تجاذب بين لون العينين كستنائي ولون الشعر بني (1.325) وكستنائي مع أحمر (1.166). وبين لون العينين عسلي ولون الشعر بني (1.48) وأحمر (1.224)، ولون العينين أخضر تتجاذب مع لون الشعر أحمر وأشقر.
- مؤشر التجاذب والتنافر يساوي (0.489) بين لون العينين عسلي ولون الشعر أشقر، وكذلك الحال بالنسبة للون العينين أخضر ولون الشعر بني وأزرق مع بني ومع أحمر، مما يعني وجود تنافر بين تلك الخصائص.

6. اختبار كاي تربيع (*CHI-SQUARE*) للإستقلالية:

على اعتبار أنه ليس لدينا أي معلومة حول العلاقة بين المتغيرين X وY، إذ يمكن أن يكونا مستقلين، فيكون استخدامنا للتحليل التقابلي غير مجد. وعلى ذلك، علينا البحث في وجود استقلاية بين المتغيرين من عدمه.

ويمكننا ترجمة هذا الكلام في ظل فرضية الاستقلالية H_0.

P(X=i et Y=j)= P(X=i) x P(Y=j).

وعلى العكس، الفرضية البديلة H_1

P(X=i et Y=j) # P(X=i) x P(Y=j).

بعبارة أخرى سنقوم باختبار فرضية العدم:

H_0 ; $f_{ij}= f_{i.}*f_{.j}$

في مقابل الفرضية البديلة:

$H_1 ; f_{ij} \# f_{i.} * f_{.j}$

وعند دراسة ظاهرة كيفية متعلقة بمتغيرتين نوعيين، يمكن البحث عن وجود اعتمادية أو استقلالية باستخدام اختبار كاي تربيع للاستقلالية، المعطى بالعلاقة:

$$X^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(nij-(ni.*n.j/n))^2}{ni.*n.j/n} = \frac{(fij-fi.*f.j)^2}{fi.*f.j}$$

ملاحظة: يستخدم التحليل التقابلي جدول الاحتمالات، لذلك فهو لا يقول أي شيء عن الدلالة الإحصائية، فقط يتصور طبيعة الارتباط بين المتغيرين.

ويتم حساب قيمته ومن ثم مقارنته مع القيمة الجدولية الحرجة بدرجة حرية،

$$r=(p-1)(q-1).$$

((عدد الأسطر−1)(عدد الأعمدة−1)).

الفرضية الصفرية للاختبار (عند مستوى دلالة 0.05 فأقل) المتغيرين مستقلين، أمّا الفرضية البديلة فتقول بأنّهما غير مستقلين عن بعضهما. ولحساب كاي تربيع، نحسب التكرار المتوقع لكل خلية في الجدول المتقاطع حسب العلاقة الموالية.

$$\text{Expected frequency } E_{r,c} = \frac{(\text{Sum of row } r) \times (\text{Sum of column } c)}{\text{Sample size}}.$$

نضرب مجموع عناصر السطر في مجموع عناصر العمود الذي تنتمي له الخلية، ونقسم الناتج على المجموع الكلي. أو عن طريق العلاقة في معادلة حساب كاي تربيع $n_{i.} * n_{.j} / n_{..}$.

بالعودة إلى مثالنا، سنحسب في الخطوة الأولى جدول التكرارات المتوقعة بالاعتماد على جدول التقاطع.

نضرب مجموع عناصر السطر في مجموع عناصر العمود الذي تنتمي إليه الخلية، ونقسم الناتج على المجموع الكلي. أو عن طريق العلاقة في معادلة حساب كاي تربيع.

$n_{i.} * n_{.j} / n_{..}$

جدول التكرارات النظرية (المتوقعة)

	بني	أحمر	أشقر
كستنائي	89.818	22.297	39.884
عسلي	46.090	11.442	20.466
أخضر	34.863	8.654	15.481
أزرق	115.227	28.605	51.167

مثال: التكرار المتوقع للخلية الناتجة من تقاطع السطر الأول مع العمود الأول:

$$n_{11}= \frac{152*286}{484}= 89.818$$

جدول حساب كاي تربيع

	بني	أحمر	أشقر	المجموع
كستنائي	9.481	0.614	27.112	37.207
عسلي	1.357	0.571	5.352	7.28
أخضر	0.985	3.302	0.017	4.304
أزرق	8.462	4.708	35.856	49.026
المجموع	20.285	9.195	68.337	97.817

$$X^2_{11}= \frac{(89.818-(119))^2}{89.818}= 9.481$$

نحسب درجات الحرية من العلاقة:

$$(r-1)(c-1)= (4-1)(3-1)=6$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48

من الجدول نلاحظ أنّ قيمة كاي تربيع المحسوبة تساوي 97.817، نقارنها بكاي تربيع الجدولية عند مستوى دلالة 0.05 فأقل ودرجات حرية 6. القيمة الجدولية لكاي تربيع 12.59، وبما أنّ قيمة كاي تربيع المحسوبة تقع ضمن منطقة رفض فرضية العدم، أكبر من القيمة الجدولية، فإنّ هناك ارتباط بين الأسطر والأعمدة، أي بين لون العينين ولون الشعر.

ملاحظات: حتى تكون النتائج التي تم التوصل إليها في اختبار كاي تربيع موثوقة، لا بد من توفر بعض الشروط.

– أن تكون 80 في المائة على الأقل من التكرارات النظرية أكبر من 5، لأنّ القيم الصغيرة تؤدي إلى تضخيم قيمة كاي تربيع.

– يجب أن يكون حجم العينة كبيرا نوعا ما، لأنّ درجات الحرية لا تأخذ بعين الاعتبار حجم العينة.

7. **مساهمة كل خلية من خلايا جدول التقاطع في قيمة كاي تربيع:** يمكن معرفة مساهمة كل خلية من خلايا جدول التقاطع في قيمة كاي تربيع المحصلة من خلال العلاقة التالية:

$$C_{pq} = \frac{r^2_{pq}}{X^2}= \frac{(n_{pq}-e_{pq})^2}{e_{pq}}= \frac{(f_{ij}-f_{i.}f_{.j})^2}{f_{i.}f_{.j}}$$

$$r_{ij} = \frac{(n_{ij}-e_{ij})}{\sqrt{e_{ij}}}$$

حيث: r_{ij} هي البواقي المعيارية، C_{pq} هي المساهمة النسبية في قيمة كاي تربيع (المساهمة في المعلومة التي تحملها كل خانة من خانات الجدول).

توزيع البواقي يتبع التوزيع الطبيعي تقريبا، في حالة كانت قيمة الخطأ المعياري أكبر من القيمة المطلقة لـ 2 فهي دالة احصائيا.

$$r_{11} = \frac{(n_{11}-e_{11})}{\sqrt{e_{11}}} = \frac{(119-89.818)}{\sqrt{89.818}} = 3.079$$

جدول البواقي المعيارية

	بني	أحمر	أشقر
كستنائي	3.079	0.784	-5.207
عسلي	1.165	0.756	-2.313
أخضر	-0.993	1.817	0.131
أزرق	-2.909	-2.169	5.988

يمكننا كذلك بالاعتماد على الجدول تحديد الخاصيات التي تتجاذب وتتنافر من خلال الإشارة، فعندما تكون الإشارة موجبة فهي تتجاذب، وعندما تكون سالبة تتنافر.

حساب المساهمة النسبية لكل خلية في قيمة كاي تربيع.

$$C_{11} = \frac{r_{11}^2}{X^2} = \frac{(3.079)}{97.817} =$$

جدول المساهمة النسبية لكل خلية في قيمة كاي تربيع

	بني	أحمر	أشقر	المجموع
كستنائي	0.031	0.006	0.277	0.314
عسلي	0.013	0.005	0.054	0.072
أخضر	0.010	0.033	0.000	0.043
أزرق	0.086	0.048	0.366	0.5
المجموع	0.14	0.092	0.697	0.929

8. القصور الذاتي الكلي (التباين الكلي): يقصد به كمية المعلومات المتوفرة في البيانات. والقصور الذاتي الكلي في الفضاء R^p هو العزم الكلي لسحابة النقط، وبالتالي فهو المتوسط المرجح لمربع المسافات لكل النقاط بالنسبة لمركز البيانات، ونكتب:

$$I_t = \sum_{i=1}^{p} f_{i.}\, d_{x^2}^2(i,G) = \sum_{i=1}^{p} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - f_{.j}\right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(fij - fi.*f.j)^2}{fi.*f.j}$$

وتعطى صيغة القصور الذاتي الكلي لسحابة نقط الأسطر والأعمدة مقارنة بمركز ثقل كل منهما، بالعلاقتين:

$$\text{Inertie } (L, G_L) = \sum_{i=1}^{p} f_{i.}\, d_{x^2}^2(i, G_L)$$

$$\text{Inertie } (L, G_C) = \sum_{i=1}^{q} f_{.j}\, d_{x^2}^2(j, G_c)$$

لدينا:

$d^2_{X^2}$(i, G_L)= $\sum_{j=1}^{q} \frac{1}{f.j} (\frac{fij}{fi.} - f.j)^2$

بالتعويض نجد:

Inertie (L, G_L)= $\sum_{i=1}^{p} f_{i.}\, d^2_{x^2}$(i,$G_L$)= $\sum_{i=1}^{p} f_{i.} * \sum_{j=1}^{q} \frac{1}{f.j} (\frac{fij}{fi.} - f.j)^2$

$=\sum_{i=1}^{p}\sum_{j=1}^{q} \frac{f_{i.}}{f.j} (\frac{fij}{fi.} - f.j)^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{f_{i.}}{f.j} (\frac{fij - fi.*f.j}{fi.})^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{f_{i.}}{f.j * f_{i.}^2} (fij - fi.* f.j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{(fij - fi.*f.j)^2}{fi.*f.j} = \frac{X^2}{n_{..}}$

وعليه، فالتباين الكلي –فاي مربّع– لسحابة الأسطر يمكن حسابه بقسمة قيمة مربع كاي على حجم العينة.

$Ph^2 = \frac{X^2}{n_{..}} = \frac{97.817}{484} = 0.202$

ملاحظة: القصور الذاتي لسحابة الأسطر تساوي القصور الذاتي لسحابة الأعمدة.

Inertie (L, G_L)= $\sum_{i=1}^{p} f_{i.}\, d^2_{x^2}$(i,$G_L$)= Inertie (L, G_C)= $\sum_{i=1}^{q} f_{.j}\, d^2_{x^2}$(j,$G_c$)= Ph^2

1.8. حساب القصور الذاتي الذي تحمله كل خاصية في جدول الأسطر:

القصور الذاتي = وزن الخاصية x المسافة بينها والسطر المتوسط.

القصور الذاتي الذي تحمله كل خاصية = وزنها x بعدها عن السطر المتوسط.

Inertie(كستنائي) =poids(كستنائي) x d ²(كستنائي)= 0.314 x 0.245= 0.133

ملاحظة: ثقل كل خاصية في جدول الأسطر هو نفسه قيمتها في هامش السطر.

	DISTO²	Poids	INERTIE
كستنائي	0.245	0.314	0.07688
عسلي	0.093	0.161	0.01505
أخضر	0.073	0.122	0.00889
أزرق	0.251	0.403	0.10129
		المجموع	0,202

القصور الذاتي الكلي هو 0.202، وهو كمية المعلومات المتوفرة في البيانات. (هذا المؤشر مهم في عملية التحليل)، يستخدم في تقييم جودة التمثيل الرسومي للبيانات.

2.8. حساب القصور الذاتي الذي تحمله كل خاصية في جدول الأعمدة:

	بني	أحمر	أشقر	
DISTO²	0.071	0.130	0.538	
Poids	0.591	0.147	0.262	المجموع
Inertie	0.042	0.019	0.141	0.202

يمكننا ملاحظة أن أصحاب الشعر الأشقر مختلفين في لون العينين عن بيقة الأفراد. كما أنّ خاصية الشعر الأشقر هي الأكثر حملا للمعلومات من بقية الخصائص.

9. تحديد المركبات الأساسية: تتمثل طريقة التحليل العاملي للتوفيقات في تلخيص الروابط الرئيسية الموجودة بين خصائص المتغيرين X و Y، وتهدف إلى تقليل الأبعاد. ويمكن اعتبار التحليل العاملي للتوفيقات على أنه تحليل للمكونات الرئيسية مزدوج، أحدهما على الصفوف والآخر على الأعمدة.

1.9. مصفوفة المعلومات: بنفس الطريقة التي نستخدمها في التحليل بالمركبات الرئيسية، نشكل مصفوفة المعلومات، مصفوفة التباين والتباين المشترك V، بين خصائص المتغير Y (محور الفضاء R^q). ونعتمد على جدول التكرارات النسبية للأسطر في ايجاد التباين المشترك بين الخاصيتين j و j'. غير أننا وإن كنا في التحليل بالمركبات الرئيسية، نستخدم المسافة الأقليدية لحساب المسافة بين أي نقطتين، فلا نقوم بذلك في التحليل التقابلي قبل اجراء تحويل مسبق لاحداثيات نقط سحابة الأسطر أو الأعمدة، هذا التحويل يكتب في الفضاء R^q على النحو التالي:

$$B_{ij} = \frac{1}{\sqrt{f_{.j}}}\frac{f_{ij}}{f_{i.}}$$

فيكون المتوسط الحسابي المرجح:

$$\bar{X}_j = \sum_i f_{i.} B_{ij}$$

$$\bar{X}_j = \sum_i f_{i.} \frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} = \sum_i \frac{f_{ij}}{\sqrt{f_{.j}}} = \frac{1}{\sqrt{f_{.j}}} \sum f_{ij} = \frac{1}{\sqrt{f_{.j}}} f_{.j}$$

وهذا يعني أنّ:

$$\bar{X}_j = \sqrt{f_{.j}}$$

وتكون مصفوفة التباين والتباين المشترك للسحابة B(I) بالنسبة للمحور j على الشكل التالي:

$$V = \begin{bmatrix} v_{11} & v_{12} & \dots\dots & v_{1q} \\ v_{21} & v_{22} & \dots\dots & v_{2q} \\ \dots & \dots & \dots\dots & \dots\dots \\ v_{q1} & v_{q2} & \dots\dots & v_{qq} \end{bmatrix}$$

– التباين V_{jj} الذي يميز تشتت سحابة النقط على طول المحور j يعطى بالعلاقة:

$$V_{jj}=\sum_i f_{i.}\left(B_{ij}-\sqrt{f_{.j}}\right)^2=\sum_i f_{i.}\left(\frac{f_{ij}}{\sqrt{f_{.j}f_{i.}}}-\sqrt{f_{.j}}\right)^2$$

– التباين المشترك:

$$V_{jj'}=\sum_i f_{i.}\left(B_{ij}-\sqrt{f_{.j}}\right)\left(B_{ij'}-\sqrt{f_{.j'}}\right)$$

ونسمي التباين الكلي للسحابة B(I) بأثر المصفوفة V، ونكتب:

$$V_B=\text{trace }(V)=\sum_j V_{jj}$$

ويتم إسقاط السحابة B(I) بشكل متعامد على محور ذو متجه وحدوي u، بطريقة تجعل المعلومات المفقودة من الإسقاط في حدها الأدنى. وكما هو الحال في التحليل بالمركبات الرئيسية نبحث عن تعظيم العلاقة u^TVu، تحت الشرط $u^Tu=1$، وذلك عبر البحث عن القيمة الذاتية الأكبر للمصفوفة V.

لدينا:

$$V_{jj'}=\sum_i f_{i.}\left(B_{ij}-\sqrt{f_{.j}}\right)\left(B_{ij'}-\sqrt{f_{.j'}}\right)$$

وبتعويض B_{ij} بقيمتها نجد:

$$V_{jj'}=\sum_i f_{i.}\left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}}-\sqrt{f_{.j}}\right)\left(\frac{f_{ij'}}{f_{i.}\sqrt{f_{.j'}}}-\sqrt{f_{.j'}}\right)$$

ويمكن كذلك كتابتها على النحو التالي:

$$V_{jj'}=\sum_i f_{i.}\left(\frac{f_{ij}-f_{i.}f_{.j}}{\sqrt{f_{i.}f_{.j'}}}\right)\left(\frac{f'_{ij}-f_{i.}f_{.j}}{\sqrt{f_{i.}f_{.j'}}}\right)$$

نضع:

$$\left(\frac{f_{ij}-f_{i.}f_{.j}}{\sqrt{f_{i.}f_{.j'}}}\right)=r_{ij}\quad\ldots\ldots\ldots\ldots\ldots(i=1\ldots p,\ j=1\ldots q)$$

r_{ij} هو العنصر المولّد للمصفوفة R، حيث:

$$V=R^TR$$

كما يمكن كذلك استخدام الصيغة المصفوفية الآتية لحساب المصفوفة R، والتي من خلالها يمكننا حساب مصفوفة المعلومات V.

$$R=D_r^{\frac{-1}{2}}(N-C_mL_m^T)D_c^{\frac{-1}{2}}$$

$$V=R^TR$$

بالعودة إلى مثالنا سنحسب المصفوفة V بكل الطرق المتاحة، لنتمكن من معرفة أيها أيسر وأقل جهدا في حساب مصفوفة المعلومات.

– باستخدام صيغتي التباين والتباين المشترك للمصفوفة V.

لدينا ثلاثة خصائص للمتغير Y (لون الشعر: بني، أحمر، أشقر).

$$V=\begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix}$$

نستخدم صيغة التباين:

$$V_{jj}=\sum_i f_{i.}\,(B_{ij}-\sqrt{f_{.j}})^2$$

لحساب:

V_{11} ، V_{22}، V_{33}

$$V_{11}= f_{1.}(B_{11}-\sqrt{f_{.1}})^2+f_{2.}(B_{21}-\sqrt{f_{.1}})^2+f_{3.}(B_{31}-\sqrt{f_{.1}})^2+f_{4.}(B_{41}-\sqrt{f_{.1}})^2=$$
$$0.314\,(\frac{0.246}{0.314\sqrt{0.591}}-\sqrt{0.591})^2+0.161(\frac{0.692}{0.161\sqrt{0.591}}-\sqrt{0.591})^2+0.122(\frac{0.492}{0.122\sqrt{0.591}}-\sqrt{0.591})^2+0.403(\frac{0.174}{0.403\sqrt{0.591}}-\sqrt{0.591})^2=\ 0.042$$
$$V_{22}=0.314\,(\frac{0.054}{0.314\sqrt{0.147}}-\sqrt{0.147})^2+0.161(\frac{0.029}{0.161\sqrt{0.147}}-\sqrt{0.147})^2+0.122(\frac{0.029}{0.122\sqrt{0.147}}-\sqrt{0.147})^2+0.403(\frac{0.035}{0.403\sqrt{0.147}}-\sqrt{0.147})^2=\ 0.019$$
$$V_{33}=0.314\,(\frac{0.014}{0.314\sqrt{0.262}}-\sqrt{0.262})^2+0.161(\frac{0.021}{0.161\sqrt{0.262}}-\sqrt{0.262})^2+0.122(\frac{0.033}{0.122\sqrt{0.262}}-\sqrt{0.262})^2+0.403(\frac{0.194}{0.403\sqrt{0.262}}-\sqrt{0.262})^2=\ 0.140$$

ونستخدم العلاقة:

$$V_{jj'}=\sum_i f_{i.}\left[\left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}}-\sqrt{f_{.j}}\right)\left(\frac{f_{ij}'}{f_{i.}\sqrt{f_{.j'}}}-\sqrt{f_{.j}'}\right)\right]$$

لحساب التباين المشترك بين الثنائيات:

V_{12}, V_{13}, V_{23}

$$V_{12}=V_{21}=0.314(\frac{0.246}{0.314\sqrt{0.591}}-\sqrt{0.591})(\frac{0.054}{0.314\sqrt{0.147}}-\sqrt{0.147})+0.161(\frac{0.692}{0.161\sqrt{0.591}}-\sqrt{0.591})(\frac{0.029}{0.161\sqrt{0.147}}-\sqrt{0.147})+0.122(\frac{0.492}{0.122\sqrt{0.591}}-\sqrt{0.591})(\frac{0.029}{0.122\sqrt{0.147}}-\sqrt{0.147})+0.403(\frac{0.174}{0.403\sqrt{0.591}}-\sqrt{0.591})(\frac{0.035}{0.403\sqrt{0.147}}-\sqrt{0.147})=0.016.$$

وبعد اتمام جميع العمليات نجد:

$$V=\begin{bmatrix} 0.042 & 0.016 & -0.075 \\ 0.016 & 0.019 & -0.039 \\ -0.075 & -0.039 & 0.140 \end{bmatrix}$$

– نحسب المصفوفة V باستخدام المصفوفة R.

لدينا:

$$r_{ij} = \left(\frac{f_{ij}-f_{i.}f_{.j}}{\sqrt{f_{i.}f_{.j}}}\right)$$

$$r_{11}=\left(\frac{0.264-0.314*0.591}{\sqrt{0.314*0.591}}\right)= 0.183,\ r_{21}=\left(\frac{0.112-0.161*0.591}{\sqrt{0.161*0.591}}\right)=0.055$$

$$r_{31}=\left(\frac{0.060-0.122*0.591}{\sqrt{0.122*0.591}}\right)=-0.044,\ r_{41}=\left(\frac{0.174-0.403*0.591}{\sqrt{0.403*0.591}}\right)= -0.131$$

وهكذا حتى نكمل بقية الأعمدة:

$$R=\begin{bmatrix} 0.183 & 0.037 & -0.237 \\ 0.055 & 0.039 & -0.102 \\ -0.044 & 0.082 & 0.011 \\ -0.131 & -0.098 & 0.274 \end{bmatrix}$$

$$V=R^TR=\begin{bmatrix} 0.183 & 0.055 & -0.044 & -0.131 \\ 0.037 & 0.039 & 0.082 & -0.098 \\ -0.237 & -0.102 & 0.011 & 0.274 \end{bmatrix}*\begin{bmatrix} 0.183 & 0.037 & -0.237 \\ 0.055 & 0.039 & -0.102 \\ -0.044 & 0.082 & 0.011 \\ -0.131 & -0.098 & 0.274 \end{bmatrix}$$

$$V=\begin{bmatrix} 0.042 & 0.016 & -0.075 \\ 0.016 & 0.019 & -0.039 \\ -0.075 & -0.039 & 0.140 \end{bmatrix}$$

باستخدام الصيغة المصفوفية:

$$R=D_r^{\frac{-1}{2}}(N - C_m L_m^T)D_c^{\frac{-1}{2}}$$

$$V=R^TR$$

$$R=\begin{bmatrix} \frac{1}{\sqrt{0.314}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{0.161}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{0.122}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{0.403}} \end{bmatrix}\left(\begin{bmatrix} 0.246 & 0.054 & 0.014 \\ 0.112 & 0.029 & 0.021 \\ 0.060 & 0.029 & 0.033 \\ 0.174 & 0.035 & 0.194 \end{bmatrix}-\right.$$

$$\left.\begin{bmatrix} 0.314 \\ 0.161 \\ 0.122 \\ 0.403 \end{bmatrix} * \begin{bmatrix} 0.591 & 0.147 & 0.262 \end{bmatrix}\right)\begin{bmatrix} \frac{1}{\sqrt{0.591}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{0.147}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{0.262}} \end{bmatrix}=$$

$$\begin{bmatrix} 0.140 & 0.037 & -0.238 \\ 0.055 & 0.035 & -0.103 \\ -0.045 & 0.083 & 0.006 \\ -0.131 & -0.100 & 0.272 \end{bmatrix}$$

$$V=\begin{bmatrix}0.140 & 0.055 & -0.045 & -0.131\\ 0.037 & 0.035 & 0.083 & -0.100\\ -0.238 & -0.103 & 0.006 & 0.272\end{bmatrix}*\begin{bmatrix}0.140 & 0.037 & -0.238\\ 0.055 & 0.035 & -0.103\\ -0.045 & 0.083 & 0.006\\ -0.131 & -0.100 & 0.272\end{bmatrix}$$

$$V=\begin{bmatrix}0.042 & 0.016 & -0.075\\ 0.016 & 0.019 & -0.039\\ -0.075 & -0.039 & 0.141\end{bmatrix}$$

2.9. تحديد القيم والأشعة الذاتية:

بعد تحديد مصفوفة المعلومات، فإنّ المرحلة الموالية هي حساب القيم والمتجهات الذاتية لهذه المصفوفة، من خلال المحدد، كما نفعل مع التحليل بالمركبات الرئيسية، غير أنّه في التحليل العاملي للتوفيقات تكون أحد القيم الذاتية لمصفوفة التباين معدومة، والشعاع الذاتي الوحدوي المرافق لها مركباته تساوي المتوسطات الهامشية على المحاور. هذه القيمة المعدومة تستثنى من الدراسة.

$$\det(V-\lambda I)=\begin{bmatrix}0.042-\lambda & 0.016 & -0.075\\ 0.016 & 0.019-\lambda & -0.039\\ -0.075 & -0.039 & 0.141-\lambda\end{bmatrix}$$

وبحل كثير الحدود نحصل على قيم الجذور المميزة، كما في الجدول الموالي:

القيمة الذاتية	نسبة التمثيل	النسبة التجميعية
0.191	95	95
0.011	5	100
0.000	0	100

تعبر القيم الذاتية عن مدى التشتت على كل عامل والتي من خلالها يمكننا تقييم جودة التمثيل على العامل، فنسبة التمثيل على المحور الأول بلغت 95 في المائة، وهي نسبة جيدة. وكلما كان مجموع القيم الذاتية التي نأخذها في التحليل قريبة من الكثافة الكلية للبيانات كانت جودة التمثيل أفضل.

ملاحظة: إذا كانت القيمة الذاتية معدومة فإنّ الشعاع الذاتي المرافق لها هو شعاع تافه ولا يؤخذ بعين الاعتبار، أي لا يطلب تحديده.

تحديد عدد المحاور:

عند اجراء طريقة التحليل العاملي بالتوفيقات من خلال جدول التقاطع، الذي يحوي p صف و q عمود. فإذا كان عدد الأسطر أكبر من عدد الأعمدة (p>q)، فإنّ العدد الأقصى للمحاور الذي يمكن الاحتفاظ به هو q−1، وبالنسبة للحالة الثانية (q>p)، فالعدد الأقصى للمحاور هو p−1.

$$H_{max}=\min(q-1,p-1)$$

واعتمادا على قاعدة كايزر، فإننا نحتفظ بالمحور الذي يوفر نسبة كثافة أعلى من $\frac{1}{H_{max}}$.

فبالنسبة لمثالنا عدد الأسطر أكبر من عدد الأعمدة، والعدد الأقصى للمحاور هو q−1 (3−1=2)، وهذا يعني أننا سنحتفظ بالمحور الذي يوفر على الأقل 50 في المائة من المعلومة، ولذلك نحتفظ بالمحور الأول وجوبا، والمحور الثاني نحتفظ به فقط لعرض التمثيل البياني.

الأشعة الذاتية الوحدوية للمصفوفة V:

U_1	U_2
−0.46	0.47
−0.24	−0.88
0.86	0.009

10. حساب المركبات الرئيسية لسحابة الأسطر وسحابة الأعمدة:

1.10. حساب المركبات الرئيسية لسحابة الأسطر: تمر المحاور الأساسية من مركز البيانات g. وعليه، فاحداثيات النقطة i لسحابة النقط B(I) على المحور α ذو شعاع التوجيه الوحدوي U_α هو جداء احداثيات النقطة i في الفضاء R^q ذو المبدأ g مع مركبات الشعاع الذاتي الوحدوي U_α.

$$F_\alpha(i) = XU_\alpha = \left(\left[\frac{f_{i1}}{f_{i.}\sqrt{f_{.1}}} - \sqrt{f_{.1}}\right], \dots \left[\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} - \sqrt{f_{.j}}\right], \dots, \left[\frac{f_{ip}}{f_{i.}\sqrt{f_{.q}}} - \sqrt{f_{.q}}\right]\right)\begin{pmatrix} U_{\alpha 1} \\ \dots \\ U_{\alpha j} \\ \dots \\ U_{\alpha p} \end{pmatrix}$$

$$F_\alpha(i) = \sum_{j=1}^{q} U_{\alpha j}\left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} - \sqrt{f_{.j}}\right) = \sum_{j=1}^{q}\left(\frac{U_{\alpha j} f_{ij}}{f_{i.}\sqrt{f_{.j}}}\right) - \sum_{j=1}^{q} U_{\alpha j}\sqrt{f_{.j}}$$

لدينا:

$$\sum_{j=1}^{q} U_{\alpha j}\sqrt{f_{.j}} = 0$$

فتكون احداثيات النقطة i على المحور α هي:

$$F_\alpha(i) = \sum_{j=1}^{q}\left(\frac{U_{\alpha j} f_{ij}}{f_{i.}\sqrt{f_{.j}}}\right)$$

ويمكن كذلك استخدام الصيغة المصفوفية الآتية في حساب المركبات الرئيسية لملف الأسطر.

$$F_\alpha = H * U_\alpha$$

وبالعودة إلى مثالنا، تكون المركبات الرئيسية لملف الأسطر (لون العينين) كما يلي:

$$F_1(1) = \frac{U_{11} f_{11}}{f_{1.}\sqrt{f_{.1}}} + \frac{U_{12} f_{12}}{f_{1.}\sqrt{f_{.2}}} + \frac{U_{13} f_{13}}{f_{1.}\sqrt{f_{.3}}} = \frac{(-0.46)0.783}{\sqrt{0.591}} + \frac{(-0.24)0.171}{\sqrt{0.147}} + \frac{(0.86)0.046}{\sqrt{0.262}} = -0.498$$

ملاحظة: يعتمد في الحساب على جدول التكرارات النسبية أو على جدول التكرارات النسبية للأسطر، نحصل على نفس النتيجة.

