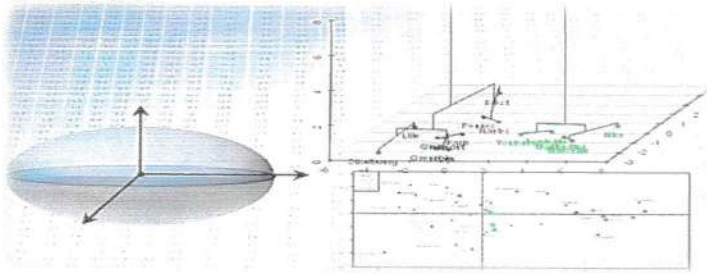
وزارة التعليم العالي والبحث العلمي
جامعة الجيلالي بونعامة -خميس مليانة-
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
ميدان التكوين في العلوم الاقتصادية والتجارية وعلوم التسيير

جامعة خميس مليانة
المجلس العلمي
كلية العلوم الاقتصادية و التجارية و علوم التسيير

مطبوعة بعنوان:

محاضـــرات وتمـاريـن محـلـولـة في مقيـــاس التطبيقـات الاوليـة لتحليـل المعطيـات



الشعبة: علـوم التسيير
التخصص: ادارة الاعمال وادارة الموارد البشرية
المستوى: السنة الثالثة

من إعداد: د. صدقـاوي صوريـة

السنة الجامعية: 2018-2019

ملخــص: " الهـــدف مـــن المقيـــاس "

هذه المطبوعة، والموجهة لطلبة السنة الثالثة علوم التسيير، تخصص ادارة الاعمال وادارة الموارد البشرية، وهي عبارة عن عرض تقديمي وعملي لمختلف التقنيات والاساليب المعتمدة في تحليل المعطيات. حيث تشمل هذه المطبوعة على العديد من الأمثلة والتمارين التطبيقية، والتي يراد بها تمكين الطلبة من التعرف على هذه الاساليب والتقنيات وكيفية استخدامها لتحليل الظواهر محل الدراسة.

ويمكن الاعتماد على هذه الاساليب والتقنيات في تصميم نموذج اقتصادي قياسي وتقديره مثلا، أو في تحليل البيانات التي تم جمعها من مختلف الدراسات الاستقصائية، الخ. لأجل ذلك لا بد من رسم نهج إحصائي يمكن الاعتماد عليه في معالجة وتفسير و وصف المعطيات، ،وبالتالي فهم الظاهرة المدروسة وتحليلها بطريقة مثلى.

والغرض من هذا النهج الإحصائي التحليلي أو الاستكشافي هو تجميع واستخلاص المعلومات الواردة في البيانات وتلخيصها وهيكلتها في شكل جداول أو تمثيلات بيانية أو مؤشرات رقمية.

وتستخدم هذه الأساليب الإحصائية، باختلاف طبيعتها ونهجها، في معظم المجالات العلمية والعملية، لذلك توجب التعرف عليها ومعرفة كيفية تطبيقها والاعتماد عليها في عملية تحليل البيانات. وتعتبر هذه المصبوعة، لما تحتويه من معلومات وتطبيقات عملية، وسيلة تساعد الطالب على:

- فهم أساسيات تحليل المعطيات والتعرف على طرق التحليل، مراحلها ومنطقها؛
- التعرف على كيفية التعامل مع المعطيات، ترتيبها وتنظيمها ووصفها؛
- كيفية تحليل هذه المعطيات والنهج الذي يجب اتباعه؛
- القدرة على المقارنة بين مختلف الطرق وبالتالي سهولة اختيار الطريقة المثلى في التحليل، حسب طبيعة البيانات؛
- إعداد تقييم البيانات لإجراء التحليل العاملي المناسب، مع استيعاب العملية المطلوبة لذاك؛
- اكتساب المعارف التي تساعد في شرح الرسومات البيانية الناتجة عن عملية التحليل؛
- التعرف على كيفية اجراء التحليلات باستخدام برنامج معالجة البيانات SPSS (الاصدار 28).

وغيرها من المعارف التي تساعد الطالب في اكتشاف الإمكانات الكاملة لمختلف تطبيقات تحليل المعطيات.

د. صدقاوي صورية

قائمـة المحتويـــــات

الصفحة

ملخـص: " الهـدف مـن المقيـاس"

قائمة المحتويات

قائمة الجداول

قائمة الاشكال

مقدمة "مدخـل لتحليل المعطيات" أ-د

1. **تمهيد** ب
2. الطريقة البيداغوجية المتبعة ت
3. ماهية تحليل المعطيات ث
4. البيانات وأنواعها ح
5. محتوى المطبوعة خ

الفصل الاول: وصف البيانات وطرق تلخيصها 1

مقدمة الفصل 2

1.1. الجداول الاحصائية 3

1.1.1. البيانات المنفصلة والبيانات النوعية 4

2.1.1. البيانات المتصلة 7

2.1. التمثيلات البيانية 9

1.2.1. الاعمدة البيانية 9

2.2.1. المدرج التكراري 10

3.2.1. الدائرة، العمود المجزأ والأعمدة المستطيلة 13

4.2.1. تمثيلات بيانية اخرى 15

3.1. تلخيص البيانات 17

1.3.1. مقاييس النزعة المركزية 17

2.3.1. مقاييس التشتت 19

4.1. تمرين تطبيقي باستخدام SPSS 21

الفصل الثاني: طريقة التحليل العاملي (Factor Analysis) 27
مقدمة الفصل 28
1.2. الاطار العام للتحليل العاملي 29
1.1.2. تعريف 30
2.1.2. ماهية العوامل 30
3.1.2. أهداف التحليل العاملي 32
4.1.2. شروط التحليل العاملي ومجالات تطبيقه 34
2.2. مختلف الطرق المستخدمة في التحليل العاملي وأنواعه 35
1.2.2. نبذة عن تطور طرق التحليل العاملي 35
2.2.2. أنواع التحليل العاملي 38
3.2.2. طرق التحليل العاملي 39
3.2. اسس تطبيق التحليل العاملي: المبدأ العام 40
1.3.2. تمثيل الأفراد في فضاء المتغيرات 43
2.3.2. تمثيل المتغيرات في فضاء الأفراد 44
3.3.2. العلاقة بين محاور القصور الذاتي وعوامل الافراد والمتغيرات 45
4.3.2. إعادة تشكيل البيانات 46
الفصل الثالث: التحليل الى مركبات أساسية (PCA) 48
مقدمة الفصل 49
1.3. طريقة التحليل الى مركبات أساسية 50
1.1.3. التعريف والاهداف 50
2.1.3. تطبيق طريقة التحليل الى مركبات أساسية 52
2.3. خطوات التحليل الى مركبات اساسية 53
1.2.3. بناء مصفوفة البيانات وحساب المتوسطات والانحرافات 53
2.2.3. حساب مصفوفة احداثيات الافراد ومصفوفة معاملات الارتباط 55
3.2.3. تحديد القيم الذاتية والاشعة الذاتية 56
4.2.3. تمثيل الافراد والمتغيرات على المحاور 59
3.3. التحليل الى مركبات اساسية باستخدام برنامج SPSS 61
1.3.3. ادراج البيانات في الجدول 62

2.3.3. إجراء التحليل الى مركبات اساسية 64
3.3.3. تفسير المخرجات وتحليل النتائج 66
الفصل الرابع: التحليل بالمعاملات للتوفيقات (CA) 74
مقدمة الفصل 75
1.4. طريقة التحليل بالمعاملات للتوفيقات 76
1.1.4. النقاط الرئيسية 76
2.4. التحليل بالمعاملات للتوفيقات: النظرية والتطبيق 78
1.2.4. مراحل التحليل 78
2.2.4. تحويل جدول البيانات (الاسطر / الاعمدة) 81
3.2.4. مثال توضيحي 83
3.4. تحليل CA باستخدام برنامج SPSS 88
1.1.4. تحضير وادخال البيانات 88
2.1.4. تحليل المعاملات للتوفيقات في SPSS 92
3.1.4. التعليق على النتائج 95
الفصل الخامس: التصيف التسلسلي (Hierarchical Classification) 101
مقدمة الفصل 102
1.5. المبدأ الاساسي للتصنيف التسلسلي 103
1.1.5. الفكرة الاساسية 103
2.1.5. أنواع التحليل التصنيفي 105
3.1.5. طريقة عمل التصنيف التسلسلي 108
2.5. خطوات التصنيف التسلسلي 111
1.2.5. تشكيل المجموعات (العناقيد) 112
2.2.5. التمثيل البياني 116
3.5. التصنيف التسلسلي باستخدام برنامج SPSS 118
1.3.5. مراحل التصنيف التسلسلي في SPSS 118
2.3.5. تفسير مخرجات SPSS 122

الفصل السادس: تمارين محلولة وتمارين مقترحة 127

مقدمة الفصل 128

1.6. تمارين محلولة 129

2.6. تمارين مقترحة 140

قائمة المراجع

الملاحق

ملحق 01: اسس اعداد الاستبيان

ملحق 02: الارتباط وطرق قياسه

ملحق 03: التذكير بالجبر الخطي والمصفوفات

ملحق 04: مراحل التحليل الوصفي باستخدام SPSS

قائمـة الجـداول

الرقم	العنوان	الصفحة
01-1	توزيع الطلبة حسب نوع الهاتف	5
02-1	توزيع عدد الاطفال المتمدرسين	6
03-1	النسبي المئوي والتكرار المتجمع الصاعد	9
01-2	علامات الطلاب في ثلاث مقاييس	31
02-2	الغرض من تطبيق مختلف طرق التحليل العاملي	39
01-3	قيم المتوسطات الحسابية والانحرافات المعيارية	54
02-3	أهم اختبارات الصلاحية	58
03-3	جدول تفسير التباين	68
04-3	احداثيات المتغيرات قبل وبعد التدوير	59
05-3	نتائج الارتباطات	70
01-4	مؤشرات تحليلي الابعاد	77
02-4	جدول البيانات	78
03-4	القيم الذاتية ونسبة التباين	84
04-4	مساهمات الصفوف والاعمدة في تشكيل المحاور	85
05-4	جودة التمثيل	85
06-4	الوزن والمسافة في تحليل CA	86
07-4	الإحداثيات	87
01-5	وصف طرق حساب المسافة بين الافراد والمجموعات	107
02-5	نموذج جدول بيانات التصنيف التسلسلي	108
03-5	مصفوفة التباعد	112
04-5	تصنيف المجموعات في SPSS	122

قائمـة الأشكـال

الرقم	العنوان	الصفحة
01–1	توزيع الأساتذة حسب عدد الأطفال المتمدرسين (الأعمدة البسيطة)	10
02–1	المدرج التكراري لتوزيع المراقبة التقنية بالأيام	11
03–1	المدرج التكراري والمضلع التكراري لتوزيع المراقبة التقنية بالأيام	12
04–1	منحنى التكرارات المتجمعة الصاعدة والنازلة	13
05–1	شكل توضيحي لمخطط الصندوق	15
01–2	تمثيل السحابة النقطية للعلامات في فضاء الطالبين الاول والخامس	29
02–2	الفرق بين اساليب التحليل العاملي واساليب التصنيف	40
03–2	تعديل سحابة نقاط الطلاب في فضاء المقاييس	41
04–2	سحابات النقاط	42
01–3	العلاقة بين المتغيرات الأصلية والعوامل	51
02–3	تأثير تطبيق PCA على مجموعة بيانات	52
03–3	اختيار المركب الأساسي الأول	52
04–3	منحنى المكونات بعد التدوير	71
05–3	تمثيل الافراد	73
01–4	التحويل جدول البيانات	82
02–4	تحليل تركيب القصور الذاتي	83
03–4	التمثيل البياني	87
01–5	طريقة عمل التصنيف التسلسلي	109
02–5	العلاقة بين علامات الطلبة	109
03–5	مخطط الشجرة الهيكلية	124

مقدمـة

مدخـل لتحليـل المعطيـات

1. تمهيد

في مجال الإدارة والأعمال، يمكن جمع قدر كبير من البيانات أو المعطيات ذات طبيعة مختلفة ومتنوعة. يمكن أن تكون بيانات توضح نسبة المبيعات، سعر شراء المواد الخام، أرباح المؤسسة على مدى عدة سنوات، تفضيلات المستهلكين، بيانات عن الموظفين، برامج التكوين، ... عموما البيانات الخام ليست سهلة التفسير، فهي عبارة عن اعداد أو رموز تتطلب التحليل والمعالجة. وبالتالي سنحتاج إلى طرق وأساليب تساعدنا في عملية معالجة وتحليل المعطيات.

وتهدف هذه الطرق، التي تعد تحليلية بشكل أساسي، إلى وصف البيانات وتنظيمها وتصنيفها وتوضيحها من خلال مراعاة العديد من وجهات النظر والدراسة والكشف عن السمات الرئيسية أو الروابط أو أوجه التشابه أو الاختلافات بين المتغيرات أو مجموعة من المتغيرات. وهذا هو الغرض من تحليل المعطيات بصفة عامة، والذي يجمع بين:

- الجوانب الإدارية ، المالية، والتسويقية لصياغة الأسئلة ذات الصلة ، جمع البيانات ، وتفسير الإجابات المقدمة.
- الجوانب الرياضية، الإحصائية والتقنية لإجراء التحليلات ذات الصلة وتنفيذها.

والغرض من هذه المطبوعة هو التعرف على مختلف هذه الاساليب المستخدمة بشكل شائع في النهج الإحصائي التحليلي للتعامل مع البيانات وتمثيلها في جداول أو رسومات بيانية او خرائط. حيث يكون الهدف من هذا التحليل هو استخراج أكبر قدر ممكن من المعلومات. ويتم التركيز فيها على العرض النظري لهذه الطرق وعلى المعالجة العملية للبيانات وتطبيقات طرق التحليل المختلفة، من خلال عدة أمثلة توضيحية كأساس لمقارنة الأساليب المستخدمة.

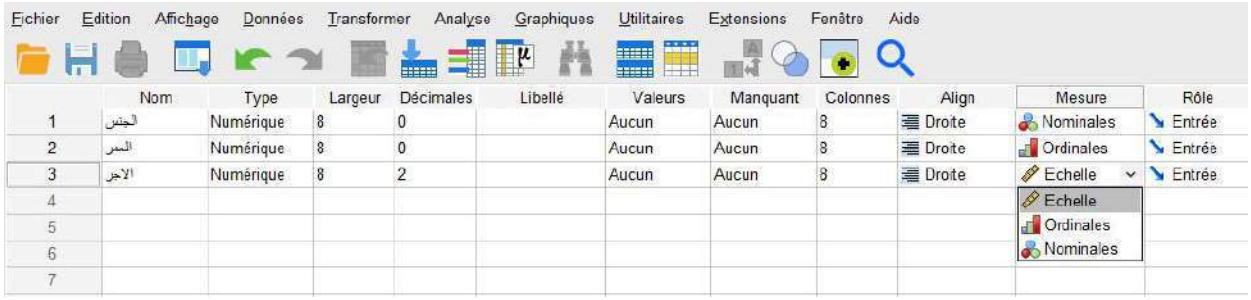
ملاحظة

سنقوم، بطبيعة الحال، بالاعتماد على أدوات اوتوماتيكية، وبصفة خاصة برامج الحزمة الاحصائية SPSS (الاصدار 28)، والتي طورت خصيصا لتسهيل عملية التعامل مع البيانات وتحليلها ومعالجتها. هذا البرنامج ساهم بشكل ملحوظ في تطوير وتعميم العديد من الطرق الإحصائية.

وقبل التطرق الى ماهية واهمية تحليل المعطيات يجب التطرق الى ثلاث عناصر هامة تتمثل في: البيانات والمتغيرات، مجمع الدراسة والعينة، والاستبيان. بالنسبة للبيانات فسوف نتطرق اليها بالتفصيل في العنصر الرابع من هذه المقدمة.

أ. **المتغيرات**: الخاصية او الصفة في عينة الدراسة او الظاهرة المدروسة، وهي قبالة للتغير. ويمكن تصنيفها الى متغيرات متصلة ومنفصلة، متغيرات مستقلة وتابعة، ومتغيرات كمية ونوعية. لذلك ينبغي تعريف طبيعة

المتغير من خلال برنامج SPSS (الاصدار 28) وذلك من خلال عدة عناصر من بينها تحديد طبيعة القياس كما توضحه الصورة الموالية:



ب. **المجتمع والعينة:** المجتمع (N) هو جميع العناصر التي تتعلق بها مشكلة البحث كطلبة السنة الثالثة تخصص ادارة الاعمال بكلية العلوم الاقتصادية بجامعة خميس مليانة. أما العينة فهي جزء من المجتمع ويشترط فيها التجانس حتى نقول انها تمثل المجتمع، كطلبة الفوج الأول للسنة الثالثة ادارة الاعمال. ويتم تحديد العينة بطريقتين إما طريقة عشوائية أو غير عشوائية.

ملاحظة

العينة غير العشوائية: تكون العينات انتقائية ولا تمثل المجتمع تمثيلا صحيحا، وإنما تتم وفق اختيار الباحث، ولذلك لا تكون هناك فرصة متساوية لأفراد المجتمع في الظهور في العينة (لا تستخدم قانون السحب العشوائي). ومنها: العينة الملائمة، العينة الحصصية، وعينة كرة الثلج.

العينة العشوائية: تسمح طرق اختيار العينات العشوائية بالحصول على عينات ممثلة للمجتمع باعتبارها تتبع توزيعا طبيعيا، ويكون احتمال سحب أي عنصر معروفا ويمكن حسابه، منها العينة العشوائية البسيطة، والعينة العشوائية المنتظمة وغيرها

ت. **الاستبيان:** وهو من اهم الأدوات المستخدمة في جمع البيانات والمعلومات المتعلقة بشكلة الدراسة. وتستخدم بيانات الاستبيان للتعرف على توجهات العينة ودراسة السلوكيات الخاصة بها. ويستعرض الملحق رقم 01 اهم الجوانب المتعلقة بالاستبيان.

2. الطريقة البيداغوجية المتبعة

لا يمكن لأي أستاذ باحث في مختلف التخصصات أن يتمكن من إعداد المقياس المدّرس إلا إذا كان ملما بالاطار النظري لهذا المقياس و متمكنا من مختلف الطرق، الاساليب والتقنيات المتعلقة بالمعالجات الاحصائية للبيانات، ومتمكنا بطبيعة الحال من خطوات المنهج العلمي. مما يعني أن الطريقة البيداغوجية المعتمدة في اعداد هذه المطبوعة تجمع بين الاطار النظري، المنهجي والتطبيقي لمقياس تحليل المعطيات.

بطريقة أخرى، فإن الطريقة البيداغوجية المتبعة تشمل عرض نظري لكل طريقة على حدى، متبوعة بمثال تطبيقي يساعد الطالب على استيعاب أهم الاسس المتعلقة بكل طريقة. بالإضافة إلى مجموعة من التمارين المحلولة

باستخدام برنامج التحليل الاحصائي SPSS (الاصدار 28)، مع توضيح مختلف المراحل المتبعة في التحليل ومناقشة مخرجات البرنامج، وهذا بغرض تعزيز قدرات الطالب، وتمكينه من تطبيق مختلف الطرق التي سنتناولها في هذه المطبوعة.

وتوفر هذه المطبوعة لطلبة السنة الثالثة، علو التسيير تخصص ادارة الاعمال وادارة الموارد البشرية بصفة خاصة وحتى لطلبة التخصصات الأخرى التابعة للعلوم الاقتصادية والتجارية وعلوم التسيير، حزمة متكاملة من المفاهيم، والاساليب والأدوات الأساسية في تحليل المعطيات وفق المقرر الرسمي لوزارة التعليم العالي والبحث العلمي. نقدم من خلالها دروس مبسطة ومختصرة وسهلة الفهم مدعمة بالعديد من الأمثلة والتمارين التي تعرض بحلول نموذجية.

و لإعداد هذه المطبوعة، قمنا بمراعاة العناصر التالية:

- عرض الأسس والقواعد العلمية التي تقوم عليها أهم طرق تحليل المعطيات دون الغوص في جوانب نظرية أخرى
- الأخذ بعين الاعتبار درجة الاستيعاب لكل طالب للوظيفة التحليلية لمختلف الأساليب.
- إعطاء أمثلة تطبيقية مع شرح مفصل للنتائج.
- تدعيم المفاهيم النظرية في كل فصل بتطبيقات مع استخدام برنامج التحليل الاحصائي SPSS.
- اقتراح مشاريع على الطلبة (خلال حصص الاعمال الموجهة) وذلك بتوفير بعض قواعد البيانات المستمدة من الواقع. هذا العنصر الاخير يساعد الطالب على توظيف المعارف المكتسبة وتنمية قدراته.

3. ماهية تحليل المعطيات

تعبر عملية تحليل المعطيات عن تنظيم وترتيب البيانات بغرض اخراجها وابرازها على شكل معلومات يتم استخدامها للإجابة على أسئلة معينة أو تحقيق هدف معين. وتأتي مرحلة تحليل المعطيات بعد عملية جمع البيانات وتنظيمها بشكل يسهّل عملية التحليل (كترتيبها في جداول،...).

ويعتبر مصطلح "تحليل المعطيات" المصطلح الشائع والذي يستخدم لوصف التحليلات الإحصائية الوصفية المتعددة الأبعاد، هذه التحليلات تجمع بين مجموعة من التقنيات ذات الأسس الرياضية وتسمح لنا بفهم بنية البيانات التي نريد تحليلها.

وتعد تحليل المعطيات أحد الأساليب القديمة نسبيا (1900- 1935) بفضل أعمال Spearman، Pearson و Hotelling، الذين ساهموا في رسم المبادئ والأسس الرئيسية لأساليب تحليل المعطيات. ومع التقدم التكنولوجي وتطور وسائل الاعلام وبرامج التحليل، توسعت وانتشرت تقنيات تحليل المعطيات.

وعموما نقوم بتحليل المعطيات بهدف:

- اختيار الاسلوب المناسب أو الطريقة المناسبة التي تساعد في تفسير المتغيرات المؤثرة على ظاهرة معينة.
- الوقوف على مدى جوهر تأثير المتغيرات على الظاهرة المدروسة.

- تقدير البيانات وذلك استنادا الى واقع العينات المستمدة من المجتمع المدروس.

وتعبّر تقنيات تحليل المعطيات عن مجموعة من التقنيات التي يمكن الاعتماد عليها لاكتشاف البنية المعقدة المحتملة لمجموعة من البيانات المتعددة الأبعاد وترجمتها إلى بنية أبسط، يمكن عادة تمثيلها بيانيا، وتلخص البيانات بشكل أفضل. ويمكن تعريفها على أنها عبارة عن مجموعة من الطرق الاساليب الوصفية التي تهدف إلى تلخيص وترتيب وتحليل المعطيات بصورة تسمح بفهم أفصل للظاهرة المدروسة.