وبعد حساب جميع المركبات، يمكن تلخيصها في الجدول الموالي.

	F_1	F_2
كستنائي	-0.491	0.057
عسلي	-0.305	-0.016
أخضر	0.018	-0.269
أزرق	0.500	0.043

1.1.10. حساب المركبات القياسية لسحابة الأسطر: بالاستعانة بعبارة الانتقال بين التحليلين R^P و R^q، تعتبر العبارة $\frac{U_{\alpha j}}{\sqrt{f_{.j}}}$ احداثية العمود j لسحابة النقط B(J) على المحور α، نرمز لها بالرمز $\widetilde{G_\alpha}(j)$، فتصبح العلاقة السابقة كما يلي:

$$F_\alpha(i) = \sum_{j=1}^{q}(\frac{f_{ij}}{f_{i.}})\,\widetilde{G_\alpha}(j)$$

وحتى تتحقق عبارة الانتقال بين التحليلين نقسم العبارة أعلاه على $\sqrt{\lambda_\alpha}$، فتصبح العبارة النهائية لاحداثيات النقطة i من سحابة النقط B(I) في الفضاء R^q على المحور α كالآتي:

$$\widetilde{F}_\alpha(i)= \frac{1}{\sqrt{\lambda_\alpha}}\,F_\alpha(i)= \frac{1}{\sqrt{\lambda_\alpha}}\sum_{j=1}^{q}(\frac{f_{ij}}{f_{i.}})\,\widetilde{G_\alpha}(j)$$

حيث أنّ $\widetilde{F}_\alpha(i)$ تمثل الاحداثيات القياسية (Coordonnées Standards) للنقطة i من سحابة النقط B(I) في الفضاء R^q على المحور α.

وعليه، يمكننا حساب المركبات القياسية لجدول الأسطر في مثالنا كما يلي:

$$\widetilde{F}_1(1)= \frac{1}{\sqrt{\lambda_1}}\,F_1(1)= \frac{1}{\sqrt{0.191}}\,(-0.498)=-1.123$$

ونلخص النتائج في الجدول الموالي:

	F_1	F_2
كستنائي	-1.123	0.556
عسلي	-0.697	-0.154
أخضر	0.042	-2.608
أزرق	1.142	0.417

2.10. حساب المركبات الرئيسية لسحابة الأعمدة: وبنفس خطوات التحليل السابقة يمكننا استنتاج العبارة النهائية لحساب الاحداثيات القياسية للنقطة j من سحابة النقط B(J) على المحور α في الفضاء R^p.

$$\widetilde{G_\alpha}(j) = \frac{1}{\sqrt{\lambda_\alpha}}\sum_{i=1}^{p}(\frac{f_{ij}}{f_{.j}})\,\widetilde{F}_\alpha(i)$$

وبنفس النمط السابق، فإنّ:

$$\widetilde{G_\alpha}(j)= \frac{1}{\sqrt{\lambda_\alpha}}\,G_\alpha(j)$$

وتكون عبارة حساب المركبات الأساسية لسحابة الأعمدة على النحو التالي:

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}}\sum_{i=1}^{p}\left(\frac{f_{ij}}{f_{.j}}\right) F_\alpha(i)$$

ويمكن كذلك حساب المركبات الرئيسية لسحابة الأعمدة باستخدام الصيغة المصفوفية الآتية:

$$G_\alpha(j)=\sqrt{\lambda_\alpha}*U_\alpha$$

وبالعودة إلى مثالنا، نحسب المركبات الرئيسية لسحابة الأعمدة:

$$G_1(1) = \frac{1}{\sqrt{\lambda_1}}\left(\frac{0.264}{0.591}(-0.491) + \frac{0.112}{0.591}(-0.305) + \frac{0.060}{0.591}(0.018) + \frac{0.174}{0.591}(0.500)\right) = -0.259$$

ملاحظة: نستخدم جدول التكرارات النسبية، أو جدول التكرارات النسبية للأعمدة وسنصل إلى النتيجة نفسها.

	G_1	G_2
بني	-0.259	0.060
أحمر	-0.267	-0.241
أشقر	0.734	-0.001

ومنه يمكننا استنتاج الاحداثيات القياسية لسحابة الأعمدة.

	G_1	G_2
بني	-0.593	0.584
أحمر	-0.611	-2.333
أشقر	1.677	-0.010

11. المساهمة النسبية في تشكيل المحاور:

1.11. المساهمة النسبية للأسطر في تشكيل المحور α: لحساب مساهمة السطر i في تشكيل المحور α نستخدم العلاقة الموالية:

$$C_i^\alpha=\frac{f_{i.}(F_\alpha(i))^2}{\lambda_\alpha}$$

$$C_1^1=\frac{f_{1.}(F_1(1))^2}{\lambda_1}=\frac{0.314*(-0.491)^2}{0.191}=0.3960$$

وبعد اتمام جميع العمليات نلخص النتائج في الجدول الموالي:

	المحور الأول	المحور الثاني
كستنائي	39.60	9.7
عسلي	7.8	0.4
أخضر	0	82.9
أزرق	52.6	7
المجموع	100	100

يساعدنا حساب نسبة مساهمة كل سطر في تشكيل المحورين الأول والثاني على معرفة أهمية كل خاصية لدينا في إنشاء المحور العاملي. وكقاعدة عامة، إذا كانت مساهمة كل سطر في تشكيل المحور تفوق القيمة $\frac{1}{p}$ ، (25 في المائة) فإنّ هذه الخاصية مهمة في تشكيل هذا العامل. ويظهر من النتائج أنّ خاصيتي اللون الكستنائي للعين واللون الأزرق مهمتين في تشكيل المحور الأول، بينما ليس لهما أهمية في تشكيل المحور الثاني، وخاصية اللون الأخضر مهمة جدا في تشكيل المحور الثاني وليس لها أي أهمية في تشكيل المحور الأول. بينما خاصية اللون العسلي غير مهمة بالنسبة للمحورين معا.

2.11. المساهمة النسبية للأعمدة في تشكيل المحور α: لحساب مساهمة العمود j في تشكيل المحور α نستخدم العلاقة الموالية:

$$C_j^{\alpha}=\frac{f_{.j}(G_{\alpha}(j))^2}{\lambda_{\alpha}}$$

$$C_1^1=\frac{f_{.1}(G_1(1))^2}{\lambda_1}=\frac{0.591*(-0.259)^2}{0.191}=0.207$$

وبعد اتمام جميع العمليات نلخص النتائج في الجدول الموالي:

	المحور الأول	المحور الثاني
بني	20.7	20.1
أحمر	5.5	79.9
أشقر	73.8	0
المجموع	100	100

بنفس المنطق السابق، ومقارنة بالنسبة $\frac{1}{q}$ (33.33 في المائة)، فإن خاصية الشعر الأشقر مهمة في تشكيل المحور الأول وليس لها أي أهمية أو مساهمة في تشكيل المحور الثاني. بينما العكس بالنسبة لخاصية الشعر الأحمر.

12. نسبة التمثيل على المحاور:

1.12. نسبة تمثيل الأسطر على المحور α: لحساب نسبة تمثيل السطر i على المحور α نستخدم الصيغة الموالية (مربع جيب التمام):

$$cos^2\,\theta_{i\propto}=\frac{(F_{\propto}(i))^2}{\sum_{\propto=1}^{p}(F_{\propto}(i))^2}$$

$$cos^2\,\theta_{11}=\frac{(F_1(1))^2}{\sum_{\propto=1}^{p}(F_{\propto}(1))^2}=\frac{(-0.491)^2}{(-0.491)^2+(0.057)^2}=0.9877$$

وبعد اتمام جميع العمليات نلخص النتائج في الجدول الموالي:

	$cos^2\,\theta_{i1}$	$cos^2\,\theta_{i2}$	المجموع
كستنائي	98.7	1.3	100
عسلي	99.7	0.3	100
أخضر	0.5	99.5	100
أزرق	99.3	0.7	100

يظهر من النتائج أنّ نسبة تمثيل كل من خاصية اللون الكستنائي والعسلي والأزرق جيدة على المحور الأول وعلى العكس نسبة تمثيل خاصية اللون الأخضر جيدة على المحور الثاني.

2.12. نسبة تمثيل الأعمدة على المحور α: لحساب نسبة تمثيل العمود j على المحور α نستخدم الصيغة الموالية (مربع جيب التمام):

$$cos^2\,\theta_{j\propto}=\frac{(G_{\propto}(j))^2}{\sum_{\propto=1}^{q}(G_{\propto}(j))^2}$$

$$cos^2\,\theta_{11}=\frac{(G_1(1))^2}{\sum_{\propto=1}^{q}(G_{\propto}(1))^2}=\frac{(-0.259)^2}{(-0.259)^2+(0.060)^2}=0.9571$$

وبعد اتمام جميع العمليات نلخص النتائج في الجدول الموالي:

	$cos^2\,\theta_{j1}$	$cos^2\,\theta_{j2}$	المجموع
بني	94.90	5.1	100
أحمر	55.1	44.9	100
أشقر	100	0	100

يظهر من النتائج أنّ نسبة تمثيل كل من خاصيتي لون الشعر بني وأشقر جيدة على المحور الأول وعلى العكس بالنسبة للمحور الثاني، فيما نسبة تمثيل خاصية الشعر الأشقر جيدة على المحور الأول والثاني.

13. النقطة الإضافية: في حالة أردنا إضافة نقطة سطر أو عمود لم ندخلها في التحليل، يكفي فقط حساب الاسقاط باستخدام المقياس المختار تحت الفضاء الجزئي.

1.13. إضافة نقطة في العمود: لتكن النقطة K في العمود التالي:

$$X_{jk}=\begin{bmatrix}\frac{n_{1k}}{n_{.k}}\\ \cdots\\ \frac{n_{ik}}{n_{.k}}\\ \cdots\\ \frac{n_{pk}}{n_{.k}}\end{bmatrix}$$

لتمثيل هذا العمود، نحسب المركبة الرئيسية للنقطة المضافة k، باستخدام العلاقة:

$$G_{jk}^{\alpha}=\frac{1}{\sqrt{\lambda_\alpha}}\sum_i \frac{n_{ik}}{n_{.k}}\, F_\alpha(i)$$

2.13. إضافة نقطة سطرية: لتكن النقطة z في السطر التالي:

$$X_{iz}=\begin{bmatrix}\frac{n_{z1}}{n_{z.}}\\ \cdots \\ \frac{n_{iz}}{n_{z.}}\\ \cdots \\ \frac{n_{zq}}{n_{z.}}\end{bmatrix}$$

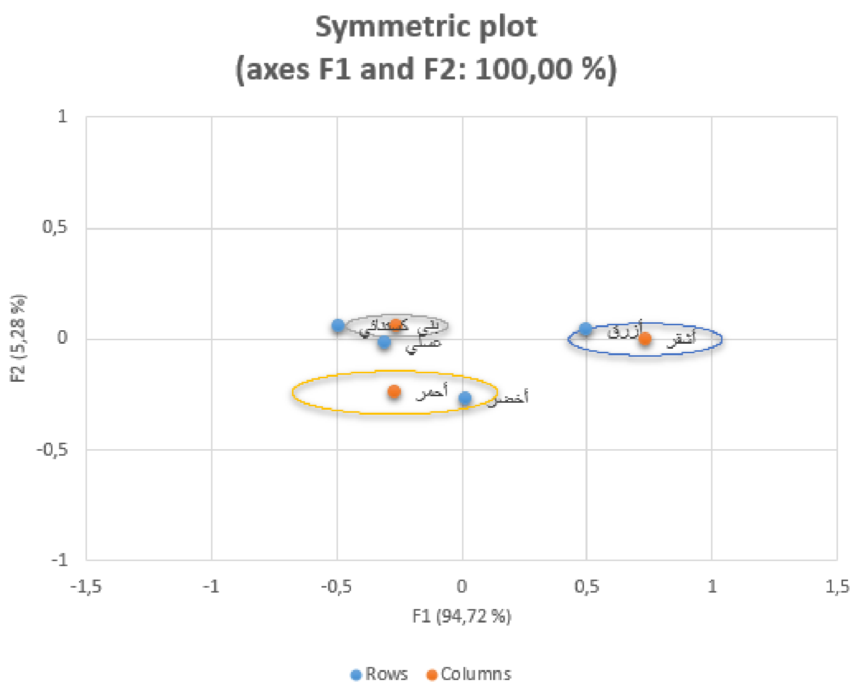
لتمثيل هذا السطر، نحسب المركبة الرئيسية للنقطة المضافة z، باستخدام العلاقة:

$$F_{iz}^{\alpha}=\frac{1}{\sqrt{\lambda_\alpha}}\sum_i \frac{n_{zj}}{n_{z.}}\, F_\alpha(j)$$

14. التمثيل البياني:

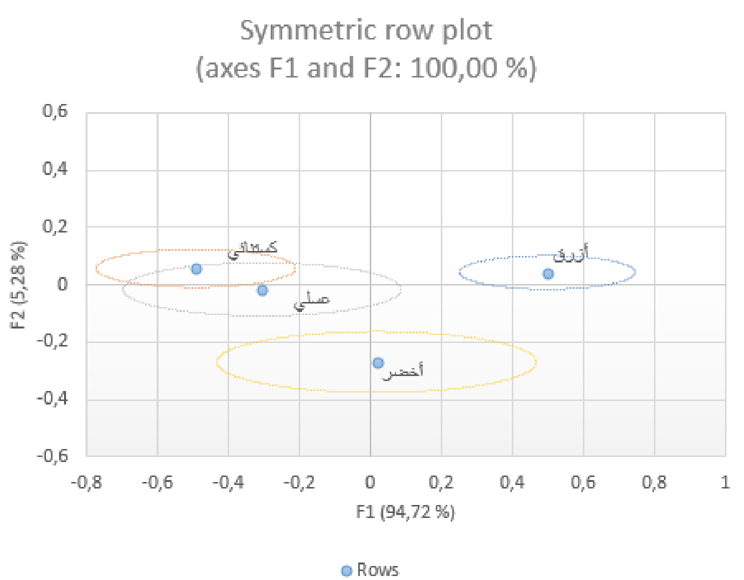
بسبب تماثل خصائص المتغيرين، يمكننا بفضل خاصية التمثيل الثنائي في طريقة التحليل العاملي بالتوفيقات دمج سحابتي الأسطر والأعمدة في تمثيل بياني واحد، فيكون لهما مركز البيانات (g) نفسه. كما أن الطريقة توفر العديد من التمثيلات والعروض البيانية، التي تستخدم وفقا للحاجة من إنشائها.

1.14. تمثيل بياني متناظر للأسطر والأعمدة: وهو الأكثر استخداما، إذ يتم تمثيل خصائص المتغيرين معا على نفس المستوى، مستعينين باحداثيات الأسطر والأعمدة، والمسافات بين النقاط تكون على أساس مصفوفة القياس لكاي تربيع. كما يمكن اضافة علامات إلى الرسم على شكل قطع ناقص (Confidence ellipses)(يمكن تحديد هذا الخيار في برنامج xlstat) لخصائص المتغيرين. وإذا كان المركز يقع داخل هذا القطع، فذلك يعني أنّ هذه الخاصية لا تساهم في علاقة الارتباط بين المتغيرين.



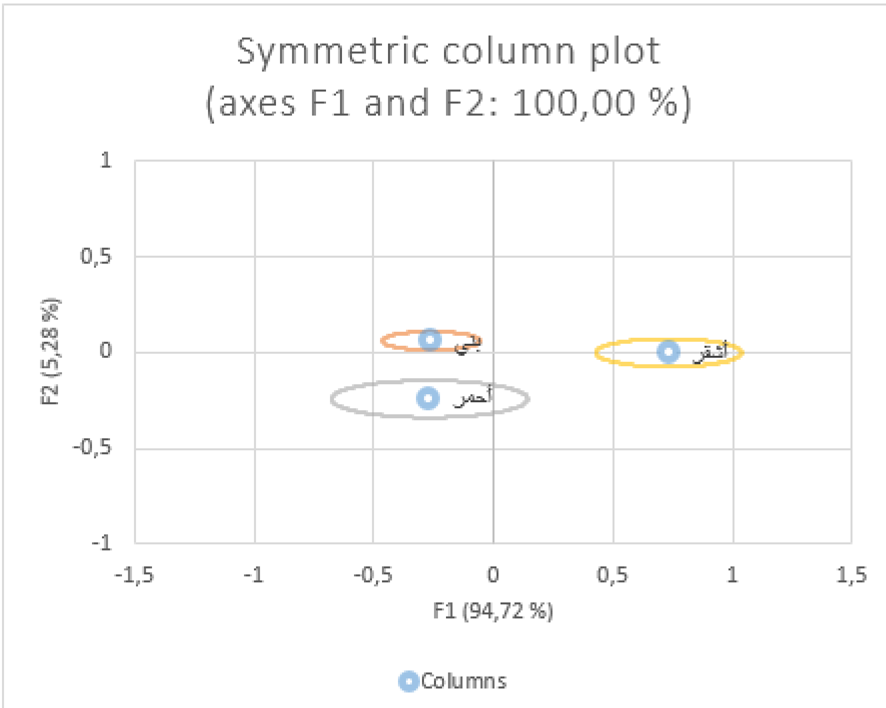
يظهر من الرسم البياني أعلاه تداخل بين خاصيتي اللون الكستنائي والعسلي للعينين، مما يشير إلى تشابه لون الشعر للأفراد الذين يملكون هاتين الخاصيتين. كما ترتبط خاصية اللون الأزرق للعينين مع اللون الأشقر للشعر، واللون الأخضر للعينين يرتبط عكسيا مع اللون الأحمر للشعر. وتجدر الإشارة هنا إلى أنّه لا يمكن تفسير التقارب بين الأسطر والأعمدة مباشرة من هذا الرسم.

2.14. التمثيل البياني المتناظر للأسطر:



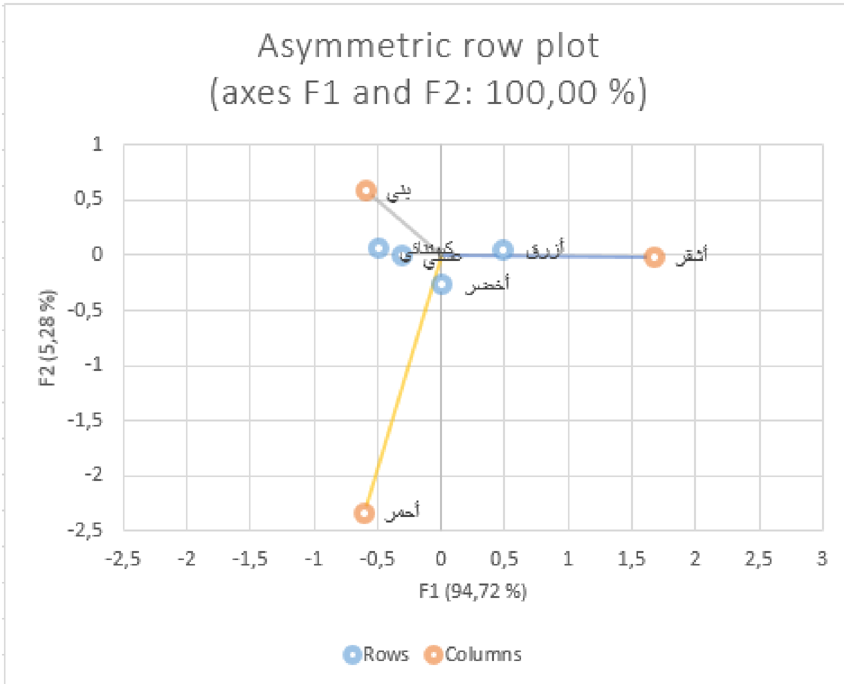
تساعدنا علامات القطع الناقص في تحديد الخاصيات التي تساهم في الارتباط بين المتغيرين، فإذا كان المبدأ يقع ضمن علامات القطع الناقص لخاصية ما فهذا يعني أنها لا تساهم في الارتباط. نلاحظ أنّ خاصية اللون الأخضر للعينين لا تساهم في الارتباط بين المتغيرين، في حين تساهم خاصية اللون الأزرق بشكل كبير في الارتباط بين المتغيرين.

3.14. التمثيل البياني المتناظر للأعمدة:



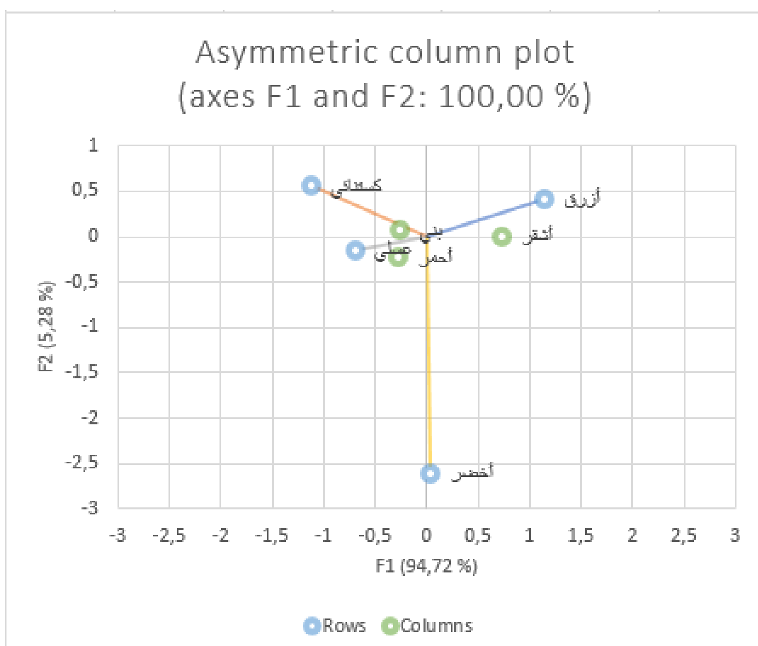
يظهر أنّ خاصية اللون الأشقر للشعر تساهم بشكل جيد في الارتباط بين المتغيرين، على عكس خاصية اللون الأحمر للشعر.

4.14. تمثيل غير متناظر للأسطر:



يتم في هذا الرسم تمثيل خصائص المتغير X، أي الأسطر، باستخدام المركبات الأساسية وتمثيل خصائص المتغير Y باستخدام المركبات القياسية. ويساعدنا ذلك في تحليل خصائص المتغير X بالنسبة لخصائص المتغير Y. ويظهر من الرسم أنّ اللون الأزرق للعينين يرتبط ارتباطا قويا مع اللون الأشقر للشعر، بمعنى أنّ الأفراد الذين لديهم أعين زرقاء يملكون في أغلبهم شعرا أشقرا.

5.14. تمثيل غير متناظر للأعمدة:

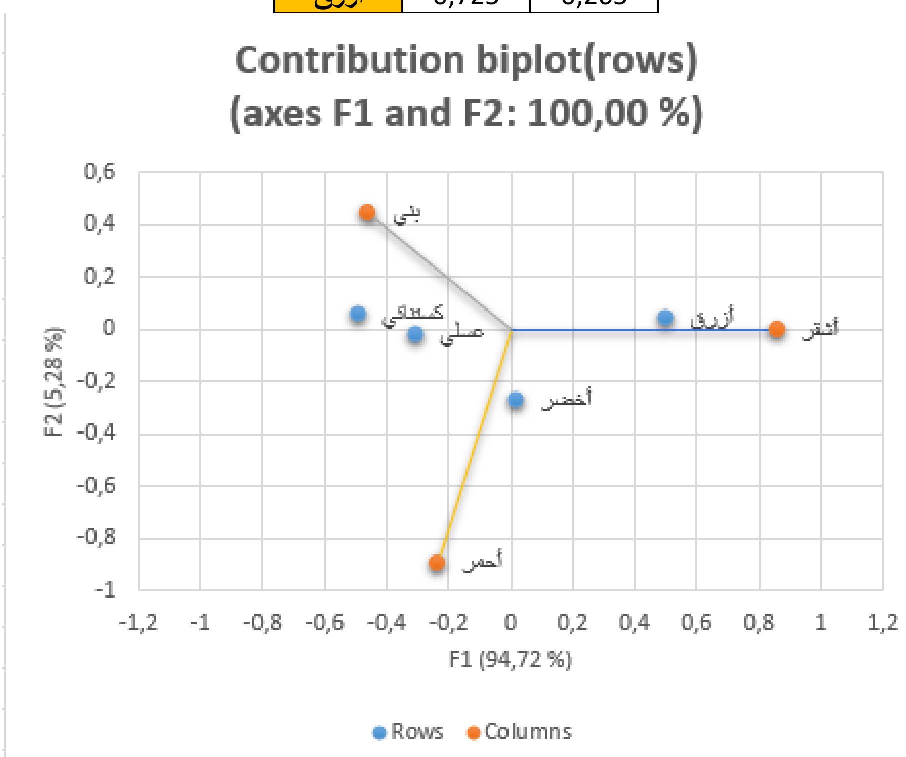


يتم في هذا الرسم تمثيل خصائص المتغير Y، أي الأعمدة، باستخدام المركبات الأساسية، وتمثيل خصائص المتغير X باستخدام المركبات القياسية. ويساعدنا ذلك في تحليل خصائص المتغير Y بالنسبة لخصائص المتغير X.

6.14. التمثيل البياني لنسب مساهمة الأسطر:

يتم في هذا الرسم تمثيل خصائص المتغير X بالمركبات الرئيسية، وتمثل خصائص المتغير Y باحداثيات نسب المساهمة. وتستخرج احداثيات نسب المساهمة للأسطر بقسمة المركبات القياسية على الجذر التربيعي لثقل الخاصية المحددة، وهي المتوسط الهامشي $\left(\frac{\tilde{F}_{\alpha}}{f_{i.}}\right)$.

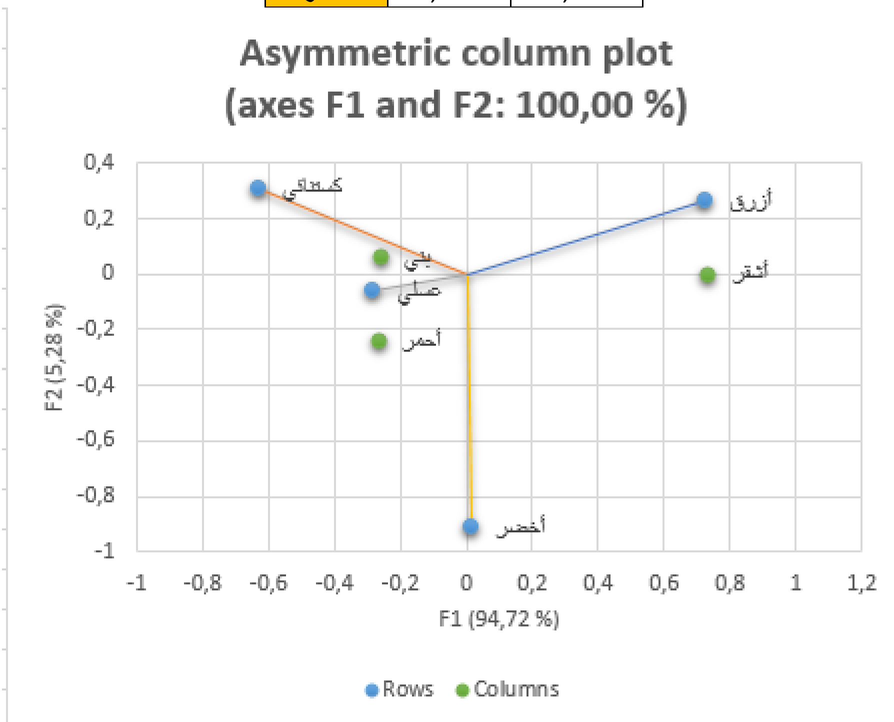
	F1	F2
كستنائي	-0,629	0,311
عسلي	-0,280	-0,062
أخضر	0,015	-0,910
أزرق	0,725	0,265



7.14. التمثيل البياني لنسب مساهمة الأعمدة:

يتم في هذا الرسم تمثيل خصائص المتغير Y بالمركبات الرئيسية، وتمثل خصائص المتغير X باحداثيات نسب المساهمة. وتستخرج احداثيات نسب المساهمة للأعمدة بقسمة المركبات القياسية على الجذر التربيعي لثقل الخاصية المحددة، وهي المتوسط الهامشي $\left(\frac{\tilde{G}_{\alpha}}{f_{.j}}\right)$.

	F1	F2
بني	-0,456	0,449
أحمر	-0,234	-0,894
أشقر	0,859	-0,005



تمارين محلولة

التمرين الأول: يحتوي الجدول الموالي على متغيرين لكل منهما ثلاثة خصائص، يعبّر المتغير الأول عن مجموعة من التجهيزات المكتبية (خزانة، مكتب، كرسي)، فيما يعبر المتغير الثاني عن ألوان هذه التجهيزات (رمادي، بني، أسود).

	رمادي	بني	أسود	المجموع
خزانة	1	3	5	9
مكتب	1	8	6	15
كرسي	4	1	1	6
المجموع	6	12	12	30

المطلوب:

1. حساب جدول التكرارات النسبية (مصفوفة الاحتمالات N)،
2. حساب جدول التكرارات النسبية للأسطر (X_r) والسطر المتوسط (L_m)، والمصفوفة القطرية للسطر المتوسط، ثم حساب المسافة بين الاسطر وبين الأسطر والسطر المتوسط، والكثافة لكل خلية، مع التعليق على النتائج.
3. حساب جدول التكرارات النسبية للأعمدة (X_j) والعمود المتوسط (C_m)، والمصفوفة القطرية للعمود المتوسط، ثم حساب المسافة بين الاعمدة وبين الأعمدة والعمود المتوسط، والكثافة لكل خلية، مع التعليق على النتائج.
4. حساب مؤشر التجاذب والتنافر لهذا الجدول، والتعليق على النتائج.
5. هل يمكن القول أنّ هناك ارتباط بين الأسطر والاعمدة، أي بين نوع التجهيز واللون.
6. حساب مصفوفة المعلومات V، باستخدام الطرق المتاحة.
7. حساب القيم والمتجهات الذاتية لجدول الأسطر، وتحديد عدد المحاور التي تأخذ في التحليل.
8. حساب المركبات الرئيسية والقياسية لسحابتي الأسطر والأعمدة.
9. حساب المساهمة النسبية للأسطر والأعمدة في تشكيل المحاور.
10. حساب نسبة التمثيل لكل من الأسطر والأعمدة على المحاور.
11. مثل الأسطر والأعمدة في المستوى وعلّق على الرسم.

الحل:

1. جدول التكرارات النسبية (مصفوفة الاحتمالات):

	رمادي	بني	أسود	المتوسط
خزانة	0,033	0,100	0,167	0.3
مكتب	0,033	0,267	0,200	0.5
كرسي	0,133	0,033	0,033	0.2
المتوسط	0.2	0.4	0.4	1

يظهر من الجدول أنّ نسبة الخزائن الرمادية اللون في العينة صغيرة 3.33%، وكذلك نسبة المكاتب الرمادية، والكراسي البنية والسوداء. في حين انّ نسبة المكاتب البنية قد بلغت 26.7%، وهي النسبة الأعلى....الخ.

وتكتب مصفوفة الاحتمالات على الشكل التالي:

$$N= \frac{1}{n}N^{*}= \frac{1}{30}\begin{bmatrix}1 & 3 & 5\\ 1 & 8 & 6\\ 4 & 1 & 1\end{bmatrix}=\begin{bmatrix}0.033 & 0.1 & 0.167\\ 0.033 & 0.267 & 0.2\\ 0.133 & 0.033 & 0.033\end{bmatrix}$$

كما يمكننا استنتاج الوزن النسبي لكل خاصية في الجدول الموالي، فالوزن النسبي لخاصية اللون رمادي في متغير اللون يساوي 20%، والوزن النسبي للخاصية كرسي في متغير التجهيز 20%. ويمكن تلخيص الأوزان في الجدول الموالي:

متغير اللون		متغير التجهيز	
الوزن	الخاصية	الوزن	الخاصية
0,2	رمادي	0.3	خزانة
0,4	بني	0.5	مكتب
0,4	أسود	0.2	كرسي

يتم حساب وزن كل خاصية بقسمة مجموع تكراراتها في كل سطر أو عمود على المجموع الكلي (حجم العينة).

2. جدول التكرارات النسبية للأسطر:

$$f_{ij}= \frac{n_{ij}}{n_{i.}}$$

	رمادي	بني	أسود	المجموع
خزانة	0,111	0,333	0,556	1
مكتب	0,067	0,533	0,400	1
كرسي	0,667	0,167	0,167	1
المجموع	0.2	0.4	0.4	1

السطر المتوسط:

$$L_m= \begin{bmatrix}0.2\\ 0.4\\ 0.4\end{bmatrix},\ D_j= \begin{bmatrix}0.2 & 0 & 0\\ 0 & 0.4 & 0\\ 0 & 0 & 0.4\end{bmatrix}$$

تحليل جدول الأسطر:

– تحليل عام: حساب المسافة بين كل خاصية في جدول الأسطر والسطر المتوسط (الموقع بالنسبة للمتوسط).

$$d^2_{X^2}(\text{i, L}_\text{m})= \sum_{j=1}^{q} \frac{1}{f.j}\left(\frac{fij}{fi.} - f.j\right)^2$$

$d^2(\text{خزانة}) = \frac{1}{0.2}(0.111-0.2)^2 + \frac{1}{0.4}(0.333-0.4)^2 + \frac{1}{0.4}(0.556-0.4)^2 = 0.111$

$d^2(\text{مكتب}) = \frac{1}{0.2}(0.067-0.2)^2 + \frac{1}{0.4}(0.533-0.4)^2 + \frac{1}{0.4}(0.4-0.4)^2 = 0.133$

$d^2(\text{كرسي}) = \frac{1}{0.2}(0.667-0.2)^2 + \frac{1}{0.4}(0.167-0.4)^2 + \frac{1}{0.4}(0.167-0.4)^2 = 1.36$

وعليه، يكون جدول المسافات كما يلي:

	DISTO²
خزانة	0.111
مكتب	0.133
كرسي	**1.360**

يمكننا ملاحظة أنّ لون الكراسي يختلف عن لون بقية التجهيزات.

– تحليل خاص (بين كل نقطتين): حساب المسافة بين كل سطرين في جدول الأسطر، وفقا للعلاقة الموالية:

$$d^2_{x^2}(\text{i, i'})= \sum_{j=1}^{p} \frac{1}{f_{.j}}\left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}}\right)^2$$

وبالتعويض نجد:

$d2(\text{خزانة, مكتب})= \frac{1}{0.2}(0.111 - 0.067)^2 + \frac{1}{0.4}(0.333 - 0.533)^2 + \frac{1}{0.4}(0.556 - 0.4)^2 = 0.165$

$d2(\text{خزانة, كرسي})= \frac{1}{0.2}(0.111 - 0.667)^2 + \frac{1}{0.4}(0.333 - 0.167)^2 + \frac{1}{0.4}(0.556 - 0.167)^2 = 1.991$

$d2(\text{مكتب, كرسي})= \frac{1}{0.2}(0.067 - 0.667)^2 + \frac{1}{0.4}(0.533 - 0.167)^2 + \frac{1}{0.4}(0.400 - 0.167)^2 = 2.269$

يمكننا استنتاج أن لون الخزائن أقرب إلى لون المكاتب. وبعد اتمام جميع العمليات يكون جدول المسافات كما يلي:

	خزانة	مكتب	كرسي
خزانة	0	0.165	1.991
مكتب		0	2.269
كرسي			0

يمكننا من نتائج الجدول ملاحظة أن لون الخزائن والمكاتب متشابه إلى حد كبير، في حين يختلف لون الخزائن عن الكراسي، ومن جانب آخر يختلف لون المكاتب عن لون الكراسي اختلافا كبيرا.

– حساب القصور الذاتي (التباين): كمية المعلومات التي تحملها كل خاصية في جدول الأسطر.

كمية المعلومات التي تحملها كل خاصية = وزنها x بعدها عن السطر المتوسط.

Inertie(خزانة) =poids(خزانة) x d ²(خزانة)= $\frac{1825}{3784}$ x 0.0693= 0.0334

	DISTO²	Poids	INERTIE	INERTIE RELATIVE
خزانة	0.111	0.3	0.0333	3.33
مكتب	0.133	0.5	0.0665	6.65
كرسي	1.360	0.2	0.2720	27.2
		المجموع	0.3718	37.18

يتضح من الجدول أعلاه أنّ الخاصية كرسي تحمل أكبر قدر من المعلومات، بينما الخاصية خزانة تحمل قدرا قليلا من المعلومات مقارنة بالخاصية مكتب التي رغم العدد الصغير للكراسي (6)، إلاّ أنها تبرز المعلومات بشكل أفضل. كما أنّ كمية المعلومات المتوفرة في البيانات تساوي 0.3718. هذا المؤشر مهم في عملية التحليل.

3. جدول التكرارات النسبية للأعمدة:

$$f_{ij}=\frac{n_{ij}}{n_{.j}}$$

	رمادي	بني	أسود	المجموع
خزانة	0,167	0,250	0,417	0.3
مكتب	0,167	0,667	0,500	0.5
كرسي	0,667	0,083	0,083	0.2
المجموع	1	1	1	1

العمود المتوسط:

$$C_m=\begin{bmatrix}0.3\\0.5\\0.2\end{bmatrix},\ D_i=\begin{bmatrix}0.3 & 0 & 0\\0 & 0.5 & 0\\0 & 0 & 0.2\end{bmatrix}$$

تحليل جدول الأعمدة:

– تحليل عام: حساب المسافة بين كل خاصية في جدول الأعمدة والعمود المتوسط (الموقع بالنسبة للمتوسط).

$$d^2_{x^2}(j,\ c_m)=\sum_{j=1}^{p}\frac{1}{f_{i.}}\left(\frac{f_{ij}}{f_{.j}}-f_{i.}\right)^2$$

d^2(رمادي) $=\frac{1}{0.3}(0.167-0.3)^2+\frac{1}{0.5}(0.167-0.5)^2+\frac{1}{0.2}(0.667-0.2)^2=1.369$

d^2(بني)$=\frac{1}{0.3}(0.250-0.3)^2+\frac{1}{0.5}(0.667-0.5)^2+\frac{1}{0.2}(0.083-0.2)^2=0.131$

d^2(أسود)$=\frac{1}{0.3}(0.417-0.3)^2+\frac{1}{0.5}(0.500-0.5)^2+\frac{1}{0.2}(0.083-0.2)^2=0.113$

حساب التباين:

Inertie(رمادي) =poids(رمادي) x d ²(رمادي)= $\frac{6}{30}$ x 1.369= 0.273

	رمادي	بني	أسود	
DISTO²	1.369	0.131	0.113	
Poids	0.2	0.4	0.4	المجموع
Inertie	0.2738	0.0524	0.0452	0.3714

يمكننا ملاحظة أن انتشار اللون الرمادي في التجهيزات المكتبية مختلف عن بقية الألوان، كما أنّ خاصية اللون الرمادي هي الأكثر حملا للمعلومات من بقية الخصائص.

للتحليل المعمق للعلاقة بين الأسطر نحسب المسافة بين كل عمودين في هذا الجدول وفقا للعلاقة الموالية:

$$d^2_{\chi^2}\ (j, j')= \sum_{i=1}^{n} \frac{1}{f_{i.}} (\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}})^2$$

d2(رمادي, بني)= $\frac{1}{0.3}(0.167-0.250)^2 + \frac{1}{0.5}(0.167-0.667)^2 + \frac{1}{0.2}(0.667-0.083)^2 = 2.207$

d2(رمادي, أسود)= $\frac{1}{0.3}(0.167-0.417)^2 + \frac{1}{0.5}(0.167-0.500)^2 + \frac{1}{0.2}(0.667-0.083)^2 = 1.946$

d2(بني, أسود)= $\frac{1}{0.3}(0.250-0.417)^2 + \frac{1}{0.5}(0.667-0.500)^2 + \frac{1}{0.2}(0.083-0.083)^2 = 0.06$

يمكننا استنتاج أن اللون البني أقرب إلى اللون الأسود. وبعد اتمام جميع العمليات يكون جدول المسافات كما يلي:

	رمادي	بني	أسود
رمادي	0	2.207	1.946
بني		0	0.06
أسود			0

4. مؤشر التجاذب والتنافر: نحسب مؤشر التجاذب والتنافر بين شروط المتغيرين بالاعتماد على جدول التكرارات النسبية، باستخدام العلاقة:

$$t_{ij} = \frac{f_{ij}}{f_{i.}f_{.j}}$$

	رمادي	بني	أسود
خزانة	0.55	0.83	1.391
مكتب	0.33	1.335	1
كرسي	3.335	0.412	0.412

يظهر من النتائج أنّ الخاصيتين كرسي ورمادي تتجاذب، وكذلك الخاصيتين مكتب واللون بني، وخزانة واللون الاسود. كما أنّ الخاصية مكتب واللون الأسود مستقلتين، أمّا بقية الخصائص فهي تتنافر.

5. كاي تربيع للاستقلالية: سنحسب في الخطوة الأولى جدول التكرارات الفرضية (المتوقعة) بالاعتماد على جدول التقاطع.

	رمادي	بني	أسود
خزانة	1.8	3.6	3.6
مكتب	3	6	6
كرسي	1.2	2.4	2.4

نضرب مجموع عناصر السطر في مجموع عناصر العمود الذي تنتمي له الخلية، ونقسم الناتج على المجموع الكلي. أو عن طريق العلاقة في معادلة حساب كاي تربيع.

$n_{i.}*n_{.j}/n$

مثال: التكرار المتوقع للخلية الناتجة من تقاطع السطر الأول مع العمود الأول:

Expected frequency $n_{11}= \frac{9*6}{30}= 1.8$

ولو حسبناه بالطريقة الثانية نجد نفس النتيجة:

$n_{1.}*n_{.1}/n=9*6/30=1.8$

نستخدم الصيغة الموالية لحساب كاي تربيع.

$$X^2= \sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(nij-(ni.*n.j/n))^2}{ni.*n.j/n}$$

التكرار النظري أو المتوقع $ni.*n.j/n$

$$X^2_{11}= = \frac{(nij-(ni.*n.j/n))^2}{ni.*n.j/n}= \frac{(1-(1.8))^2}{1.8}= 0.355$$

	رمادي	بني	أسود	المجموع
خزانة	0.355	0.1	0.544	0.999
مكتب	1.333	0.667	0	2
كرسي	6.533	0.816	0.816	8.165
المجموع	8.221	1.583	1.36	11.164

نحسب درجات الحرية من العلاقة:

$(r-1)(c-1)= (3-1)(3-1)=4$

من الجدول نلاحظ أنّ قيمة كاي تربيع المحسوبة تساوي 11.164، نقارنها بكاي تربيع الجدولية عند مستوى دلالة 0.05 فأقل ودرجات حرية 4. القيمة الجدولية لكاي تربيع 9.49، وبما أنّ قيمة كاي تربيع المحسوبة تقع ضمن منطقة رفض فرضية العدم، أكبر من القيمة الجدولية، فإنّ هناك ارتباط بين الأسطر والأعمدة، أي بين التجهيزات والألوان.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09

ملاحظة مهمة: من الشروط الضرورية لاختبار كاي تربيع أن تحتوي 80 في المائة من الخلايا على تكرارات متوقعة تفوق 5. لأنّ القيم المنخفضة للتكرارات المتوقعة تؤدي إلى تضخيم قيمة كاي تربيع، مما قد ينتج عنه رفض فرضية العدم وهي صحيحة.

ملاحظة: يمكن حساب التباين الكلي بقسمة قيمة كاي تربيع على حجم العينة.

$$\text{PHI-2}=\frac{X^2}{n}=\frac{11.164}{30}=0.3721$$

وهي كمية المعلومات المتاحة من الارتباط بين التجهيزات والألوان.

يمكننا كذلك معرفة العلاقات (بين خاصيات كل متغير) التي تساهم بأكبر قدر من المعلومات، باستخدام الأخطاء المعيارية.

يتم حساب الأخطاء المعيارية من العلاقة التالية:

$$R_{ij}=\frac{nij-eij}{\sqrt{eij}}$$

يتبع التوزيع الطبيعي تقريبا، في حالة كانت القيمة المعيارية أكبر من القيمة المطلقة لـ 2 فهي دالة احصائيا.

جدول الأخطاء المعيارية

	رمادي	بني	أسود
خزانة	-0.596	-0.316	0.778
مكتب	-1.54	0.8	0
كرسي	2.577	-0.903	-0.903

يمكننا كذلك بالاعتماد على الجدول تحديد الخاصيات التي تتجاذب وتتنافر من خلال الإشارة، فعندما تكون الإشارة موجبة فهي تتجاذب، وعندما تكون سالبة تتنافر. وعندما تكون مساوية للصفر فهناك استقلالية بين السطر والعمود.

حساب المعلومات التي تحملها كل خلية من العلاقة:

$$C=\frac{R^2}{X^2}$$

نسبة مساهمة كل خلية في المعلومات %

	أسود	بني	رمادي
خزانة	5.42	0.89	3.181
مكتب	0	5.73	21.24
كرسي	7.30	7.30	59.485

6. حساب مصفوفة التباين والتباين المشترك V.

1.6. حساب مصفوفة المعلومات (التباين والتباين المشترك) وفقا للعلاقة المصفوفية:

$$H=D_r^{\frac{-1}{2}}(N-C_mL_m^T)D_c^{\frac{-1}{2}}$$

$$V=H^TH$$

$$H=\begin{bmatrix}\frac{1}{\sqrt{0.3}} & 0 & 0\\ 0 & \frac{1}{\sqrt{0.5}} & 0\\ 0 & 0 & \frac{1}{\sqrt{0.2}}\end{bmatrix}\left(\begin{bmatrix}0.033 & 0.100 & 0.167\\ 0.033 & 0.267 & 0.200\\ 0.133 & 0.033 & 0.033\end{bmatrix}-\begin{bmatrix}0.3\\0.5\\0.2\end{bmatrix}*[0.2 \quad 0.4 \quad 0.4]\right)\begin{bmatrix}\frac{1}{\sqrt{0.2}} & 0 & 0\\ 0 & \frac{1}{\sqrt{0.4}} & 0\\ 0 & 0 & \frac{1}{\sqrt{0.4}}\end{bmatrix}$$

$$H=\begin{matrix}-0.110 & -0.058 & 0.136\\ -0.212 & 0.150 & 0\\ 0.465 & -0.166 & -0.166\end{matrix}$$

$$V=\begin{matrix}-0.110 & -0.212 & 0.465\\ -0.058 & 0.150 & -0.166\\ 0.136 & 0 & -0.166\end{matrix}*\left(\begin{matrix}-0.110 & -0.058 & 0.136\\ -0.212 & 0.150 & 0\\ 0.465 & -0.166 & -0.166\end{matrix}\right)=\begin{bmatrix}0.273 & -0.103 & -0.092\\ -0.103 & 0.053 & 0.020\\ -0.092 & 0.020 & 0.046\end{bmatrix}$$

نحسب القيم الذاتية للمصفوفة V باستخدام المحدد، فتكون كما يلي:

$\lambda_1 = 0.34,\ \lambda_2=0.03,\ \lambda_3=0$

2.6. حساب مصفوفة المعلومات باستخدام العلاقة:

$$\left(\frac{f_{ij}-f_{i.}f_{.j}}{\sqrt{f_{i.}f_{.j}}}\right)=r_{ij} \quad ....(i=1....p,\ j=1.....q)$$

العبارة r_{ij} هي العبارة العامة للمصفوفة R، حيث:

$V=R^TR$

R= $\begin{bmatrix} \frac{0.033-0.3*0.2}{\sqrt{0.3*0.2}} & \frac{0.1-0.3*0.4}{\sqrt{0.3*0.4}} & \frac{0.167-0.3*0.4}{\sqrt{0.3*0.4}} \\ \frac{0.033-0.5*0.2}{\sqrt{0.5*0.2}} & \frac{0.267-0.5*0.4}{\sqrt{0.5*0.4}} & \frac{0.2-0.5*0.4}{\sqrt{0.5*0.4}} \\ \frac{0.133-0.2*0.2}{\sqrt{0.2*0.2}} & \frac{0.033-0.2*0.4}{\sqrt{0.2*0.4}} & \frac{0.033-0.2*0.4}{\sqrt{0.2*0.4}} \end{bmatrix} = \begin{bmatrix} -0.110 & -0.057 & 0.135 \\ 0.104 & 0.149 & 0 \\ 0.465 & -0.166 & -0.166 \end{bmatrix}$

V= $R^T.R= \begin{bmatrix} -0.110 & 0.104 & 0.465 \\ -0.057 & 0.149 & -0.166 \\ 0.135 & 0 & -0.166 \end{bmatrix} * \begin{bmatrix} -0.110 & -0.057 & 0.135 \\ 0.104 & 0.149 & 0 \\ 0.465 & -0.166 & -0.166 \end{bmatrix}$

V= $\begin{bmatrix} 0.237 & -0.055 & -0.092 \\ -0.055 & 0.053 & 0.020 \\ -0.092 & 0.020 & 0.046 \end{bmatrix}$

3.6. حساب مصفوفة المعلومات باستخدام عبارتيي التباين والتباين المشترك:

V= $\begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix}$

نستخدم صيغة التباين:

$$V_{jj}= \sum_i f_{i.} (\frac{f_{ij}}{\sqrt{f_{.j}} f_{i.}} - \sqrt{f_{.j}})^2$$

لحساب:

V_{11} ، V_{22}، V_{33}

$$V_{11}= 0.3(\frac{0.033}{\sqrt{0.2}*0.3} - \sqrt{0.2})^2+ 0.5(\frac{0.033}{\sqrt{0.2}*0.5} - \sqrt{0.2})^2+ 0.2(\frac{0.133}{\sqrt{0.2}*0.2} - \sqrt{0.2})^2=0.236$$

ونستخدم صيغة التباين المشترك:

$$V_{jj'} = \sum_i f_{i.} \left[\left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} - \sqrt{f_{.j}}\right)\left(\frac{f_{ij}'}{f_{i.}\sqrt{f_{.j'}}} - \sqrt{f_{.j}'}\right)\right]$$

لحساب:

V_{12}، V_{13}، V_{23}.

V= $\begin{bmatrix} 0.237 & -0.055 & -0.092 \\ -0.055 & 0.053 & 0.020 \\ -0.092 & 0.020 & 0.046 \end{bmatrix}$

4.6. حساب مصفوفة المعلومات من البيانات الأصلية (غير المحولة).

$V=H^T*H$

H= $\frac{n_{ij}}{\sqrt{n_{i.}*n_{.j}}}$

$$H=\begin{bmatrix}\frac{1}{\sqrt{9*6}} & \frac{3}{\sqrt{9*12}} & \frac{5}{\sqrt{9*12}}\\ \frac{1}{\sqrt{15*6}} & \frac{8}{\sqrt{15*12}} & \frac{6}{\sqrt{15*12}}\\ \frac{4}{\sqrt{6*6}} & \frac{1}{\sqrt{6*12}} & \frac{1}{\sqrt{6*12}}\end{bmatrix}=\begin{bmatrix}0.136 & 0.288 & 0.481\\ 0.105 & 0.596 & 0.447\\ 0.667 & 0.117 & 0.117\end{bmatrix}$$

$$V=\begin{bmatrix}0.136 & 0.105 & 0.667\\ 0.288 & 0.596 & 0.117\\ 0.481 & 0.447 & 0.117\end{bmatrix}*\begin{bmatrix}0.136 & 0.288 & 0.481\\ 0.105 & 0.596 & 0.447\\ 0.667 & 0.117 & 0.117\end{bmatrix}=\begin{bmatrix}0.474 & 0.180 & 0.190\\ 0.180 & 0.452 & 0.419\\ 0.190 & 0.419 & 0.445\end{bmatrix}$$

7. نحسب القيم الذاتية للمصفوفة V من خلال المحدد.

$$\det(W-\lambda I)=\begin{bmatrix}0.474-\lambda & 0.180 & 0.190\\ 0.180 & 0.452-\lambda & 0.419\\ 0.190 & 0.419 & 0.445-\lambda\end{bmatrix}=0$$

وبحل كثير الحدود نحصل على قيم الجذور المميزة، كما في الجدول الموالي:

القيمة الذاتية	C	C(I)
1	القيمة الذاتية 1	تعادل القيمة الذاتية 0
0.342	λ_1	تعادل القيمة الذاتية الأولى
0.003	λ_2	تعادل القيمة الذاتية الثانية

الأشعة الذاتية للمصفوفة C:

U	U_1	U_2
0.706	−0.894	−0.022
1	0.333	−0.699
1	0.297	0.714

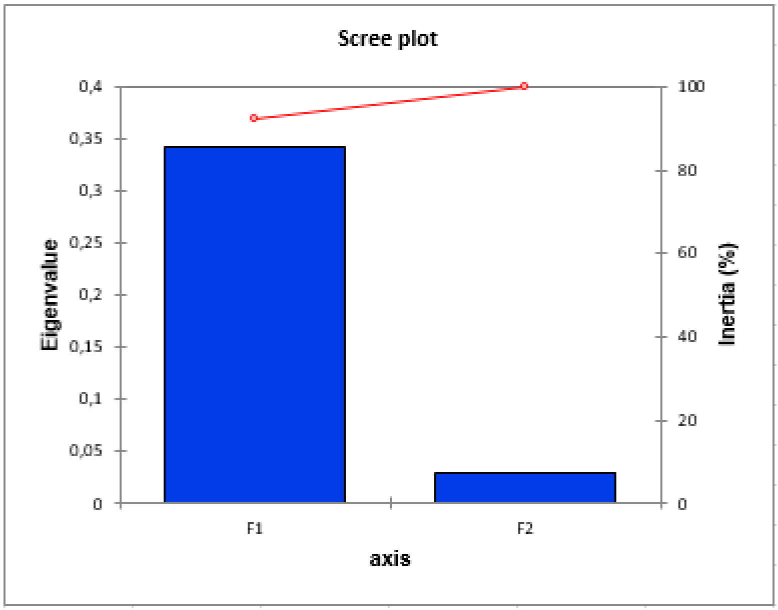
تحديد عدد المحاور:

$$H_{max}=\min(q-1,p-1)$$

واعتمادا على قاعدة كايزر، فإننا نحتفظ بالمحور الذي يوفر نسبة كثافة أعلى من $\frac{1}{H_{max}}$.

لدينا عدد الأسطر يساوي عدد الأعمدة، والعدد الأقصى للمحاور هو q−1 (3−1=2)، وهذا يعني أننا سنحتفظ بالمحور الذي يوفر على الأقل 50 في المائة من المعلومة، ولذلك نحتفظ بالمحور الأول وجوبا، والمحور الثاني نحتفظ به فقط لعرض التمثيل البياني.

معيار التمثيل البياني:



نسبة المعلومات المتاحة التي يتم استرجاعها بواسطة المحور العاملي الأول والثاني:

القيمة الذاتية	W	R
0.34	0.3428/0.3722	92.10
0.03	0.0294/0.3722	7.90
المجموع	0.3428+0.0294/0.3722	100

8. حساب المركبات الرئيسية.