الغرض من تحليل المعطيات اذن هو وصف المعطيات باستخدام التقنيات المتاحة. وتتوافق المعطيات غالبا مع العديد من المتغيرات العشوائية والعديد من الأفراد أو العناصر. واعتمادا على طبيعة الاشكالية المطروحة، ووفقا لطبيعة البيانات، هناك عدة طرق ممكنة. وتعد تحليل المعطيات أحد أكثر فروع الإحصاءات حيوية، يغطي بشكل أساسي مجموعتين من التقنيات:

- التحليل العاملي: والذي ينتمي إلى الهندسة الإقليدية ويؤدي إلى استخراج القيم الذاتية والأشعة الذاتية. و تسعى هذه الطريقة إلى تقليل عدد المتغيرات وتلخيصها لعدد صغير من المكونات بالاعتماد على أدوات الجبر الخطي بشكل أساسي وإعطاء تمثيلات بيانية يتم فيها تحويل العناصر المراد وصفها إلى نقاط على محاور.
- نماذج التصنيف، حيث تهدف إلى تقليل حجم الأفراد من خلال تشكيل مجموعات متجانسة، وذلك باختيار مؤشر القرب وخوارزمية التجميع أو التصنيف التي تتيح الحصول على شجرة تصنيف.

وسوف نركز، من خلال محتويات هذه المطبوعة، من جهة، على التحليل العاملي وبشكل أساسي على وصف الطريقتين الأكثر استخداما في التحليل، والتي تسمح بربط العديد من المتغيرات ببعضها البعض: **طريقة التحليل بالمركبات الأساسية** (PCA) وطريقة **التحليل بالمعاملات للتوفيقات** (CA). كما سيتم من جهة أخرى، تقديم **طريقة التصنيف التسلسلي** (Hierarchical Classification) كوسيلة مساعدة لتفسير تحليل العوامل، مما يسمح باستكمال وإثراء نتائجه.

وتوفر هذه الطرق المتعددة ا تمثيلات بيانية توفر أفضل ملخص ممكن للمعلومات الموجودة في جداول البيانات.

ملاحظة

تعرف تقنيات التحليل العاملي، والتي تشتمل على جزء من أساس عام مشترك بين الجميع، بما يسمى بالتحليل العام، والذي يعتمد على الأفكار التي طورها سابقا كل من Eckart و Yong (1936). هذه التقنيات تم تطويرها في السنوات الأخيرة أكثر من الناحية النظرية، لا سيما من وجهة نظر التحليل باستخدام البرامج الاحصائية المتطورة.

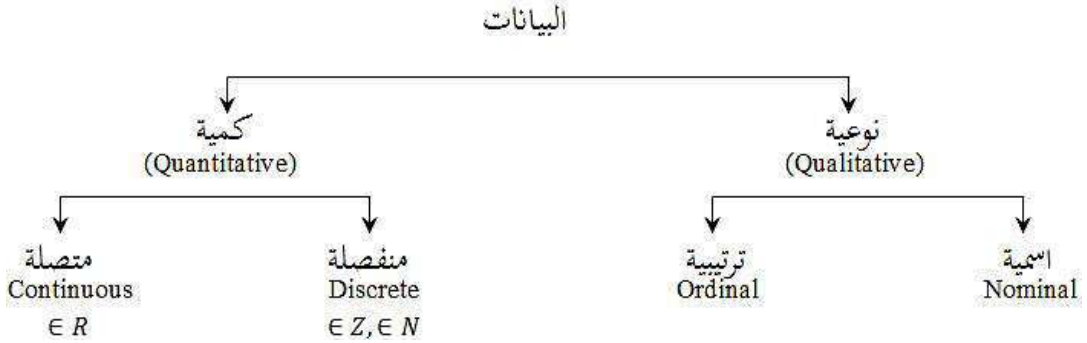
بالإضافة الى ذلك، سوف نقوم بإبراز **الطرق الاحصائية الكلاسيكية** التي تساعدنا في تحليل المعطيات وفهمها. ويتعلق الأمر بمختلف الطرق الوصفية الاحصائية، التي تخص المعالجات الاحصائية وتساهم في وصف الظاهرة المدروسة أو المتغير المدروس.

4. البيانات وأنواعها

تتمثل المادة الرئيسية لتحليل المعطيات في جدول البيانات. وتعتمد طبيعة هذا الجدول على طبيعة المتغيرات التي يتكون منها. حيث يمكننا تمثيل العناصر الاحصائية (الأفراد) في صفوف، ويتم وصف ودراسة هؤلاء الأفراد بواسطة مجموعة من المتغيرات، تمثل في اعمدة الجداول الاحصائية.

ويصف المتغير الإحصائي خاصية الأفراد المختلفين المعرفين. ويتم تعريف المتغير من خلال مجموعة نسميها مجموعة من الملاحظات. بشكل أدق، المتغير عبارة عن خاصية ذات طابع إحصائي معين. هذا الطابع عبارة عن مجموعة من القيم أو القياسات والتي قد تكون في شكل أرقام، أو صفات أو رموز. وهي تعبر عن مواد خام (بيانات) يتم جمعها تحليلها وتفسيرها، وهي نوعان:

- ذات طابع نوعي غير قابلة للقياس: وهي عبارة عن بيانات غير رقمية أو تكون مرتبة رقميا في شكل مستويات أو فئات تنقسم إلى:
  - بيانات نوعية تقاس بمعيار اسمي (Nominale): تتكون من مجموعات، حيث كل مجموعة لها خصائصها تميزها عن المجموعات الأخرى. مثلا: الجنس، الحالة الاجتماعية، الخ.
  - بيانات نوعية تقاس بمعيار ترتيبي (Ordinale): تتكون من فئات يمكن ترتيبها تصاعديا أو تنازليا. مثلا: تقديرات الطلبة، المستوى الدراسي، الخ.
- ذان طابع كمي قابل للقياس: بمعنى أنه يأخذ قيمه في مجموعات رياضية مثل مجموعة الأعداد الطبيعية، الحقيقية ... وهي بيانات يعبر عنها بالأرقام وتنقسم بدورها إلى:
  - بيانات متصلة: عبارة عن قيم تعبر عن صفة يمكن قياسها القيم الصحيحة والكسرية. مثلا: درجة الحرارة، الوزن، الطول، الخ.
  - بيانات منفصلة: هي بيانات تدل على أن الصفة يمكن عدها وتأخذ قيم صحيحة. مثلا: عدد أفراد العائلة، عدد الاولاد، الخ.



الانواع السابقة تجعل من الجداول ذات أشكال مختلفة وهذا الاختلاف هو الذي يحدد بدوره طرق تحليل المعطيات التي يجب تبنيها. كما أن اختيار الطريقة الإحصائية الملائمة يعتمد على الهدف المراد تحقيقه من الدراسة بالإضافة إلى طبيعة المتغيرات وعددها.

ويعتمد غالبا في جمع البيانات على عدة وسائل من بينها: الملاحظة، القابلة، والاستبيان. ويمكن تلخيص هذه المصادر من خلال الجدول التالي:

المصادر		التعريف
التصنيف الاول	*داخلية*	يتم جمعها من المحيط الداخلي وتشمل البيانات التي نجدها في الوثائق المرتبطة الخاصية المدروسة
	خارجية	نحصل عليها من المحيط الخارجي كالدراسات التي تقوم بها المنظمات والهيئات (البنك الدولي مثلا)
التصنيف الثاني	*ثانوية*	وهي المصادر التي نحصل منها على البيانات بطريقة غير مباشرة، من خلال أجهزة، وهيئات رسمية متخصصة
	اولية	وهي المصادر التي نحصل منها على البيانات بشكل مباشر، حيث يقوم الباحث نفسه بجمع البيانات مباشرة من العينة

وتجدر الاشارة الى ان الاسلوب المستخدم في جمع البيانات يتحدد وفقا للهدف المراد تحقيقه وكذا حجم المجتمع محل الدراسة. وهنا نجد اسلوبين:

أ. **اسلوب الحصر الشامل**: يستخدم لحصر جميع عناصر المجتمع حيث يتم جمع بيانات عن كل عنصر من عناصر المجتمع، كحصر جميع طلبة كلية العلوم الاقتصادية لجامعة خميس مليانة. ويتميز هذا الأسلوب بالشمول وعدم التحيز، ودقة النتائج، ولكن يعاب عليه أنه يحتاج إلى الوقت والمجهود، والتكلفة العالية.

ب. **أسلوب المعاينة** : يتم هنا معاينة جزء من المجتمع، حيث يتم اختيار العينة بالاعتماد على الطرق المعروفة، ثم تتم عملية جمع البيانات، من خلال الاستبيان مثلا، ثم يقوم الباحث بدارسة وتحليل البيانات وتقييم النتائج.

5. محتوى المطبوعة

انطلاقا مما سبق، ونظرا لأهمية هذا المقياس بالنسبة للطلبة و حتى الباحثين، وسعيا منا إلى تقديم ما هو مقرر دراسته في هذا المقياس، قمنا بتنظيم هذه المطبوعة على النحو الذي يساهم في إثراء البحوث في هذا المجال. من أجل ذلك قمنا بتقسيم محتوى هذه المطبوعة إلى ستة فصول كما يلي:

الفصل الأول: " **وصف البيانات وطرق تلخيصها** " يلخص هذا الفصل مختلف الطرق الاحصائية الوصفية والتي تساعد الطالب في عملية تلخيص المعطيات وتنظيمها وترتيبها (كالعرض الجدولي، التمثيلات البيانية ...) بشكل يسمح بالتعرف أكثر على طبيعة البيانات المراد تحليلها واستنتاج بعض النتائج الأولية منها.

الفصل الثاني: ": **طريقة التحليل العاملي** " نقدم من خلال هذا الفصل مفهوم التحليل العاملي وتطور مختلف طرقه الذي يسمح بإسقاط بيانات كبيرة في مساحة ذات بُعد أصغر. وسوف نتعرف أكثر على هذه التقنية الاستكشافية وكيفية تطبيقها على مجموعة من المتغيرات بغرض إيجاد العوامل الأساسية (مجموعات فرعية من المتغيرات) التي تم إنشاء المتغيرات المرصودة منها.

الفصل الثالث: " **طريقة التحليل الى مركبات أساسية** " يتناول هذا الفصل تقنية التحليل العاملي التي تستعمل لتحليل المعطيات الكمية. ويتعلق الأمر بالمركبات الأساسية، اسسها ومبادئها وطريقة عملها.

الفصل الرابع: " **التحليل بالمعاملات للتوفيقات** " في هذا الفصل سنقوم بدراسة نوع آخر من التحليل العاملي الذي يتمثل هدفها في تحليل البيانات الكيفية. ويخصص هذا الفصل للتعرف على هذه الطريقة التي تعتبر في مجملها وصفية، تعمم اختبار الاستقلالية (chi-square).

الفصل الخامس: " **التصنيف التسلسلي** " ننتقل في هذا الفصل إلى مشكلة التصنيف الخاضع للإشراف، مما يجعل من الممكن تسليط الضوء على المجموعات الفرعية في البيانات. سوف نتعرّف على طريقة التصيف التسلسلي التي عادة ما تكون مكملة لتحليل المعطيات . وسوف نتطرق من خلال هذا الفصل الى عرض الطريقة وكيفية تطبيقها وتفكيك مجموعة من الافراد أو المتغيرات الى مجموعة منفصلة.

بالإضافة إلى احتواء كل فصل على امثلة تدعيمية بغرض تبسيط المفاهيم ومساعدة الطالب على استيعاب مختلف الطرق الاحصائية التحليلية وفهم أفضل للدرس مع تطبيقات باستخدام برنامج التحليل الاحصائي SPSS (الاصدار 28).

اما في الفصل السادس: " **تمارين محلولة وتمارين مقترحة**" فقد قمنا بإدراج مجموعة من التمارين مع اقتراح حلول لبعض منها باستخدام برنامج التحليل الاحصائي SPSS.

وتجدر الاشارة أيضا إلى انه تم إدراج بعض الملاحق في هذه المطبوعة تشمل بعض التذكيرات الهامة في الجبر الخطي وحساب المصفوفات، أيضا كيفية قياس العلاقة بين مختلف المتغيرات، الخ.

الفصـل الأول

وصـف البيانـات وطــرق تلخيصـــها

مقدمة الفصل

من أجل القيام بعملية التحليل يحتاج الباحث مجموعة من الأساليب الإحصائية، وعليه أن يختار منها وفق ما يتلاءم مع طبيعة البيانات التي قام بجمعها، وحسب أهداف البحث وغيرها من المعايير التي تساعده في إعداد البيانات للتحليل. يعتبر علم الإحصاء أحد أهم الأساليب الكمية العلمية الواسعة الاستخدام سواء في مجال البحث أو الإدارة أو اتخاذ القرار.

يستعمل كأداة فعالة لجمع ومعالجة و تحليل البيانات حول المشكلة المطروحة ثم الجواب عليها بشكل موضوعي. ويعد أيضا أداة لخدمة متخذي القرار والعلماء في مختلف مجالات المعرفة عن طريق تزويدهم بالمؤشرات التحليلية التي تساعدهم على اتخاذ القرارات الرشيدة بشأن المشكلات قيد الدراسة. ولهذا إذا فان مقياس الاحصاء الوصفي يدرس في مختلف الجامعات وفي جميع التخصصات نظرا لأهميته في جميع مجالات المعرفة العلمية.

ومنه فقبل التطرق الى كيفية استخدام اساليب التحليل العاملي وجب تذكير الطالب بمختلف الاساليب الاحصائية الوصفية وكيفية جمع البيانات وتنظيمها في جداول إحصائية ثم عرضها بيانيا، دون الدخول في التفاصيل النظرية. بالنظر الى محتويات هذا الفصل سيتمكن الطالب:

- معرفة كيفية جمع، تنظيم، عرض وتحليل البيانات
- كيفية عرض البيانات
- التعرف على مختلف الاشكال البيانية حسب طبيعة البيانات (كمية او نوعية)
- التعرف على بعض المقاييس الوصفية

تمهيد

في معظم الأوقات تأتي البيانات على الشكل التالي: " نقوم بتحديد، على مدى n وحدة (تعرف بالأفراد)، مجموعة من المتغيرات العددية p " (انظر الملحق). وعندما تكون كل من n و p كبيرة، نحاول تجميعها على شكل معطيات مفهومة وقابلة للتحليل.

وبالتالي تتمثل الخطوة الأولى في وصف النتائج التي تم الحصول عليها لكل متغير بشكل منفصل. وبالرغم من كونها غير كافية، تعتبر هذه المرحلة اساسية في عملية التحليل لأية دراسة احصائية.

لذلك سنقوم من خلال هذا الفصل بوصف المتغير X والذي يأخذ القيم : x_1، x_2، x_3، ... x_n. ولكن قبل ذلك يجب الاجابة على سؤال مهم هو: **لماذا نلخص البيانات؟**

نلخص البيانات بهدف تبسيطها حتى يتسنى لنا فهم طبيعتها. يتم تجميع هذه البيانات في شكل **جداول ورسومات بيانية** وملخصات **رقمية**. وهذا ما يسمى عادة "الإحصاء الوصفي". ويمكننا استخدام بعض الادوات التي تساعد في تبسيط قراءة البيانات، من أهمها: العرض الجدولي والعرض البياني.

1.1. الجداول الاحصائية

بعد جمع المعطيات تتشكل لدينا مجموعة كبيرة من البيانات غير المنظمة والتي حصلنا عليها من الاستبيان مثلا (انظر الملحق). وبالنظر الى ذلك يتعذر علينا استيعابها ويصعب استخلاص أية نتائج منها. لذا يجب تنظيمها بطريقة تسهل عملية التحليل للاستفادة منها. ويتم ذلك بتصنيفها وتقسيمها إلى مجموعات متجانسة ووضعها في صورة جداول تلخصها، ويتوقف هذا التقسيم على طبيعة البيانات وعلى الغرض والهدف من البحث.

الغرض من ذلك وضع المعطيات الخاصة بالظاهرة المدروسة في جداول تتكون من عدة اعمدة واسطر توضح قيم الظاهرة. وتأخذ هذه القيم شكل صفات أو قيم نقطية أو فئات، مع التكرارات المناظرة لها.

ومن اهم انواع الجداول الاحصائية نجد:

- جداول التوزيع التكراري البسيطة و التي تستخدم لوصف وتلخيص البيانات التي تتعلق بظاهرة واحدة فقط سواء كانت كيفية أو كمية.
- جداول التوزيع التكراري المزدوجة والتي تستعمل عند دراسة ظاهرتين

مثال 1-01:

يوضح الجدول الأول توزيع الطلبة حسب التخصص، والثاني يبين توزيع الطلبة حسب السن والتخصص:

الجدول الاول: توزيع الطلبة حسب التخصص

التخصص	عدد الطلبة
ادارة الموارد البشرية	20
تسيير المؤسسة	32
اقتصاد بنكي	18
التسويق	35
المجموع	**105**

الجدول الثاني: توزيع الطلبة حسب السن والتخصص

التخصص / السن	[20-18[	[22-20[	[24-22]	المجموع
ادارة الموارد البشرية	10	8	2	20
تسيير المؤسسة	15	10	7	32
اقتصاد بنكي	10	8	0	18
التسويق	20	12	3	35
المجموع	55	38	12	**105**

وترتبط عملية تلخيص البيانات بطبيعة المعطيات (كمية أو نوعية) وسنقوم أولا بتوضيح كيفية تسهيل قراءة البيانات النوعية والبيانات الكمية ذات الطبيعة المنفصلة أولا، ثم ننتقل الى توضيح كيفية تلخيص البيانات المتصلة.

1.1.1. البيانات المنفصلة و البيانات النوعية

لتكوين جدول توزيع للبيانات النوعية أو الكمية المفصلة نحتاج إلى إعداد جدول مكون عدة أعمدة، يخصص كل عمود لتوضيح صفة معينة، التكرارات، النسب المئوية، الخ. ثم نقوم بتفريغ المعطيات التي قمنا بجمعها في الجدول. وعموما فان الجدول الإحصائي يأخذ الشكل التالي:

x_i	n_i	f_i

وسوف نقوم بتوضيح كيفية تسهيل قراءة المعطيات النوعية او المعطيات الكمية المفصلة من خلال المثال التالي:

أ. البيانات النوعية

مثال1-02:

أردنا اجراء دراسة حول نوع الهاتف الذكي الذي يملكه كل طالب، من أجل ذلك قمنا باستجواب 50 طالبا من بين 200 طالب في تخصص العلوم الاقتصادية. وكانت النتائج المحصل عليها كما يلي:

iPhone	Samsung	Huawei	Samsung	iPhone	iPhone	Oppo	Samsung	Condor	iPhone
Oppo	Condor	iPhone	Samsung	Huawei	iPhone	Condor	iPhone	Huawei	Huawei
Condor	Huawei	Samsung	Oppo	iPhone	iPhone	Samsung	Oppo	iPhone	iPhone
Samsung	iPhone	Condor	iPhone	Samsung	Samsung	iPhone	Huawei	Samsung	Samsung
Samsung	iPhone	Samsung	Oppo	Huawei	iPhone	iPhone	Huawei	iPhone	iPhone

من خلال ما سبق نلاحظ ان هذه المعطيات ذات طبيعة كيفية (نوعية). وما يمكن استنتاجه ايضا هو ان هذه المعطيات بوضعها السابق تصعب عملية التعرف على الهاتف الذكي المفضل للطلبة، خاصة اذا كان العدد هو 500 بدلا من 50 طالب.

وبالتالي دعت الحاجة الى ضرورة ترتيب الاجابات السابقة في جدول توزيع تكراري يلخصها ويسهل عملية قراءتها، كما يلي:

الجدول رقم 1-01: توزيع الطلبة حسب نوع الهاتف الذكي

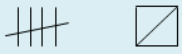
نوع الهاتف	التكرار	النسبة (%)
iPhone	19	38
Samsung	13	26
Huawei	8	16
Condor	5	10
Oppo	5	10
المجموع	**50**	**100**

يمكننا القول انه قبل ادراج هذه المعطيات وترتيبها في جدول، كان من الصعب فهمها، أو بطريقة اخرى كان يصعب علينا تحديد عدد الطلبة (أو نسبتهم) الذين يملكون هاتف معين (iPhone، Samsung، ...). ولكن بعد تفريغ البيانات في الجدول، اصبح بالإمكان معرفة ما هو نوع الهاتف الأكثر استعمالا.

إن وضع البيانات على النحو السابق (في جدول) يساعد بطريقة أكثر وضوحا على استنتاج عدة معلومات (راتول، 2005) كانت غير واضحة في الصورة الخام، فمثلا من السهل الآن معرفة عدد الطلبة الذين لديهم نفس الهاتف، أو الهاتف الأكثر انتشارا بين الطلبة. فيمكن ان نقول مثلا أن 19 طالبا من بين 50 يمتلكون هاتف من نوع iPhone. وهو ما يمثل 38% من الطلبة، متبوع بـ Samsung (26%)، الخ.

ملاحظة

يمكن وضع العلامات أو مجموعة من الحزم المكونة من خمسة خطوط لحساب التكرارات. ويمكن تمثيل الحزمة على الصور التالية:



ويمكن عرض البيانات في شكل توزيع تكراري نسبي للتعبير عن الأهمية النسبية لتكرار كل فئة بالنسبة لإجمالي التكرارات، ويحسب التكرار النسبي بالصيغة التالية:

$$f_i = \frac{n_i}{\sum n_i}$$

وأيضا من بين ايجابيات ترتيب وتنظيم المعطيات في جدول نذكر: امكانية حساب النسبة المئوية التي يمثلها كل نوع من أنواع الهاتف المستخدم، أو تحويل التكرار النسبي إلى تكرار نسبي مئوي وهذا بضربه في 100.

ملاحظة

يفيد التكرار النسبي في تقليص الشكل البياني عندما يكون عدد القيم كبيرا، بينما يفيد التكرار النسبي المئوي في إظهار الشكل البياني عندما يكون عدد القيم صغيرا.

بعد إعداد جدول التوزيع التكراري يكون من المناسب في أغلب الأحيان

ويمكن كتابة الجدول السابق على شكل افقي، كما يلي:

نوع الهاتف	iPhone	Samsung	Huawei	Condor	Oppo	المجموع
التكرار	19	13	8	5	5	50
النسبة(%)	38	26	16	10	10	100

ب. البيانات الكمية المنفصلة

لتبويب بيانات المتغيرات المنفصلة نقوم بتصنيفها إلى مجموعات متشابهة (Casin, 2013)، ثن نقوم بوضعها في جداول مكونة من أعمدة. حيث يخصص العمود الأول مثلا لقيم الظاهرة بعد ترتيبها، والعمود الثاني يخصص لتفريغ البيانات أما العمود الثالث فيخصص للتكرارات، والمثال أدناه يعطي توضيحا لذلك:

مثال 1-03:

اراد قسم الضمان الاجتماعي لجامعة ما تحديد قيمة منحة التمدرس التي على الجامعة دفعها لـ 30 استاذ. من أجل ذلك قامت بعد عدد الاطفال المتمدرسين لكل استاذ. وكانت النتائج كما يلي:

1	3	0	2	3
3	3	1	2	4
4	0	3	2	1
2	1	0	4	2
3	2	3	3	1
2	1	2	3	0

ويمكن عرض البيانات السابقة في جدول، كما يلي:

الجدول رقم 1-02: توزيع عدد الاطفال المتمدرسين

عدد الاطفال المتمدرسين	التكرار	النسبة (%)
0	4	13.33
1	6	20
2	8	26.67
3	9	30
4	3	10
المجموع	**30**	**100**

وبالتالي يمكننا، بعد تفريغ المعطيات في جدول، القول أن أكثر من25% من الاساتذة لديهم طفلين متمدرسين.

2.1.1. البيانات المتصلة

عند دراسة متغير كمي متصل يضم مجال الدراسة ما لانهاية من القيم، ونظرا لصعوبة وضع كل هذه القيم نقوم بتقسيمها إلى مجالات جزئية تسمى فئات حيث يحدد عدد الفئات حسب حجم العينة وحسب توزيع الوحدات الإحصائية على مجال الدراسة (مواري، 1991). ولتكوين جدول التوزيع التكراري لمتغير كمي ذو طبيعة متصلة نتبع الخطوات التالية:

1. حساب المدى: أو المجال الذي تنتشر فيه المعطيات، زيمكن حسابه بحساب الفرق بين أحبر قيمة أصغر قيمة.

$$E = X_{max} - X_{min}$$

2. تحديد عدد الفئات: إن استخدام عدد قليل من الفئات يؤدي إلى تسهيل العمليات الحسابية مع انخفاض الدقة، بينما يؤدي زيادة عددها إلى كثرة العمليات الحسابية غير انها تزيد من الدقة. ويتحدد عدد الفئات بظروف الظاهرة قيد الدراسة ووجهة نظر الباحث. وعلى العموم فمن الأفضل ألا يقل عدد الفئات عن 5 ولا يزيد عن 15. ويتم تحديد عدد الفئات المطلوبة لتشكيل جدول التوزيع التكراري باستخدام بعض المعادلات الرياضية، من بينها:

- معادلة ستورجس (Staurges): التي تعطى بالعلاقة التالية (بن فرحات وقطوش، 2014):

$$K = 1 + 3.322\log(n)$$

حيث: K يمثل عدد الفئات و n تمثل عدد القيم

- معادلة يول (Yule): نقوم بحساب عدد الفئات بالاعتماد على الصيغة التالية (رويسات، 2006):

$$K = 2.5\sqrt[4]{n}$$

3. تحديد طول الفئة: يتم تحديد طول الفئة بالعلاقة التالية :

$$L = \frac{E}{K}$$

ملاحظة

الفئة يجب مراعاة المتباينة التالية عند تحديد طول:

طول الفئة $\times$ عدد الفئات $\geq$ المدى

تجدر الاشارة الى انه يجب تحديد بداية ونهاية كل فئة، على أن تكون بداية الفئة الأولى أصغر من أو تساوي أصغر قيمة في المعطيات، أما الفئة الاخيرة فتكون أكبر من أكبر قيمة في المعطيات.

مثال 1-04:

لتكن المعطيات التالية والتي تمثل المراقبة التقنية للسيارات الجديدة بالأيام، وعددها 20 سيارة.

12	14	19	18	14
15	15	18	17	18
20	27	22	23	16
22	28	31	28	13

على الرغم من أن عدد المشاهدات لا يتعدى 20 مشاهدة فإنه من الصعب أن تكون لنا فكرة واضحة وسريعة عن هذه القيم، لهذا وجب ترتيبها وحصرها في فئات ثم وضعها في جدول توزيع تكراري يشمل عدد من الفئات والتكرارات. ونلاحظ أن المعطيات ذات طبيعة كمية، ولترتيبها في جدول لا بد من تحديدها في مجالات أو فئات، أي علينا اتباع الخطوات التي سبق وان شرحناها.

1. حساب المدى: من خلال المعطيات السابقة نلاحظ أن اصغر قيمة تساوي 31 وأقل قيمة تساوي 12

(12)	14	19	18	14
15	15	18	17	18
20	27	22	23	16
22	28	(31)	28	13

وبالتالي نحسب المدى على النحو التالي:

$$E = X_{max} - X_{min} = 31 - 12 \quad ; \quad E = 19$$

ومنه فإن المدى يساوي 19.

2. تحديد عدد الفئات: لتحديد عدد الفئات سنعتمد على المعادلة الأكثر استخداما وهي معادلة (Staurges):

$$K = 1 + 3.322 \log(20) = 1 + 3.322 * 1.301$$
$$K = 5.32 \approx 5$$

اذن عدد الفئات هو خمس فئات

3. تحديد طول الفئة: مما سبق يمكننا حساب طول الفئة والذي يساوي

$$L = \frac{19}{5} = 3.8$$

ومنه فان الطول النسبي للفئة هو 3.8 ولكن نأخذ 4 كطول الفئة بغرض ترتيب البيانات.

مع مراعاة المتباينة التالية

طول الفئة × عدد الفئات ≤ المدى

$$21 < 20\ (5 \times 4)$$

والآن نقوم بتفريغ المعطيات في جدول التوزيع التكراري ونحصل على الجدول التالي:

الجدول رقم 1-03: النسبي المئوي والتكرار المتجمع الصاعد

الفئة	التكرار	%	التكرار %المتجمع
]16-12]	7	35	35
]20-16]	6	30	65
]24-20]	3	15	80
]28-24]	3	15	90
]32-28]	1	1	100
المجموع	**20**	**100**	**-**

الفئة الاولى تنطلق من أقل قيمة للمعطيات (وهي 12) ويضاف إليها طول الفئة لتحديد نهايتها وبداية الفئة الثانية وهكذا:

- الفئة الأولى:]16-12[12 + 4 = 16
- الفئة الأولى:]20-16[16 + 4 = 20

ونستمر بنفس الطريقة حتى نحصل على جميع الفئات الخمس.

وقد ساعدنا اعداد هذا الجدول على فهم الحقائق وتسهيل قراءة المعطيات للوصول الى استنتاجات لا يمكن الوصول إليها من المعطيات الخام.

2.1. التمثيلات البيانية

إن استخدام التمثيل البياني يجعل المعلومات الإحصائية أكثر وضوحا وفهما، مما يساعد على أخذ فكرة شاملة وسريعة عن الظاهرة المدروسة. بالإمكان وصف وتلخيص البيانات باستخدام عدة رسومات وأشكال بيانية (Bourbonnais, 2004). إذ تمكن هذه الأخيرة من القيام بتحليل سريع للظاهرة المدروسة، ويمكن أن نستخدم أنواع مختلفة للعرض البياني حسب نوع المتغير المدروس (مواري، 1991).

وهناك مجموعة واسعة من الرسوم البيانية، والاكثر شيوعا هي:

1.2.1. الاعمدة البيانية

هي عبارة عن أعمدة بسيطة تتناسب أطوالها مع التكرار المقابل لقيمة معينة للمتغير المدروس وتسمى الأعمدة البسيطة.

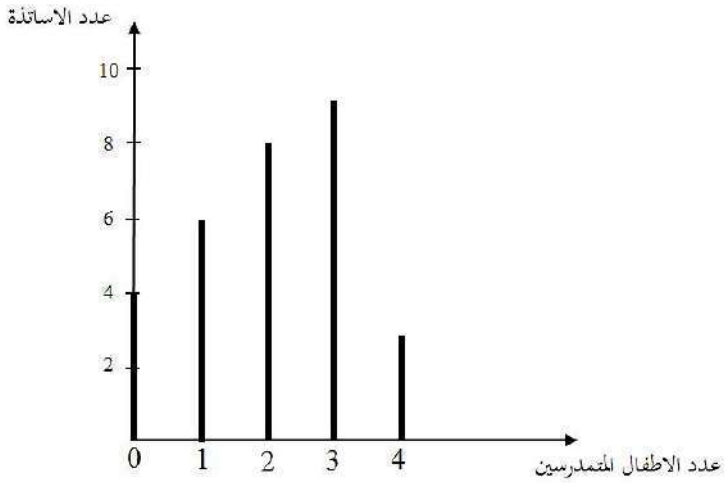
مثال 1-05:

سنقوم بعرض بيانات الجدول في المثال 1-03 بالطريقة المناسبة. حيث أن أفضل طريقة لعرض هذه البيانات هي الأعمدة البسيطة.

الجدول رقم 1-02: توزيع عدد الاطفال المتمدرسين

عدد الاطفال المتمدرسين	التكرار
0	4
1	6
2	8
3	9
4	3
المجموع	**30**

الشكل رقم 1-01: توزيع الأساتذة حسب عدد الأطفال المتمدرسين (الأعمدة البسيطة)



ملاحظة

يمكننا عرض التكرارات المتجمعة الصاعدة والنازلة بيانيا:

ت. التكرار المتجمع الصاعد: هو عبارة عن قطع مستقيمة متصاعدة حسب تصاعد التكرارات التجميعية الصاعدة المقابلة لكل قيمة من قيم المتغير الإحصائي المدروس.

ث. التكرار المتجمع النازل: هو عبارة عن قطع مستقيمة متنازلة حسب تنازل التكرارات التجميعية النازلة، حيث أن القطعة المستقيمة الأولى تقابل مجموع التكرارات وأصغر قيمة للمتغير المدروس والقطعة الثانية تقابل مجموع التكرارات ناقص التكرار البسيط الأول مع القيمة الثانية للمتغير الإحصائي وهكذا (جلاطو، 2002).

2.2.1. المدرج التكراري

وهو العرض البياني الاكثر استعمالا في حالة بيانات ذات طبيعة كمية متصلة (موراي، 1991). وهو عبارة عن مستطيلات (أعمدة) متلاصقة تمثل تكرارات أو قيم كل فئة من الفئات، حيث أن طول كل منها يتناسب مع التكرار المقابل، وقاعدة كل منها تساوي طول الفئة المقابلة، حيث توضع الفئات على محور السينات، بينما توضع التكرارات على محور العينات. وفي حالة تساوي أطوال الفئات فإن قاعدة المقارنة ثابتة ومتساوية وعليه نقوم برسم المدرج التكراري مباشرة. لتوضيح ذلك نستهل بمعطيات الجدول رقم 1-03 السابق.

مثال 1-06:

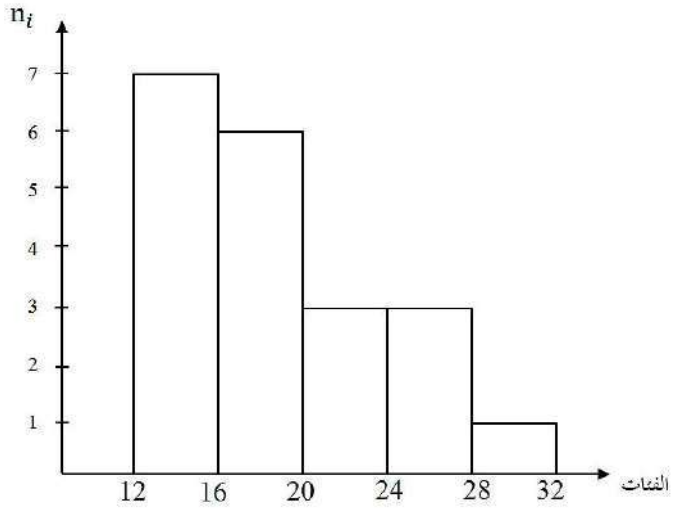
مثل بيانيا معطيات الجدول رقم 1-03

الجدول رقم 1-03: النسبي المئوي والتكرار المتجمع الصاعد

الفئة	التكرار
]16-12]	7
]20-16]	6
]24-20]	3
]28-24]	3
]32-28]	1
المجموع	**20**

بما أن الفئات متساوية الطول نقوم برسم المدرج التكراري مباشرة.

الشكل رقم 1-02: المدرج التكراري لتوزيع المراقبة التقنية بالأيام



ملاحظة

من المفيد قبل رسم المدرج التكراري ملاحظة أطوال الفئات إن كانت متساوية أم لا. وفي حال لم تكن فئات التوزيع متساوية الطول نقوم بتعديل التكرارات حتى يكون هناك تناسب بين طول الفئة والتكرار المقابل لها، لأجل ذلك نستخدم المعادلة التالية:

$$n_i' = \frac{n_i}{L_i} * L'$$

حيث:

n_i' يمثل التكرار المعدل ، n_i يمثل تكرار الفئة ، L_i يمثل طول الفئة ،

L' يمثل طول الفئة المختار (هو القاسم المشترك الأكبر لأطوال الفئات.)

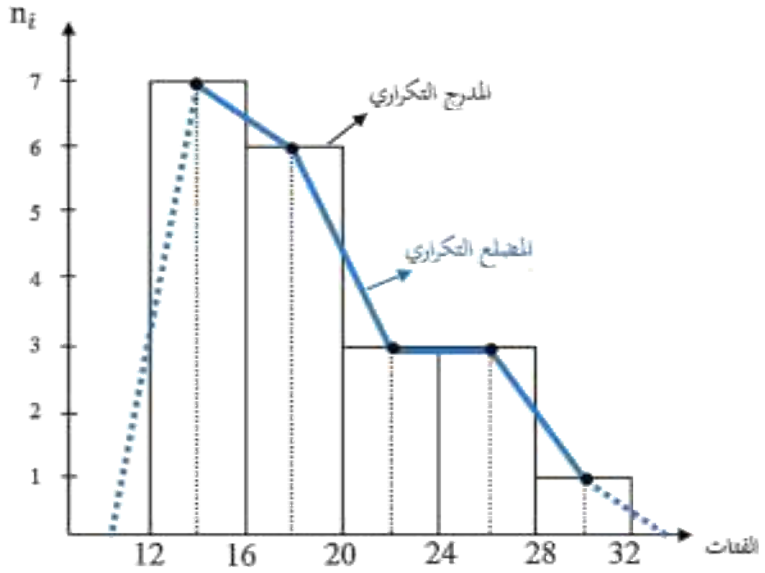
فيما يخص دائما البيانات الكمية المتصلة فيمكننا تمثيل أيضا باستخدام التمثيلات البيانية التالية:

أ. المضلع التكراري

والذي يمثل مجموعة من القطع المستقيمة المتصلة والمنكسرة تتحدد بنقاط إحداثياتها: مراكز الفئات والتكرارات المقابلة لها.

ولتوضيح ذلك نقوم برسم المضلع التكراري من المثال السابق، كما يلي:

الشكل رقم 1-03: المدرج التكراري والمضلع التكراري لتوزيع المراقبة التقنية بالأيام



ب. منحنى التكرارات المتجمعة الصاعدة والنازلة

يتم رسم منحنى التكرار المتجمع الصاعد عن طريق إيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود العليا للفئات والتكرارات المتجمعة الصاعدة المقابل لها. أما منحنى التكرار المتجمع النازل فيتم رسمه بإيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود الدنيا للفئات والتكرارات المتجمعة النازلة المقابلة لها، وتعطينا نقطة التقاطع بين المنحنيين قيمة الوسيط.

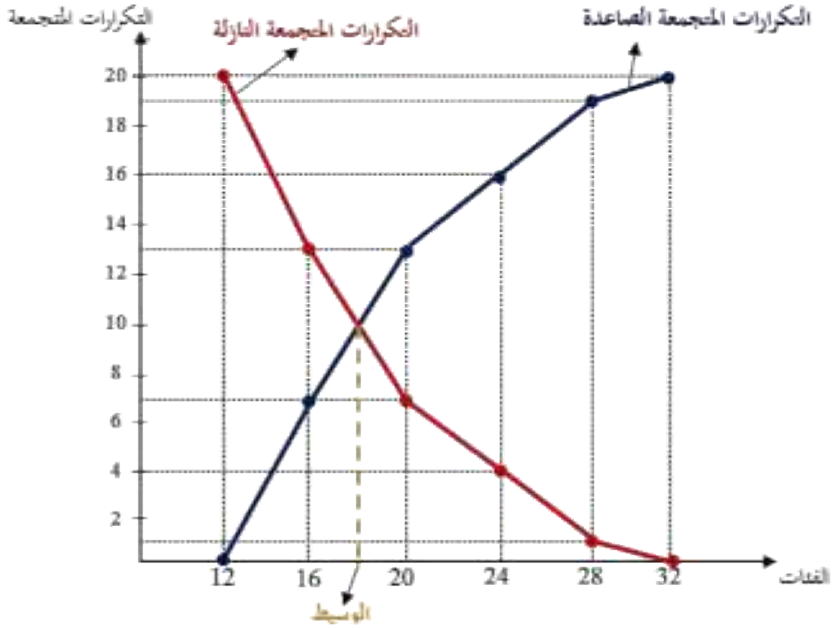
ويمكننا توضيح ذلك من المثال السابق.

الجدول رقم 1-03: النسبي المئوي والتكرار المتجمع الصاعد

الفئة	n_i	$n_i \uparrow$	$n_i \downarrow$
]16-12]	7	7	20
]20-16]	6	13	13
]24-20]	3	16	7
]28-24]	3	19	4
]32-28]	1	20	1
المجموع	**20**	-	-

ونقوم بتمثيل منحنى التكرارات المتجمعة الصاعدة والنازلة كما يلي:

الشكل رقم 1-04: منحنى التكرارات المتجمعة الصاعدة والنازلة



3.2.1. الدائرة، العمود المجزأ والأعمدة المستطيلة

للمتغيرات النوعية تمثيلات بيانية مختلفة أهمها:

- ***العرض الدائري***

مبدأ هذه الطريقة مبني على ترجمة بيانات الجدول (الأعداد أو النسب) إلى زوايا، حيث يتناسب فيها التكرار مع قيس الزاوية، ثم نقل نتائج الحسابات إلى شكل بياني متمثل في دائرة. وبالتالي نقوم بإضافة عمود إلى جدول المعطيات يحتوي على الزوايا المركزية المقابلة لكل تكرار.

ويتم حساب الزوايا المركزية باستخدام العلاقة التالية:

$$\alpha_i^{\circ} = \frac{n_i}{\sum n_i} * 360° = f_i * 360°$$

مثال 1-07:

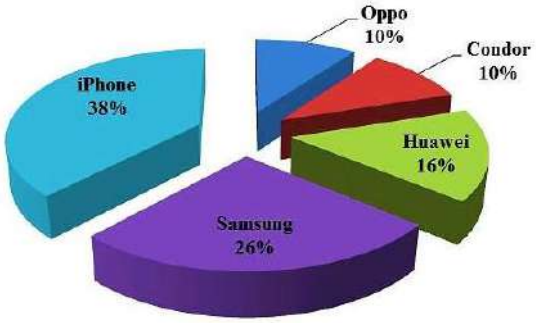
بالاستناد الى معطيات المثال 1-02 سنقوم بعرض البيانات باستخدام العرض الدائري.

الجدول رقم 1-01: توزيع الطلبة حسب نوع الهاتف الذكي

نوع الهاتف	iPhone	Samsung	Huawei	Condor	Oppo	المجموع
n_i	19	13	8	5	5	50
f_i (%)	38	26	16	10	10	100
α_i°	136.8°	93.6°	57.6°	36°	36°	360°

ويمكن تمثيل النتائج كما في الشكل الموالي:

العرض الدائري لتوزيع الطلبة حسب نوع الهاتف المستخدم

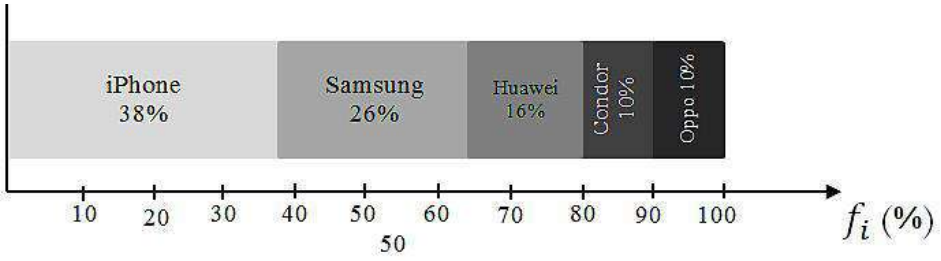


- ***العمود المجزأ***

هو عبارة عن مستطيل مقسم إلى عدة أجزاء، كل جزء منه يقابل تكرار معين للخاصية المدروسة. ، ومن الأفضل عند رسم العمود استعمال النسب المئوية المقابلة لكل تكرار حيث أن طول المستقيم يمثل 100%.

بالعودة الى معطيات المثال السابق نمثل المعطيات على شكل عمو مجزأ، كما يلي:

العمود المجزأ لتوزيع الطلبة حسب نوع الهاتف المستخدم

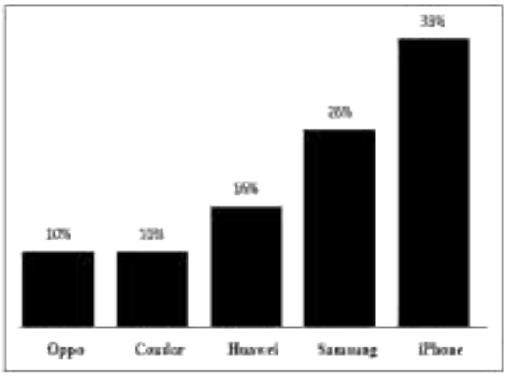


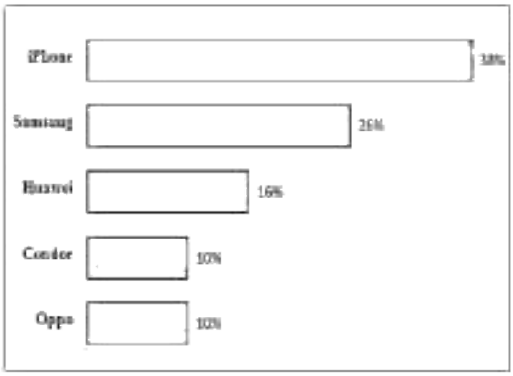
- ***الأعمدة المستطيلة***

هي عبارة عن مستطيلات متباعدة بمسافات ثابتة ولها قواعد متساوية تتناسب أطوالها مع التكرارات المقابلة لمكونات الخاصية المدروسة.