1.8. حساب المركبات الرئيسية لسحابة الأسطر:

$$F_\alpha(i) = \sum_{j=1}^{q}\left(\frac{U_{\alpha j} f_{ij}}{f_{i.}\sqrt{f_{.j}}}\right)$$

$$F_1(1) = \frac{-0.89*0.033}{0.3\sqrt{0.2}} + \frac{0.33*0.100}{0.3\sqrt{0.4}} + \frac{0.3*0.167}{0.3\sqrt{0.4}} =$$

$$F_1 = \begin{bmatrix} 0.22 \\ 0.35 \\ -1.16 \end{bmatrix},\ F_2 = \begin{bmatrix} 0.25 \\ -0.14 \\ -0.03 \end{bmatrix}$$

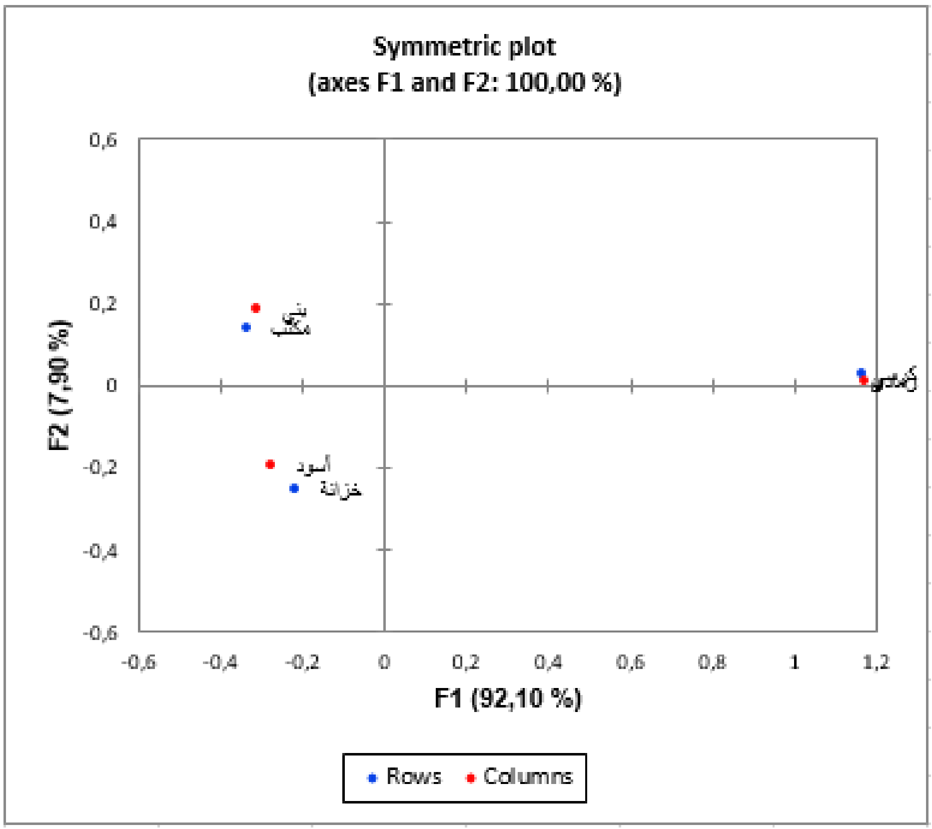
المتوسط	أسود	بني	رمادي	
0.3	0,167	0,100	0,033	خزانة
0.5	0,200	0,267	0,033	مكتب
0.2	0,033	0,033	0,133	كرسي
1	0.4	0.4	0.2	المتوسط

8.2. حساب المركبات الرئيسية لسحابة الأعمدة:

$$G_{Jj}^{\alpha} = \frac{1}{\sqrt{\lambda}} \sum_i \sqrt{\frac{K}{K.j}} * F_{Ii}^{\alpha}$$

$$G_J^{1} = \begin{bmatrix} -1.17 \\ 0.30 \\ 0.27 \end{bmatrix},\ G_J^{2} = \begin{bmatrix} -0.009 \\ 0.19 \\ -0.18 \end{bmatrix}$$

9. تمثيل الأسطر والأعمدة في المستوى:



10. المساهمة المطلقة والنسبية للأسطر والأعمدة:

10.1. المساهمة المطلقة لعناصر سحابة الأسطر على المحور α:

$$\alpha_i^{\alpha} = \frac{Ki.}{K} * \frac{(F_i^{\alpha})^2}{\lambda\alpha} * 100$$

$$\alpha_1^{1} = \frac{9}{30} * \frac{(0.21)^2}{0.34} * 100 = 3.89$$

ونلخص النتائج في الجدول الموالي:

الأسطر	α_i^1	α_i^2
فلاح	3.89	0,026
مدير	0,020	-0,027
إطار	-0,263	0,016

نلاحظ من خلال الجدول أنّ هو المحدد للمحور الأول و هو المحدد للمحور الثاني.

2.9. المساهمة النسبية للأسطر على المحور α:

$$C_i^{\alpha} = \frac{(F_i^{\alpha})^2}{\sum_{j=1}^{p}\frac{K}{K.j}(\frac{Kij}{Ki.}-\frac{K.j}{K})^2}$$

$$C_1^{1} = \frac{(0.21)^2}{\left(\frac{30}{6}(\frac{1}{9}-\frac{6}{30})^2\right)+\left(\frac{30}{12}(\frac{3}{9}-\frac{12}{30})^2\right)+\left(\frac{30}{12}(\frac{5}{9}-\frac{12}{30})^2\right)}$$

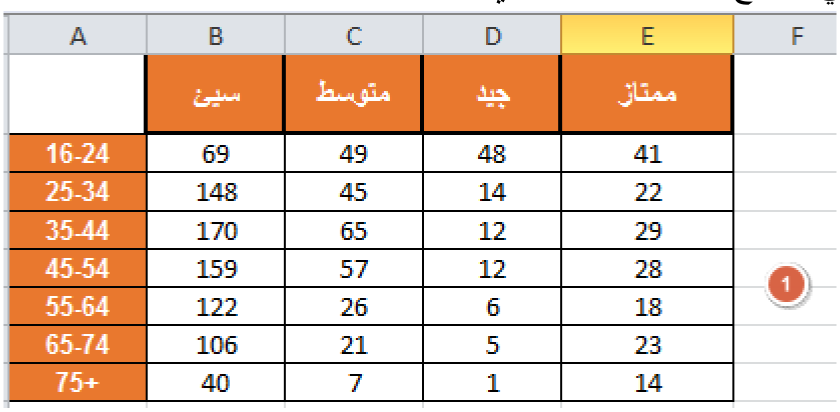
التطبيق باستخدام XLSTAT

مثال: تم استطلاع رأي رواد السينما حول فيلم شاهدوه، وطلب منهم تحديد فئة أعمارهم.

	ضعيف	متوسط	جيد	ممتاز
16-24	69	49	48	41
25-34	148	45	14	22
35-44	170	65	12	29
45-54	159	57	12	28
55-64	122	26	6	18
65-74	106	21	5	23
75+	40	7	1	14

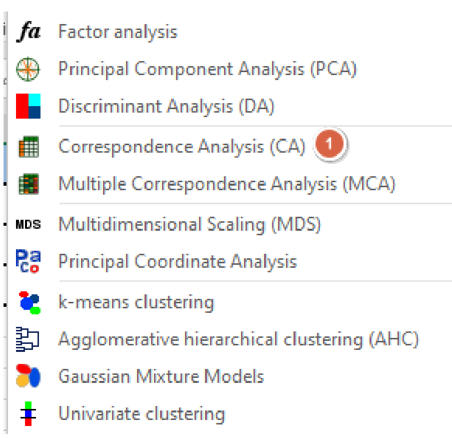
لتنفيذ التحليل العاملي بالتوفيقات في برنامج XLSTAT نتبع الخطوات الآتية:

في الخطوة الأولى ننقل جدول التقاطع في برنامج Excel بشكل عادي.

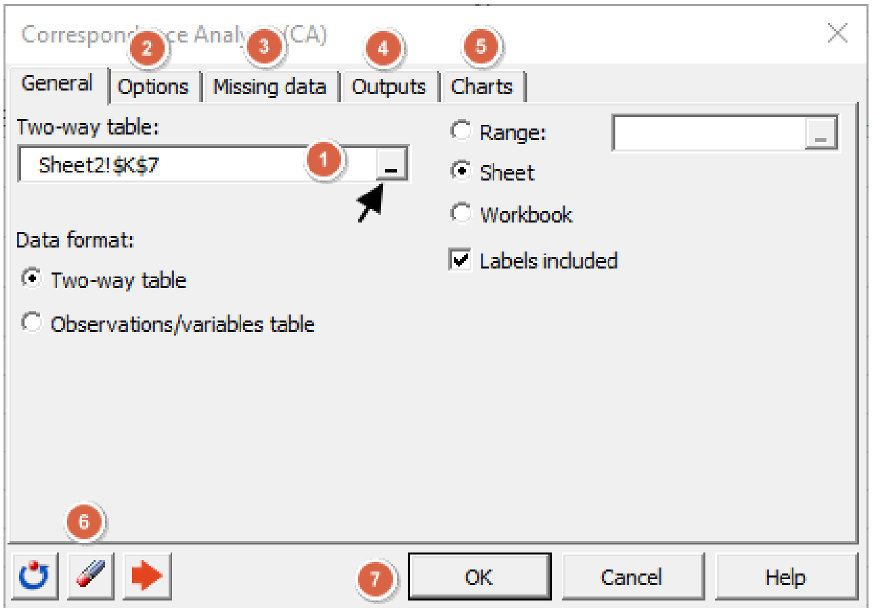


A	B	C	D	E	F
	سيئ	متوسط	جيد	ممتاز	
16-24	69	49	48	41	
25-34	148	45	14	22	
35-44	170	65	12	29	
45-54	159	57	12	28	1
55-64	122	26	6	18	
65-74	106	21	5	23	
75+	40	7	1	14	

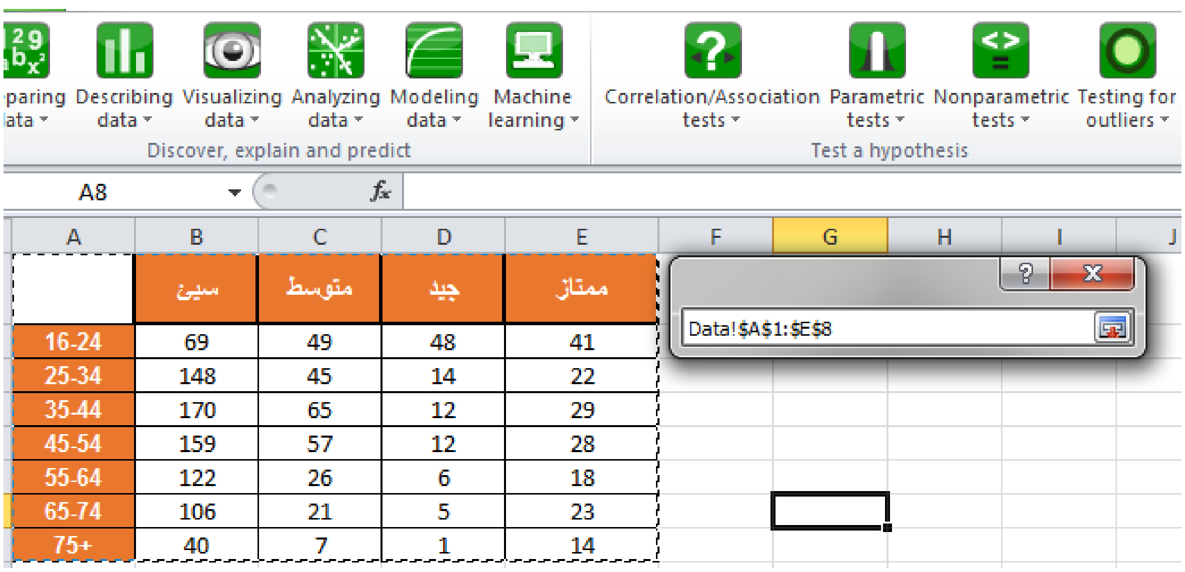
1. ننقل بيانات جدول التقاطع إلى برنامج excel.
2. نضغط على Analyzing data.



ومن القائمة المنبثقة نختار Correspondence Analysis (CA)، فتظهر النافذة الموالية:



1. نضغط على علامة – لتحديد حدود جدول التقاطع لدينا، قم بتضليل جميع خانات الجدول، بما فيها التسميات.



	سيئ	متوسط	جيد	ممتاز
16-24	69	49	48	41
25-34	148	45	14	22
35-44	170	65	12	29
45-54	159	57	12	28
55-64	122	26	6	18
65-74	106	21	5	23
75+	40	7	1	14

في مثالنا هذا تم تحديد الخلايا من A1 حتى E8، ثم اضغط المربع الأزرق للعودة إلى النافذة الرئيسية. بعد ذلك يمكنك تحديد أحد الخيارات التالية:

– جدول ثنائي الاتجاه (**Two-way table**): حدد هذا الخيار إذا كانت بياناتك تتوافق مع جدول ثنائي الاتجاه، حيث تحتوي الخلايا على التكرارات المقابلة للفئات المختلفة لمتغيرين نوعيين (في هذه الحالة يكون جدول التقاطع أكثر دقة)، أو أي نوع آخر من القيم.

– جدول المشاهدات/المتغيرات (Observations/variables table): حدد هذا الخيار إذا كانت بياناتك تتوافق مع N من المشاهدات الموصوفة في متغيرين نوعيين، ويتوافق هذا النوع من الجداول عادةً مع استطلاع يتكون من سؤالين. وأثناء العمليات الحسابية، سيقوم XLSTAT تلقائيًا بتحويل هذا الجدول إلى جدول تقاطع.

– النطاق أو المدى (Range): قم بتنشيط هذا الخيار إذا كنت تريد عرض النتائج بدءًا من خلية في ورقة عمل موجودة. ثم حدد الخلية المقابلة.

– الورقة (Sheet): قم بتنشيط هذا الخيار لعرض النتائج في ورقة عمل جديدة من المصنف النشط.

– المصنف (Workbook): قم بتنشيط هذا الخيار لعرض النتائج في مصنف جديد.

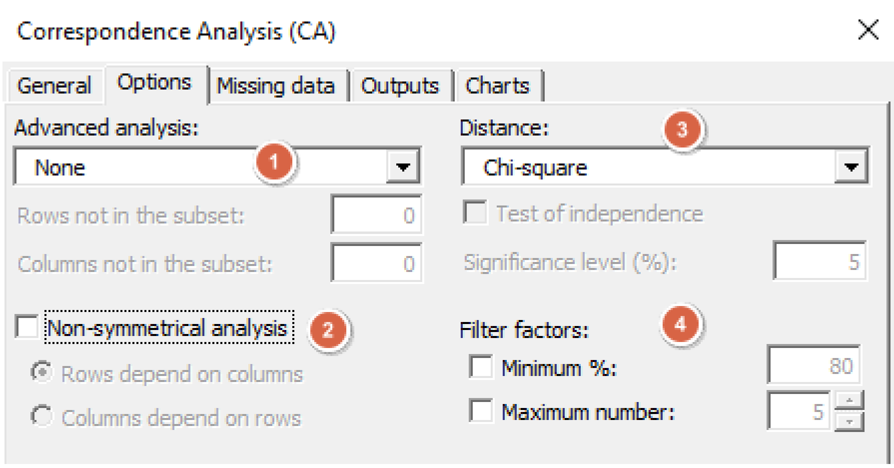
– تضمين التسميات (Labels included): هذا الخيار مهم جدا غذا كنت تريد عرض تسميات الصفوف والأعمدة، وإذا تم تضمين تسميات الأعمدة والصفوف في التحديد.

– تسميات المتغيرات (Variable labels): يكون هذا الخيار مرئيًا فقط إذا حددت تنسيق جدول المشاهدات/ المتغيرات. قم بتنشيط هذا الخيار إذا كان الصف الأول يحتوي على تسميات متغيرة (حالة جدول المشاهدات/ متغيرات) أو تسميات الفئات (حالة جدول منفصل).

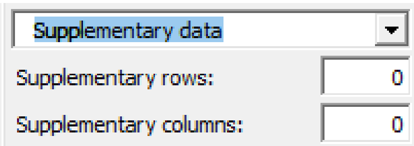
– الأوزان: يكون هذا الخيار مرئيًا فقط إذا حددت تنسيق جدول المشاهدات/المتغيرات. قم بتنشيط هذا الخيار إذا كنت تريد ترجيح المشاهدات. إذا لم تقم بتنشيط هذا الخيار، فسيتم اعتبار الأوزان مساوية لـ 1. يجب أن تكون الأوزان أكبر أو تساوي 0. إذا تم تنشيط خيار "التسميات المتغيرة" ، فتأكد من تحديد رأس التحديد أيضًا.

بالنسبة لمثالنا سنحدد جدول ثنائي الاتجاه (Two-way table)، وورقة (Sheet) لعرض النتائج في ورقة عمل جديدة في نفس المصنف النشط . ونحدد كذلك خيار تضمين التسميات (Labels included)، لعرض أسماء الأسطر والأعمدة في المخرجات.

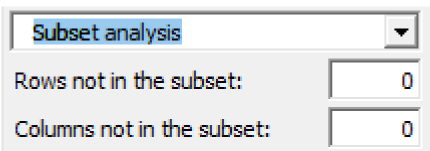
2. نضغط Options:



2.1. البيانات التكميلية (Supplementary data) (وهي ما نطلق عليه النقاط الإضافية- أنظر الصفحة....): إذا حددت هذا الخيار، فيمكنك إدخال عدد الصفوف و/ أو الأعمدة التكميلية (الصفوف والأعمدة التكميلية عبارة عن بيانات سلبية لا تؤخذ في الاعتبار عند حساب مساحة التمثيل). يتم حساب إحداثياتهم لاحقًا. لاحظ أن البيانات التكميلية يجب أن تكون الصفوف و/ أو الأعمدة الأخيرة في جدول البيانات.



2.1.1. تحليل المجموعة الفرعية (Subset analysis): إذا حددت هذا الخيار، يمكنك عندئذٍ إدخال عدد الصفوف و/أو الأعمدة لاستبعادها من تحليل المجموعة الفرعية. لاحظ أن البيانات المستبعدة يجب أن تكون الصفوف و/أو الأعمدة الأخيرة في جدول البيانات.



2.2. التحليل غير المتناظر (Non-symmetrical analysis): يسمح هذا الخيار بحساب التحليل التقابلي غير المتناظر.[1]

- الصفوف تعتمد على الأعمدة (Rows depend on columns): حدد هذا الخيار إذا كنت تعتبر أن متغير الصف يعتمد على متغير العمود وإذا كنت تريد تحليل الارتباط بين كليهما مع مراعاة هذه التبعية.
- تعتمد الأعمدة على الصفوف (Columns depend on rows): حدد هذا الخيار إذا كنت تعتقد أن متغير العمود يعتمد على متغير الصف وإذا كنت تريد تحليل الارتباط بين كليهما مع مراعاة هذه التبعية.

3.2. المسافة (Distance): يسمح لنا هذا الخيار بحساب المسافة بين الاسطر والسطر المتوسط وبين الأسطر مع بعضها مثنى مثنى، وكذلك بالنسبة للأعمدة. ويمكننا اختيار استخدام مسافة كاي تربيع (Chi-Square)، بحسب الطريقة الكلاسيكية للتحليل التقابلي، أو باستخدام مسافة Hellinger[2]. هذا الخيار لا يتوفر إذا تم تحديد الخيار غير متناظر (non-symmetrical).

وكخلاصة، تم اقتراح ثلاث طرق للتحليل:

- التحليل العاملي للتوفيقات الكلاسيكي (CA): لا تحدد خيار "التحليل غير المتناظر" وحدد مسافة Chi-Square.
- التحليل العاملي للتوفيقات غير المتناظر (NSCA): حدد خيار "التحليل غير المتناظر" وحدد مسافة Chi-Square.

---

[1] -

[2] -

– التحليل العاملي للتوفيقات باستخدام مسافة Hellinger (HD): لا تحدد خيار "التحليل غير المتناظر" وحدد مسافة Hellinger.

– اختبار الاستقلالية (**Test of independence**): قم بتنشيط هذا الخيار إذا كنت تريد أن يقوم XLSTAT بحساب اختبار الاستقلالية بناءً على إحصائية مربع كاي.

– مستوى الدلالة (٪) (**Significance level**): أدخل قيمة مستوى الدلالة للاختبار (القيمة الافتراضية: 5٪).

* بالنسبة إلى مثالنا سنعتمد على التحليل العاملي للتوفيقات الكلاسيكي (CA).

4.2. تصفية العوامل (**Filter factors**): يمكنك تفعيل أحد الخيارين التاليين لتقليل عدد العوامل المعروضة:

– الحد الأدنى٪: قم بتنشيط هذا الخيار ثم أدخل الحد الأدنى للنسبة المئوية التي يجب الوصول إليها لتحديد عدد العوامل المراد عرضها.

– العدد الأقصى: قم بتنشيط هذا الخيار لتعيين الحد الأقصى لعدد العوامل التي يجب مراعاتها عند عرض النتائج.

3. نضغط على خيار القيم المفقودة (Missing data): للتعامل مع القيم المفقودة في حالة جداول التقاطع والجداول ذات الاتجاهين، يمكنك تحديد ما يلي من الخيارات:

– عدم قبول البيانات المفقودة: قم بتنشيط هذا الخيار بحيث يمنع XLSTAT العمليات الحسابية من المتابعة إذا تم الكشف عن بيانات مفقودة.

– استبدال البيانات المفقودة بـ 0: قم بتنشيط هذا الخيار إذا كنت تعتقد أن البيانات المفقودة تعادل القيمة 0.

– استبدال البيانات المفقودة بقيمتها المتوقعة: قم بتنشيط هذا الخيار إذا كنت تريد استبدال البيانات المفقودة بالقيمة المتوقعة. وتحسب كما يلي:

$$E_{n_{ij}}=\frac{n_{i.}n_{.j}}{n}$$

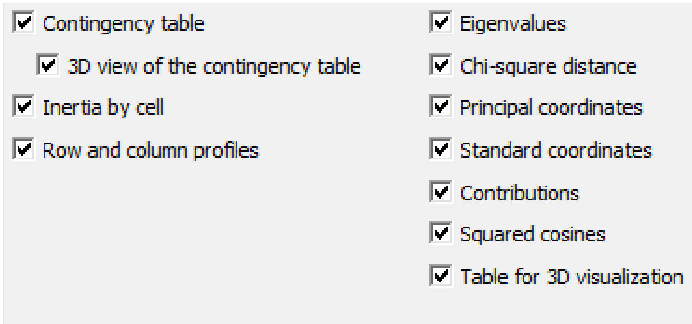
أمّا في حالة اختيار خيارات جداول المشاهدات/ المتغيرات (**Observations/variables table**):

– عدم قبول البيانات المفقودة: قم بتنشيط هذا الخيار بحيث يمنع XLSTAT العمليات الحسابية من المتابعة إذا تم الكشف عن بيانات مفقودة.

– إزالة المشاهدات: قم بتنشيط هذا الخيار لتجاهل المشاهدات التي تحتوي على بيانات مفقودة.

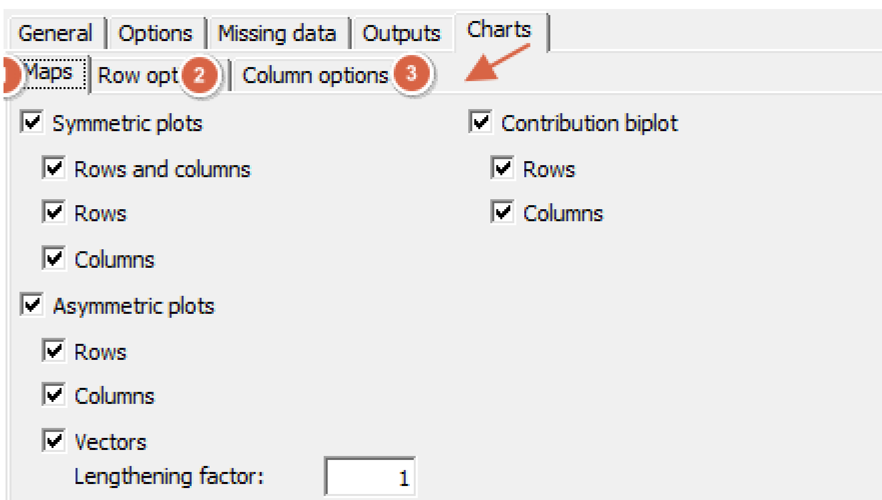
– تجميع القيم المفقودة في فئة جديدة: قم بتنشيط هذا الخيار لتجميع البيانات المفقودة في فئة جديدة من المتغير المقابل.

4. نضغط Outputs:



يمكننا اختيار اظهار المخرجات، بالترتيب: جدول التقاطع (**Contingency table**)، عرض ثلاثي الأبعاد لجدول التقاطع (**3D view of the contingency table**)، الكثافة لكل خلية في الجدول (**Inertia by cell**)، جدول التكرارات النسبية للأسطر وللأعمدة (**Row and column profiles**)، القيم الذاتية (**Eigenvalues**)، مسافة كاي تربيع (**Chi-square (or Hellinger) distances**)، المركبات الرئيسية لجدولي الأسطر والأعمدة (**Principal coordinates**)، المركبات القياسية لجدولي الأسطر والأعمدة (**Standard coordinates**)، المساهمة النسبية للأسطر والأعمدة في تشكيل المحاور العاملية (**Contributions**)، نسبة التمثيل على المحاور (جيب التمام التربيعي)(**Squared cosines**)، جدول العرض ثلاثي الأبعاد (**Table for 3D visualization**).

5. نضغط Charts:



1.5 Maps (الخرائط): كل هذه المخططات البيانية تم شرحها في الدرس بالتفصيل (أنظر الصفحة ....).

– المخططات المتناظرة (**Symmetric plots**): قم بتنشيط هذا الخيار لعرض المخططات التي تلعب فيها نقاط الصف ونقاط العمود دورًا متناظرا. تستند هذه الخرائط على الإحداثيات الرئيسية لنقاط الصفوف ونقاط الأعمدة.

– الصفوف والأعمدة: قم بتنشيط هذا الخيار لرسم مخطط يتم فيه عرض نقاط الصفوف ونقاط الأعمدة.

– الصفوف: قم بتنشيط هذا الخيار لإنشاء مخطط تُعرض فيه نقاط الصف فقط.

– الأعمدة: قم بتنشيط هذا الخيار لإنشاء مخطط تعرض فيه نقاط الأعمدة فقط.

– المخططات غير المتناظرة: قم بتنشيط هذا الخيار لعرض المخططات التي تلعب فيها نقاط الصف ونقاط الأعمدة دورًا غير متناظرا. تستخدم هذه المخططات من ناحية الإحداثيات الرئيسية ومن ناحية أخرى الإحداثيات القياسية.

– الصفوف: قم بتنشيط هذا الخيار لعرض مخطط حيث يتم عرض نقاط الصف باستخدام إحداثياتها الرئيسية، ويتم عرض نقاط الأعمدة باستخدام إحداثياتها القياسية.

– الأعمدة: قم بتنشيط هذا الخيار لعرض مخطط حيث يتم عرض نقاط الصف باستخدام إحداثياتها القياسية، ويتم عرض نقاط الأعمدة باستخدام إحداثياتها الرئيسية.

– المتجهات: قم بتنشيط هذا الخيار لعرض المتجهات للإحداثيات القياسية على المخططات غير المتناظرة.

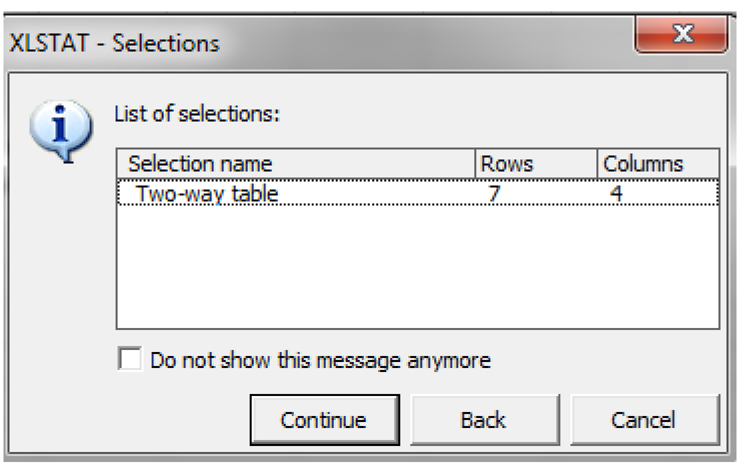
– عامل الطول: قم بتنشيط هذا الخيار لتعديل طول المتجهات.

– : قم بتنشيط هذا الخيار لعرض biplots المساهمة التي تلعب فيها نقاط الصف ونقاط العمود دورًا غير متناظر. تستخدم هذه المخططات من ناحية الإحداثيات الرئيسية ومن ناحية أخرى إحداثيات المساهمة التي تأخذ في الاعتبار الأوزان.

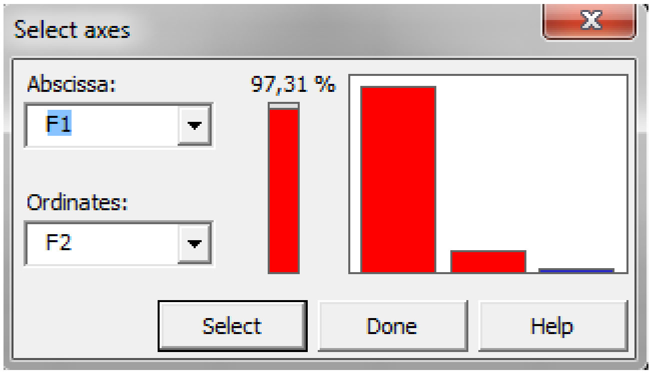
– الصفوف: قم بتنشيط هذا الخيار لعرض مخطط حيث يتم عرض نقاط الصف باستخدام إحداثياتها الرئيسية ، ويتم عرض نقاط الأعمدة باستخدام إحداثيات مساهمتها.