وباستخدام المعطيات السابقة يمكننا عرض الاعمدة المستطيلة، كما يلي:

الاعمدة المستطيلة لتوزيع الطلبة حسب نوع الهاتف المستخدم





4.2.1. تمثيلات بيانية اخرى

في علم الإحصاء الوصفي، هناك عدة أنواع التمثيلات البيانية التي يمكن أن نستخدمها في التحليل، نذكر منها:

- *مخطط الصندوق*

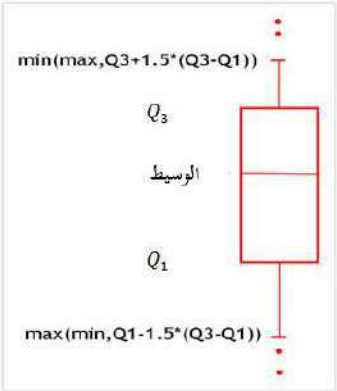
مخطط الصندوق (أو Box Plot) هو رسم بياني يوضح التوزيع الإحصائي لمجموعة البيانات، ويساعد هذا الرسم على عرض كيفية توزيع البيانات على طول خط الأعداد. من جهة أخرى، يمتاز هذا النوع من التمثيلات البيانية بأنه سهل الإعداد.

يمثل هذا المخطط، الذي قدمته J.W. Tukey,، تمثيلا بانيا عالي الكفاءة والذي يمثل بوضوح الخصائص الرئيسية للمتغير.

ويتوافق الصندوق مع الجزء المركزي من التوزيع، حيث أن نصف القيم بين الربعين الأول والثالث 1 وQ_3 (انظر الشكل الموالي). ويمتد المخطط من كلا جانبي الصندوق إلى القيم التالية:

- على اليسار حتى $Q_1 - 1.5(Q_3 - Q_1)$ إذا كانت هناك قيم أصغر وإلا يصل إلى الحد الأدنى للقيمة ؛
- على اليمين حتى $Q_1 + 1.5(Q_3 - Q_1)$ كانت هناك قيم أكبر، وإلا فستصل إلى القيمة القصوى.

الشكل رقم 1-05: شكل توضيحي لمخطط الصندوق



ملاحظة

تجدر الاشارة الى النقاط التالية:

- القيمة المركزية لمخطط الصندوق هي الوسيط، حيث هناك عدة قيم أكبر أو أقل من هذه القيمة في العينة.
- حواف الصندوق هي الربيعيات (بالنسبة للحافة السفلية، ربع الملاحظات لها قيم أصغر وثلاثة أرباع لها قيم أكبر ، والحافة العليا تتبع نفس المنطق).
- يتم حساب نهايات الصندوق (الحواف) باستخدام 1.5 أضعاف المساحة الربيعية (المسافة بين الربيع الأول والثالث).

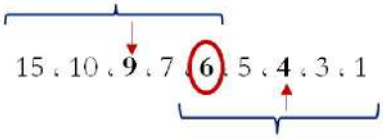
لتوضيح ذلك، نعرض مثال بسيط عن كيفية تمثيل المخطط.

مثال 1-08:

نعتبر سلسلة القيم التالية والمكونة من 9 عناصر:

$$15 \ ، 10 \ ، 9 \ ، 7 \ ، 6 \ ، 5 \ ، 4 \ ، 3 \ ، 1$$

الوسيط (والذي يمثل أيضا الربيع الثاني) Q_2 يقسم السلسلة الى مجموعتين $(Q_2 = 6)$. ويقسم الربيع الاول المجموعة الاولي(القيم الخمسة الأقل) الى قسمين، اذن $Q_1 = 4$. ويقسم الربيع الثالث المجموعة الثانية (القيم الخمسة الأكبر) الى قسمين، اذن $Q_3 = 9$.



ملاحظة

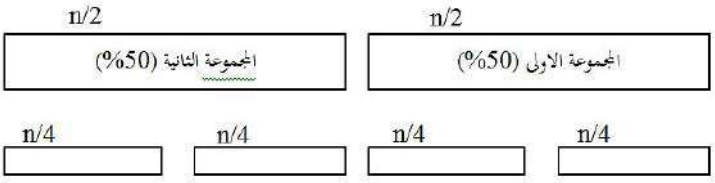
إن حساب الربيعيات يختلف اعتمادا على كون عدد القيم n زوجيا أم فرديا. وعموما نتبع الطريقة التالية:

- ترتيب المعطيات تصاعديا
- تقسيم المعطيات الى مجموعتين متساويتين

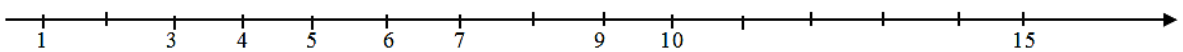
ومن ثم نحصل على مجموعة لكل منها 50% من الملاحظات، حيث: إذا كان عدد القيم فردي، فإن الوسيط يوافق القيمة الوسطى. أما إذا كان عدد القيم زوجي فان الوسيط هو متوسط النقطتين الوسطتين

ومن ثمة نقوم بحساب الوسيط لكل مجموعة، حيث أن:

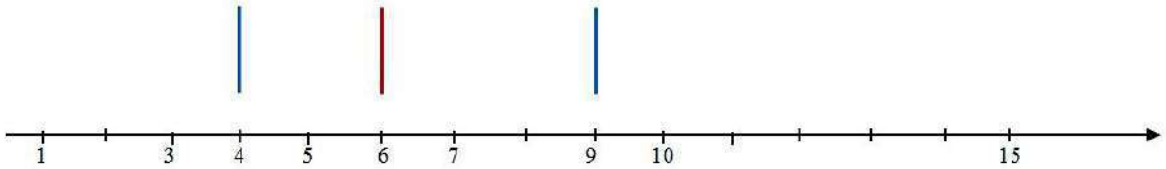
- وسيط المجموعة الاولى هو الربيع الاول (Q_1) والذي يمثل 25%من الملاحظات
- وسيط المجموعة الثانية هو الربيع الثالث (Q_3) والذي يمثل 75%من الملاحظات



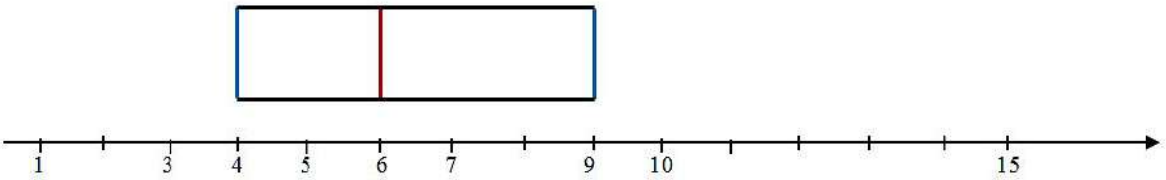
وحتى نتمكن من تمثيل مخطط الصندوق لا بد من ارسم خط التمركز أو خط العرض الإحصائي (خط الأعداد).



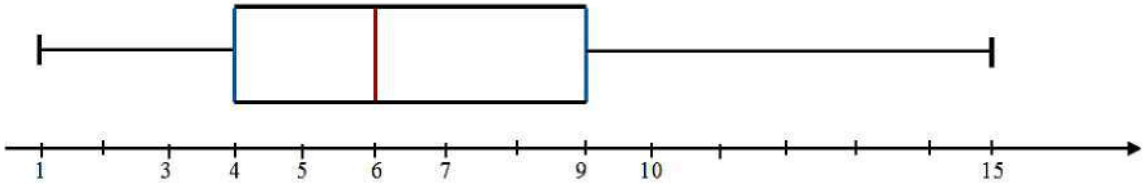
يجب أن يكون طوله مناسبا لعرض سلسلة البيانات، مع التأكد من إضافة جزء زائد على كلا الجانبين. يتم ترقيم هذا الخط بشرط التأكد من ترك مسافات متساوية بين كل رقم وآخر، وإذا كنا تتعامل مع كسور عشرية، فيجب الاخذ بعين الاعتبار ذلك وتحديدهم على خط التمركز.



ثم نقوم بتحديد الربيع الأول والثاني والثالث على خط التمركز بخط عمودي. ثم نقوم بتوصيل الحدود العليا للربيع الأول الحدود العليا للربيع الثالث مرورا بالربيع الثاني، ونوصل الحدود الدنيا بنفس الطريقة.



وبعدها نقوم بتحديد القيم المتطرفة هي القيمة الأصغر والقيمة الأكبر من بين الأعداد في سلسلة المعطيات، وهما: القيمة المتطرفة الأدنى هي1 والقيمة المتطرفة الأعلى هي 15.



ونكون بذلك قد اعددنا مخطط الصندوق.

3.1. تلخيص البيانات

بعد التطرق الى كيفية عرض البيانات الإحصائية وتلخيصها في جداول تكرارية أو رسوم بيانية، بهدف الحصول على بعض الخصائص للظاهرة محل الدراسة، سوف نستعرض فيما يلي أنواع مهمه من المقاييس الإحصائية وهي مقاييس النزعة المركزية ومقاييس *التشتت* (Bourbonnais, 2004).

حيث انه من المعروف أن التمثيلات البيانية تكون أقل دقة، لذلك يجب أن يكون لدينا مقاييس عددية تصف لنا المعطيات.

1.3.1. مقاييس النزعة المركزية

وهي عبارة عن مقاييس عددية تعين موقع التوزيع، وهى مهمة عموما للمقارنة بين التوزيعات المختلفة. ومن خلال ملاحظة البيانات الخاصة بأي ظاهرة سواء في صورها الأولية أو بعد تلخيصها وتبويبها في جداول توزيع تكراري نجد أن معظم مفرداتها تتمركز حول قيمة معينة، وهذه القيمة تمثل مركز التوزيع لذا فإن الحصول عليها ضروري ومهم في دراسة خصائص التوزيع والمقارنة بين التوزيعات التكرارية المختلفة لنفس الظاهرة،

لذلك يلجأ الباحثون بعد جمع البيانات عن الظاهرة المدروسة إلى تلخيص وتبويب البيانات في جداول توزيع تكراري ثم عرضها بيانيا، لتأتي بعدها مرحلة استخدام المقاييس الإحصائية على بيانات الظاهرة للتعبير عنها بقيمة أو أكثر تفيد في المقارنة والتنبؤ، وهذه المقاييس تسمى مقاييس النزعة المركزية.

بشكل عام ، من الضروري تلخيص سلسلة من الملاحظات من خلال مؤشرات نموذجية ، وأشهرها هو الوسط الحسابي. فيما يلي قائمة لمقاييس النزعة المركزية الأكثر استعمالا:

1. المتوسط الحسابي

يعتبر المتوسط الحسابي من أشهر وأكثر متوسطات النزعة المركزية استخداما وشيوعا في الإحصاء وهو مركز التوازن لأي ظاهرة. وهو عبارة عن حاصل قسمة مجموع قيم المعطيات على عددها ويرمز به بالرمز $\bar{x}$ ، حيث:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. الوسيط

عند ترتيب المعطيات (المشاهدات) ترتيبا تصاعديا أو تنازليا فإن الوسيط يكون البيان (المشاهدة) التي يقع 50% من البيانات قبلها في الترتيب و 50% من البيانات بعدها في الترتيب. وإذا كان عدد البيانات فرديا فإن الوسيط يكون المشاهدة التي تقع في المنتصف، وإذا كان عدد البيانات زوجيا فإن الوسيط هو متوسط المشاهدتين اللتين تقعان في المنتصف.

ويمكن تقدير قيمة الوسيط بيانيا من المنحنى المتجمع الصاعد، حيث أن اسقاط نقطة تقاطع المنحنيين على محور السينات توافق قيمة الوسيط.

3. المنوال

تعبر قيمة المنوال عن المشاهدة الأكثر تكرارا فهو بمثابة القيمة الشائعة، وقد يكون للبيانات منوال واحد ويمكن أن يكون لها أكثر من منوال، كما يمكن أن لا يوجد منوال. ويعتبر المنوال أفضل مقياس لوصف البيانات النوعية.

ولتوضيح كيفية حساب المقاييس السابقة نستخدم معطيات المثال التالي:

مثال 1-09:

نعتبر المشاهدات التالية:

$$4، 6، 8، 9، 3$$

- حساب المتوسط الحسابي

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{4+6+8+9+3}{5} = \frac{30}{5} \quad ; \quad \bar{x} = 6$$

- حساب الوسيط

لحساب الوسيط نقوم بترتيب البيانات تصاعديا كما يلي:

$$3، 4، 6، 8، 9$$

بما أن عدد القيم فردي($n = 5$) فان رتبة الوسيط تساوي:

$$\frac{n+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$$

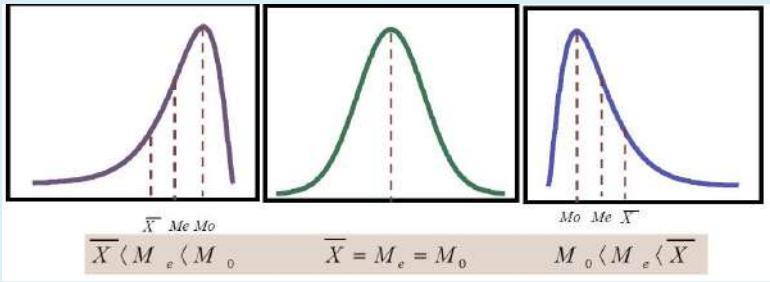
وبالتالي قيمة الوسيط هي القيمة التي ترتيبها يقابل رتبة الوسيط: $M_e = 6$

- حساب المنوال

لا يوجد في هذه السلسلة أي قيمة تكررت أكثر من مرة وعليه فإنه لا يوجد منوال لهذه المعطيات.

ملاحظة

توجد علاقة تجريبية بين المقاييس الثلاثة (الوسط الحسابي، الوسيط، المنوال) وذلك في حالة التوزيعات التكرارية أحادية المنوال وغير المتماثلة والمتماثلة وذات الالتواء البسيط. ويمكن توضيح هذه العلاقة من خلال الاشكال الموالية:



2.3.1. مقاييس التشتت

مقاييس النزعة المركزية لا تكفي لإعطاء صورة كاملة عن علاقة البيانات ببعضها البعض، فقد نجد أن لسلسلتين مختلفتين نفس المتوسط الحسابي، بينما مدى البيانات للسلسلتين مختلف، هذا الفرق في البيانات مقابل تساوي المتوسط الحسابي يجعل من الضرورة استخدام مقاييس أخرى تكمل المقاييس الأولى، وتسمى هذه المقاييس بمقاييس التشتت (معتوق، 2007).

مقاييس التشتت هي عبارة عن مقاييس إحصائية هدفها قياس مدى تشتت و تباعد البيانات عن بعضها البعض.

فمثلا لو توفرت لدينا السلسلتان الإحصائيتان التاليتان:

السلسلة الاولى: 0، 5، 10

السلسلة الثانية: 3، 5، 7

وبحساب المتوسط الحسابي لكلتا السلسلتين نجد انه يساوي 5، فإذا اكتفينا بالمتوسط الحسابي كمقياس فاننا نقر ان السلسلتين متشابهتين. ولكن في الحقيقة إن قيم السلسلة الأولى أكثر تباعد من قيم السلسلة الثانية، وهنا يأتي دور مقاييس التشتت أو الاختلاف لتضيف هذه الناحية في البيانات الإحصائية.

والتشتت غالبا ما يقصد به درجة أو مقدار التفاوت أو الاختلاف بين مفردات الظاهرة المدروسة، وتعتبر معطيات الظاهرة متجانسة عندما تكون قيمها قريبة من بعضها البعض ونقول في هذه الحالة أنها غير مشتتة. أما إذا كانت بيانات الظاهرة متباعدة وغير متجانسة فنقول أن قيم الظاهرة *مشتتة* (Saporta, 2011).

وهناك العديد من المقاييس التي يمكن استخدامها لقياس مدى تقارب أو تباعد القيم عن بعضها البعض، زمن أهمها:

1. التباين

هو عبارة عن المتوسط الحسابي لمربعات الفروق بين قيم المتغير الإحصائي ومتوسطها الحسابي، ونستخدم مربعات الفروق هنا تفاديا لاستخدام القيم المطلقة. ويرمز للتباين بالرمز σ^2 أو $V(X)$ ، ويحسب بالصيغة التالية:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

لا يوجد في هذه السلسلة أي قيمة تكررت أكثر من مرة وعليه فإنه لا يوجد منوال لهذه المعطيات.

ملاحظة

أحيانا يطرح العدد واحد من المفردات (n–1) أو من مجموع التكرارات وذلك عند تقدير مقياس التباين والانحراف المعياري للعينات الصغيرة الحجم وأن العدد واحد يمثل درجات الحرية. وتعتبر العينة كبيرة إذا زاد حجمها عن 30 مشاهدة.

2. الانحراف المعياري

الانحراف المعياري هو الجذر التربيعي للتباين ، ويعتبر من أهم المقاييس الإحصائية للتشتت وأكثرها استخداما في النظريات والقوانين الإحصائية، ويرمز له بالرمز σ.

ولتوضيح كيفية حساب التباين والانحراف المعياري نأخذ المثال التالي:

مثال 1–10:

سنقوم بحساب الانحراف المعياري والتباين لعلامات مجموعة من الطلاب مقياس تحليل المعطيات.

11 ،9، 14، 16، 5

نقوم بإجراء الحسابات الضرورية في الجدول الموالي:

n	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	11	(11 - 11) = 0	$(0)^2 = 0$
2	9	(9 - 11) = -2	$(-2)^2 = 4$
3	14	(14 - 11) = 3	$(3)^2 = 9$
4	16	(16 - 11) = 5	$(5)^2 = 25$
5	5	(5 - 11) = -6	$(-6)^2 = 36$
المجموع	**55**	**0**	**74**

- ايجاد قيمة المتوسط الحسابي

$$\bar{x}\frac{1}{n} = \sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{11+9+14+16+5}{5} = \frac{55}{5} \quad ; \quad \bar{x} = 11$$

- حساب التباين

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{74}{5} = 14.8$$
$$\sigma^2 = 14.8$$

- حساب الانحراف المعياري

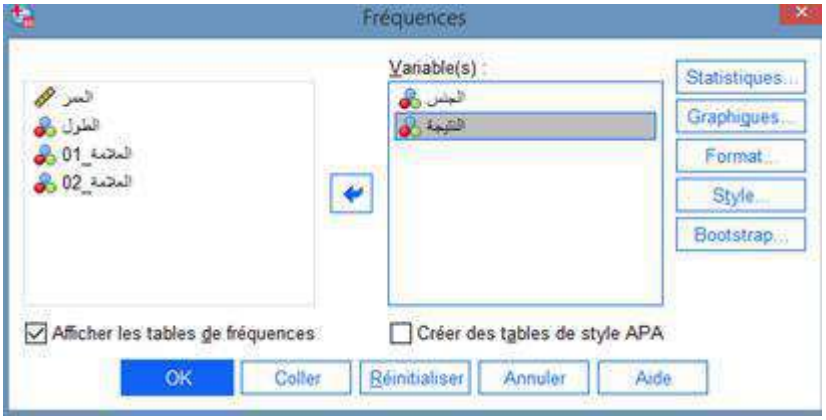
$$\sigma = \sqrt{\sigma^2} = \sqrt{14.8}$$
$$\sigma = 3.84$$

4.1. تمرين تطبيقي باستخدام SPSS

من بيانات الجدول التالي، قم بإجراء تحليل إحصائي مع حساب مختلف المقاييس الإحصائية (النزعة المركزية، التشتت) التكرارات، ثم مثل المعطيات بيانيا.

الجنس	انثى	ذكر	انثى	انثى	ذكر	انثى	ذكر	ذكر	ذكر	انثى	انثى	ذكر	ذكر	انثى	انثى
العمر	25	30	28	27	37	29	31	28	38	28	32	35	28	21	21
الطول	1.82	1.91	1.77	1.82	1.8	1.82	1.79	1.89	1.87	1.79	1.81	1.88	1.70	1.72	1.72
العلامة 1	6	5	5	7	3	5	7	5	6	4	6	7	4	3	5
العلامة 2	9	7	4	5	4	8	6	5	5	4	7	6	2	4	4
النتيجة	ناجح	ناجح	راسب	ناجح	راسب	ناجح	ناجح	ناجح	ناجح	راسب	ناجح	ناجح	راسب	راسب	راسب

سنعتمد على برنامج SPSS للتحليل الوصفي لبيانات الجدول. لذا في قائمة Analyse سنختار Fréquences (انظر الملحق رقم 4 للتعرف على كيفية اجراء التحليل الوصفي) من اجل الحصول على جداول تكرارية مع بعض التمثيلات البيانية وذلك بإدخال متغيرات نوعية هي: الجنس والنتيجة، كما يلي:



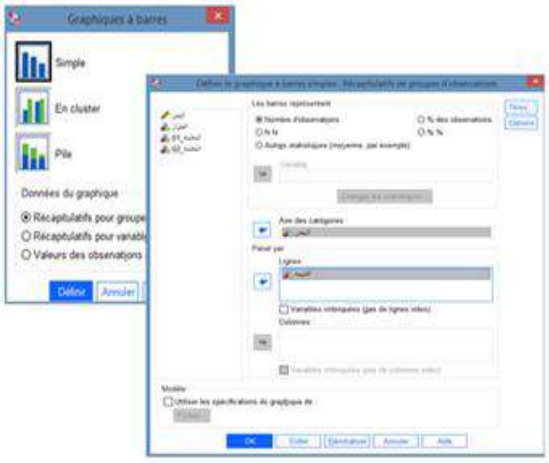
وبالضغط على OK يمكننا الحصول على جدولي التكرارات التاليين:

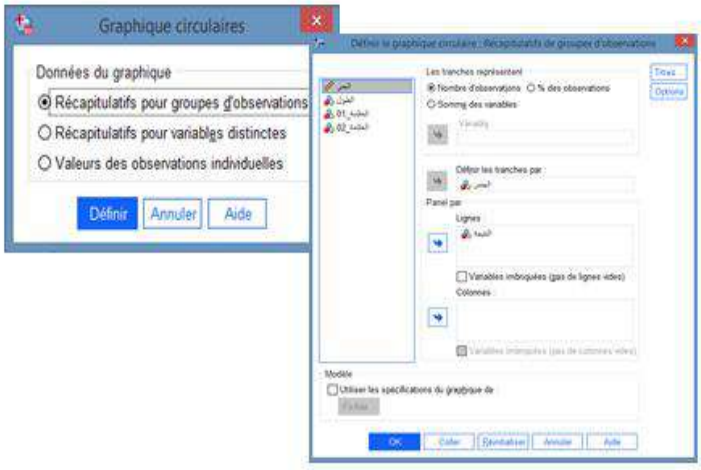
الجنس					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	ذكر	7	46,7	46,7	46,7
	انثى	8	53,3	53,3	100,0
	Total	15	100,0	100,0	

النتيجة					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	ناجح	9	60,0	60,0	60,0
	راسب	6	40,0	40,0	100,0
	Total	15	100,0	100,0	

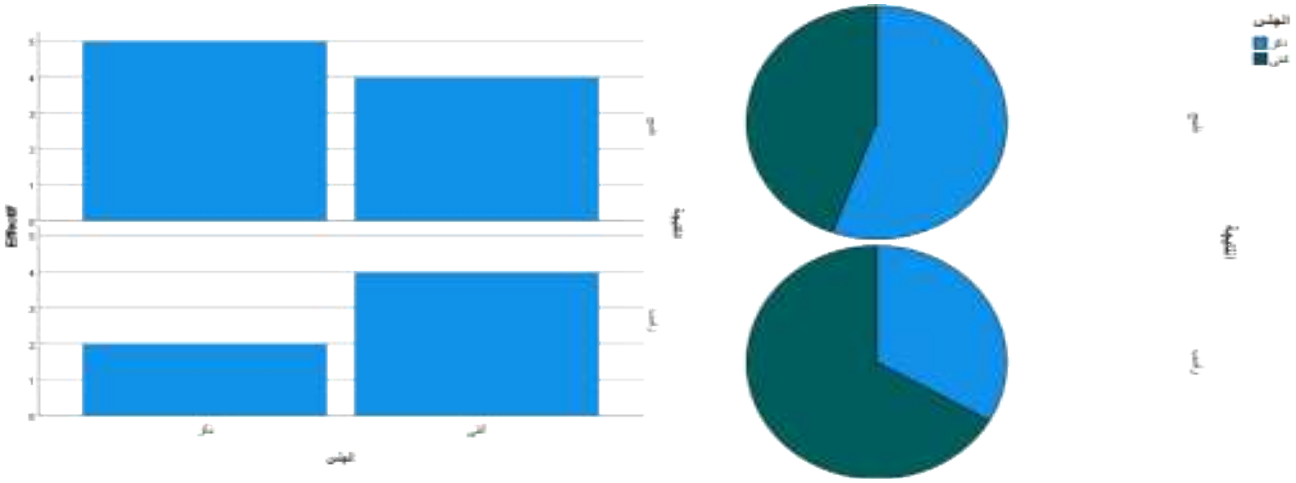
من الجدولين نلاحظ أن 53.3% من العينة هم اناث بينما 46.7% ذكور. كما ان 60% من المشاركين ناجحين.

يمكننا تمثيل ذلك بيانيا، لذا من القائمة الرئيسية Graphiques نختار Barres وذلك بعد تنشيط Simple لمجموعات الحالة، كما تبينه الصورة الموالية على يسار الصورة، علما انه يتم تنشيطه أوتوماتيكيا في SPSS (Carricano et al., 2010). يمكننا ايضا تمثيل البيانات عن طريق الدائرة وذلك من خلال اختيار Circulairesفي القائمة في مربع الحوار، وفي مربع الحوار الناتج نقوم بتفعيل Récapitulatifs pour groupes d'observations ثم Définir.





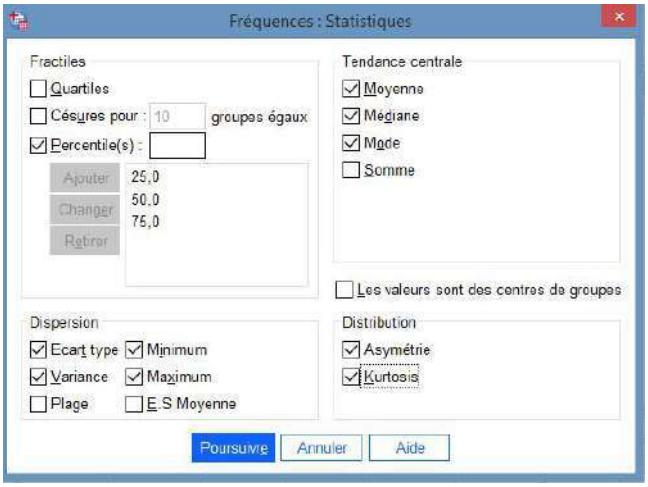
وبالضغط على OK نحصل على التمثيلين البيانيين التاليين:



نشير هنا الى انه في محرر الرسم البياني الذي يتم عرضه بالنقر المزدوج على الرسم البياني في SPSS هناك العديد من الأزرار التي تسمح بخيارات تحرير مختلفة، فمثلا يمكننا تغيير اللون، وتغيير خط الكتابة ، وتحريك أساطير المحور، وغيرها.

الان باستخدام المتغيرات الكمية سنقزم بإنشاء جداول تكرارية ، تمثيلات رسومية (مدرج تكراري) ، وحساب مقاييس النزعة المركزية، والتشتت.

من قائمة Analyse نختار Descriptives ثم Fréquences وبعدها نقوم بإدخال المتغيرات الكمية في نافذة المتغيرات ثم نختار نضغط على Statistiques ونقوم بتفعيل مقاييس النزعة المركزية ومقاييس التشتت، والشكل، ...الخ، كما في الصورة الموالية:



ثم نضغط على Poursuivre ثم على OK ونحصل على النتائج التالية:

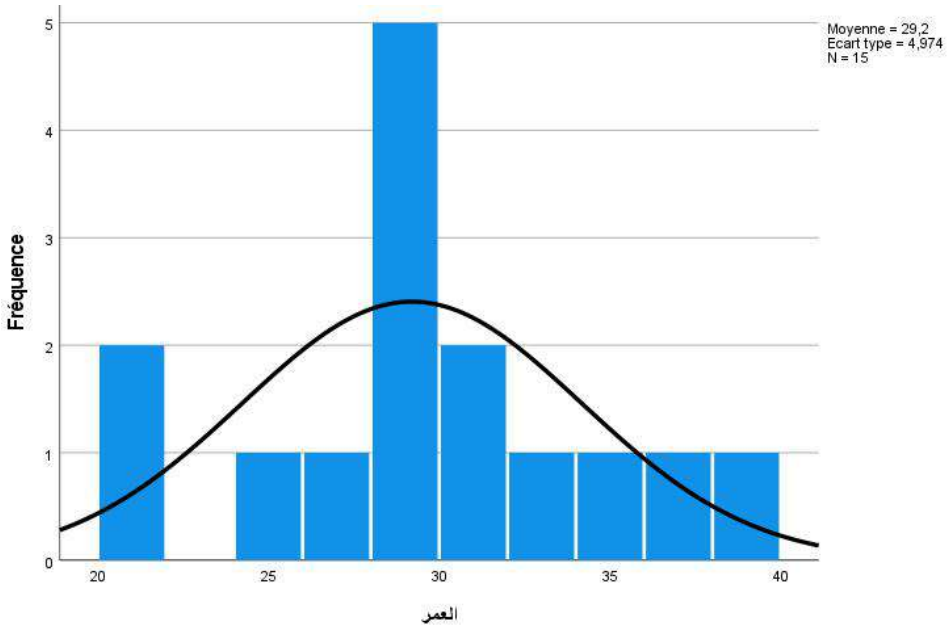
Statistiques					
		العمر	01_العلامة	02_العلامة	الطول
N	Valide	15	15	15	15
	Manquant	0	0	0	0
Moyenne		29,20	5,20	5,33	180,73
Médiane		28,00	5,00	5,00	181,00
Mode		28	5	4	182
Ecart type		4,974	1,320	1,839	6,319
Variance		24,743	1,743	3,381	39,924
Asymétrie		,124	-,206	,386	-,114
Erreur standard d'asymétrie		,580	,580	,580	,580
Kurtosis		-,128	-,740	-,054	-,650
Erreur standard d'aplatissement		1,121	1,121	1,121	1,121
Minimum		21	3	2	170
Maximum		38	7	9	191
Percentiles	25	27,00	4,00	4,00	177,00
	50	28,00	5,00	5,00	181,00
	75	32,00	6,00	7,00	187,00

من بين مخرجات SPSS ايضا نجد جداول التكرار لكل متغير وننقل مثال لهذه الجداول:

01_العلامة					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	3	2	13,3	13,3	13,3
	4	2	13,3	13,3	26,7
	5	5	33,3	33,3	60,0
	6	3	20,0	20,0	80,0
	7	3	20,0	20,0	100,0
	Total	15	100,0	100,0	

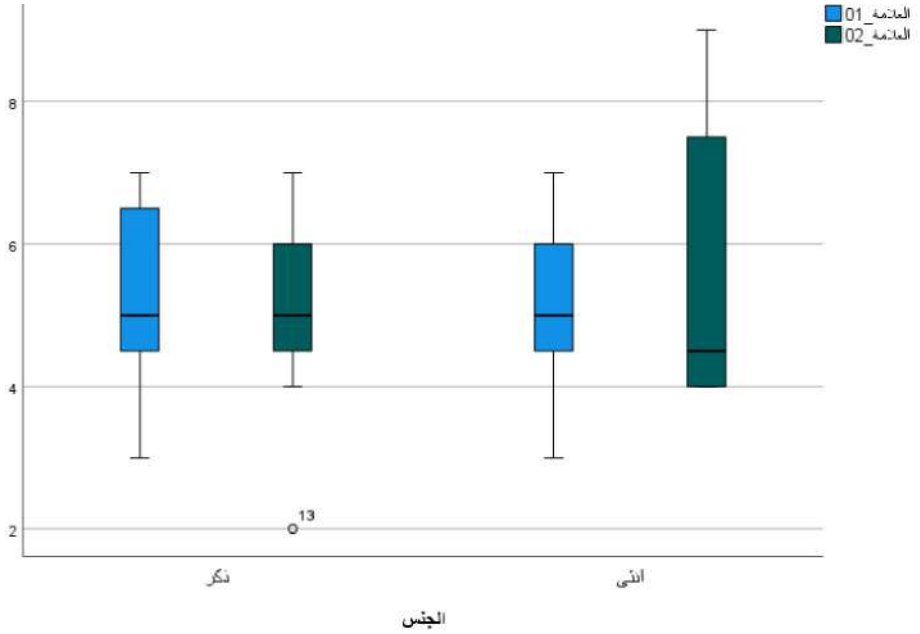
والان من قائمة Graphiques سنختار Histogramme وندرج متغير العمر مع تفعيل Afficher la courbe gaussienne ، ثم نضغط على OK ذلك كما يلي:

يمكن أيضا اجراء هذه العملية من خلال قائمة Analyse ثم اختيار Descriptives ثم Fréquences ثم تفعيل الرسم البياني.



يمكننا ايضا انشاء مخطط الصندوق لإجمالي عدد العناصر وتصنيفها حسب الجنس للمتغيرين التاليين: العلامة 1 والعلامة 2. للحصول على مخطط الصندوق، نختار في قائمة Graphiques التمثيل البياني Boites à moustaches ثم نختار تنفيذ مخطط الصندوق البسيط (Simple) وندرج المتغيرين المختارين في المجموعة للمجموعة ثم نضغط على OK.

ونحصل على التمثيل البياني التالي:



الفصـل الثـاني

التـحـليـل العامـلي

مقدمة الفصل

ينطبق التحليل الإحصائي الوصفي على الظواهر التي ندرس فيها متغير واحد أو اثنين على مجموعة من الأفراد أو الوحدات الاحصائية. وكما راينا في الفصل السابق فان تحليل مختلف الخصائص الوصفية للظاهرة تعتبر مرحلة أساسية في عملية تحليل المعطيات ولكنها غير كافية. فغالبا ما يكون من الضروري تحليل المعطيات مع مراعاة طبيعتها المتعددة الأبعاد. لان في معظم التطبيقات، نلاحظ وجود عددا كبيرا من المتغيرات، وليس متغير واحد فقط، لكل فرد.

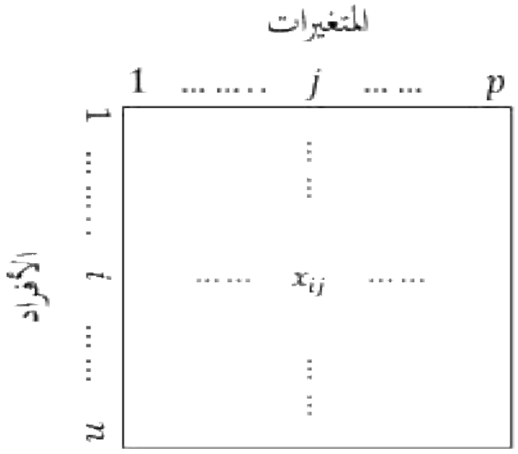
ويعد التحليل العاملي من أهم الطرق المستعملة بشكل خاص لاستكشاف بنية هذه المعطيات. كما أنها تعتبر أساس معظم الطرق الوصفية متعددة الأبعاد. فقد ساهمت مختلف تقنيات التحليل العامل، باعتبارها أحد الأساليب الرياضية الهامة، في تنظيم وتصنيف الظواهر في عدة مجالات. ومع التقدم الملموس في البرامج المتخصصة للتحليل الاحصائي، ونتيجة لتطور طرق تحليل المعطيات واتساع استخداماتها أصبح التحليل العاملي من الاساليب البالغة الاهمية في البحوث العلمية.

لذا إن لم يكن الطالب ملما بأساليب التحليل العاملي فإنه لا يستطيع التعرف على المشكلة التي يصلح فيها استخدام هذا النوع من التحليل الاحصائي ، فستخرج له نتائج ضخمه لن يستطيع تفسيرها واستخلاص دلالتها من الاطار النظري الاساسي والمنطق العلمي الذي انطلق منه.

1.2. الاطار العام للتحليل العاملي

غالبا ما تتضمن أساليب تحليل المعطيات تنظيما معينا للمعطيات يتمثل في شكل يسمح بجميع هذه المعطيات وترتيبها في جدول كما سبق ان تطرقنا اليه في الفصل السابق. وبالتالي فإن فكرة جدولة المعطيات لا تزال موجودة في جميع طرق التحليل الاحصائي للمعطيات.

الشكل رقم 2-01: التمثيل الجدولي للمعطيات



حيث نقوم بتمثيل الأفراد أو الوحدات الإحصائية في الصفوف ونقوم بوصف كل فرد بخاصية تعرف بالمتغيرات

ملاحظة

وسوف نحتفظ بالتقييمات كما في الجدول السابق، حيث تمثل x_{ij} قيمة المتغير j بالنسبة للفرد i، مع:

$$i = 1, \dots n \quad \text{و} \quad k = 1, \dots p$$

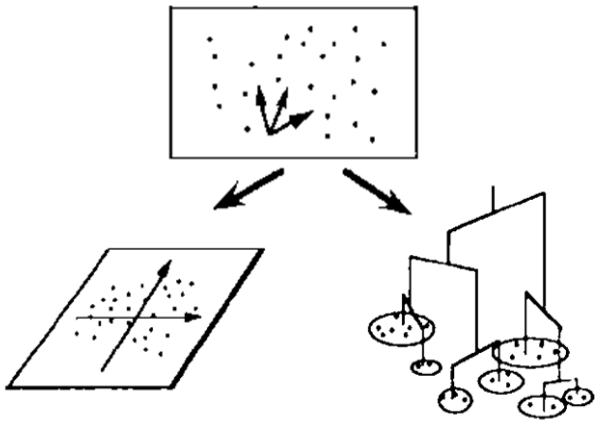
وذلك حتى نتمكن من تبسيط العلاقات. إذن n تمثل في نفس الوقت عدد الأفراد ومجموع الأفراد ($\{1, \dots i, \dots . n\}$). وبنفس الطريقة فإن p تمثل عدد المتغيرات ومجموع المتغيرات ($\{1, \dots j, \dots . p\}$).

يمكن أن يسهل هذا العرض الجدولي للمعطيات قراءة البيانات التي تم جمعها. وتكون هذه القراءة أكثر سهولة عندما تكون لدينا كمية قليلة من المعطيات. ولكن في الكثير من الاحيان يكون حجم الجدول كبير، ويحدث ذلك غالبا عندما نبحث عن علاقات بين أكثر من عنصرين احصائيين أو أكثر من متغيرين. في هذه الحالة فإن هذا التمثيل والتقنيات البسيطة للإحصاء الوصفي ليست كافية للتحليل.

يعتبر التحليل العاملي من الاساليب الإحصائية التي تستخدم في دراسة الظواهر بهدف إرجاعها إلى العوامل المؤثرة فيها، وهو عملية رياضية تستهدف تفسير معاملات الارتباط ذات الدلالة الاحصائية بين مختلف المتغيرات (Bouroche et Saporta, 1994). والهدف الرئيسي هو تبسيط الارتباطات بين مختلف المتغيرات التي تدخل في التحليل وصولا الى العوامل المشتركة التي تصنف العلاقة بين هذه المتغيرات وتفسرها.

1.1.2. تعريف

توفر طرق التحليل العاملي تحليلا مفصلا لمجموعات كبيرة من القيم العددية، تكون عادة في شكل رسومات بيانية. وللقيام بذلك ، نحاول تقليل أبعاد جدول البيانات من خلال تمثيل الارتباطات (أو التبعيات) بين الأفراد وبين المتغيرات على هذه البيانات.



ومنه يمكن تعريف التحليل العاملي على أنه:

"الاسلوب الإحصائي الذي يستخدم في تحليل معطيات متعددة ارتبطت فيما بينها بدرجات مختلفة من الارتباط ، لتلخص في صورة تصنيفات مستقلة قائمة على أسس نوعية للتصنيف" .

ويتولى الباحث فحص هذه الأسس التصنيفية واستشفاف ما بينها من خصائص مشتركة وفقا للإطار النظري والمنطق العلمي الذى بدأ به. ويمكن تعريفه أيضا على أنه:

"مجموعة من الأساليب الإحصائية، التي تهدف إلي تخفيض عدد المتغيرات أو المعطيات المتعلقة بظاهرة معينة ".

أي انه يعبر عن مجموعة من العمليات الرياضية بغرض تبسيط الارتباطات بين مختلف المتغيرات الداخلة في التحليل وصولا إلى العوامل المشتركة التي تصف العلاقة بين هذه المتغيرات وتفسيرها

وهو يعبر عن طريقة إحصائية متعددة المتغيرات تستخدم في تحليل البيانات أو مصفوفات الارتباط ، أو مصفوفات التباينات للمتغيرات وحواصل ضربها. ويكون الهدف هو توضيح العلاقات بين تلك المتغيرات ينتج عنها عدد من المتغيرات الجديدة أو المفترضة تسمى بالعوامل.

2.1.2. ماهية العوامل

إن العوامل الناتجة عن عملية التحليل العاملي هي عبارة عن متغيرات مثل المتغيرات الأخرى، ولكن مع وجود فرق بسيط هو ان جل المتغيرات يمكن قياسها بشكل مباشر، أما العوامل فهي متغيرات افتراضية أو متغيرات كامنة مشتقة من مجموعة من متغيرات تم قياسها قياسا مباشرا، ومعنى ذلك أن العوامل تنبع من داخل مجموعة من العالقات بين المتغيرات.

والعنصر الرئيسي لتحليل المعطيات هو جدول المعطيات. تعتمد طبيعة هذا الجدول على طبيعة المتغيرات التي يتكون منها. فالفرد هو عنصر في مجموعة محدودة تسمى مجموعة الأفراد، يتم وصفهم بواسطة المتغيرات.

ببساطة يقوم التحليل العاملي على دراسة علاقات بين عدد من المتغيرات:

$$x_1, x_2, x_3 \dots x_p$$

بدلالة مجموعة من العوامل:

$$F_1, F_2 \dots F_X$$

تسمى عوامل مشتركة وهي التي نريدها ونعتمد عليها.

وهذه العوامل تعتمد في تركيبها على أسس إحصائية مثل االإنحدار، ... ويكون عددها أقل من عدد المتغيرات الأصلية ($X < p$) وتساعدنا على فهم طبيعة العلاقات بين المتغيرات الاصلية.

كما راينا في الفصل السابق، تعرض المعطيات في شكل جداول، تحتوي على:

- صف i ($i = 1, \dots n$) والذي يمثل n فرد، كأسماء الكلبة مثلا.
- عمود j ($j = 1, \dots p$) والذي يمثل p متغير، والتي التي يمكن أن تكون علامات لمقاييس يدرسها الطلبة (الأفراد).

ومن أجل فهم مبدأ الأساليب الإحصائية القائمة على تحليل أبعاد العوامل، من الجيد أن نمثل هندسيا الانتشار النقطي لمجموعة من الأفراد n (الصفوف) ومجموعة المتغيرات p (الأعمدة). حيث كل مجموعة تصف المجموعة الاخرى. ثم نقوم بعد ذلك بتحديد المسافات بين نقاط الخطوط ونقاط الأعمدة التي تترجم الارتباطات بين الأفراد والمتغيرات.

مثال 2-01:

يمثل الجدول التالي علامات ستة طلاب في ثلاث مقاييس

جدول رقم 2-01: علامات الطلاب في ثلاث مقاييس

الطالب / المقياس	الاقتصاد	الرياضيات	الاحصاء
1	08	11	11
2	10	11	12
3	13	12	13
4	10	09	13
5	12	09	05
6	09	08	08

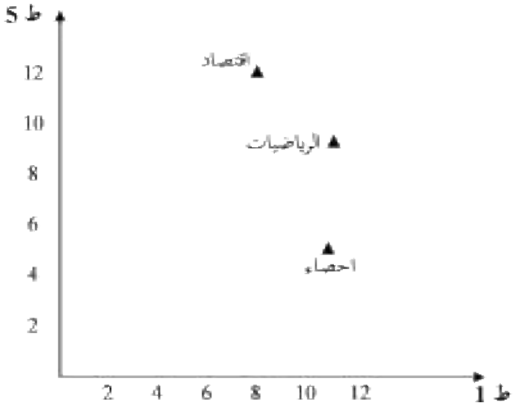
$$n = 6$$
$$p = 3$$

من وجهة نظر هندسية، فان المقياس (أو المتغير) هو نقطة يكون إحداثياتها هي العلامات التي تحصل عليها الطلبة (الافراد). على نحو مماثل، الفرد هو نقطة تكون إحداثياتها هي الملاحظات المخصصة للمقاييس.