– الأعمدة: قم بتنشيط هذا الخيار لعرض مخطط حيث يتم عرض نقاط الصف باستخدام إحداثيات مساهمتها ، ويتم عرض نقاط الأعمدة باستخدام إحداثياتها الرئيسية.

2.5. اضغط موافق، يظهر مربع الحوار التالي.



لدينا جدول متقاطع بسبعة أسطر وأربعة أعمدة. اضغط مواصة، فيظهر مربع الحوار التالي.



حدد عدد المحاور، لاحظ انّ المحورين الأول والثاني يفسران معا 97 في المائة من التباين (القصور الذاتي) في سحابة النقط. لذلك، سنكتفي بهما، نضغط Done.

1. اختبار كاي تربيع للاستقلالية بين الأسطر والأعمدة (بين الفئة العمرية والحكم على جودة الفلم):

Chi-square (Observed value)	148,268
Chi-square (Critical value)	28,869
DF	18
p-value	<0.0001
alpha	0,050

يظهر من الجدول أن قيمة كاي تربيع للقيم المرصودة (المشاهدة) تساوي 148.268، وهي أكبر من قيمة كاي تربيع للقيم المتوقعة (النظرية) (28.869)، عند درجات حرية تساوي 18 ((7−1)(4−1))، ومستوى دلالة 0.0001، وهو أقل من مستوى الدلالة الحرج، الذي عنده نقبل أو نرفض فرضية العدم. وعليه، نرفض فرضية العدم ونقبل الفرضية البديلة، توجد علاقة بين الفئات العمرية ومستوى الحكم على جودة الفلم.

2. القصور الذاتي الكلي:

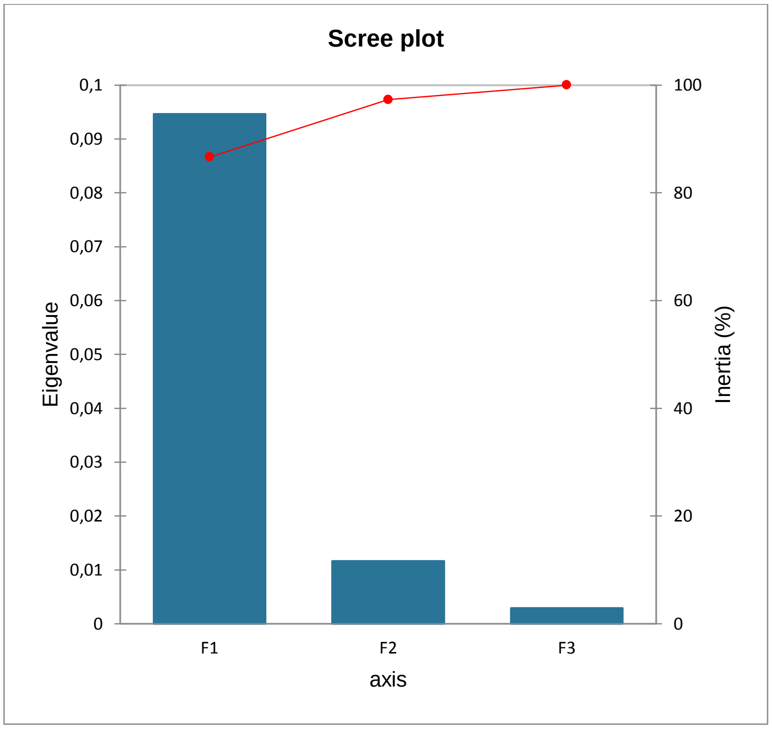
Total inertia:	0,109

3. القيم الذاتية ونسبة المساهمة في تشكيل المحاور، والنسبة التجميعية المتصاعدة.

	F1	F2	F3
Eigenvalue	0,095	0,012	0,003
Inertia %	86,640	10,674	2,685
Cumulative %	86,640	97,315	100,000

تتوافق قيم eigenvalues مع التباين المستخرج بواسطة كل عامل (بُعد). يمكن تقييم جودة التحليل من خلال العودة إلى جدول قيم eigenvalues أو الرسم البياني أدناه. يعتبر التحليل في هذا المثال ذا نوعية جيدة حيث أن مجموع قيمتي eigenvalues يضيف ما يصل إلى 97٪ من إجمالي القصور الذاتي.

4. الرسم البياني للقيم الذاتية:



لاحظ أنّ شكل المرفق تشكل عند القيمة الذاتية الثانية.

5. تحليل جدول الأسطر:

1.5. الوزن، المسافات، المسافات المربّعة إلى الأصل، والقصور الذاتي، والقصور الذاتي النسبي.

	Weight (relative)	Distance	Sq-Distance	Inertia	Inertia relative
16-24	0,153	0,718	0,516	0,079	0,720
25-34	0,169	0,117	0,014	0,002	0,021
35-44	0,203	0,152	0,023	0,005	0,043
45-54	0,189	0,124	0,015	0,003	0,027
55-64	0,127	0,235	0,055	0,007	0,064
65-74	0,114	0,239	0,057	0,007	0,060
75+	0,046	0,396	0,157	0,007	0,066

الفئات العمرية 25-34 و35-44 و45-54 لها أقصر مسافة إلى الأصل، مما يشير إلى أن ملفات تعريف هذه الفئات العمرية قريبة من السطر المتوسط.

2.5. جدول التكرارات النسبية للاسطر (ملف تعريف الأسطر).

	سيء	متوسط	جيد	ممتاز	Sum
16-24	0,333	0,237	0,232	0,198	1,000
25-34	0,646	0,197	0,061	0,096	1,000
35-44	0,616	0,236	0,043	0,105	1,000
45-54	0,621	0,223	0,047	0,109	1,000
55-64	0,709	0,151	0,035	0,105	1,000
65-74	0,684	0,135	0,032	0,148	1,000
75+	0,645	0,113	0,016	0,226	1,000
Mean	0,608	0,184	0,067	0,141	1,000

يتم عرض ملفات تعريف الصفوف (التكرار النسبي للأسطر) بالإضافة إلى ملف التعريف المتوسط (السطر المتوسط). في مثالنا، تكون الملفات الشخصية للفئات العمرية 25-34 و35-44 و45-54 قريبة من بعضها البعض ومن الملف الشخصي المتوسط. لقد تم توقع هذا الأخير من خلال المسافة إلى الأصل.

3.5. جدول المسافات (Chi-square distances) (مسافة مربع كاي) بين الفئات العمرية.

	16-24	25-34	35-44	45-54	55-64	65-74	75+
16-24	0	0,810	0,832	0,821	0,937	0,910	0,943
25-34	0,810	0	0,119	0,093	0,165	0,232	0,440
35-44	0,832	0,119	0	0,034	0,227	0,273	0,448
45-54	0,821	0,093	0,034	0	0,202	0,244	0,424
55-64	0,937	0,165	0,227	0,202	0	0,131	0,365
65-74	0,910	0,232	0,273	0,244	0,131	0	0,235
75+	0,943	0,440	0,448	0,424	0,365	0,235	0

تعطي المسافات بين الصفوف معلومات حول التشابه بين الفئات. مرة أخرى، يبدو أن الفئات العمرية 25-34 و35-44 و45-54 متشابهة، مع وجود مسافة أقل من 0.2.

4.5. الإحداثيات الرئيسية (المركبات الأساسية) (الأسطر):

	F1	F2	F3
16-24	0,718	0,025	0,008
25-34	-0,087	-0,056	0,056
35-44	-0,103	-0,097	-0,055
45-54	-0,097	-0,069	-0,036
55-64	-0,216	0,042	0,083
65-74	-0,183	0,152	0,019
75+	-0,162	0,342	-0,118

5.5. المركبات القياسية للأسطر:

	F1	F2	F3
16-24	2,332	0,234	0,139
25-34	-0,282	-0,516	1,027
35-44	-0,334	-0,902	-1,007
45-54	-0,315	-0,640	-0,666
55-64	-0,702	0,392	1,535
65-74	-0,595	1,411	0,358
75+	-0,527	3,164	-2,176

الإحداثيات القياسية هي إحداثيات رئيسية مقسومة على الجذر التربيعي لمعامل القيمة الذاتية المقابلة. ومجموع المربعات المرجح للإحداثيات القياسية يساوي 1 لكل عامل.

مثلا، القيمة الذاتية المقابلة للمركبة الرئيسية الأولى، هي 0.095، نقسم كل مركبة رئيسية على هذه القيمة نحصل المركبة القياسية (0.718/0.308=2.332).

6.5. نسب مساهمة الاسطر في تشكيل المحاور:

	Weight (relative)	F1	F2	F3
16-24	0,153	0,830	0,008	0,003
25-34	0,169	0,013	0,045	0,178
35-44	0,203	0,023	0,165	0,206
45-54	0,189	0,019	0,077	0,084
55-64	0,127	0,062	0,020	0,298
65-74	0,114	0,040	0,227	0,015
75+	0,046	0,013	0,457	0,216

تتوافق المساهمات مع أهمية كل فئة لكل عامل (بعد)، ومجموع المساهمات يساوي 1 لكل عامل. وكقاعدة عامة، إذا كانت المساهمة أكبر من واحد مقسوما على عدد الأسطر، فإن الفئة مهمة للعامل المحدد. في مثالنا، تعتبر المجموعة 16–24 مهمة للعامل F1، والمجموعتين 65–74 و 75+ مهمة للعامل F2.

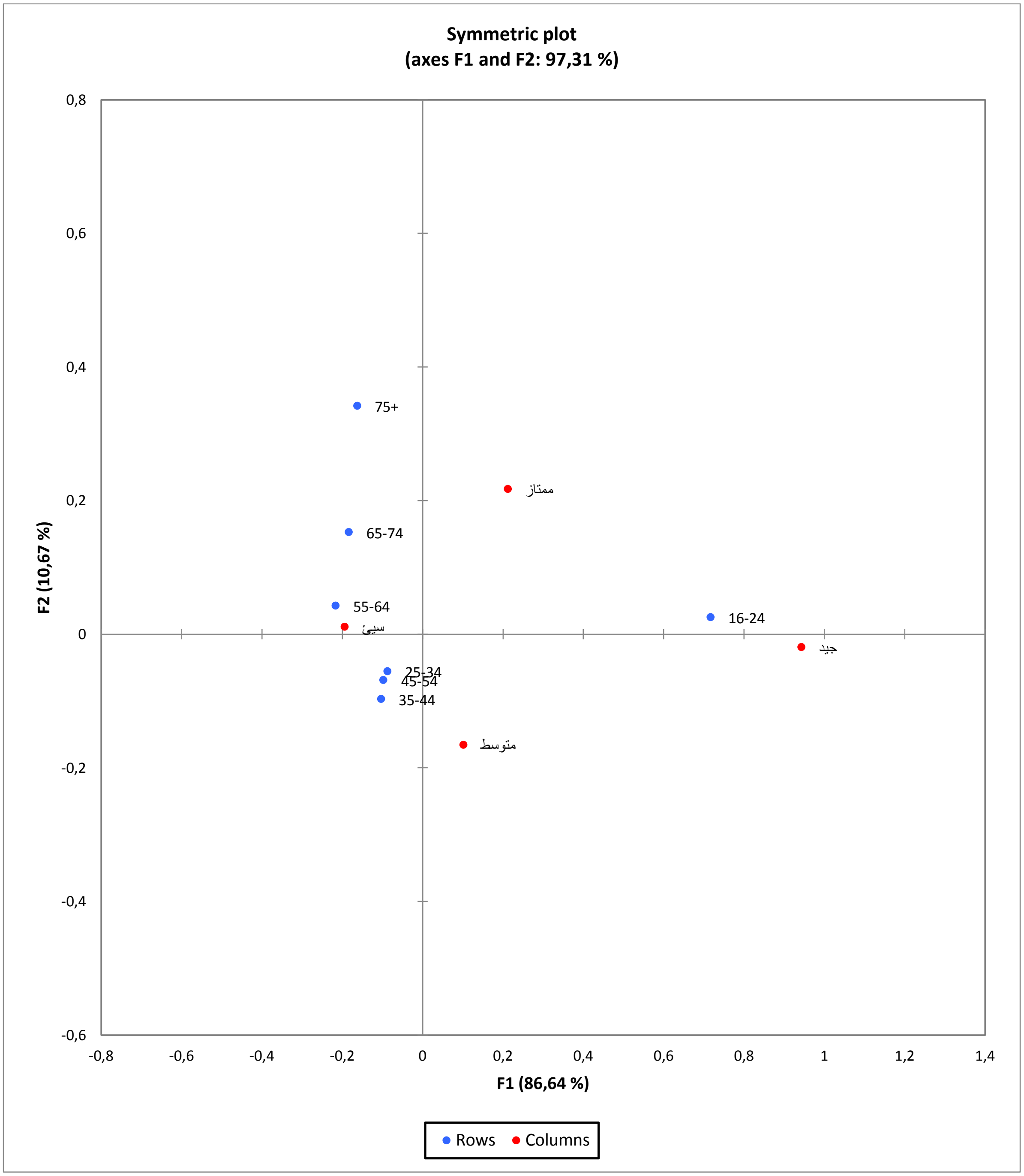
7.5. جودة تمثيل الأسطر:

	F1	F2	F3
16-24	0,999	0,001	0,000
25-34	0,549	0,226	0,225
35-44	0,458	0,412	0,129
45-54	0,607	0,309	0,084
55-64	0,843	0,032	0,125
65-74	0,587	0,407	0,007
75+	0,167	0,744	0,089

يعرض الجدول أعلاه جيب التمام التربيعي للصفوف. يمثل جيب التمام التربيعي أهمية كل عامل لكل فئة، ومجموع جيب التمام التربيعي يساوي 1 لفئة معينة. في مثالنا، يُعزى كل التباين في المجموعة 16–24 تقريبًا إلى العامل F1.

ملاحظة: يتم تحليل جدول الأعمدة بالطريقة نفسها.

6. التمثيل البياني:



التطبيق باستخدام برنامج SPSS

مثال: يحتوي الجدول الموالي على مجموعة من الشركات وتقييم خصائص بطاقات فيزا كارد التي تقدمها كل شركة.

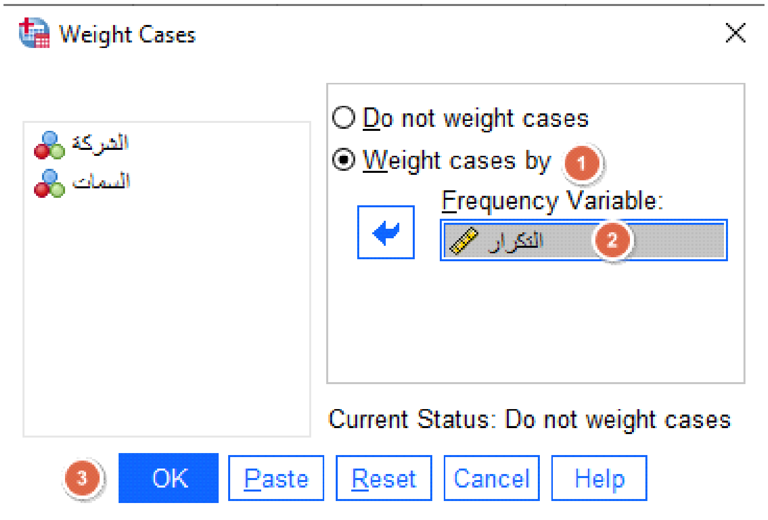
	Amazon	Flipcart	Myntra	Naaptol	Japong	Snapdeal	Loacalbania	المجموع
سهولة الاستعمال	56	65	15	11	12	8	8	175
الدفع الآمن	48	53	8	3	3	34	25	174
عروض خاصة	64	51	36	28	12	16	17	224
خدمة ما بعد البيع	41	35	3	65	8	8	46	206
موعد التسليم	32	19	5	31	25	14	13	139
سعر تنافسي	51	37	13	5	9	15	14	144
المجموع	292	260	80	143	69	95	123	1062

نقوم بنقل البيانات من الجدول إلى برنامج SPSS.

	الشركة	السمات	التكرار
1	1	1	56
2	2	1	65
3	3	1	15
4	4	1	11
5	5	1	12
6	6	1	8
7	7	1	8
8	1	2	48
9	2	2	53
10	3	2	8
11	4	2	3
12	5	2	3
13	6	2	34
14	7	2	25
15	1	3	64
16	2	3	51
17	3	3	36
18	4	3	28
19	5	3	12
20	6	3	16
21	7	3	17
22	1	4	41
23	2	4	35
24	3	4	3
25	4	4	65
26	5	4	8
27	6	4	8
28	7	4	46
29	1	5	32
30	2	5	19
31	3	5	5
32	4	5	31
33	5	5	25
34	6	5	14
35	7	5	13
36	1	6	51
37	2	6	37
38	3	6	13
39	4	6	5
40	5	6	9
41	6	6	15
42	7	6	14

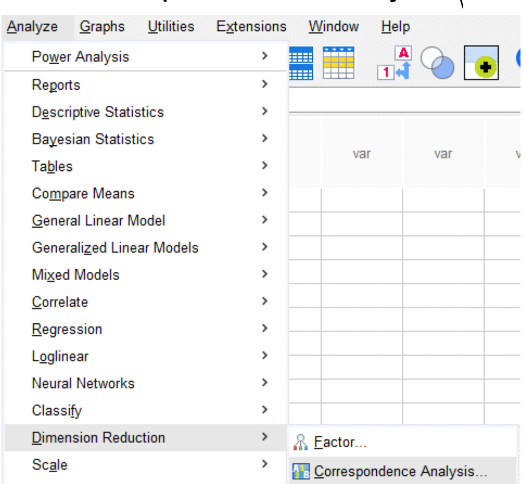
في الخطوة الأولى علينا أن نقوم بترجيح أوزان الحالات بواسطة التكرارات من خلال الأمر Weight Cases الموجود بالقائمة Data.

Data> Weight Cases…

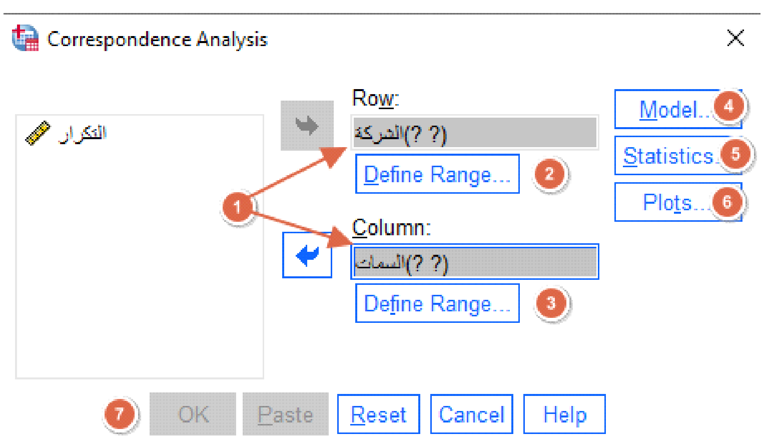


نؤشر على الخيار Weight cases by ثم ننقل التكرار إلى الخانة Frequency Variable ونضغط ok سيتم ترجيح أوزان المشاهدات، ولن يظهر أي تغيير على البيانات، لأنّ العملية تتم في الخلفية.

ثم من القائمة الرئيسية Analayze اختر Dimension Reduction ثم Correspondence Analysis

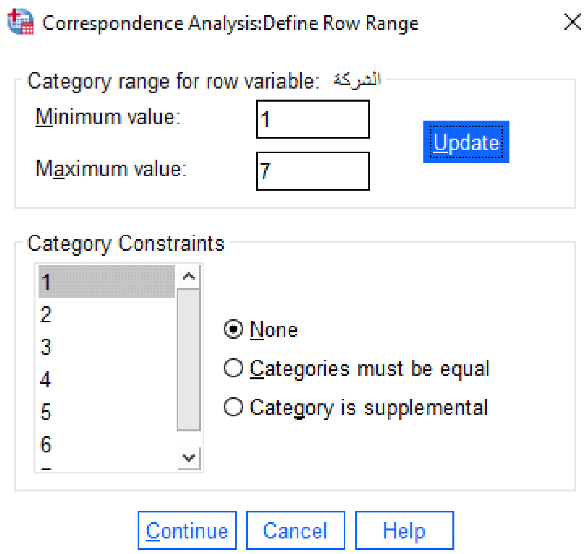


Analyze <Dimension Reduction<Correspondence Analysis…



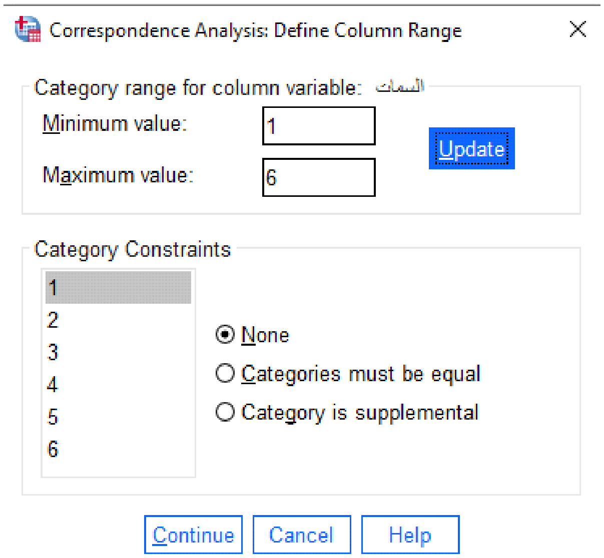
1. ننقل المتغير الاسمي (الشركة) إلى خانة الصفف (Row)، والمتغير الاسمي السمات إلى خانة العمود (Column).

2. نضغط Define Range ونحدد المدى، في حالتنا هذه توجد 7 شركات، يعني من 1 إلى 7.

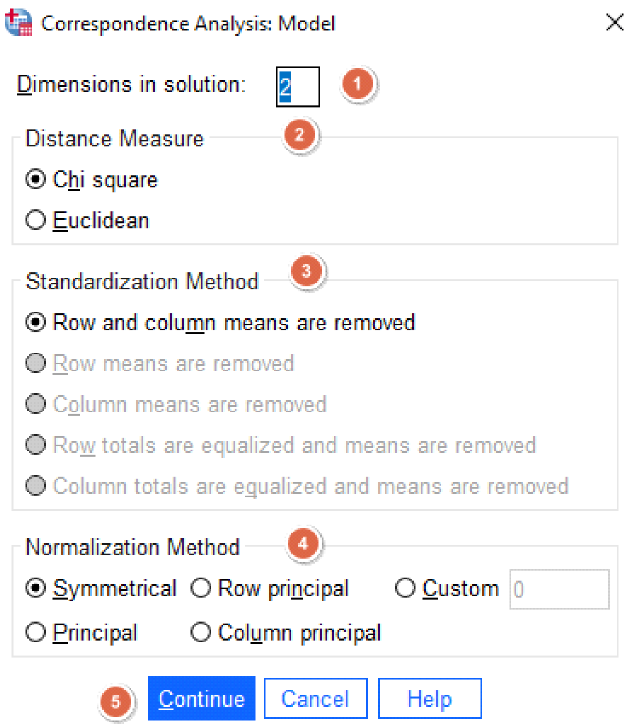


نضع 1 في أدنى قيمة، و7 في أعلى قيمة، ثم نضغط Update، ومن ثمّ Continue.

3. نضغط Define Range مرة أخرى ونحدد المدى، في حالتنا هذه توجد 6 سمات، يعني من 1 إلى 6.



4. نضغط النموذج (Model):



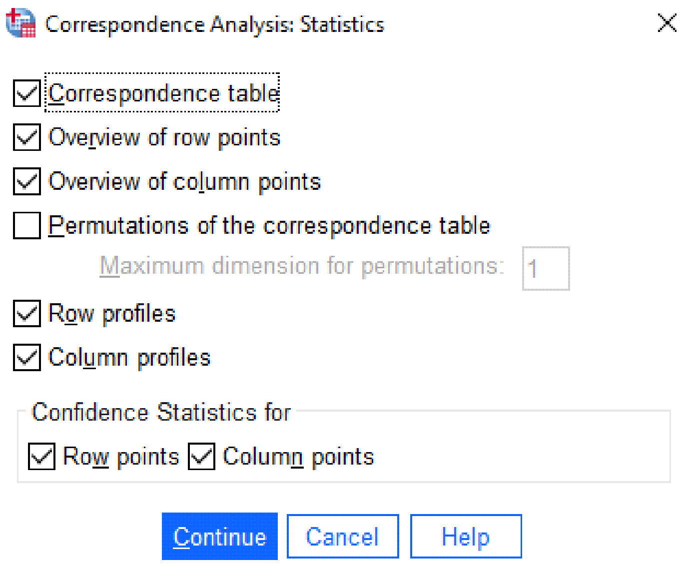
1. يمكنك تحديد عدد الأبعاد في الحل، افتراضيا تجدها 2.
2. نختار طريقة حساب المسافة (Distance Measure)، إمّا عن طريق كاي تربيع أو بطريقة المسافة الإقليدية.
3. نحدد طريقة الترجيح (طريقة التوحيد).

– Row and column means are removed: يتم توسيط الصفوف والأعمدة، هذه الطريقة تستخدم في التحليل التقابلي القياسي.

4. نختار طريقة التماثل (Symmetrical) أو أي طريقة أخرى من الطرق الخمسة.

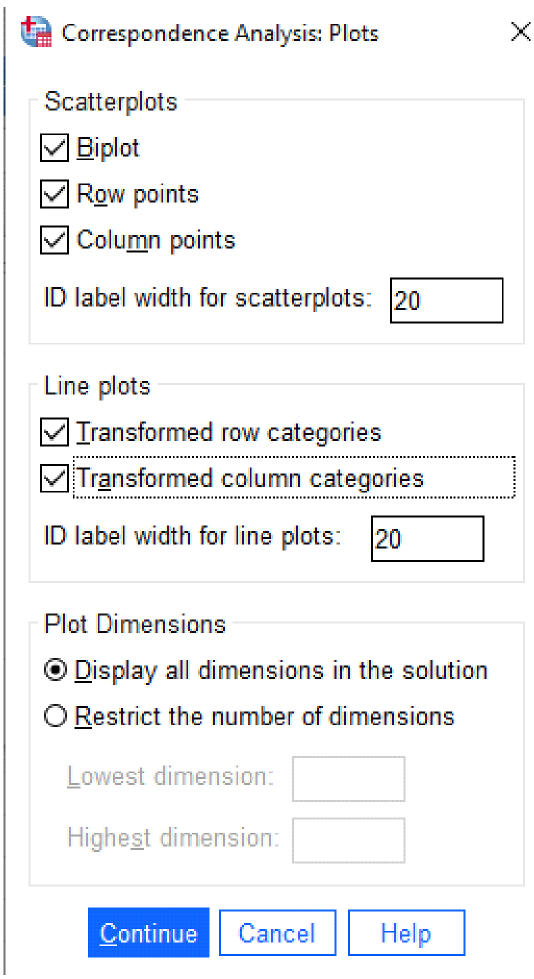
بالنسبة لطريقة التماثل: لكل بُعد، تكون درجات الصف هي المتوسط المرجح لدرجات العمود مقسومة على القيمة المفردة المطابقة، ودرجات العمود هي المتوسط المرجح لدرجات الصف مقسومة على القيمة المفردة المطابقة، استخدم هذه الطريقة إذا كنت تريد فحص الاختلافات أو أوجه التشابه بين فئات المتغيرين.

5. نضغط Continue.
6. نضغط Statistics:



يتيح لنا هذا الخيار عرض جدول التقاطع، وجدولي التكرارات النسبية للأسطر والأعمدة. وكذلك إلقاء نظرة عامة على نقاط الصف والعمود، مثل القصور الذاتي ومساهمة كل صف أو عمود في القصور الذاتي الكلي...الخ.

7. نضغط على Plots ونختار الرسوم البيانية التي نريدها، ونضغط Continue.



نعود لمربّع الحوار الرئيسي ونضغط ok.

الجدول الأول: جدول التقاطع (التقابل).