ويمكن توضيح ذلك في الشكل الموالي مع تمثيل نقاط الانتشار للطالبين الأول والخامس، انطلاقا من معطيات الجدول السابق.

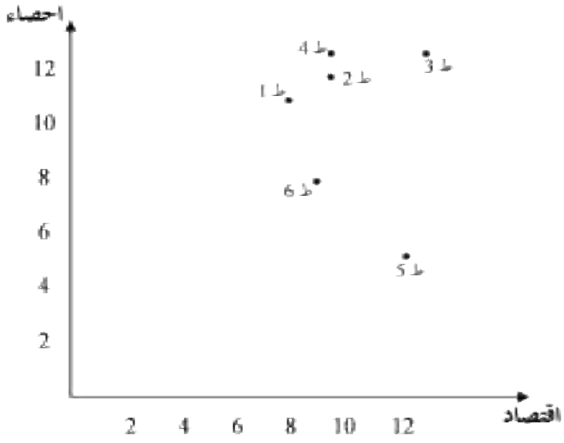
الشكل رقم 2-01: تمثيل السحابة النقطية للعلامات في فضاء الطالبين الاول والخامس

الطالب / المقياس	الاقتصاد	الرياضيات	الاحصاء
1	08	11	11
5	12	09	5



بنفس الطريقة، نقوم بإنشاء سحابة الطلبة الستة في فضاء (فضاء ذي بعدين)، المقياسين الاقتصاد والاحصاء، كما يلي:

الطالب / المقياس	الاقتصاد	الاحصاء
1	08	11
2	10	12
3	13	13
4	10	13
5	12	05
6	09	08



والتحليل العاملي يبدأ بحساب معاملات الارتباط بين عدد من المتغيرات مثل العلامات في مثالنا، وعند الحصول على مصفوفة الارتباطات بين هذه المتغيرات لعينة البحث التي اجريت عليها الدراسة، نقوم بتحليل تحليل هذه المصفوفة تحليلا عامليا بغرض الحصول على أقل عدد ممكن من العوامل (أو المحاور) والتي تمكننا من التعبير عن أكبر قدر من التباين بين هذه المتغيرات.

بطريقة أخرى ،ساعدنا هذه العملية على تحليل مجموعة من معاملات الارتباط بين عدة متغيرات واختزالها إلى عدد أقل من العوامل، أي يساعد على فهم تركيب مصفوفة الارتباط أو التباين المشترك من خلال عدد أقل من العوامل.

3.1.2. أهداف التحليل العاملي

تشكل اساليب التحليل العاملي أهم اساليب تحليلات المعطيات. وهي تستند إلى مبدأ واحد وهو تقليص المتغيرات الى عوامل، وهذا هو السبب الذي يجعلنا نتحدث عن التحليل العاملي (Escoffier et Pagès,

1998). يعتمد هذا المبدأ على حقيقة أن السحابتين اللتان تمثلان الصفوف والأعمدة في الجدول المدروس على التوالي يتم بناؤها وتمثيلهما على الرسوم البيانية. و تساعد هذه الأشكال على إجراء تحليل أكثر سهولة للصفوف والأعمدة المرتبطة ببعضها البعض.

وحسب Cattell (1952) فإن هدف المنهج العلمي هو اكتشاف الحقائق والعلاقة المتغيرات والظواهر، واكتشاف القوانين التنبؤية. من هذا المنطلق يمكننا أن نقول أن التحليل العاملي يعتبر منهج كلي يهدف إلى اكتشاف العموميات الاساسية، الوظيفية والعضوية، وبالتالي فهو وسيلة من وسائل التبسيط والتحليل.

الغرض الرئيسي من هذا النوع من الاساليب هو المساعدة في فهم وتحليل كميات كبيرة من المعطيات. تقليص الابعاد، الحصول على تمثيلات بيانية مثلى ، البحث عن عوامل أو متغيرات كامنة ...

وبمعنى أدق حدد الفرضيات واختباراتها، وذلك بتحديد أصغر عدد من العوامل المحددة التي يمكن أن تفسر العلاقات التي نلاحظها بين عدد كبير من الظواهر الواقعية وإلى أي مدى يؤثر كل من هذه العوامل في كل متغير.

و لقد بين أيزنك (Eyzanck, 1953) أن للتحليل العاملي ثلاثة أهداف أساسية مرتبطة بثلاث عناصر هي: طبيعة العوامل، واختلاف طرق التحليل (عدد كبير من طرق استخراج العوامل) والتدوير ، وهى الأهداف ذاتها لأي فرع من فروع الاحصاء وهى :

1. الوصف .
2. اختبار فرضيات الدراسة
3. اقتراح فرضيات انطلاقا من المعطيات الأولية

وقد ناقش العديد من الباحثين استخدامات التحليل العاملي على هذه المستويات الثلاثة، مع تعريف العامل في كل مستوى. فبالنسبة للهدف الأول فإن العامل احصاء مختصر يهدف إلى اقتصاد في الوصف، ويصف علاقات مستقيمة بين مجموعة من المتغيرات، ولا يتضمن العامل تحديدا لأي معنى سيكولوجي أو أسباب، ولا يقترح فروضا أو يثبتها.

وعموما يمكن تلخيص مختلف الاهداف في النقاط التالية:

- تلخيص المتغيرات الى عدد أقل من العوامل الرئيسية التي يمكن أن تفسر الظاهرة.
- إبراز مجموعة العناصر الكامنة التي يصعب الكشف عنها والتي يمكن أن يكون لها دور في تفسير العلاقات بين عدد كبير من المتغيرات.
- الحصول على مجموعة جديدة من المتغيرات (العوامل) وبعدد أقل لتحل جزئيا أو كليا محل المتغيرات.
- التعرف على المتغيرات التي لها دلالة إحصائية هامة والتي تتطلب مزيدا من عمليات التحليل الأخرى كالانحدار.
- يعتبر التحليل العاملي من الاساليب التي تساعد في تقليص العلاقات المعقدة بين مجموعة من المتغيرات إلى صورة خطية بسيطة نسبيا

- الكشف عن العالقات غير المتوقعة.
- حل مشكلة المتغيرات التفسيرية مثل مشكلة الارتباطات العالية بين المتغيرات المستقلة والتي تؤدي إلى عدم ثبات قيم معاملاتها الانحدارية المعيارية في تحليل الانحدار.

وانطلاقا من كل ما سبق يمكننا القول أن أوضح وظيفة للتحليل العاملي تتمثل في خفض أو اختزال مكونات جداول الارتباطات إلى اقل عدد ممكن ليسهل تفسيرها.

4.1.2. شروط التحليل العاملي ومجالات تطبيقه

تحاول اساليب التحليل العاملي للإجابة على السؤال التالي: مع الأخذ في الاعتبار أوجه التشابه بين الأفراد والروابط بين المتغيرات ، هل من الممكن تلخيص جميع البيانات من خلال عدد صغير من القيم دون فقدان المعلومات الهامة؟ في الواقع من خلال السعي لتقليل عدد المتغيرات التي تصف البيانات، لا يمكن تقليل كمية المعلومات إلا في أفضل الأحوال. الدافع وراء هذا الخفض في عدد القيم هو أن القليل من القيم يسهل تمثيلها هندسيا وبيانيا (أحد أغراض تحليل البيانات).

لذا يتطلب استخدام التحليل العاملي بعض الافتراضات والشروط التي يجب أن تتوفر في المعطيات المطلوب تحليلها وكيفية معالجتها، ويمكن تقسيم هذه الشروط الى قسمين هما:

أ. الشروط العامة

ومن أهمها:

- يشترط أن تكون المتغيرات موزعة توزيعا طبيعيا
- ينبغي ألا تكون العينة صغيرة الحجم او غير ممثلة للمجتمع المستهدف، وألا تكون متحيزة أيضا
- يفضل تجنب استخدام متغيرات غير مستقلة من الوجهة التجريبية والمتغيرات التي لا تتميز بالبساطة بالتحليل العاملي.
- وجود علاقات خطية بين المتغيرات
- استقلالية الأخطاء في كل متغير

ب. شروط أثناء عملية التحليل

- يجب أن تعتبر العوامل الناتجة من التحليل العاملي عن متغيرات واقعية يستطيع الباحث تفسيرها في ضوء إطار نظري أو نظرية معينه تؤكد وجود عوامل في الواقع.
- تعتمد عملية تفسير العوامل على عدد المتغيرات المتشبعة تشبع لا يقل عددها على ثلاثة متغيرات.

ملاحظة

تشير مختلف برامج الاحصاء إلى ان التشبع الدال لا يقل عن 0.6

يتم اليوم استخدام أساليب التحليل العاملي في عدد كبير من المجالات التي لا يمكن تعدادها، نذكر منها:

- في التسويق على سبيل المثال حيث تستخدم هذه الاساليب على نطاق واسع لإدارة العملاء (لاقتراح عروض مستهدفة جديدة ، ...).
- كما أنها تسمح بتحليل الدراسات الاستقصائية، وذلك عن طريق تفسير الظواهر (حيث يتعين أخذ الكثير من البيانات النوعية في عين الاعتبار).
- يمكننا أيضا الاستشهاد بالبحث الوثائقي الذي يعد مهما للغاية وخاصة مع تطور وسائل الاتصال كالأنترنت (حيث يتعلق الأمر هنا بنوع البيانات النصية أو غيرها).
- وتعد المعطيات المتوفرة بكميات كبيرة في مجال الأرصاد الجوية، أيضا من بين أحد الدوافع الأولى لتطوير طرق التحليل العاملي.

ففي الواقع ، فإن أي مجال من مجالات العلوم يجب أن يتعامل مع كمية كبيرة من المعطيات بمختلف انواعها بالاعتماد على هذه الأساليب. وكذلك الامر بالنسبة لمختلف المجالات الاقتصادية (التأمين ، البنوك ، الاتصالات، إلخ).

2.2. مختلف الطرق المستخدمة في التحليل العاملي وأنواعه

تعتبر الأسس النظرية للأساليب التحليل العاملي قديمة، حيث انبثقت من أعمال الباحثان الأمريكيان: سبيرمان (Spearman, 1904) و كل من ثورستون (Thurstone) وبيرت (Burt) (1931 ، 1947) في التحليل العاملي بصفة عامة. ثم اقترح كل من هوتيلينج (Hotteling, 1935) طريقة التحليل بالمركبات الاساسية، وهيرشفيلد (Hirschfeld, 1935) وجوتمان (Guttman) (1941 ، 1959) طريقة التحليل بالمعاملات للتوفيقات.

من الناحية العملية، انتشر استخدام هذه الاساليب على نطاق واسع فقط، خاصة مع تطور برامج التحليل الاحصائي للمعطيات (Husson et al., 2017)، تحت اسم "التحليل متعدد المتغيرات" (Multivariate Analysis).

وتستند جميع هذه الأساليب على الأدوات الكلاسيكية للهندسة الإقليدية التي تم تطويرها في سياق الجبر الخطي.

1.2.2. نبذة عن تطور طرق التحليل العاملي

على الرغم من أن دراسة بنية الكميات الكبيرة من المعطيات تعتبر حديثة العهد في وقتنا الحالي، إلا أن المبادئ والأسس التي تستند إليها طرق تحليل المعطيات قديمة.

فيما يتعلق بالتحليل العاملي، فإن أسسها مبنية على أعمال سبيرمان (Spearman) والتي تعود لسنة 1904، أين قدم Spearman مفهوم العامل لأول مرة. حيث أراد أن يبحث عن متغير توضيحي الذي يختفي خلف العلامات التي حصل عليها مجموعة من الطلبة في عدة مقاييس. حيث يفسر هذا المتغير العامل العام للقدرة أو ما يعرف بـ: " تحليل العوامل ".

و خلال سنوات الثلاثينيات (30) من القرن الماضي، ظهرت مشكلة البحث عن عدة عوامل، بفضل العمل الذي قام به كل من بيرت (Burt) و ثورستون (Thurstone). حيث بحثا عن عاملين، ثم عن عدة عوامل كالذاكرة، الذكاء ، إلخ. حيث ان هذه العوامل لا يمكن ملاحظتها بشكل مباشر ولكن من المحتمل أن تشرح وتفسر احصائيا العديد من العلامات التي حصل عليها الطلاب. وتتعلق هذه المسألة بتلخيص عدد من العوامل بمعلومات متعددة الأبعاد انطلاقا من تحليل عدد صغير من العوامل.

بعد ذلك ، تم اقتراح طريقة التحليل بالمركبات الأساسية (Principal Components Analysis, PCA) التي طورها Hotelling (1933) والتي يمكننا أيضا اسنادها أيضا إلى أعمال بيرسون (Pearson, 1901). حيث تعتبر هذه الطريقة الأفراد الذين تم تمثيلهم في الجدول، الذي سنقوم بتحليله، بمثابة متجهات (أشعة) لسحابة نقاط ذات p من الأبعاد.

وقد قام Hotelling باقتراح تقليص بُعد المسافة من خلال إسقاط سحابة النقاط الفردية على فضاء البعد k (k صغير ثابت) يساعد على تعديل وضبط السحابة بشكل أفضل.

وتعتبر هذه الطريقة من أكثر طرق التحليل العاملي دقة وشيوعاً واستخداما، نظرا لدقة نتائجها بالمقارنة ببقية الطرق، زما لها مزايا عدة منها:

- تؤدي إلى تشعبات دقيقة
- كل عامل يستخرج أقصى كمية من التباين
- تنتج أقل قدر ممكن من البواقي
- المصفوفة الارتباطية تختزل إلى أقل عدد من العوامل المتعامدة غير المرتبطة

من وجهة نظر حديثة، فإن تحليل المركب الأساسي هو:

" تقنية لتمثيل المعطيات ذات الطابع المثالي وفقا لمعايير جبرية وهندسية معينة، ويستخدم بشكل عام دون الرجوع إلى فرضيات ذات طابع إحصائي أو لنموذج معين " (Lebart et Piron, 2000)

كما أن هناك طريقتان مشتقتان منها، ويتعلق الامر بـ:

- **طريقة التحليل التمييزي** (Discriminant Analysis) للعوامل، الذي بدأه فيشر (Fisher) في عام 1936 ، والذي يصف العلاقة بين المتغير النوعي ومجموعة من المتغيرات الكمية.

- **التحليل القانوني** (Canonical Analysis) الذي قدمه Hotelling في عام 1936 ، الذي كان هدفه الأولي هو التعبير ، بمساعدة عدد صغير من توليفة من المتغيرات، أن الصلة بين مجموعتين من البيانات الكمية. وهي تعتمد على نفس مجموعة النتائج الرياضية التي سيتم التطرق اليها لاحقا في هذا الفصل (المبدأ العام لأسس تطبيق التحليل العاملي

وتجدر الاشارة أيضا إلى أن اعمال Hirschfeld (1935) و Guttman (1941 ، 1959) جاءت لتقترح طريقة اخرى للتحليل العاملي تعرف بطريقة " التحليل بالمعاملات للتوفيقات " (Correspondence Factor Analysis, CFA). حيث توفر هذه الطريقة، دون افتراضات مسبقة تبسيط التمثيلات إلى حد كبير مما يساعد في عملية في التفسير.

إن استعمال هذه الطريقة يسمح لنا من التأكد من صحة الارتباطات بين المتغيرات ومختلف خصائصها من جهة والتمثيل التركيبي للعلاقات أو الارتباطات بين مختلف الأنواع تبعا لخصائصها من جهة أخرى.

وقد استعملت هذه الطريقة في عدة بحوث، وهي من اساليب التحليل العاملي التي تساعد على تصنيف وترتيب المعطيات. وفي هذا المجال، قام بن زكري (Benzecri) (1963، 1973) باقتراح طريقة أخرى المتعلقة بالتحليل العاملي للمتغيرات الكمية المتعددة أو طريقة التحليل بالمعاملات المتعددة للتوفيقات (Analyse des Correspondances Multiple, ACM).

وتعتبر هذه الطريقة عملية أساسية في البحث العلمي. وقد وضعت للكشف عن مختلف الارتباطات بين العوامل المتعددة، ولعا عدة أهداف، أهمها اختصار عدد المتغيرات التي تخص عامل ما بغرض تعريفه.

وفي الحقيقة، يصعب حصر تطور طرق التحليل العاملي فقط في أعمال الباحثين الذين أشرنا اليهم، لأنه هناك العديد من الباحثين الذين ساهموا في تطويرها نذكر منهم: بوفون (Buffon, 1749) ، أدانسون (Adanson, 1757) وليني (Linné, 1758).

ونذكر أيضا من خلال هذه النبذة التاريخية تحليل المعطيات غير الرقمية (النوعية) الذي قدمته مدرسة جديدة للإحصائيين الأمريكيين تحت اسم " القياس متعدد الأبعاد " (Multidimensional Scaling, MDS) من خلا اعمال كل من: كارول (Carrol)، كروسكال (Kruskal)، ويش (Wish) وغيرهم. وتتمثل أساليبها الرئيسية في:

- تحليل القرب او الموقع
- تحليل التفضيلات
- تحليل القياس المشترك (مع امكانية شرح متغير نوعي ترتيبي باستخدام المتغيرات الاسمية)

وقد استخدمت هذه الطرق بشكل كبير وفي عدة مجالات، خاصة في التسويق وهي من الاساليب التي تقوم بتحليل العلاقات بين المتغيرات بالاعتماد على التمثيل البياني (Visual).

وبالتالي فإن المبدئ الرئيسي لهذه الطريقة يتوقف على القدرة على ملاحظة أو مراقبة نفس مجموعة المعطيات بيانياً من خلال اختلاف "المنظور البصري" وكذلك لمقارنة تمثيلاتها. يتم تعريف كل منظور بالطريقة التي نقيس بها المسافات أو الاختلافات بين الأشياء.

2.2.2. أنواع التحليل العاملي

عموما، يمكن تقسيم التحليل العاملي إلى نوعين هما:

- *التحليل العاملي الاستكشافي* (Exploratory Factor Analysis, EFA)

يستخدم هذا النوع في الحالات التي تكون فيها العلاقات بين المتغيرات والعوامل الكامنة غير معروفة، وبالتالي فإن التحليل العاملي بهدف اكتشاف العوامل باعتبارها فئات من هذه المتغيرات، وبالتالي فيمكنا وصف هذه المتغيرات.

ويعتبر التحليل العاملي الاستكشافي من الاساليب التي تسمح باستكشاف البنية الكامنة لمجموعة من المعطيات. حيث تعبر هذه البينية عن وجود عدد من العوامل الأساسية التي يمكنها شرح سبب ارتباط بعض المتغيرات مع بعضها البعض. ولا يمكن ملاحظة هذه العوامل بشكل مباشر، ولكن يتم استنتاجها من خلال مراعاة نمط الارتباط الذي لوحظ بين المتغيرات.

- *التحليل العاملي التوكيدي* (Confirmatory Factor Analysis, CFA)

يستخدم لاختبار الفرضيات التي تفترض وجود أنماط أو عوامل خاصة من العلاقات في المعطيات يمكن على أساسها تصنيف المتغيرات، اي اختبار الفرضيات المتعلقة بوجود أو عدم وجود علاقة بين المتغيرات. كما يستخدم هذا النوع من التحليل أيضا لتقييم قدرة نموذج العوامل على التعبير عن المعطيات الفعلية وكذلك للمقارنة بين عدة نماذج للعوامل.

باعتباره امتدادا للتحليل العاملي الاستكشافي (EFA)، يركز التحليل العاملي التوكيدي (CFA) على المتغيرات الكامنة التي تقوم عليها تنظيم الاختلافات بين الأفراد التي لوحظت على مجموعة من الاختبارات. ما يصيفه هذا النوع من التحليل، بالمقارنة مع التحليل الاستكشافي، هو امكانية اختبار مدى كفاية المعطيات بالنسبة للنموذج النظري.

بطريقة أخرى فان هذا التحليل، وبفضل المنهج الاستنتاجي الافتراضي الذي يعمل عليه، يقوم مسبقا بوضع الافتراضات المتعلقة بالمتغيرات الكامنة.

ويعتبر التحليل العاملي التوكيدي حالة خاصة لنمذجة بالاعتماد على المعادلة الهيكلية (Structural Equation Modelling)، حيث يتم أولا تحديد نموذج يحدد: عدد العوامل، العلاقات المحتملة بين هذه العوامل، والعلاقات بين هذه العوامل والمتغيرات، درجة الخطأ المرتبط بكل متغير والارتباطات المحتملة بينها.

3.2.2. طرق التحليل العاملي

يتم تصنيف طرق التحليل متعدد المتغيرات بالنظر إلى الغرض او الهدف المراد الوصول اليه من خلال تطبيق أساليب التحليل العاملي (تقليل الحجم أو التصنيف) ونوع المعطيات المطلوب تحليلها (الكمية أو النوعية أو كلاهما معا). ويمكن تلخيص ذلك من خلال الجدول الموالي:

الجدول رقم 2-02: الغرض من تطبيق مختلف طرق التحليل العاملي

الغرض	الطريقة	حالة الاستخدام
للوصف أو تقليص الأبعاد	التحليل بالمركبات الأساسية (PCA)	p من المتغيرات الكمية
	التحليل التمييزي (Discriminant)	p من المتغيرات الكمية ومتغير نوعي واحد
	التحليل بالمعاملات للتوفيقات (CFA)	متغيرين نوعيين
	التحليل بالمعاملات المتعددة للتوفيقات (ACM)	p من المتغيرات النوعية
	التحليل القانوني (Canonical)	p و q من المتغيرات الكمية
	القياس متعدد الأبعاد (MDS)	تحديد المواقع (متعدد الأبعاد)
التصنيف	التصيف التسلسلي (Hierarchical classification)	تصنيف الأفراد في أقسام متجانسة

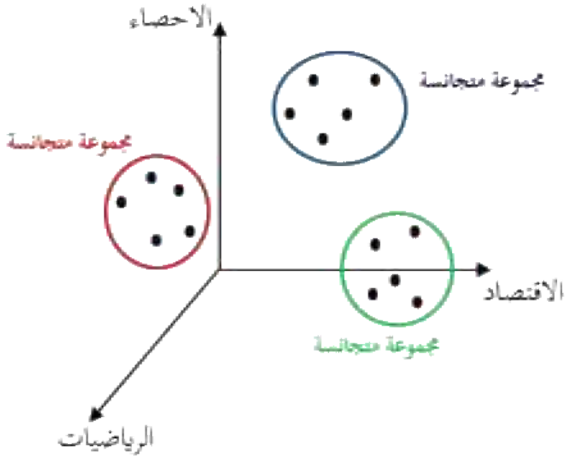
ويمكن توضيح ما سبق من خلال العودة الى المثال السابق (مثال 2-01) والذي يوضح معطيات حول علامات 6 طلاب في 3 مقاييس.