Correspondence Table

	السمات						
الشركة	سهولة الاستعمال	الدفع الآمن	عروض خاصة	خدمة ما بعد البيع	وقت التسليم	سعر تنافسي	Active Margin
Amazon	56	48	64	41	32	51	292
Flipkart	65	53	51	35	19	37	260
Myntra	15	8	36	3	5	13	80
Naaptol	11	3	28	65	31	5	143
Jabong	12	3	12	8	25	9	69
Snapdeal	8	34	16	8	14	15	95
Localbania	8	25	17	46	13	14	123
Active Margin	175	174	224	206	139	144	1062

الجدول الثاني: جدول التكرارات النسبية للأسطر.

Row Profiles

	السمات						
الشركة	سهولة الاستعمال	الدفع الآمن	عروض خاصة	خدمة ما بعد البيع	وقت التسليم	سعر تنافسي	Active Margin
Amazon	,192	,164	,219	,140	,110	,175	1,000
Flipkart	,250	,204	,196	,135	,073	,142	1,000
Myntra	,188	,100	,450	,038	,063	,163	1,000
Naaptol	,077	,021	,196	,455	,217	,035	1,000
Jabong	,174	,043	,174	,116	,362	,130	1,000
Snapdeal	,084	,358	,168	,084	,147	,158	1,000
Localbania	,065	,203	,138	,374	,106	,114	1,000
Mass	,165	,164	,211	,194	,131	,136	

الجدول الثالث: جدول التكرارات النسبية للاعمدة.

Column Profiles

	السمات						
الشركة	سهولة الاستعمال	الدفع الآمن	عروض خاصة	خدمة ما بعد البيع	وقت التسليم	سعر تنافسي	Mass
Amazon	,320	,276	,286	,199	,230	,354	,275
Flipkart	,371	,305	,228	,170	,137	,257	,245
Myntra	,086	,046	,161	,015	,036	,090	,075
Naaptol	,063	,017	,125	,316	,223	,035	,135
Jabong	,069	,017	,054	,039	,180	,063	,065
Snapdeal	,046	,195	,071	,039	,101	,104	,089
Localbania	,046	,144	,076	,223	,094	,097	,116
Active Margin	1,000	1,000	1,000	1,000	1,000	1,000	

الجدول الرابع: ملخّص.

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	,382	,146			,575	,575	,029	-,031
2	,231	,053			,210	,784	,028	
3	,188	,035			,139	,923		
4	,129	,017			,065	,988		
5	,054	,003			,012	1,000		
Total		,254	269,570	<,001[a]	1,000	1,000		

a. 30 degrees of freedom

هذا الجدول مهم جدا في عملية التحليل، فهو يحتوي على قيمة كاي تربيع للاستقلالية، ومستوى الدلالة، ونلاحظ أنّ قيمة كاي تربيع تساوي 269.57 عند مستوى دلالة 0.0001، وهو أقل من مستوى الدلالة الحرج 0.05، وعليه، يمكن القول أنّ متغيرات الصفوف والأعمدة متغيرات مرتبطة وليست مستقلة، بمعنى آخر أنّه توجد علاقة بين الشركات والسمات المقاسة. كما نلاحظ في العمود الثالث قيمة القصور الذاتي المفسر بواسطة كل مركبة أساسية، ويتضح أنّ المركبتين الأولى والثانية يفسران معا 92.3 في المائة من التباين الكلي. مما يعني أنّ التمثيل جيد جدا.

الجدول الخامس: نسبة مساهمة الأسطر في تشكيل المحاور.

Overview Row Points[a]

الشركة	Mass	Score in Dimension		Inertia	Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
		1	2		1	2	1	2	Total
Amazon	,275	-,268	,077	,009	,052	,007	,798	,040	,838
Flipkart	,245	-,412	-,109	,024	,109	,013	,654	,028	,682
Myntra	,075	-,640	,773	,035	,081	,195	,335	,296	,631
Naaptol	,135	1,290	,198	,088	,586	,023	,972	,014	,985
Jabong	,065	,305	1,042	,035	,016	,305	,066	,467	,533
Snapdeal	,089	-,451	-,694	,031	,048	,187	,225	,322	,547
Localbania	,116	,600	-,734	,031	,109	,270	,509	,460	,968
Active Total	1,000			,254	1,000	1,000			

a. Symmetrical normalization

من العمود السادس نلاحظ أنّ شركة Amazone ساهمت في تشكيل المحور الأول بـ 0.052، أي بنسبة 5.2 بالمائة، وكانت مساهمتها في تشكيل المحور الثاني 0.7 بالمائة، مما يعني أنّ مساهمتها في تشكيل المحور الأول كانت أفضل. وأعلى إسهاما كان لشركة Naaptol في المحور الأول بنسبة 58 بالمائة. ثم شركة Jabong في المحور الثاني بنسبة 30 بالمائة. ثم شركة Localbania في المحور الثاني بنسبة 27 بالمائة.

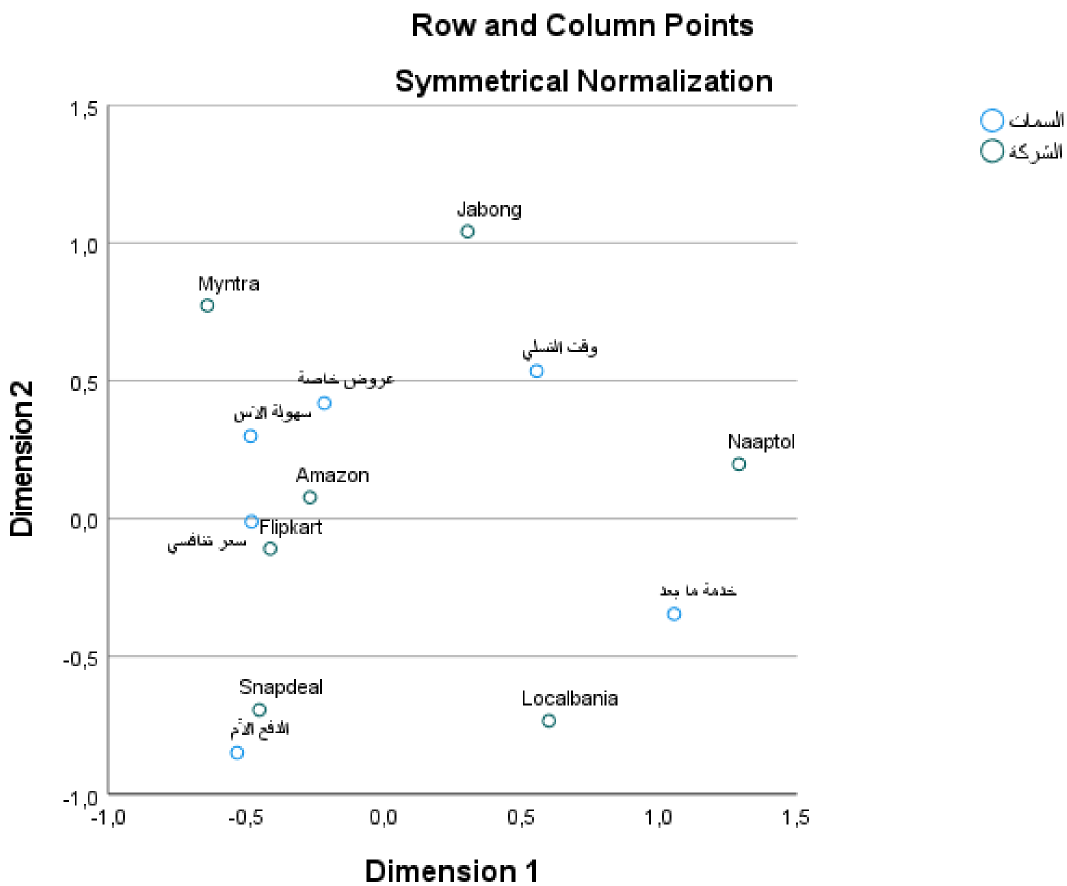
الجدول السادس: نسبة مساهمة الأعمدة في تشكيل المحاور.

Overview Column Points[a]

		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
السمات	Mass	1	2	Inertia	1	2	1	2	Total
سهولة الاستعمال	,165	-,483	,299	,029	,101	,064	,505	,117	,622
الدفع الآمن	,164	-,532	-,849	,048	,121	,512	,365	,563	,929
عروض خاصة	,211	-,215	,419	,025	,026	,161	,149	,342	,492
خدمة ما بعد البيع	,194	1,054	-,346	,092	,565	,101	,894	,058	,953
وقت التسليم	,131	,556	,535	,045	,106	,163	,344	,193	,537
سعر تنافسي	,136	-,480	-,011	,014	,082	,000	,828	,000	,829
Active Total	1,000			,254	1,000	1,000			

a. Symmetrical normalization

يوضّح الجدول نسبة مساهمة كل سمة في المحورين الأول والثاني. نلاحظ أنّ خدمة ما بعد البيع هي الأعلى مساهمة بنسبة 56 بالمائة في المحور الأول، ونسبة مساهمة الدفع الآمن في البعد الثاني كانت 51 بالمائة.



يتضح أنّ سمة خدمة ما بعد البيع مرتبطة بشركة Naaptol، وسمة الدفع الآمن موجودة أكثر لدى شركة Snapdeal.

الفصل الخامس:

التصنيف التسلسلي الهرمي

AHC

طريقة التصنيف التسلسلي

Classification Hiérarchique

هو أسلوب إحصائي يستخدم لتحديد كيف يمكن تجميع الوحدات المختلفة، مثل الأشخاص أو المجموعات أو المجتمعات معًا، بسبب الخصائص المشتركة بينها. ويُعرف أيضًا باسم التجميع، وهو عبارة عن أداة استكشافية لتحليل البيانات تهدف إلى تصنيف كائنات مختلفة إلى مجموعات بطريقة معينة، بحيث عندما يكونون منتمين إلى نفس المجموعة، يكون لديهم درجة قصوى من الارتباط وعندما لا ينتمون إلى نفس المجموعة، درجة الارتباط تكون في الحد الأدنى. ويمكن من خلاله تجميع البيانات في مجموعات عنقودية، وكل عنقود عبارة عن مجموعة من المشاهدات المتجانسة فيما بينها، بحيث تكون كل العناصر المكونة لكل عنقود متشابهة، وفي نفس الوقت مختلفة مع العناصر المكونة لعنقود آخر.

وغالبًا ما يستخدم التصنيف بالموازاة مع التحليلات الأخرى (مثل التحليل بالمركبات الاساسية). ويجب أن يكون الباحث قادرًا على تفسير التصنيف بناءً على فهمه للبيانات، لتحديد ما إذا كانت النتائج التي ينتجها التحليل ذات مغزى.

كيف تعمل؟

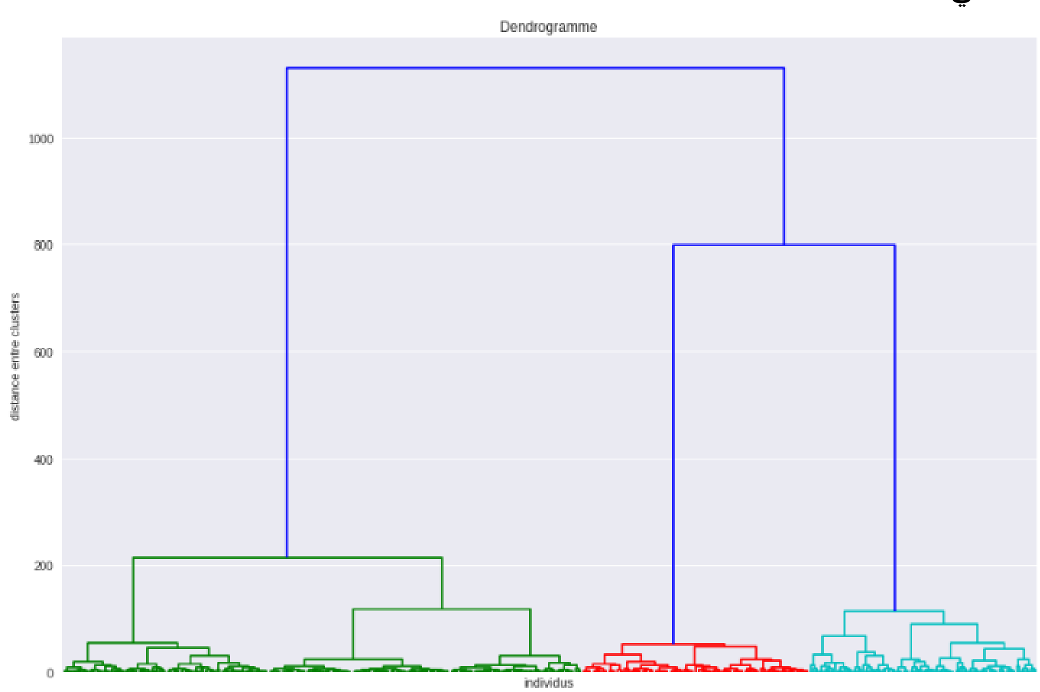
أنظر إلى هذه البيانات: كم عدد المجموعات التي تراها؟

ربما سيقول البعض ثلاثة مجموعات، والبعض الآخر 4 أو حتى 5. في الواقع، هناك 5 مجموعات مميزة، لكن اثنتين منها متقاربة جدًا، لذلك يمكننا اعتبار أنها تشكل مجموعة واحدة والإجابة هي 4. هاتان المجموعتان المتقاربتان جدًا هما أيضًا قريبتان نسبيًا من مجموعة ثالثة، إذا قمنا بتجميعها معًا، فإننا نعتبر أن

هناك 3 مجموعات. لو كانت إجابتك 3 فنظرتك عامة جدًا، بينما لو كانت 5، فذلك يعني أنّك تفضل الحصول على تحليل أدق وأكثر تعمقًا. وبالتالي، فإن مستوى التعمق هو الذي يحدد التسلسل الهرمي.

في التصنيف التسلسلي يتم تمثيل البيانات بواسطة مخطط شجري يدعى ديندوغرام (Dendrogramme)، ويعتمد التصنيف في كل مرحلة من التجزئة على جمع العناصر الأكثر اقترابا مثنى مثنى. وبالنسبة للمتغيرات يمكن تصنيفها من أجل تقليص عددها، إذ يمكن أن يكون عددها كبيرا ومتكررا، أي أنّ هناك متغيرات تفسر بنفس القدر، وبالتالي يسمح التصنيف باختبار المتغيرات الأكثر تمثيلا.

التصنيف التصاعدي والتنازلي: توجد طريقتين للتصنيف التسلسلي، التصاعدي والتنازلي، ففي التصاعدي (يسمى أيضًا التجميع التكتلي) نعتبر أولاً أن كل نقطة عبارة عن كتلة، لذلك هناك عدد من المجموعات يساوي عدد النقاط، ثم نبحث عن أقرب مجموعتين، ونجمعهم في كتلة واحدة، وتتكرر هذه الخطوة حتى يتم تجميع جميع النقاط في مجموعة كبيرة واحدة؛ وفي التنازلي (يسمى أيضًا التكتل الانقسامي) نبدأ بمجموعة كبيرة تحتوي على جميع النقاط، ثم نقسمها على التوالي حتى نحصل على أكبر عدد ممكن من المجموعات كما هو الحال في النقاط.



والمفتاح في عملية التصنيف هو المسافة بين المجموعات على المحور ص، فكلما زاد حجمها، زاد عدد التجمعات. يوجد في الجزء العلوي من هذا المحور مجموعة واحدة فقط ، وفي أسفل هذا المحور يوجد عدد من العناقيد يساوي عدد الأفراد. فإذا أردنا الآن إنشاء قسم به مجموعتان، علينا قطع الشجرة على ارتفاع 1000. إذا أردنا 3 مجموعات، فإننا نقطع عند 600، 4 مجموعات عند 200، إلخ.

ملاحظة: إذا كان لدينا العديد من النقاط، فقد يكون الجزء السفلي من مخطط الديندوقرام غير قابل للقراءة، وهذا هو سبب تمثيل الجزء العلوي من الشجرة فقط (في بعض الحالات).

المسافة بين الأفراد في نفس المجتمع: لتجميع الأفراد المتشابهين، لا بد من معيار للتجميع، وللقيام بذلك علينا فحص جميع المعلومات المتعلقة بالأفراد، مثلا في حالة المرضى، درجة الحرارة، ضغط الدم،...إلخ (x_i, y_i .....)، ولدينا مجموعة من الأفراد، ونفترض أنّ كل فرد يمثل نقطة في الفضاء:

M_i=$(x_i, y_i, z_i, \ldots.)$

فإذا كان هناك متغيرين فقط (x_i, y_i)، نحصل على سحابة من النقط في المستوى:

NP= $\{M_I, i = 1 \ldots \ldots . . n\}$

المسافة الأقليدية بين فردين (M_i, M_j) في هذا الفضاء تعطى بالعلاقة التالية:

$$d^2(M_i, M_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

هذه المسافة تكون صغيرة إذا كان الفردين متشابهين، وتكون كبيرة إذا كانا مختلفين. ويمكننا ربط كل سحابة من الأفراد بمصفوفة نسميها مصفوفة المسافات:

$$D = (d_{ij})/\ 0 \le i \le n, 0 \le j \ \le \ n = \left((d^2)M_i, M_j\right).$$

وهي مصفوفة بعدد n صفوف و n أعمدة، معاملاتها موجبة ومتماثلة وصفرية القطر ، لأنّ:

$d^2(M_i, M_j)$=$d^2(M_j, M_i)$..................$d^2(M_i, M_i)$=0.

لذلك من أجل سحابة بعدد n من الأفراد، يوجد $\frac{n(n-1)}{2}$ مسافة يتم حسابها. وإلى جانب المسافة الإقليدية، يمكننا تحديد مسافات أخرى (وبالتالي مصفوفات أخرى للمسافات)، مثلا:

$d^1(M_i, M_j)$= $|x_i - x_j| + |y_i - y_j|$

$d^\infty(M_i, M_j)$= $Max\ \{|x_i - x_j|, |y_i - y_j|\}$

المسافة الإقليدية بين نقطتين تحسب من العلاقة:

$$d(i, i') = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2}$$

المسافة الإقليدية مربّعة، تحسب من العلاقة:

$$d^2(i, i') = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2$$

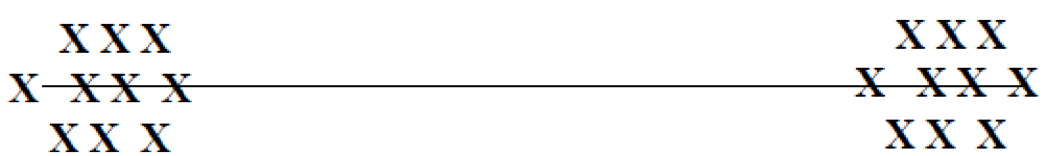
الاختلافات بين الطبقات (Ecarts entre classes): نفترض أنّ السحابة NP تتكون من عدة مجموعات أو طبقات ($NP_1, NP_2, \ldots\ldots, NP_k$)، ونسمي مركز الثقل لكل مجموعة G ($G_1, G_2, \ldots\ldots, G_k$). ووزن كل

مجموعة P $(P_{1,}\ P_2,\ldots\ldots,\ P_k)$. فإذا كان لكل الأفراد نفس الوزن $(\frac{1}{n})$، فإن وزن المجموعة NP_J يساوي مجموع أفرادها مقسوما على n. ويكون مجموع الوزن لجميع المجموعات يساوي 1. ولحساب التقارب والتباعد بين طبقتين، توجد العديد من الطرائق، من أهمها:

1. مؤشر التجميع لأدنى مسافة (ربط بسيط): المسافة بين مجموعتين هي المسافة بين أقرب نقطتين. هذا يعادل القول بأنّ مجموعتان متقاربتان إذا كانت اثنتان على الأقل من نقاطهما متقاربة. ويتم فيها تجميع عناصر كل زوج في مجموعة جديدة، وتكرر العملية إلى غاية تجميع كل الأصناف في مجموعة واحدة.

$$\mathbf{(H_1\ ,H_2)=\ Min\ d^2(X_i\ ,X_j)}$$

2. **مؤشر التجميع لأقصى مسافة (الربط الكامل):** تعتبر المسافة بين مجموعتين هي المسافة بين أبعد نقطتين، وهذا يعادل بالتالي قولنا "مجموعتان متقاربتان إذا كانت جميع نقاطهما متقاربة".



$$(H_1,\ H_2)\ =\ Max\ d^2(X_i,\ X_j)$$

3. مؤشر التجميع للمسافة المتوسطة (متوسط الارتباط): تعتبر المسافة بين مجموعتين هي متوسط جميع المسافات بين نقاط مجموعة واحدة ونقاط المجموعة الأخرى. لحسابها، نقوم بتعداد جميع أزواج النقاط الممكنة من مجموعة إلى أخرى، ثم نحسب مسافة كل زوج، ثم نحسب المتوسط.

4. الرابط المركزي (): المسافة بين مجموعتين تعتبر المسافة بين النقطتين الوسطى: تجعل طرق الارتباط من الممكن ضمان فصل المجموعات جيدًا، لأن المسافة بين المجموعات تؤخذ في الاعتبار. ومع ذلك، فهم لا يضمنون أن التجمعات ضيقة على نفسها (هذا هو مفهوم القصور الذاتي داخل الطبقة). لحل هذه

المشكلة، هناك طريقة وارد التي، في كل تكرار، أي في كل مرة يتم فيها تجميع مجموعتين في 1، تسعى إلى تقليل الزيادة في القصور الذاتي داخل الطبقة بسبب تجميع المجموعتين.

اقترح وارد هذه الطريقة سنة 1963 وتسمح من الحصول على مجموعات بطريقة تصغر من فقدان المعلومات الناتجة من كل تجميع حيث أنها تعتمد على تحليل التباين، حيث يتمثل في تدنئة أو تصغير مجموع المربعات لكل الأزواج الممكن تشكيلها في كل مرحلة ويتم تصنيف المجموعات عن طريق وارد كما يلي:

– تحسب مصفوفة التجميع من العلاقة:

$$d(NP_m, NP_l)=\frac{p_m p_l}{p_m+p_l} d^2(G_m, G_l)$$

– يتم تجميع المجموعتين A وB مع العنصر C وفقا للعلاقة:

$$\delta((A, B) :C)=\frac{(n_A+n_C)\delta(A,C)+(n_B+n_C)\delta(B,C)-n_C\delta(A,B)}{n_A+n_B+n_C}$$

القصور الذاتي داخل الطبقات والقصور الذاتي بين الطبقات: نسمي القصور الذاتي الكلي (التباين الكلي) لسحابة النقاط NP المجموع المرجح لمربعات مسافات نقاطها من مركز ثقل السحابة. لذلك، أذا كانت G نقطة مركز ثقل سحابة النقاط، ولجميع النقاط نفس الوزن، فإنّ القصور الذاتي الكلي لسحابة النقاط:

$$I(NP) = \frac{1}{n}(d^2(M_1, G)^2 + d^2(M_2, G)^2) + \cdots + d^2(M_n, G)^2$$

وإذا كانت سحابة النقط تتكون من عدة طبقات، تفكك مثنى مثنى، لتكون أكثر تجانسا من التباين الكلي لكل طبقة. مجموع التباينات لكل الطبقات يسمى التباين داخل الطبقات.

$$I_{intra}=I(NP_1) + I(NP_2) + \cdots + I(NP_k)$$

إجمالي القصور الذاتي للسحابة بشكل عام لا يساوي مجموع القصور الذاتي للفئات التي تتكون منها، أي القصور الذاتي داخل الطبقة (باستثناء الحالة التي تكون فيها مراكز الثقل من بين جميع الفئات مشوشة) لأنه من الضروري أيضًا مراعاة تشتت الفئات فيما يتعلق بمركز ثقل السحابة. ويحدد القصور الذاتي بين الطبقات بواسطة:

$$I_{inter}=\bar{p}_1 d^2(G_1, G)^2+\bar{p}_2 d^2(G_2, G)^2+\ldots+\bar{p}_k d^2(G_k, G)^2$$

حيث: $\bar{p}$ الوزن الكلي للطبقة NP_j.

ملاحظة: الجمود الكلي لسحابة من النقاط مكونة من فئات مختلفة مفككة اثنين اثنين هو مجموع القصور الذاتي داخل الطبقات والقصور الذاتي بين الطبقات، أي:

$$I(t) = I_{intra} + I_{inter}$$

تحسب المسافة الاقليدية مربّعة بين نقطتين من العلاقة:

$$d^2(i,i') = \sum_{j=1}^{p}(x_{ij} - x_{i'j})^2$$

وفي حالة كانت المتغيرات غير متجانسة ، علينا مركزة البيانات وترجيحها.

التمرين رقم 1: لدينا الجدول التالي:

	X	Y
M1	2	0
M2	0	1
M3	0	2
M4	3	4
M5	5	4

1. أحسب المسافات بين الأفراد (مصفوفة القرابة) مستخدما المسافة الإقليدية مربّعة.

2. صنّف الأفراد باستخدام مؤشر التجميع لأدنى مسافة، وارسم الديندوغرام.

3. صنّف الأفراد باستخدام مؤشر التجميع لأقصى مسافة، وارسم الديندوغرام.

4. تصنيف الأفراد وفقا لمؤشر المسافة المتوسطة، وارسم الديندوغرام.

5. صنّف الأفراد وفقا لمؤشر وارد (Ward)، وارسم الديندوغرام.

الحل:

1. جدول المسافات.

$$d^2(i,i') = \sum_{j=1}^{p}(x_{ij} - x_{i'j})^2$$

$d^2(M1, M2) = (2-0)^2 + (0-1)^2 = 5$

وبعد اتمام جميع الحسابات يكون جدول القرابة على النحو التالي:

	M1	**M2**	**M3**	**M4**	**M5**
M1	0	5,000	8,000	17,000	25,000
M2	5,000	0	1,000	18,000	34,000
M3	8,000	1,000	0	13,000	29,000
M4	17,000	18,000	13,000	0	4,000
M5	25,000	34,000	29,000	4,000	0

2. تصنيف الأفراد باستخدام مؤشر التجميع لأدنى مسافة.

ننظر في أدنى مسافة في جدول المسافات، وهي بين M2 و M3 (1). نقوم بوضع M2 و M3 في مجموعة واحدة، نسميها A مثلا. ثم في الخطوة الثانية نحسب المسافة بين المجموعة الجديدة A وبقية الأفراد، كما يلي:

δ(A, M1)= $Min[d^2(\mathrm{M2,M1});\ d^2(\mathrm{M3,M1})] = d^2(\mathrm{M2,M1})$ =5

لحساب المسافة بين A و M1، نحسب المسافة بين M2 و M1، وهي 5، ثم بين M3 و M1، 8، نأخذ أدنى مسافة، وهي بين M2 و M1.

δ(A, M4)= $Min[d^2(\mathrm{M2,M4});\ d^2(\mathrm{M3,M4})] = d^2(\mathrm{M3,M4})$ =13

δ(A, M5)= $Min[d^2(\mathrm{M2,M5});\ d^2(\mathrm{M3,M5})] = d^2(\mathrm{M3,M5})$ =29

نحصل على المصفوفة الموالية:

	M1	M4	M5	A
M1	0	17,000	25,000	5
M4	17,000	0	4,000	13
M5	25,000	4,000	0	29
A	5	13	29	0

أدنى مسافة في هذه المصفوفة بين M5 و M4، نكرر العملية، ونقوم بتجميع M4 مع M5 في مجموعة جديدة، نسميها B. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

δ(B, M1)= $Min[d^2(\mathrm{M4,M1}\ );\ d^2(\mathrm{M5,M1})] = d^2(\mathrm{M4,M1})$ =17

δ(B, A)= $Min[d^2(\mathrm{M4,A});\ d^2(\mathrm{M5,A})] = d^2(\mathrm{M4,A})$ =13

وتصبح المصفوفة كما يلي:

	M1	A	B
M1	0	5	17
A	5	0	13
B	17	13	0

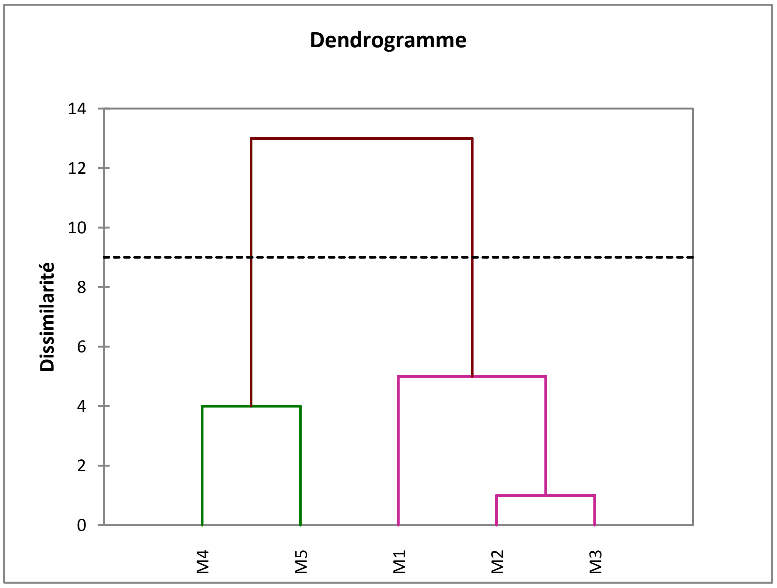
أدنى مسافة في هذه المصفوفة بين A و M1، نكرر العملية، ونقوم بتجميع A مع M1 في مجموعة جديدة، نسميها C. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

δ(C, B)= $Min[d^2(\mathrm{M1,B}\ );\ d^2(\mathrm{A,B})] = d^2(\mathrm{A,B})$ =13

وتصبح المصفوفة كما يلي:

	B	C
B	0	13
C	13	0

ويكون شكل الديندوغرام على النحو التالي:



3. تصنيف الأفراد وفقا لأقصى مسافة، ورسم الديندوغرام.

	M1	M2	M3	M4	M5
M1	0	5,000	8,000	17,000	25,000
M2	5,000	0	1,000	18,000	34,000
M3	8,000	1,000	0	13,000	29,000
M4	17,000	18,000	13,000	0	4,000
M5	25,000	34,000	29,000	4,000	0

ننظر في أدنى مسافة في جدول المسافات، وهي بين M2 و M3 (1). نقوم بوضع M2 و M3 في مجموعة واحدة، نسميها A مثلا. ثم في الخطوة الثانية نحسب المسافة بين المجموعة الجديدة A وبقية الأفراد، كما يلي:

δ(A, M1)= $Max[d^2(\text{M2}, \text{M1});\ d^2(\text{M3}, \text{M1})] = d^2(\text{M3}, \text{M1})$ =8

لحساب المسافة بين A و M1، نحسب المسافة بين M2 و M1، وهي 5، ثم بين M3 و M1، 8، نأخذ أقصى مسافة، وهي بين M3 و M1.

δ(A, M4)= $Max[d^2(\text{M2}, \text{M4});\ d^2(\text{M3}, \text{M4})] = d^2(\text{M3}, \text{M4})$ =18

δ(A, M5)= $Max[d^2(\text{M2}, \text{M5});\ d^2(\text{M3}, \text{M5})] = d^2(\text{M3}, \text{M5})$ =34

نحصل على المصفوفة الموالية:

	M1	M4	M5	A
M1	0	17,000	25,000	8
M4	17,000	0	4,000	18
M5	25,000	4,000	0	34
A	8	18	34	0

أدنى مسافة في هذه المصفوفة بين M5 و M4، نكرر العملية، ونقوم بتجميع M4 مع M5 في مجموعة جديدة، نسميها B. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$\delta(B, M1) = Max[d^2(M4, M1); d^2(M5, M1)] = d^2(M4, M1) = 25$

$\delta(B, A) = Max[d^2(M4, A); d^2(M5, A)] = d^2(M4, A) = 18$

وتصبح المصفوفة كما يلي:

	M1	A	B
M1	0	8	25
A	8	0	18
B	25	18	0

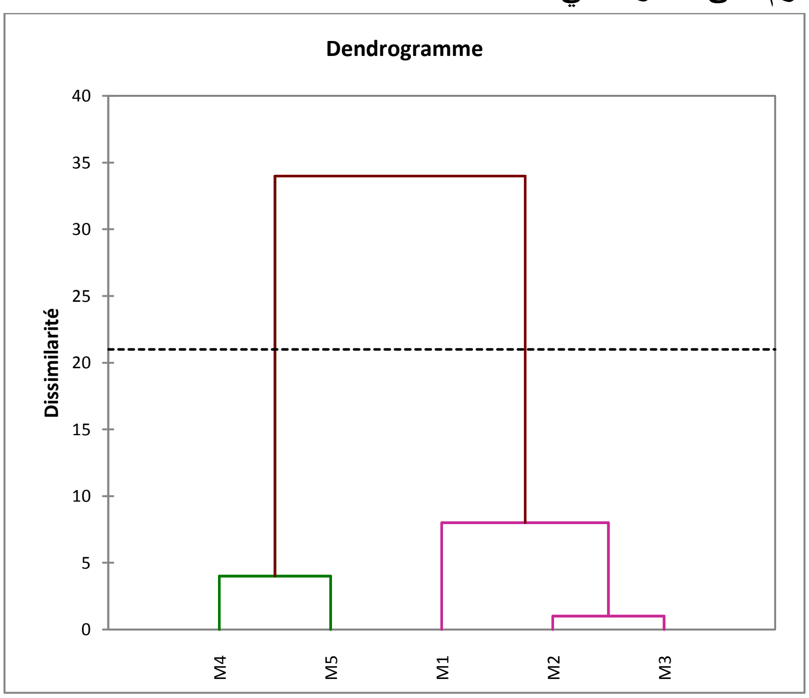
أدنى مسافة في هذه المصفوفة بين A و M1، نكرر العملية، ونقوم بتجميع A مع M1 في مجموعة جديدة، نسميها C. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$\delta(C, B) = Min[d^2(M1, B); d^2(A, B)] = d^2(A, B) = 13$

وتصبح المصفوفة كما يلي:

	B	C
B	0	25
C	25	0

ويكون شكل الدينوغرام على النحو التالي:



لقد اختلف شكل الديندوغرام، فقد تم تجميع M2 مع M3 لتشكيل المجموعة A ثم تجميع M4 و M5 لتشكيل المجموعة B، وتكونت المجموعة C من تجميع M1 مع المجموعة A، وتجميع المجموعة C مع المجموعة B يشكل المجموعة الكلية.

4. تصنيف الأفراد وفقا لمؤشر المسافة المتوسطة.

	M1	M2	M3	M4	M5
M1	0	5,000	8,000	17,000	25,000
M2	5,000	0	1,000	18,000	34,000
M3	8,000	1,000	0	13,000	29,000
M4	17,000	18,000	13,000	0	4,000
M5	25,000	34,000	29,000	4,000	0

ننظر في أدنى مسافة في جدول المسافات، وهي بين M2 و M3 (1). نقوم بوضع M2 و M3 في مجموعة واحدة، نسميها A مثلا. ثم في الخطوة الثانية نحسب المسافة بين المجموعة الجديدة A وبقية الأفراد، كما يلي:

δ(A, M1)= $Moy[d^2(\text{M2},\text{M1}\ );\ d^2(\text{M3},\text{M1})] = Moy[5;8\ ]$ =6.5

δ(A, M4)= $Moy[d^2(\text{M2},\text{M4}\ );\ d^2(\text{M3},\text{M4})] = Moy[18;13\ ]$ =15.5

δ(A, M5)= $Moy[d^2(\text{M2},\text{M5}\ );\ d^2(\text{M3},\text{M5})] = Moy[34;29\ ]$ =31.5

وتصبح المصفوفة كما يلي:

	M1	M4	M5	A
M1	0	17,000	25,000	6.5
M4	17,000	0	4,000	15.5
M5	25,000	4,000	0	31.5
A	6.5	15.5	31.5	0

أدنى مسافة في هذه المصفوفة بين M5 و M4، نكرر العملية، ونقوم بتجميع M4 مع M5 في مجموعة جديدة، نسميها B. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

δ(B, M1)= $Moy[d^2(\text{M4},\text{M1}\ );\ d^2(\text{M5},\text{M1})] = Moy[17;25\ ]$ =21

δ(B, A)= $Moy[d^2(\text{M4},\text{A});\ d^2(\text{M5},\text{A})] = Moy[15.5;31.5\ ]$ =23.5

وتصبح المصفوفة كما يلي:

	M1	A	B
M1	0	6.5	21
A	6.5	0	23.5
B	21	23.5	0

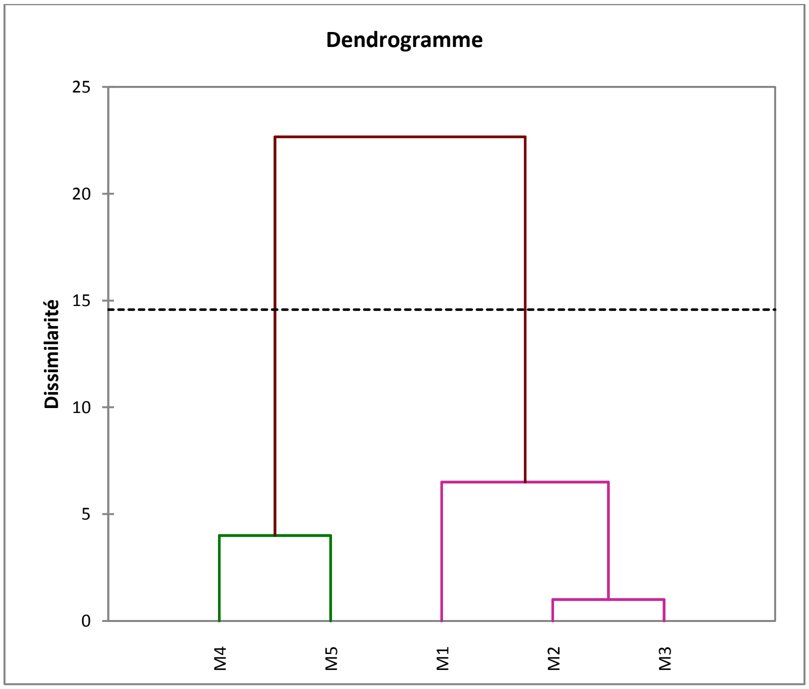
أدنى مسافة في هذه المصفوفة بين A و M1، نكرر العملية، ونقوم بتجميع A مع M1 في مجموعة جديدة، نسميها C. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$$\delta(C, B) = Moy[d^2(M1, B);\ d^2(A, B)] = Moy[21; 23.5] = 22.25$$

وتصبح المصفوفة كما يلي:

	B	C
B	0	22.25
C	22.25	0

ويكون شكل الديندوغرام على النحو التالي:



5. تصنيف الأفراد وفقا لمؤشر التجميع لوارد.

	M1	M2	M3	M4	M5
M1	0	5,000	8,000	17,000	25,000
M2	5,000	0	1,000	18,000	34,000
M3	8,000	1,000	0	13,000	29,000
M4	17,000	18,000	13,000	0	4,000
M5	25,000	34,000	29,000	4,000	0

نحسب مصفوفة التجميع وفقا للعلاقة:

$$d(NP_m, NP_l) = \frac{p_m p_l}{p_m + p_l} d^2(G_m, G_l)$$

$d\,(M1, M2) = \frac{1*1}{1+1} d^2(5) = 2.5$, $d(M1, M3) = \frac{1*1}{1+1} d^2(8) = 4$

وهكذا حتى نحصل على المصفوفة الموالية:

	M1	M2	M3	M4	M5
M1	0	2.5	4	8.5	12.5
M2	2.5	0	0.5	9	17
M3	4	0.5	0	6.5	14.5
M4	8.5	9	6.5	0	2
M5	12.5	17	14.5	2	0

ننظر في أدنى مسافة في الجدول أعلاه، وهي بين M2 و M3 (0.5). نقوم بوضع M2 و M3 في مجموعة واحدة، نسميها A مثلا. ثم في الخطوة الثانية نحسب المسافة بين المجموعة الجديدة A وبقية الأفراد، كما يلي:

$$\delta((A, B) :C)=\frac{(n_A+n_C)\delta(A,C)+(n_B+n_C)\delta(B,C)-n_C\delta(A,B)}{n_A+n_B+n_C}$$

$$\delta(A(M2, M3) :M1)=\frac{(n_A+n_C)\delta(M2,M1)+(n_B+n_C)\delta(M3,M1)-n_C\delta(M2,M3)}{n_A+n_B+n_C}$$

$$= \frac{(1+1)(2.5)+(1+1)(4)-1(0.5)}{1+1+1}= 4.166$$

$$\delta(A(M2, M3) :M4)= \frac{(1+1)(9)+(1+1)(6.5)-1(0.5)}{1+1+1}=10.166$$

$$\delta(A(M2, M3) :M5)= \frac{(1+1)(17)+(1+1)(14.5)-1(0.5)}{1+1+1}=20.833$$

وتصبح المصفوفة كما يلي:

	M1	A	M4	M5
M1	0	4.17	8.5	12.5
A	4.17	0	10.17	20.833
M4	8.5	10.17	0	2
M5	12.5	20.833	2	0

أدنى مسافة في هذه المصفوفة بين M5 و M4، نكرر العملية، ونقوم بتجميع M4 مع M5 في مجموعة جديدة، نسميها B. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$$\delta(B(M4, M5) :M1)= \frac{(1+1)(8.5)+(1+1)(12.5)-1(2)}{1+1+1}=13.333$$

$$\delta(B(M4, M5) :A)= \frac{(1+1)(10.17)+(1+1)(20.833)-1(2)}{1+1+1}=20$$

وتصبح المصفوفة كما يلي:

	M1	A	B
M1	0	4.17	13.333
A	4.17	0	20
B	13.333	20	0

أدنى مسافة في هذه المصفوفة بين A و M1، نكرر العملية، ونقوم بتجميع A مع M1 في مجموعة جديدة، نسميها C. ونحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$\delta(C(M1, A) : B) = \frac{(1+1)(13.333)+(1+1)(20)-1(4.17)}{1+1+1} = 20.83$

وتصبح المصفوفة كما يلي:

	B	C
B	0	20.83
C	20.83	0

التمرين الأول: لديك الجدول الموالي.

	X	Y
a	2	10
b	2	5
c	8	4
d	5	8
e	7	5
f	6	4
g	1	2
h	4	9

1. أحسب مصفوفة المسافات بين الأفراد باستخدام المسافة الإقليدية.
2. استخدم طريقة أدنى مسافة في رسم الدينودوغرام.

الحل:

1. حساب مصفوفة المسافات.

$d(i, i') = \sqrt{}$

$d(a, b) = \sqrt{(2-2)^2 + (10-5)^2} = 5$

	a	b	c	d	e	f	g	h
a	0	5,000	8,485	3,606	7,071	7,211	8,062	2,236
b	5,000	0	6,083	4,243	5,000	4,123	3,162	4,472
c	8,485	6,083	0	5,000	1,414	2,000	7,280	6,403
d	3,606	4,243	5,000	0	3,606	4,123	7,211	1,414
e	7,071	5,000	1,414	3,606	0	1,414	6,708	5,000
f	7,211	4,123	2,000	4,123	1,414	0	5,385	5,385
g	8,062	3,162	7,280	7,211	6,708	5,385	0	7,616
h	2,236	4,472	6,403	1,414	5,000	5,385	7,616	0

2. استخدم طريقة أدنى مسافة في رسم الدينودوغرام.

أدنى مسافة في الجدول السابق بين (d,h)، (c,e)، (f,e).

$d(c, e) = d(f, e) = d(d, h) = 1.414$

سنقوم بتجميع (c,e,f) في مجموعة واحدة نسميها G_1. وتجميع (d,h) في مجموعة نسميها G_2 ثم نحسب المسافة بين المجموعتين الجديدتين وبقية الأفراد.

$\delta(G_1, a) = Min[d(c,a) ; d(f,a) ; d(e,a)] = d(e,a) = 7.071$

$\delta(G_1, b) = Min[d(c,b) ; d(f,b) ; d(e,b)] = d(f,b) = 4.123$

$\delta(G_1, g) = Min[d(c,g) ; d(f,g) ; d(e,g)] = d(f,b) = 5.385$

$\delta(G_2, a) = Min[d(h,a) ; d(d,a)] = d(h,a) = 2.236$

$\delta(G_2, b) = Min[d(h,b) ; d(d,b)] = d(h,a) = 4.243$

$\delta(G_2, g) = Min[d(h,g) ; d(d,g)] = d(h,a) = 7.211$

$\delta(G_2, G_1) = Min[d(h,G1) ; d(d,G1)] = d(d,G1) = 3.606$

ويصبح الجدول كما يلي:

	a	b	G_1	G_2	g
a	0	5	7.071	2.236	8.062
b	5	0	4.123	4.243	3.162
G_1	7.071	4.123	0	3.606	5.385
G_2	2.236	4.243	3.606	0	7.211
g	8.062	3.162	5.385	7.211	0

أدنى مسافة في هذا الجدول بين G_2 و a، نقوم بتجميع (G_2, a) في مجموعة واحدة نسميها G_3. ثم نحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$\delta(G_3, b) = Min[d(G2,b) ; d(a,b)] = d(G2,b) = 4.243$

$\delta(G_3, G1) = Min[d(G2,G1) ; d(a,G1)] = d(G2,b) = 3.606$

$\delta(G_3, g) = Min[d(G2,g) ; d(a,g)] = d(G2,g) = 7.211$

ويصبح الجدول كما يلي:

	b	G_1	g	G_3
b	0	4.123	3.162	4.243
G_1	4.123	0	5.385	3.606
g	3.162	5.385	0	7.211
G_3	4.243	3.606	7.211	0

أدنى مسافة في هذا الجدول بين g و b، نقوم بتجميع (g, b) في مجموعة واحدة نسميها G_4. ثم نحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$\delta(G_4, G_1) = Min[d(b,G1) ; d(g,G1)] = d(b,G1) = 4.123$

$\delta(G_4, G_3) = Min[d(b,G3) ; d(g,G3)] = d(b,G3) = 4.243$

ويصبح الجدول كما يلي:

	G_1	G_4	G_3
G_1	0	4.123	3.606
G_4	4.123	0	4.243
G_3	3.606	4.243	0

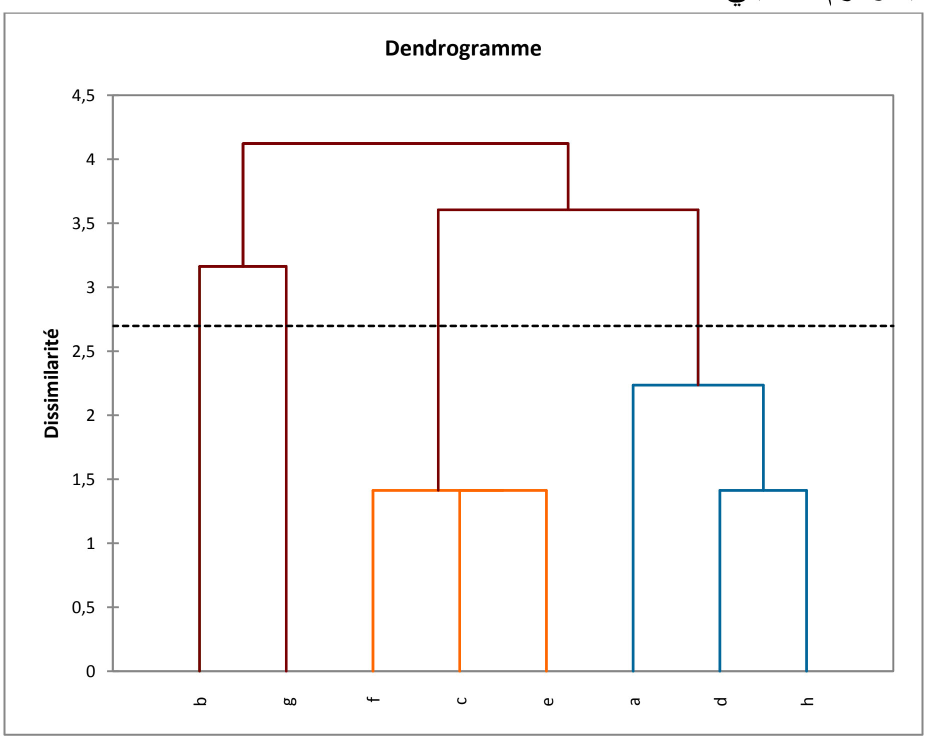
أدنى مسافة في هذا الجدول بين G_1 و G_3، نقوم بتجميعهما في مجموعة واحدة نسميها G_5. ثم نحسب المسافة بين المجموعة الجديدة وبقية الأفراد.

$$\delta(G_5, G_4) = Min[d(G1, G4) ; d(G3, G4)] = d(G1, G4) = 4.123$$

ويصبح الجدول كما يلي:

	G_5	G_4
G_5	0	4.123
G_4	4.123	0

ويكون شكل الديندوغرام كما يلي:

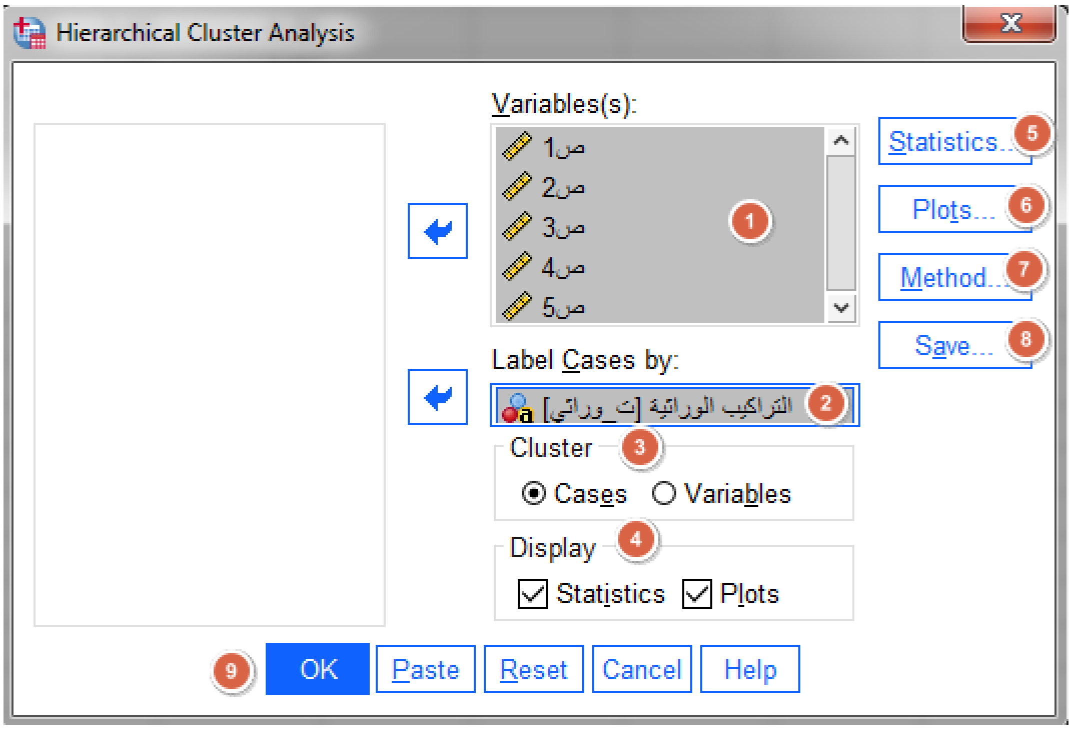


التطبيق باستخدام برنامج SPSS

مثال: لدينا بعض التراكيب الوراثية والصفات المقاسة لمجموعة من الأفراد.

ت_وراثي	ص1	ص2	ص3	ص4	ص5
muna	68,00	80,00	33,20	4,05	3,95
azol	64,00	83,25	40,00	4,65	4,50
jumbo	66,00	87,00	37,30	3,60	4,30
azegar	66,00	83,25	28,50	3,65	2,85
cameltooh	72,00	83,25	37,50	3,90	2,70

Analyze>Classify>Hierarchical Cluster…



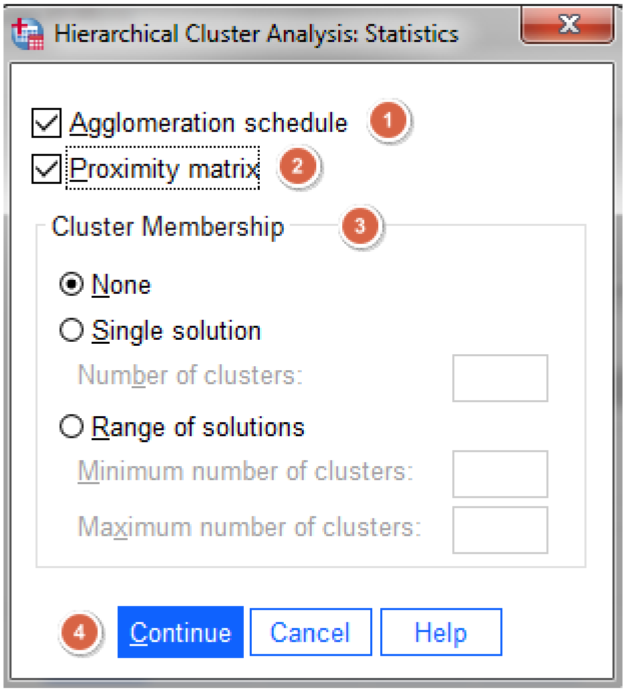
1- ننقل المتغيرات (الصفات الوراثية) إلى مربّع Variables.

2- ننقل المتغير الاسمي، التراكيب الوراثية إلى خانة Label Cases By.

3- نختار كيفية إنشاء العناقيد، بواسطة المشاهدات (Cases) أو المتغيرات (Variables).

4- نحدد العرض (Display)، إحصاءات ورسوم بيانية (Statistics, Plots)، أو إحصاءات فقط أو رسوم فقط.

5- نضغط Statistics:

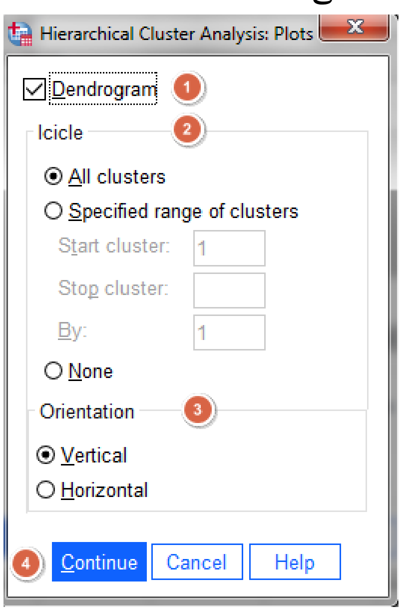


1.5. نحدد هذا الخيار لعرض جدول الطبقات (Agglomeration schedule).

2.5. نحدد هذا الخيار لحساب مصفوفة القرابة (جدول المسافات) (Proximity matrix).

3.5. في خانة أعضاء العناقيد (Cluster Membership) يمكنك اختيار الحل الوحيد (Single solution) وتحدد فيه عدد العناقيد، أو تختار مدى (Range of solutions) تحدد فيه الحد الأدنى والأعلى لعدد العناقيد.

4.5. نضغط على Continue، للعودة إلى صندوق الحوار الرئيسي ونضغط على Plots:



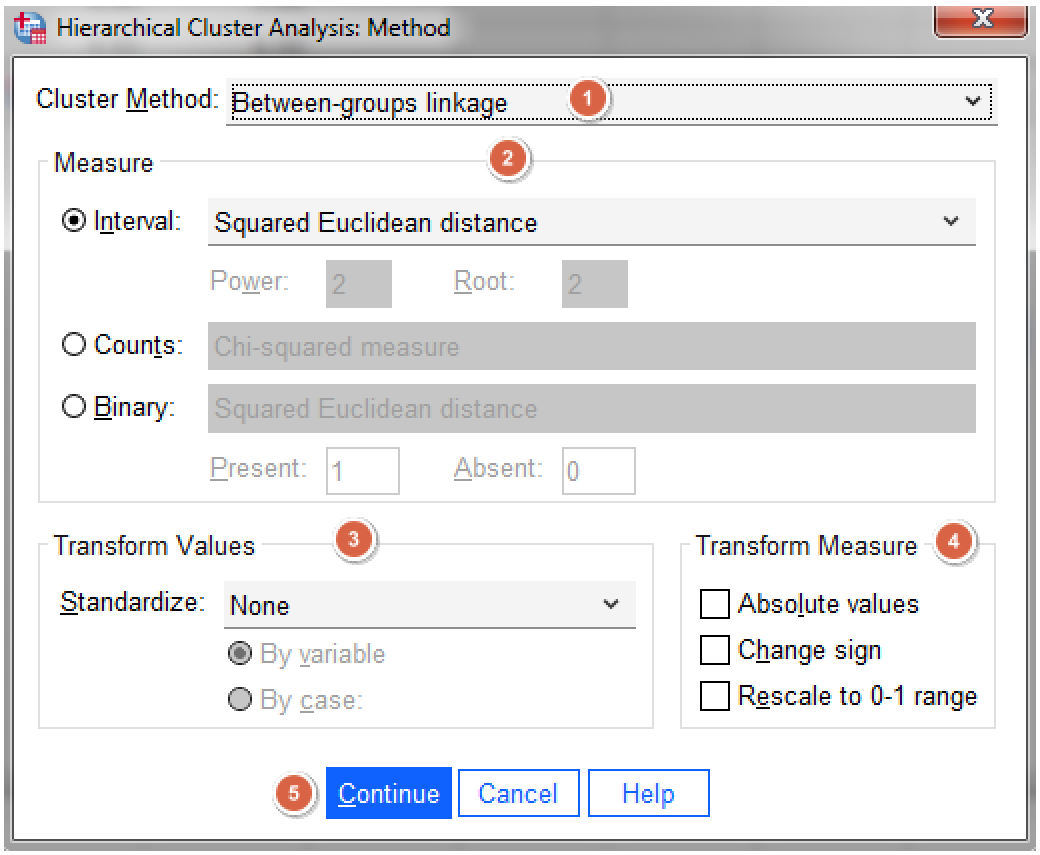
1.6. نختار رسم الشجرة (الديندوغرام)(Dendrogram).