حيث حتى وإن تمكنا من حساب المسافات بين الأسطر (الطلاب) والمسافات بين الأعمدة (المقاييس) في الجدول، فلن يكون بمقدورنا تصورها بيانيا مباشرة. هذا يعني أن التمثيلات الهندسية المرتبطة بها تتضمن عادة مسافات بأكثر من بعدين أو ثلاثة أبعاد، لذا فمن الضروري إجراء تحويلات وتقديرات للحصول على تمثيل البياني.

يمكن وصف جدول المسافات المرتبطة بهذه الأشكال الهندسية، والتي يمكن اعتبارها بسيطة من حيث المبدأ، ولكنها معقدة بسبب وجود عدد الكبير من الأبعاد، من خلال مجموعتين رئيسيتين من الطرق، وهما أساليب التحليل العاملي وأساليب التصنيف.

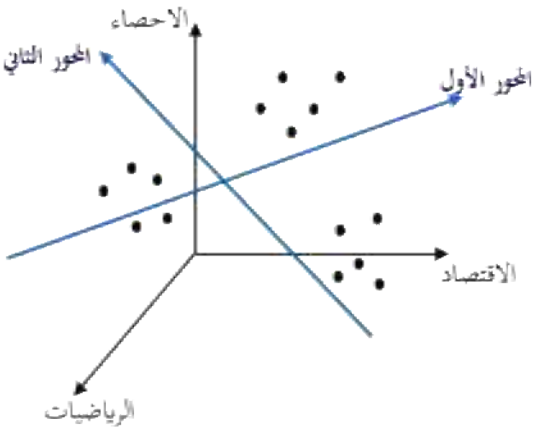
حيث تهدف الأولى إلى البحث عن الاتجاهات الرئيسية التي تنحرف فيها النقاط عن متوسط النقاط بصورة أكبر. أما الثانية فتبحث عن مجموعات أو فئات من الأفراد تكون أكثر تجانسا. ويمكن توضيح هذه الفكرة من خلال الشكل الموالي:

الشكل رقم 2-02: الفرق بين اساليب التحليل العاملي واساليب التصنيف



نماذج التصنيف

(البحث عن مجموعات متجانسة)



التحليل العاملي

(البحث عن اتجاهات رئيسية)

وتستخدم هذه الاساليب (التحليل العاملي ونماذج التصنيف) المعطيات بنفس الطريقة، أي الأفراد ممثلة في الصفوف والمتغيرات في الاعمدة. كما أن الاختلاف في عملية التحليل يساعد على اثراء تفسير العلاقات والارتباطات (Bouroche et Saporta, 1994).

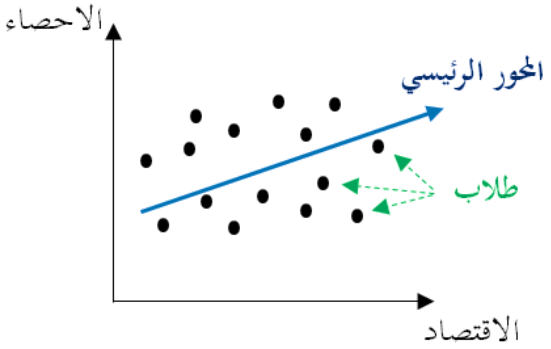
3.2. اسس تطبيق التحليل العاملي: المبدأ العام

يساعد التحليل العاملي على اجراء عمليات على كميات كبيرة من المعطيات وفي نفس الوقت يوضح طبيعة الارتباطات بين مختلف المتغيرات، وذلك من خلال أساليب تحقق نوعا من التقليص في المعطيات بغرض اشتقاق البنية الداخلية في شكل رسوميات بيانية.

الهدف من ذلك هو قياس مسافات فرعية ذات أبعاد منخفضة تتناسب بشكل أفضل مع سحابة نقاط الأفراد وسحابة نقاط لمتغيرات، بحيث هذه القياسات (التقارب في هذه المسافات الفرعية) تعكس قدر الإمكان التقارب الحقيقي. نحصل بالتالي على فضاء للتمثيل (أو فضاء عاملي) والذي تحدده المحاور الرئيسية للقصور الذاتي ويمثل نقاط السحابة في نظام المحاور هذا.

حيث تقوم هذه المحاور بإجراء تعديلات لمجموعة النقاط باستخدام طريقة المربعات الصغرى والتي تقوم بتقليل مجموع مربعات الفروقات بين النقاط والمحاور (أنظر الشكل الموالي).

الشكل رقم 2- 03: تعديل سحابة نقاط الطلاب في فضاء المقاييس



يتوافق المحور الأول مع أقصى خط امتداد سحابة النقاط، والمحور الثاني يقوم بتعظيم نفس المعيار السابق حتى يتعامد مع المحور الأول. وتستمر العملية بالنسبة للمحاور التالية، والتي تكون متعامدة مع بعضها البعض. وهذا التعامد يعكس استقلالية المحاور والتي تشير في الواقع الى عدو الارتباط.

وإجراءات التعديل هي نفسها لكلتا السحابتين، حيث نقوم بإثبات وجود علاقات بسيطة تربط المحاور المحسوبة في الفضاءين (الفضاء المتعلق بالأفراد و الفضاء المتعلق بالمتغيرات).

وفيما يلي شرح مفصل للمبدأ الذي تقوم عليه أساليب التحليل العاملي.

أ. مبدأ التحليل العاملي

يعتمد المبدأ العام للتحليل العاملي على فرضية مزدوجة، تفرض وجود شعاع صف u_1 بـ p من المركبات (أو المكونات) وشعاع عمود v_1 بـ n مركب (أو مكون).

فكما في الشكل (1-01) فإن المصفوفة X، حيث $X = \{x_{ij}\}$، تكتب على الشكل التالي:

$$X = v_1 u_1'$$

حيث: u_1' هو مقلوب الشعاع u_1

بالإضافة إلى $(n+p)$ من قيم الاشعة u_1 و v_1، فإننا نقوم بإيجاد قيم $(n.p)$ للمصفوفة X.

هذا التقليص يكون بالغ الأهمية اذا كانت كل من n و p كبيرة. بالإضافة إلى ذلك فإن هذا التقليص لا يؤدي إلى فقدان المعلومات خلال عملية التحليل.

ويمكن أن نقول بطريقة اخرى أن الهدف هنا هو البحث عن تقدير تقريبي للمصفوفة X من الرتبة S. وهذا يعني أن هذه التحليلات تسعى إلى كتابة المصفوفة X بحيث:

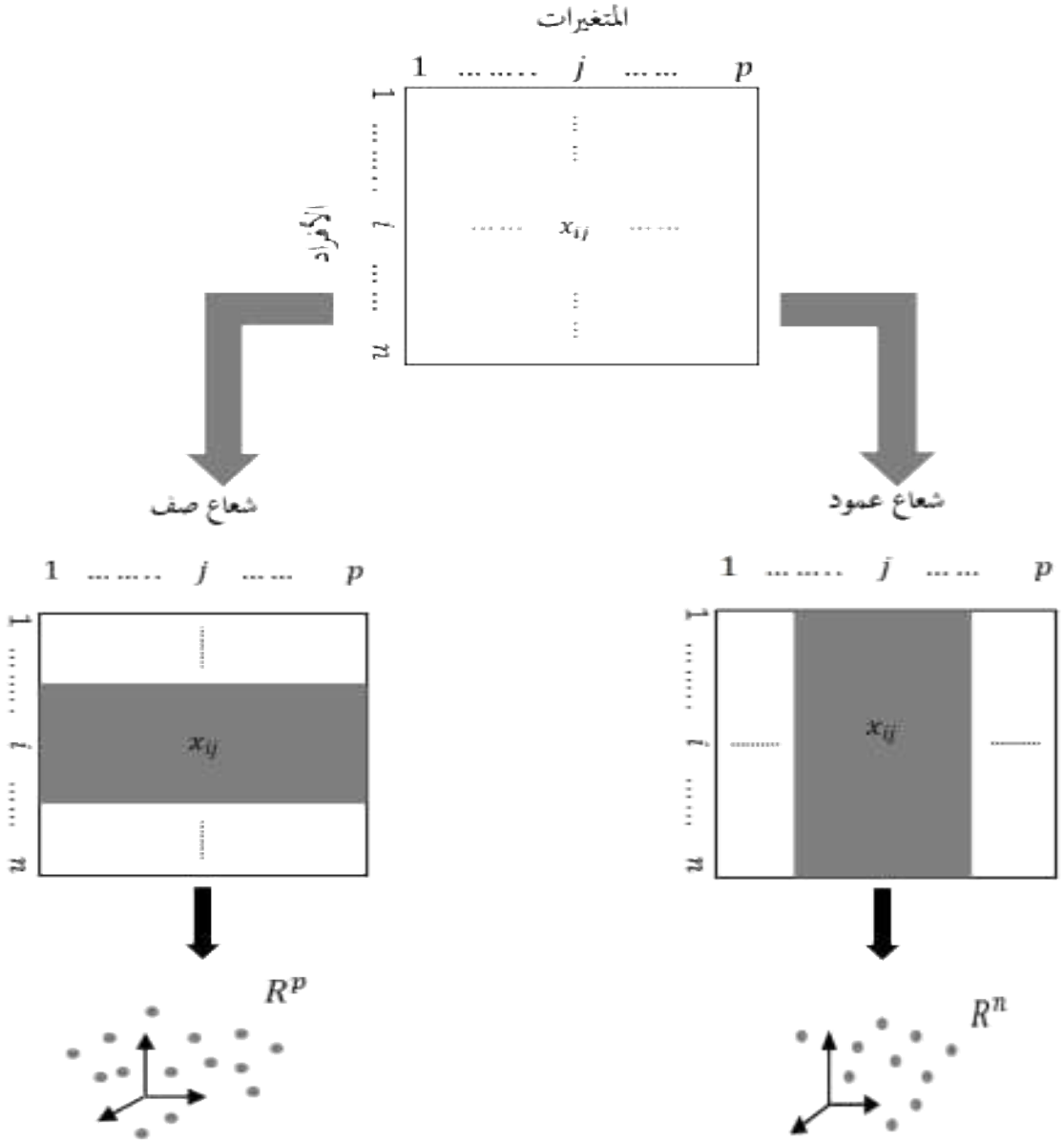
$$X = v_1 u_1' + v_2 u_2' + \cdots + v_S u_S' + E$$

حيث تمثل E مصفوفة متكونة من n سطر و p عمود للقيم التي لم تدخل في التحليل وتعرف أيضا بـ: " مصفوفة البواقي ". كما أنه يتم اعادة بناء القيم الأولية $(n.p)$ للمصفوفة X بطريقة مثلى بواسطة قيم $S.(n+p)$ لأشعة S $(u_q$ و $v_q)$.

وبالتالي تعتبر المعطيات إما كأفراد يتم وصفهم بواسطة المتغيرات (p) باستخدام الأشعة u_q بـ p من المكونات، أو كمتغيرات يتم وصفهم بواسطة الأفراد (n) من خلال الأشعة v_q بـ n من المكونات.

ويمكن توضيح ذلك بيانيا كما يلي:

الشكل رقم 2- 04: سحابات النقاط



وبالتالي نحصل على نقاط n في الفضاء R^p و p في الفضاء R^n. وتجدر الاشارة إلى أن هناك عدة أشكال مميزة لسحابات سواء بالنسبة لفضاء الأفراد أو المتغيرات.

وبالتالي سوف نقوم بتعديل سحابة الأفراد في فضاء المتغيرات، وسحابة المتغيرات في فضاء الأفراد.

1.3.2. تمثيل الأفراد في فضاء المتغيرات

سنسعى في هذه المرحلة إلى تقديم صور تقريبية لسحابة الأفراد في R^p ونرمز لهم بالرمز I_n. لأجل ذلك نفرض أن السحابة I_n محتواة داخل فضاء فرعي شعاعي ذو البعد S، (حيث : $S < p$). مما يعنى أننا افترضنا أن مصفوفة البواقي معدومة. بشكل أعم ، نفترض أنه أعيد تشكيل السحابة I_n على نحو أفضل في فضاء فرعي ذو البعد S. يمكننا بالتالي إعادة تشكيل الأفراد n وكذا المصفوفة X من خلال احداثيات الأفراد على المحاور الجديدة X.

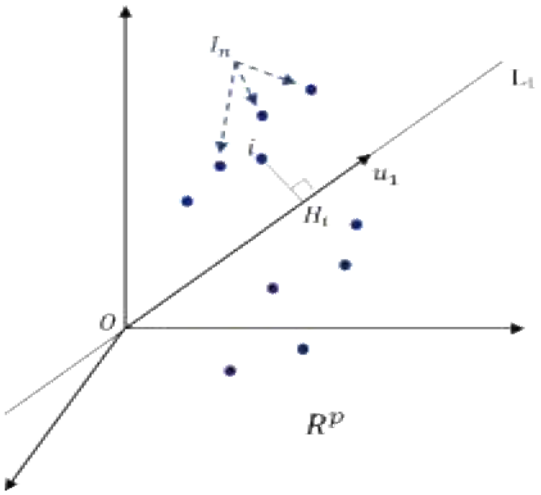
وبالتالي فإننا نقوم باستبدال قيم المصفوفة X $(n.p)$ بـ:

$$n.S\ (\text{احداثيات}) + p.S(\text{مكون})$$

فمثلا إذا كان لدينا 500 طالب وكل طالب حصل على 20 علامة (في كل المقاييس)، وإذا كانت $S = 10$، هذا يعني انه إذا كانت سحابة نقاط 500 طالب محتواة داخل فضاء فرعي ذو البعد 10، فإننا نقوم بتقليص المعطيات (500 * 20 = 10^4 من قيم المصفوفة X) من خلال دراسة 5200 قيمة .

$$500 * 10 + 20 * 10 = 5200$$

وبالتالي فأن الخطوة الاولى تقتضي البحث عن فضاء شعاعي أحادي البعد، أي ايجاد خط L_1 يمر على المبدأ ويعدل بطريقة أفضل السحابة I_n. وبالتالي نأخذ الحالة التي يكون فيها $S = 1$.



إن الإسقاط الموجود على الخط L_1 والذي يعدل بصورة أفضل سحابة النقاط يعطي الحد الاقصى للتشتت أو القصور الذاتي على طول الخط. ونحن نعرف أنه لتعظيم التشتت على ينبغي تقليص المسافات بين نقاط السحابة والخط L_1. مما يعني أن تعظيم التشتت يقتضي تعظيم المجموع مربعات الانحرافات (نظرية فيتاغورس):

$$\sum_{i \in n} Oi^2 = \sum_{i \in n} OH_i^2 + \sum_{i \in n} iH_i^2$$

وبالتالي لتعظيم التشتت يجب تعظيم: $u_1' X' X u_1$ حيث يعبر u_1 عن الشعاع الوحدوي لـ L_1. وبالتالي نحصل على المساواة التالية:

$$\sum_{i \in n} OH_i^2 = (Xu_1)'(Xu_1) = u_1'X'Xu_1$$

والتي تمثل القصور الذاتي على طول الخط L_1.

ومنه يمكننا ان نقول أن المشكل يتوقف على ايجاد u_1 الذي يعظم $u_1'X'Xu_1$، مع الاخذ بعين الاعتبار القيد $u_1'u_1 = 1$. وبالتالي فإن الفضاء الفرعي ذو البعد الأمثل (أعظم قصور ذاتي) هو المحور L_1 المعرف بواسطة الشعاع u_1 والذي يمثل حل لهذه المشكلة.

وبنفس الطريقة، لتحديد الفضاء الفرعي ثنائي الأبعاد التي يتناسب مع سحابة النقاط I_n، نأخذ بعين الاعتبار الحالة $S = 1$.

الفضاء الفرعي ثنائي الأبعاد يتميز بوجود محورين L_1 و L_2 المعرفين بالشعاعين u_2 و u_1 حيث u_2 متعامد مع u_1، تحت القيود التالية:

$$\left.\begin{array}{r} \text{معظمة } u_2'X'Xu_2 \\ u_2'u_2 = 1 \text{ (شرط التوزيع الطبيعي)} \\ u_2'u_1 = 0 \text{ (شرط التعامد)} \end{array}\right]$$

في حالة التي يكون فيها $S \geq 2$، بالتراجع، فإن الفضاء الفرعي ذو البعد S الذي يتناسب مع سحابة النقاط I_n بشكل أفضل يحتوي على الأشعة $u_1, u_2, \dots u_{S-1}$. هذا الفضاء الفرعي تم إنشاؤه من خلال الفضاء الفرعي $(u_1, u_2, \dots u_{S-1})$ ذو البعد $S-1$ والشعاع u_S المتعامد مع هذا الفضاء الفرعي (مع كل الأشعة u_q) ويحقق ما يلي:

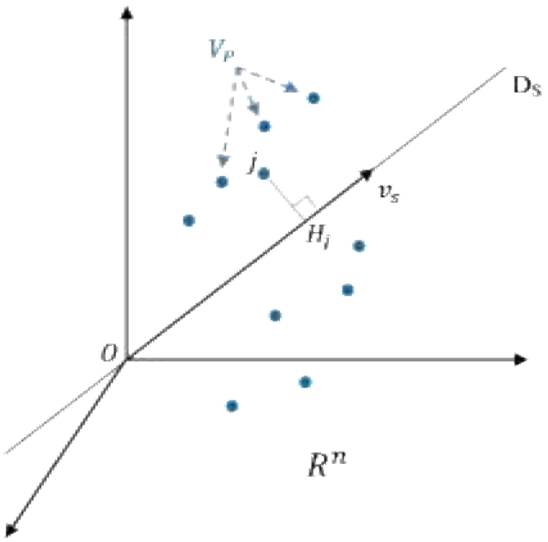
$$\left.\begin{array}{r} \text{معظمة } u_s'X'Xu_s \\ u_s'u_s = 1 \end{array}\right]$$

ملاحظة

يشير الشعاع u_S إلى اكبر قيمة ذاتية للمصفوفة $X'X$ ، و $u_s'X'Xu_s = \lambda u_s'u_s = \lambda$ هي القصور الذاتي الذي تم اسقاطه على المحور L_S.

2.3.2. تمثيل المتغيرات في فضاء الأفراد

كما في سحابة الأفراد، فإننا سنبحث، وبنفس الطريقة، عن صورة لسحابة المتغيرات والتي نرمز لها بالرمز V_p في R^n. المنهج مماثل لمنهج سحابة الأفراد فقط نأخذ بعين الاعتبار X' بدلا من X.



وبالتالي فإن القصور الذاتي على طول الخط D_S يكتب على الشكل التالي:

$$(X'v_s)(Xv_s) = v_s'XX'v_s$$

بالإضافة إلى أن المحور العاملي ممثل بواسطة الشعاع v_s مع:

$$\left.\begin{array}{l} v_s'XX'v_s \text{ معظمة} \\ v_s'v_s = 1 \text{ (التوزيع الطبيعي)} \\ v_s'v_q = 0 \text{ من أجل كل } q = \{1, 2, \ldots., S-1\} \text{ (شرط التعامد)} \end{array}\right\}$$

ونحصل على الفضاء الفرعي المعدل بنفس الطريقة كما في حالة الأفراد.

3.3.2. العلاقة بين محاور القصور الذاتي وعوامل الافراد والمتغيرات

الغرض هنا هو توضيح العلاقات، والتي تعرف أيضا بالعلاقات الانتقالية، بين التعديلات في الفضاءين، حيث:

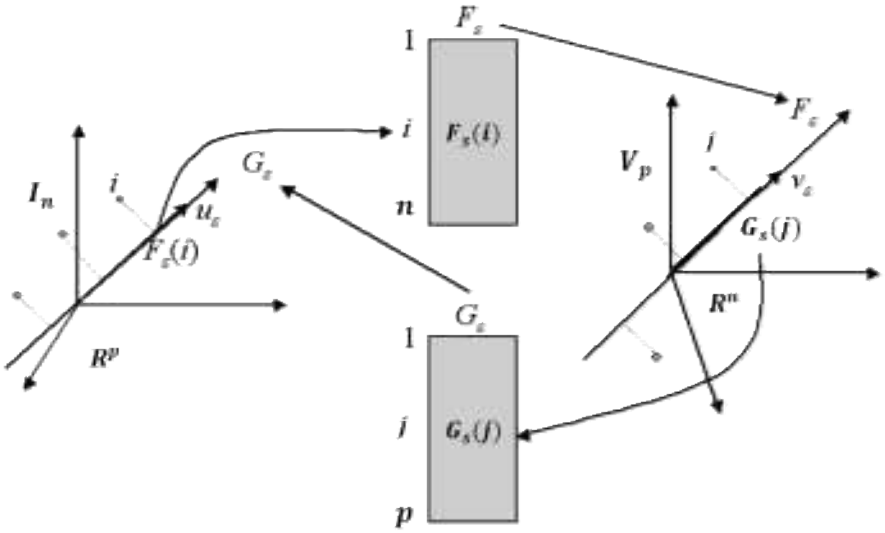
- تعبر كل من $\lambda L_s = u_s'X'Xu_s$ و $\lambda D_s = v_s'XX'v_s$ على التوالي عن القصور الذاتي للمحور L_s والمحور D_s.
- تعبر كل من $F_s = Xu_s$ و $G_s = X'v_s$ عن العامل من الدرجة S على التوالي لـ I_n و V_p، وبالتالي يمثل F_s الشعاع الناتج عن اسقاط السحابة I_n على المحور رقم S في R^p، أما G_s فيمثل الشعاع الناتج عن اسقاط السحابة V_p على المحور رقم S في R^n.

القصور الذاتي λL_s على طول الخط L_s تساوي القصور الذاتي λD_s على طول الخط D_s، نرمز لها بالرمز λ_s.

علاقات الانتقال بين الفضاءين R^p و R^n تعطى على الشكل التالي:

$$\begin{cases} v_s = \dfrac{F_s}{\sqrt{\lambda_s}} \\ u_s = \dfrac{G_s}{\sqrt{\lambda_s}} \end{cases}$$

,يمكن تمثيل العلاقات الانتقالية بين الفضاءين بواسطة الشكل الموالي، والذي يمثل العلاقات بين **محاور** القصور الذاتي للسحابة و**عوامل** السحابة الأخرى.