2.6. تحت Icicle (الالواح الجليدية) نختار رسم الألواح الجليدية لجميع الطبقات او نحدد نطاق معين. كما تعرض مخططات Icicle معلومات حول كيفية دمج الحالات في مجموعات عند كل تكرار للتحليل.

3.6. يتيح لك تحديد اتجاه المخطط، رأسي أو أفقي.

4.6. ثم Continue.

7. نضغط Method (الطريقة):



1.7. من Cluster Method نختار طريقة التجميع، وتوجد عدة طرق:

– الارتباط بين المجموعات (between-groups linkage)، الارتباط داخل المجموعات(within-groups linkage)، أدنى مسافة (nearest neighbor)، أقصى مسافة (furthest neighbor)، المسافة المتوسطة (centroid clustering)، المسافة الوسيطة (median clustering)، طريقة وارد (Ward's method).

2.7. من Measure نختار مقياس المسافة الإقليدية مربّعة لحساب مصفوفة القرابة بين المتغيرات. توجد العديد من المقاييس، بحسب نوع قياس المتغيرات:

– المسافة (Interval): المسافة الاقليدية (Euclidean distance)، المسافة الإقليدية مربّعة (squared Euclidean distance)، جيب التمام (cosine)، ارتباط بيرسون (Pearson correlation)، Chebychev، block، Minkowski، المخصصة (customized).

– العد (Counts): المسافة الإقليدية مربّعة، ومقياس فاي مربّع.

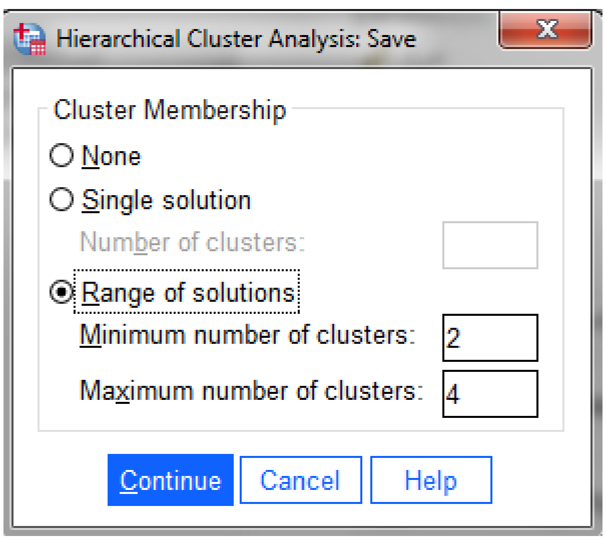
– الثنائية (Binary): توجد العديد من المقاييس.

3.7. تحويل القيم (Transform Values): يمكنك تحويل القيم بتوحيد قيم البيانات لأي من الحالات أو القيم قبل حساب التقريب (غير متاح للبيانات الثنائية). طرق التوحيد المتاحة هي درجات z والمدى من 1 إلى 1، والمدى من 0 إلى 1، والحد الأقصى للحجم 1، والمتوسط 1، والانحراف المعياري 1.

4.7. تحويل القياس (Transform Measures): يسمح لك بتحويل القيم الناتجة عن قياس المسافة. يتم تطبيقها بعد حساب قياس المسافة. البدائل المتاحة هي القيمة المطلقة وعلامة التغيّر وإعادة التدوير إلى النطاق 0–1.

5.7. ثم نضغط Continue.

8. نضغط Save:



8– يسمح لنا بتحديد عدد العناقيد (عضويات المجموعات) لحل واحد أو مجموعة من الحلول.يمكنك تحديد عدد العناقيد مباشرة، أو اختيار مدى مثلا بين 2 و4. ويمكن بعد ذلك استخدام المتغيرات المحفوظة في التحليلات اللاحقة لاستكشاف الاختلافات الأخرى بين المجموعات. ثم Continue.

9– نضغط ok فتظهر النتائج الموالية:

الجدول الأول: ملخص المشاهدات.

Case Processing Summary[a]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
5	100,0	0	,0	5	100,0

a. Ward Linkage

خاص بعدد المشاهدات والقيم المفقودة، وطريقة التجميع (Ward Linkage).

الجدول الثاني: مصفوفة القرابة بين المتغيرات.

Proximity Matrix

	Squared Euclidean Distance				
Case	1:muna	2:azol	3:jumbo	4:azegar	5:cameltooh
1:muna	,000	73,465	70,135	38,023	46,637
2:azol	73,465	,000	26,495	139,973	74,053
3:jumbo	70,135	26,495	,000	93,607	52,753
4:azegar	38,023	139,973	93,607	,000	117,085
5:cameltooh	46,637	74,053	52,753	117,085	,000

This is a dissimilarity matrix

يوضح الجدول مصفوفة المسافة بين التركيبات الوراثية، حيث يظهر أن أدنى مسافة بين التركيب الوراثي azol والتركيب الوراثي jumbo3.

الجدول الثالث: طريقة التجميع.

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	3	13,248	0	0	3
2	1	4	32,259	0	0	4
3	2	5	70,111	1	0	4
4	1	2	146,445	2	3	0

يظهر الجدول أعلاه طريقة التجميع في العناقيد والخطوات التي تم بها التحليل الهرمي.

الجدول الرابع: أعضاء العناقيد.

Cluster Membership

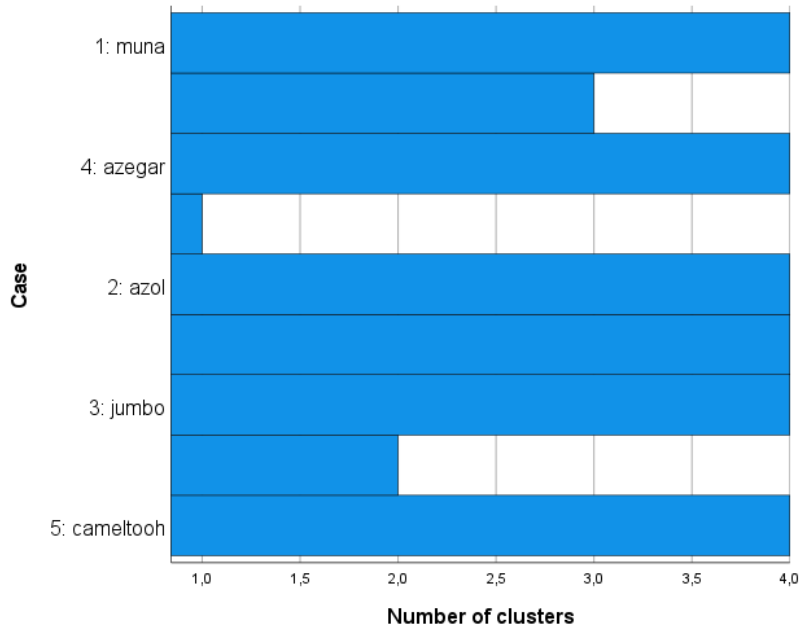
Case	4 Clusters	3 Clusters	2 Clusters
1:muna	1	1	1
2:azol	2	2	2
3:jumbo	2	2	2
4:azegar	3	1	1
5:cameltooh	4	3	2

يبين عضوية كل تركيبة وراثية في العناقيد، بداية من 4 عناقيد وثلاثة ثم عنقودين. حيث نلاحظ أنه قد تم تجميع التراكيب الوراثية على ثلاثة خطوات: في الخطوة الأولى تم انشاء أربعة عناقيد، حيث تم وضع التركيبة الوراثية muna و cameltooh في العنقود الأول، ووضعت التركيبة الوراثية azol في العنقود الثاني، والتركيبة الوراثية jumbo وضعت في العنقود الثالث، والتركيبة azegare وضعت في العنقود الرابع. وفي الخطوة الثانية تم انشاء ثلاثة عناقيد، وتم وضع التركيبتين muna و cameltooh في العنقود الأول، والتركيبتين azol و jumbo في العنقود الثاني، والتركيبة azegare في العنقود الثالث. وفي الخطوة الثالثة تم إنشاء عنقودين، حيث وضعت التركيبات الوراثية muna,azegare,cameltooh في العنقود الأول، ووضعت التركيبتين azol و jumbo في العنقود الثاني.

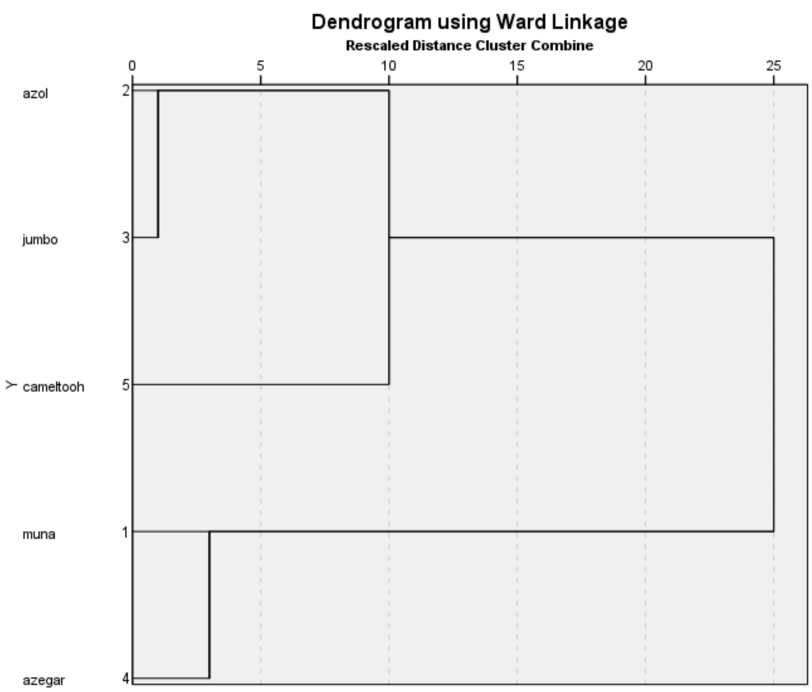
نلاحظ مثلا أنّ التركيبة الوراثية azegar وضعت في الخطوة الأولى في العنقود الرابع، ثم في الخطوة الثانية دمجت في العنقود الثالث، وفي الخطوة الثالثة تم دمجها مع العنقود الأول.

وتتم هذه العملية بناءً على مصفوفة القرابة بين المتغيرات، فالمتغيرات التي يكون الارتباط بينها كبيرا يتم وضعها في عنقود واحد. لاحظ من مصفوفة القرابة أنّ الارتباط بين التركيبة azegar و muna يساوي 0.997 وبين azegar و cameltooh يساوي 0.994، وبين التركيبتين cameltooh و muna بلغ 0.999 ، (الفارق بين 0.02 و 0.05). بينما التركيبتين azol و jumbo، فإنّ معامل الارتباط أو القرابة بينهما كان 0.999.

رسم الألواح الجليدية: توضّح تجميع التركيبات داخل العناقيد.



رسم الشجرة: يوضح كذلك تجميع العناصر أو التركيبات في العناقيد.



توضح مراحل التجميع الثلاثة، التي رأيناها في الجدول الرابع.

ولو عدنا إلى ملف الداتا لدينا سنجد البرنامج قد أنشأ ثلاث متغيرات جديدة (CLU4_1, CLU3_1, CLU2_1).

CLU4_1	CLU3_1	CLU2_1
1	1	1
2	2	2
3	2	2
4	3	1
1	1	1

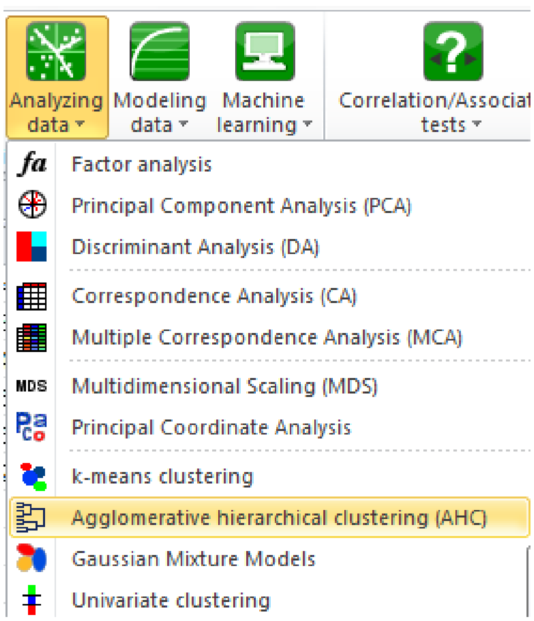
حيث يمكننا أن نلاحظ خطوات عملية التجميع كما شرحنا سابقا.

التطبيق باستخدام برنامج XLSTATS

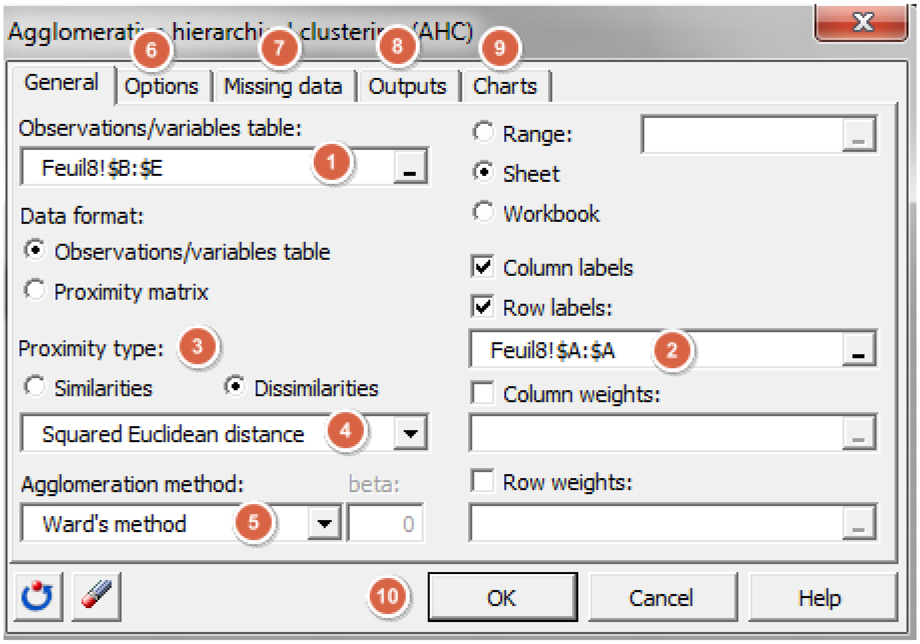
مثال: بالتطبيق على المثال السابق، التراكيب الوراثية والصفات المقاسة لمجموعة من الأفراد.

muna	68	80	33,2	3,95
azol	64	83,25	40	4,5
jumbo	66	87	37,3	4,3
azegar	66	83,25	28,5	2,85
cameltooh	72	83,25	37,5	2,7

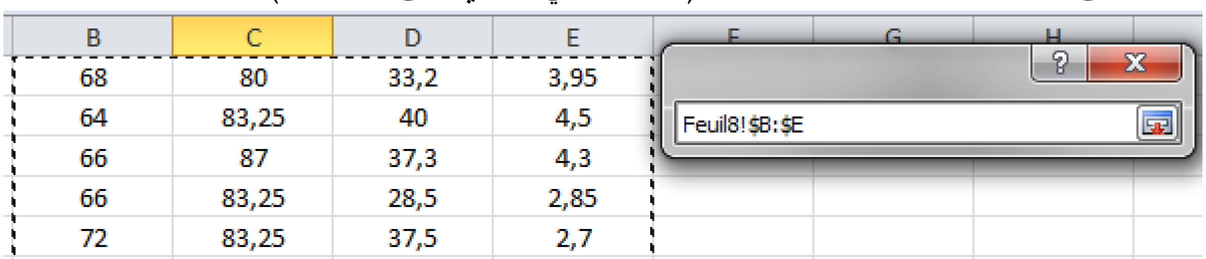
1. من Analyzing data نختار AHC.



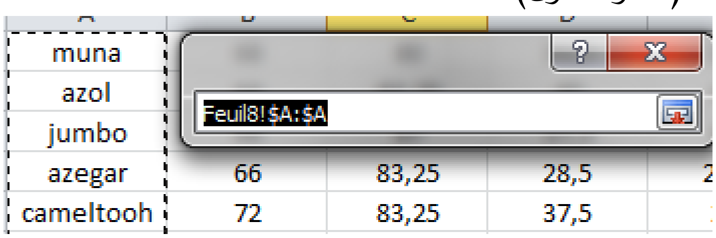
تظهر معنا النافذة الموالية:



1. نضغط على علامة – لتحديد نطاق البيانات (الأعمدة التي تحتوي على المتغيرات).



2. نضغط على علامة – لتحديد عمود التسميات (العمود الأول).

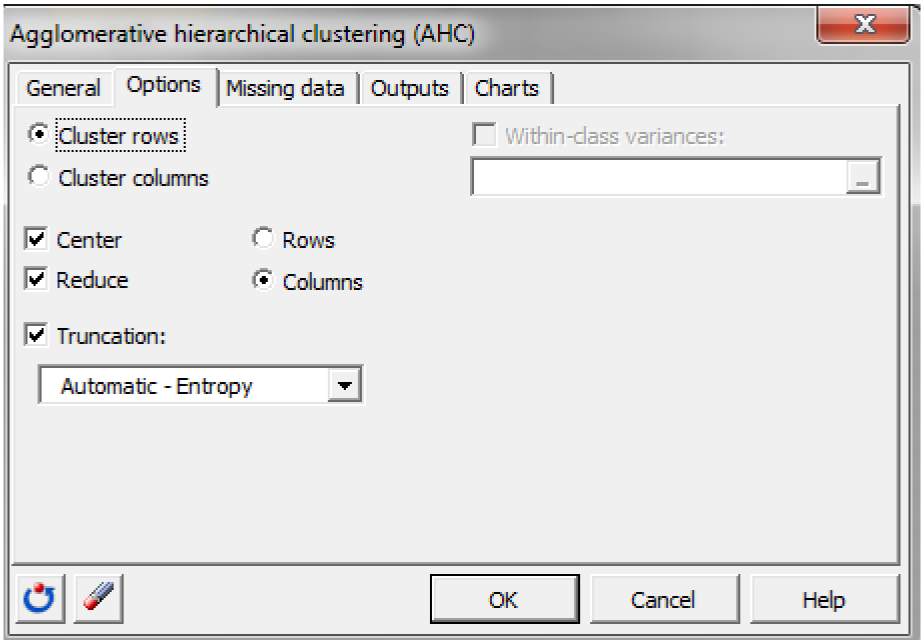


3. نحدد نوع التقارب (Proximity type): (similarities / dissimilarities) أوجه التشابه أو الاختلاف، يحدد نوع البيانات ونوع التقارب، قائمة الفهارس الممكنة لحساب مصفوفة التقارب.

4. نحدد طريقة حساب المسافة، في مثالنا استخدمنا المسافة الإقليدية مربّعة (squard Euclidean distance).

5. نختار مؤشر التجميع الذي نريد استخدامه، في مثالنا استخدمنا مؤشر وارد (ward's method).

6. نضغط Options:



يمكنك الاختيار بين تجميع الصفوف (Cluster rows) (الأفراد) أو الأعمدة (Cluster columns) (المتغيرات). كما يمكنك مركزة البيانات وترجيحها في حالة كانت وحدات المتغيرات غير متجانسة. وقم بتنشيط الخيار (Truncation) إذا كنت تريد من XLSTAT أن يقوم تلقائيًا بتعريف مستوى الاقتطاع، وبالتالي عدد الفئات التي تريد الاحتفاظ بها، أو إذا كنت تريد تحديد عدد الفئات المراد إنشاؤها (يمكنك السماح لعدد الفئات بالتنوع بين حدين)، أو المستوى الذي سيتم عنده اقتطاع مخطط الديندوغرام.

7. يمكنك كذلك تحديد خيارات القيم المفقودة.

8. تحديد المخرجات.

9. تحديد الرسوم البيانية.

10. اضغط ok.

1. الملخص الاحصائي:

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
ص1	5	0	5	64,000	72,000	67,200	3,033
ص2	5	0	5	80,000	87,000	83,350	2,479
ص3	5	0	5	28,500	40,000	35,300	4,516
ص4	5	0	5	2,700	4,500	3,660	0,833

يلخص هذا الجدول عدد المشاهدات لكل متغير، وعدد الحالات المفقودة، أكبر وأصغر قيمة، والمتوسط والانحراف المعياري لكل متغير.

2. مصفوفة القرابة (جدول المسافات):

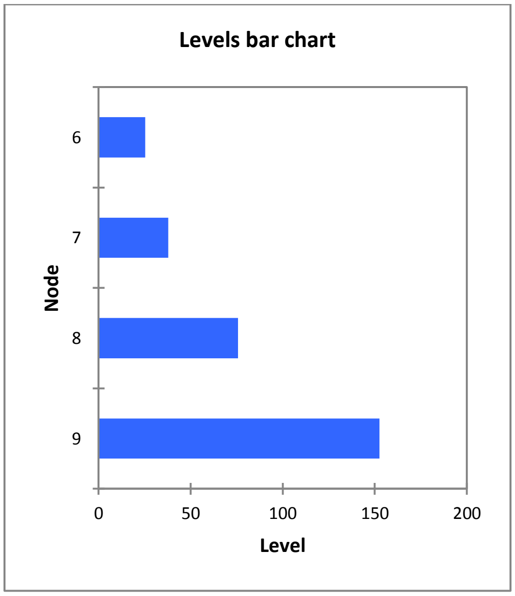
	muna	azol	jumbo	azegar	cameltooh
muna	0	73,105	69,932	37,863	46,615
azol	73,105	0	25,393	138,973	73,490
jumbo	69,932	25,393	0	93,605	52,663
azegar	37,863	138,973	93,605	0	117,023
cameltooh	46,615	73,490	52,663	117,023	0

يظهر من الجدول أنّ أدنى مسافة بين التركيبتين azol و jumbo.

3. إحصاء العقد:

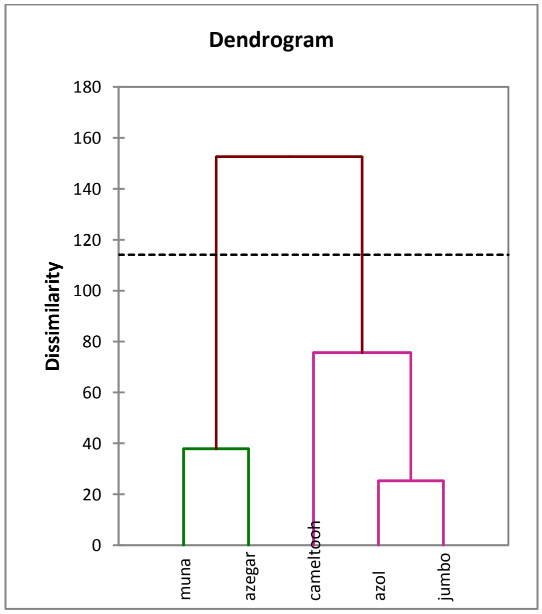
Node	Level	Weight	Objects	Left son	Right son
9	152,572	5	5	7	8
8	75,638	3	3	5	6
7	37,863	2	2	1	4
6	25,393	2	2	2	3

4. الرسم التخطيطي لمستويات العقد:



ينقل شكل الجدول أعلاه معلومات حول بنية البيانات، فعندما نلاحظ قفزات كبيرة، معنى ذلك أنّه لدينا مجموعة من الهياكل غير المتجانسة.

5. الديندوغرام:



يمثل بوضوح كيفية تقدم الخوارزمية لتجميع الأفراد ثم المجموعات الفرعية. وفي النهاية، جمعت الخوارزمية تدريجياً كل الأفراد.

6. النتائج حسب الفئة.

Class	1	2
Objects	3	1
Sum of weights	3	1
Within-class variance	25,258	0,000
Minimum distance to centroid	3,030	0,000
Average distance to centroid	4,022	0,000
Maximum distance to centroid	5,021	0,000
	azol	azegar
	jumbo	
	cameltooh	

من الجدول أعلاه، يمكننا أن نرى توزيع كل تركيبة وراثية في الفئتين (المجموعتين). والتباين داخل الفئة، للفئة 2 (يسارًا على مخطط الديندوغرام) أقل من الفئة 1 (على اليمين) ، مما يؤكد تجانسًا أكبر للفئة 2.

7. النتائج حسب الأفراد.

Observation	Class
azol	1
jumbo	1
azegar	2
cameltooh	1

يظهر الجدول أعلاه الفئة التي ينتمي لها كل فرد.

قائمة المراجع

1. Aluja, T., Morineau, A., Sanchez, G. (2018) Principal Component Analysis for Data Science. https://pca4ds.github.io
2. Ousmane THIARE, Analyse Factorielle des Correspondances (AFC) , othiare@ugb.edu.sn
3. Anderson, T. W. (1963). Asymptotic theory for principal component analysis. **"Annals of Mathematical Statistics**", 34, 122–148.
4. Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. "**Journal of educational psychology**" 24, 417–441.
5. Jolliffe, I. T. (1986). Principal component analysis. Springer-verlag.
6. - Michael E. Tipping, christopher m. Bishop (1999). Probabilistic principal component analysis microsoft research.
7. Brigitte Esco.er, Jérôme Pagès (2008). Analyses factorielles simples et multiples : Objectifs, méthodes et interprétation.
8. Charlotte Baey (2019). Analyse de données. https://baeyc.github.io/teaching/
9. Mohamed ElMerouani (2019), https://elmerouani.jimdofree.com/analyse-des-donn%C3%A9es/
10. Alboukadel Kassambara, Analyse de données, (2019). http://www.sthda.com/french/
11. François Husson, Sébastien Lêet Jérôme Pagès (2009). Analyse des données avec R, **Presses Universitaires de Rennes**, 224 p. (ISBN 978-2-7535-0938-2).
12. Herve´Abdi and Lynne J. Williams, Principal component analysis, "**WIREs Computational Statistics"**, Volume 2, July/August 2010.
13. D. Chessel, J. Thioulouse & A.B. Dufour, Introduction à la classification Hiérarchique, Biostatistique / Fiche stage7.doc / Page 3 / 07/09/04 /, http://pbil.univ-lyon1.fr/R/stage/stage7.pdf
14. Jean-Marc Lasgouttes, Cours d'analyse de donnees, 2019-2020, INRIA Paris.
15. Anne B Dufour, Analyse en Composantes Principales, Octobre 2013, https://pbil.univ-lyon1.fr/R/pdf/add1.pdf
16. Gilbert Saporta, Ndèye Niang, Analyse en composantes principales, HAL Id: hal-02507732 https://hal-cnam.archives-ouvertes.fr/hal-02507732, Submitted on 13 Mar 2020
17. Jean-Marc Lasgouttes, Variables quantitatives : analyse en composantes principales http://ana-donnees.lasgouttes.net/
18. MICHEL MAURIN, Sur les qualités des métriques en ACP, **Statistique et analyse des données**, tome 12, no 3 (1987), p. 58-74 http://www.numdam.org/item?id=SAD_1987___12_3_58_0
19. Mohamed El Merouani, Analyse des donnees: Les methodes factorielles, 2014, http:\\elmerouani.jimdo.com
20. Ricco RAKOTOMALALA, Analyse Factorielle des Correspondances, Université Lumière Lyon 2.
21. Phillip Yelland, An Introduction to Correspondence Analysis, January 2010, https://www.researchgate.net/publication/228417045
22. Dominique LAFFLY, INTRODUCTION À L'ANALYSE FACTORIELLE DES CORRESPONDANCES, Université de Pau, PAU.
23. Rémi Bachelet, L'AFC pour les nuls, http://rb.ec-lille.fr
24. François Husson, Analyse en Composantes Principales, https://www.youtube.com/watch?v=KrNbyM925wI&list=PLnZgp6epRBbRn3FeMdaQgVsFh9Kl0fjqX&index=1&pp=iAQB
25. Analyse de données, https://www.youtube.com/@Comprendrelinformatique