علاقات الانتقال بين العوامل تكتب على الشكل:

$$\begin{cases} F_s(i) = \sum_{j \in p} x_{ij} u_s(j) = \dfrac{\sum_{j \in p} x_{ij}\, G_s(j)}{\sqrt{\lambda_s}} \\ G_s(j) = \sum_{i \in n} x_{ij} v_s(i) = \dfrac{\sum_{i \in n} x_{ij}\, F_s(i)}{\sqrt{\lambda_s}} \end{cases}$$

هاتين الصيغتين تثبتان ن عوامل السحابتين يجب أن تفسران معا. وبالتالي يمكننا القول أن التحليل العاملي يسعى لتحليل السحابة I_n و السحابة V_p معا.

4.3.2. إعادة تشكيل البيانات

من الممكن إعادة بناء جدول المعطيات (المصفوفة X) من خلال تحليل قيم المصفوفة X. في الواقع، بما أن v_s يمثل الشعاع الذاتي من الدرجة S للمعيار 1 من المصفوفة $X'X$ الموافق للقيمة الذاتية λ_s و v_s يمثل الشعاع الذاتي من الدرجة S للمعيار 1 من المصفوفة XX' الموافق لنفس القيمة الذاتية، فيكون لدينا:

$$Xu_s = \sqrt{\lambda_s} v_s$$

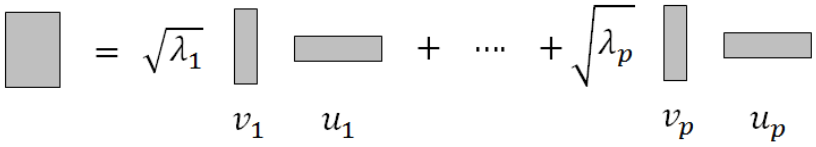
أين:

$$X \sum_{s \in p} u_s\, u'_s = \sum_{s \in p} \sqrt{\lambda_s} v_s u'_s$$

الأشعة الذاتية متعامدة ومن المعيار 1، ومنه نحصل على:

$$X = \sum_{s \in p} \sqrt{\lambda_s} v_s \, u'_s$$

يمكن توضيح هذه الصيغة من خلال تحليلات القيمة الفردية من القيم الذاتية (والتي تمثل أيضا القصور الذاتي)، والأشعة الذاتية المرتبطة بها في الشكل الموالي:



إذن فإن هذه العملية تفرض الحصول على القيم $n.p$ الموجودة في p من الأشعة الذاتية.

للتذكير فقط، يحتوي الملحق رقم 3 على طريقة حساب القيم الذاتية والاشعة الذاتية.

الفصـل الثالث

التحـليـل الى مـركبـات أساسيـة

مقدمة الفصل

غالبا ما تعرف طريقة التحليل الى مركبات أساسية (أو كما يطلق عليها بالإنجليزية Principal Component Analysis PCA) على انها طريقة تساعد في تقليل البيانات أو تقنية تقليل الأبعاد. ويقصد هنا انه في الأساس نبدأ بمجموعة من المتغيرات، لنقل n، وفي نهاية عملية التحليل نحصل على عدد أقل ولكنه لا يزال يعكس نسبة كبيرة من المعلومات الواردة في مجموعة البيانات الأصلية. والهدف هو إنشاء العلاقات المتبادلة التي تحدث بين مجموعة من المتغيرات بغرض تقليل التعقيد الذي يعكس ما هو أساسي أو أكثر أهمية وبالتالي تحديد بعض العوامل التركيبية القليلة التي تزيد من التمييز بين الأفراد.

ويقدم هذا الفصل تفاصيل عن طريقة التحليل الى مركبات اساسية والتي يمكن أن تساعد الطالب في إعادة هيكلة بياناته على وجه التحديد عن طريق تقليل عدد المتغيرات. حيث سنتطرق الى مجموعة الإجراءات المحددة في تحليل العوامل من أجل الحصول على نظرة عامة أولية لطبيعة هذه الطريقة. كما اننا سنعرض الميزات الأساسية لهذه الطريقة وصياغتها الرياضية، والتمثيلات الرسومية، وجميع العناصر اللازمة لقراءة أو تفسير نتائج هذا النوع من التحليل. كما سنعرض أمثلة توضيحية تساهم في فهم كريقة عمل PCA. وفي نهاية الفصل سيتم تفصيل كيفية الحصول على هذه النتائج من خلال تطبيق برنامج SPSS.

1.3. طريقة التحليل الى مركبات أساسية

تم تصميمها لأول مرة بواسطة كارل بيرسون في عام 1901 ، وتم دمجه في الإحصاء الرياضي بواسطة Hotelling في عام 1933 (صواليلي، 2011)، ولم يتم استخدام طريقة التحليل الى مركبات الرئيسية إلا منذ ظهور وسائل الحوسبة الحالية واتساع استخدامها. ويمكن تقديم دريقة التحليل الى المركب الاساسي من خلال وجهات نظر مختلفة:

- فبالنسبة للإحصائي الكلاسيكي تتمثل الطريقة في البحث عن المحاور الرئيسية التي تشير إلى التوزيع الطبيعي متعدد الأبعاد، ويتم تقدير هذه المحاور من خلال عينة معينة.
- بالنسبة لخبراء العوامل يعبر ذلك عن حالة خاصة لطريقة التحليل العاملي للمعتمدة في القياس النفسي.
- أما من وجهة نظر محللي البيانات فيتعلق الأمر بتقنية تمثيل البيانات ذات الطابع الأمثل وفقا لمعايير جبرية وهندسية معينة والتي يستخدمها الباحث بشكل عام دون الرجوع إلى فرضيات ذات طبيعة إحصائية ولا إلى نموذج معين. وتعتبر وجهة النظر هذه الأكثر انتشارا حاليا.

ويرتبط أصل وتطور تحليل العوامل بشكل أساسي بالعمل المنجز في مجال علم النفس في مواجهة مشكلة قياس عوامل الذكاء البشري، ووجد الباحثون في هذا المجال أن بعض الاختبارات النفسية واختبارات الذكاء المختلفة كانت مرتبطة ارتباطا وثيقا ببعضها البعض، وبالتالي لم يقيسوا جوانب أو عوامل أو أبعاد مختلفة للذكاء. وهنا لاحظ سبيرمان أنه نتيجة للاختبارات المختلفة كان هناك عامل عام (عامل G) للذكاء بالإضافة الى عدد من العوامل المحددة. وهنا تم تطوير هذه الطريقة.

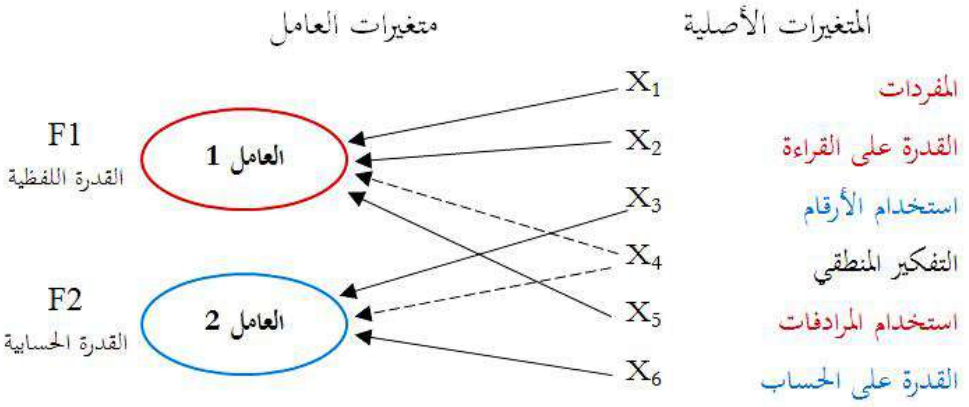
1.1.3. التعريف والاهداف

تحليل PCA هو أسلوب إحصائي مبسط متعدد المتغيرات يسمح بتحويل مجموعة من المتغيرات الأصلية المرتبطة ببعضها البعض، إلى مجموعة تركيبية من المتغيرات تسمى العوامل أو المكونات الأساسية. وتعتمد طريقة التحليل الى مركبات أساسية على إيجاد المتجهات المميزة لمصفوفة التباين والتباين المشترك للمتغيرات التوضيحية، مما يعني إيجاد الجذور المميزة لمصفوفة الارتباط. وتقدم طريقة PCA العديد من الاختلافات اعتمادا على التحولات التي تم إجراؤها على جدول البيانات (Jolliffe, 2002).

لذلك فان هذه الطريقة تضم مجموعة متنوعة من الإجراءات الإحصائية تعتبر أن هناك سلسلة من العلاقات الكامنة بين مجموعة من المتغيرات والتي تحدد ظاهرة معينة. لتحليل هذه العلاقات، أو لاكتشاف الأسباب غير الواضحة فيها وتفسير هيكلها يوفر تحليل PCA عددا أقل من الأبعاد الكامنة وراء مجموعة المتغيرات الأصلية. هذه العوامل ليست أكثر من متغيرات جديدة يتم الحصول عليها من المتغيرات الأصلية.

ويمكن تخطيط العلاقة بين المتغيرات الأصلية والعوامل من خلال المثال الموضح في الشكل الموالي:

الشكل رقم 3-01: العلاقة بين المتغيرات الأصلية والعوامل



تظهر طريقة تحليل PCA وجود متغيرات تقيس بشكل مشترك نفس الظاهرة أو العديد منها، أي أن العوامل أو الأبعاد الموجودة هي مجموعات (خطية) من المتغيرات الأصلية، وسوف يساهم كل من هذه المتغيرات في تكوين متغيرات العامل الجديدة. لذلك فإن النموذج الرياضي الذي يبنى عليه تحليل PCA هو النموذج الخطي ويحاول شرح الارتباطات والتغيرات لمجموعة من المعلومات مع عدد أقل من المتغيرات الأساسية الجديدة التي تأتي من مجموعة المتغيرات الأصلية.

من الناحية الهندسية تتوافق العوامل مع المحاور المتعامدة مع بعضها البعض والتي تناسب شكل سحابة النقاط التي تشكل مجموعة من الأفراد في فضاء المتغيرات.

ومنه يمكننا القول ان تحليل PCA هو طريقة جبرية إحصائية نحاول من خلالها تجميع وإعطاء بنية للمعلومات الموجودة في مصفوفة البيانات. ويتكون الإجراء من توحيد هذه المصفوفة في مساحة متجهية نحاول العثور فيها على محاور أو أبعاد (تسمى أيضا المكونات أو العوامل) والتي باعتبارها مزيجا خطيا من المتغيرات التي تم ادخالها:

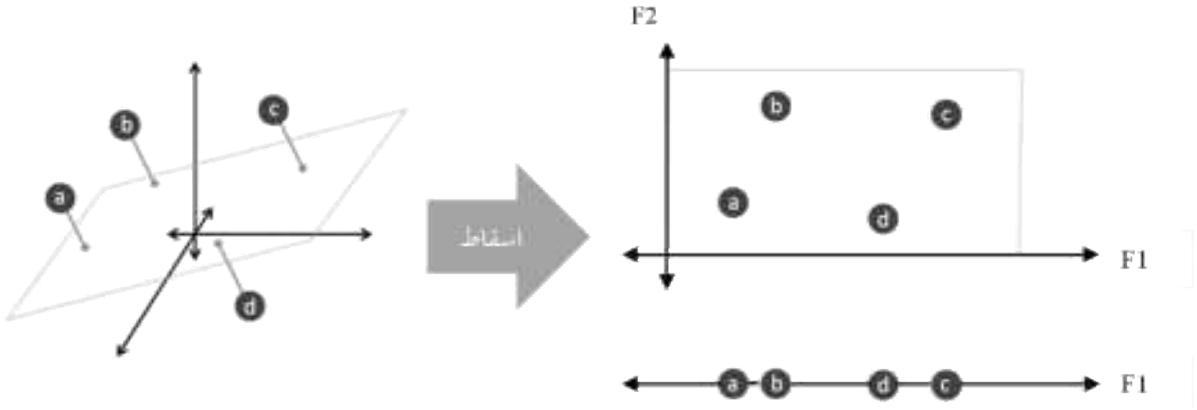
- لا تؤدي الى فقدان المعلومات الأصلية وبالتالي الحفاظ على التباين الكلي.
- لا يوجد علاقة ارتباط أي أن المتغيرات مستقلة خطيا مما يضمن هيكلة المتغيرات الأولية.
- لها أهمية تفاضلية ومعروفة في شرح التباين الكلي.

والهدف الأساسي هو تقليل عدد المتغيرات الاصلية لهذا يتم أخذ المحاور أو المركبات الموجودة كمتغيرات جديدة، وبالتالي تقوم بتبسيط وتقليل وهيكلة المعلومات الأولية. ومنه يمكن القول ان التحليل إلى مركبات أساسية يهدف الى تفسير الظواهر الاقتصادية والاجتماعية عن طريق معطيات مختصرة ومختزلة تمثل مجموعة علاقات بين عدد من المتغيرات مجمعة في مركبات أساسية.

2.1.3. تطبيق طريقة التحليل الى مركبات أساسية

يمكن توضيح ذلك من خلال الشكل الموالي:

الشكل رقم 3-02: تأثير تطبيق PCA على مجموعة بيانات

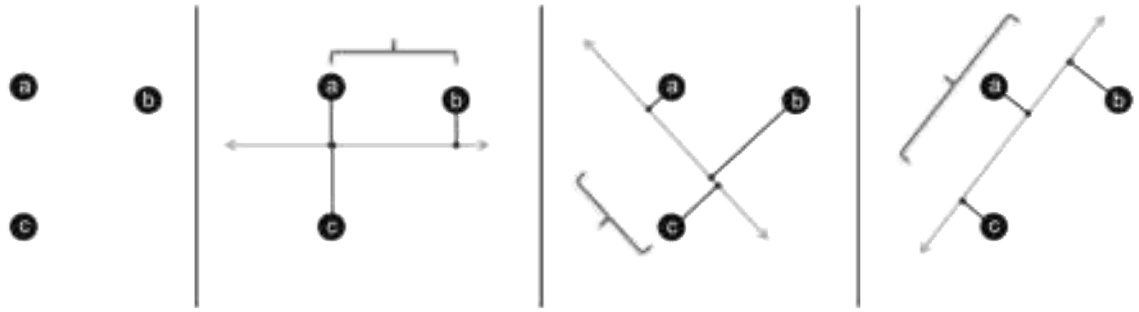


حيث توضح النقاط الأربعة الموجودة على يسار الشكل البيانات الأصلية والتي يمكن تمثيلها في فضاء ثلاثي الأبعاد. نفترض الان أنه سيتم تقليل أبعاد البيانات إلى عالين (في فضاء ثنائي الأبعاد)، حيث يتم إسقاط النقاط كما هو موضح على يمين الشكل.

يمتد هذا الفضاء الجزئي ثنائي الأبعاد بواسطة متجهات متعامدة تسمى المركب الأساسي الأول (F1) والمركب الرئيسي الثاني (F2). بالإضافة إلى ذلك ، يمكن تشكيل تمثيل اخر للبيانات الأصلية في الأبعاد بشكل أكبر عن طريق إسقاط النقاط للأسفل على خط أحادي الأبعاد محدد بواسطة F1. كل مركب اساسي هو ببساطة متجه لأرقام النقطة الثابتة التي تحدد محورا في مساحة الميزة المختصرة.

يمكن فهم كيفية تطبيق PCA على أنها مشكلة تحسينكما هو موضح في الشكل التالي:

الشكل رقم 3-03: اختيار المركب الأساسي الأول



توضح الصورة الفرعية الموجودة في أقصى اليسار عينة من ثلاث نقاط ممثلة في فضاء ثنائي الابعاد، حيث يمكن ان يكون المكون الرئيسي الأول F1 أي واحد من عدد لا نهائي من الخطوط التي تحدث في هذا الفضاء. يمكن بعد ذلك النظر إلى تحسين هذا المركب من خلال تحديد الخط الذي يقلل من إجمالي خطأ الإسقاط (المعروف أيضا بخطأ إعادة البناء) الذي ينتج عند عرض البيانات الأصلية بيانيا.

وتظهر العناصر الفرعية الثلاثة الأخرى للشكل ثلاثة خطوط متوقعة للمركب الاول، حيث يظهر الجزء الفرعي الموجود في أقصى اليمين الحل الامثل لان المسافة بين البيانات وموقعها على الخط (الاسقاط) تقلل من مجموع أخطاء الإسقاط.

والأهم من ذلك ، أن العثور على مثل هذا الخط يحل في نفس الوقت مشكلة ثانية مكافئة والمتمثلة في البحث عن الخط الذي يزيد من تباين البيانات المتوقعة.

للتذكير فان التباين يتم تعريفه على أنه قياس كمي لانتشار النقاط. عندما يتم عرض نقاط البيانات الأصلية على الخط ، يمكن حساب التباين على تلك القيم المتوقعة. وبالتالي يتم حساب خط F1 بحيث يتم تعظيم هذا التباين وبشكل مكافئ يتم تقليل خطأ الإسقاط. بمجرد حساب F1 يتم بعد ذلك حساب F2 وجميع المركبات الاساسية الاخرى.

ويمكن تقسيم عملية تحليل PCA إلى أربع (4) مراحل أساسية تتمثل في:

- **اولا:** اختيار مجموعة المتغيرات والطي يعد أمرا اساسيا، ويكون ذلك وفقا لنموذج تحليل دقيق. ويجب أن يكون هذا الاختيار متناسقا مع شروط تطبيق طريقة PCA ويجب استيفاء بعض الشروط كالارتباط بين بين المتغيرات التي يتم إنشاؤها من خلال فحص مصفوفة الارتباط.
- **ثانيا:** يتم بعد ذلك استخراج المحاور أو المركبات مع حساب القيم الذاتية أو الفروق المدرجة في كل من المحاور التي تحدد التباين الموضح من قبل كل منها ويتم تحديد عدد العوامل التي يجب الاحتفاظ بها في التحليل.
- **ثالثا:** يتم تفسير العوامل أو المكونات من حساب الارتباط بينها وبين المتغيرات الأصلية ، القواسم المشتركة وإعادة تكوين مصفوفة الارتباط ، مع إمكانية تدوير الحل العامل لتحسين إمكانية تفسير المحاور المضروبة.
- **رابعا:** بعد التأكد من صحة النتائج ، يمكننا حساب درجات أو قيم العوامل للأفراد في متغيرات العوامل للحصول على تحليلات إضافية مع إجراءات أخرى.

وسوف ندرج فيما يلي مثال نوضح فيه كيفية تطبيق طريقة PCA.

2.3. خطوات التحليل الى مركبات اساسية

يتم هذا التحليل انطلاقا من مصفوفة المعطيات والتي تمثل المتغيرات الظاهرة المدروسة وكذا عدد أفراد العينة وتتم خطوات هذا التحليل من خلال:

1.2.3. بناء مصفوفة البيانات وحساب المتوسطات والانحرافات

وهي مصفوفة البيانات التي تتكون منn من الاسطر والتي تمثل الأفراد وعدد m من الاعمدة و التي تمثل المتغيرات ومن أجل اجراء التحليل العاملي يتم تحويل هاته المتغيرات الى قيم معيارية :

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

وتشكل المصفوفة X جدول البيانات أو الملاحظات ويمثل ترتيب الأعمدة والصفوف معا 100% من المعلومات.

من هذه المصفوفة نحسب مجموعة جديدة غير مرتبطة ببعضها البعض، حيث تتناقص تبايناتها تدريجيا. يمثل كل عنصر مجموعة خطية من المتغيرات الأصلية (Saporta, 2011).

ويتم تحويل مصفوفة المعطيات الى مصفوفة معيارية لجعل المعطيات متجانسة بحيث تصبح المصفوفة X بالشكل التالي:

$$\begin{pmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1m} - \bar{x}_m \\ \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{nm} - \bar{x}_m \end{pmatrix}$$

من اجل ذلك نحتاج الى حساب المتوسطات الحسابية والانحرافات المعيارية.

ويمكن الاعتماد على معطيات مثال الفصل السابق (مثال 2-01) والذي يمثل علامات ستة طلاب في ثلاث مقاييس هي الاقتصاد، الرياضيات، والاحصاء.

المتوسط الحسابي $\bar{x}=\frac{1}{n}\sum_{i=1}^{n} X_{ij}$

الانحراف المعياري $\partial_j=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{ij}-\bar{X}_J)^2}$

نقوم بتطبيق الصيغ المبينة اعلاه لحساب قيم المتوسطات الحسابية والانحرافات المعيارية، ونحصل على النتائج المبينة في الجدول التالي:

الجدول رقم 3-01: قيم المتوسطات الحسابية والانحرافات المعيارية

الرقم	الاقتصاد	الرياضيات	الاحصاء	$(X_{ij}-\overline{X_J})^2$ الاقتصاد	$(X_{ij}-\overline{X_J})^2$ الرياضيات	$(X_{ij}-\overline{X_J})^2$ الاحصاء
01	12	11	11	1	1	1
02	10	11	12	1	1	0
03	13	12	13	4	4	1
04	10	09	13	1	1	1
05	12	09	12	1	1	0
06	09	08	11	4	4	1
المجموع	**66**	**60**	**72**	**12**	**12**	**4**
$\bar{x}$	11	10	12	2	2	0.667
∂	1.414	1.414	0.816	-	-	-

ملاحظة

الانحراف المعياري للمتغير الثالث هو الأقل وعليه فان السلسلة الإحصائية لهذا المتغير هي الأكثر استقرار أي متوسطها الحسابي يكون اكثر تعبيرا لهذه السلسلة.

2.2.3. حساب مصفوفة احداثيات الافراد ومصفوفة معاملات الارتباط

حتى نحصل على العلاقة الموجودة بين المتغيرات المدروسة نحتاج الى ترجمتها عن طريق مصفوفة التباين المشترك أو مصفوفة الارتباط وللحصول على مصفوفة التباين المشترك نقوم بحساب المصفوفة والتي تمثل احداثيات الافراد ويعبر عنها بالصيغة التالية:

$$\widehat{X_{ij}}=\frac{X_{ij}-\bar{X}_j}{\partial_j}$$

ونحصل على المصفوفة التالية:

$$\widehat{X_{ij}}=\begin{pmatrix} 0.707 & 0.707 & -1.22 \\ -0.707 & 0.707 & 0 \\ 1.414 & 1.414 & 1.22 \\ -0.707 & -0.707 & 1.22 \\ +0.707 & -0.707 & 0 \\ -1.414 & 1.414 & -1.22 \end{pmatrix}$$

بعد ذلك نقوم بحساب مصفوفة معاملات الارتباط C كما يلي:

$$C=\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

ملاحظة

تجدر الاشارة الى أن $r_{jj}=1$ هذا يعني ان: $r_{11}=r_{22}=r_{33}=1$ ، اذن عناصر القطر تساوي الواحد. بالإضافة الى ذلك فان: $r_{jk}=r_{kj}$ أي يمكننا القول مثلا $r_{12}=r_{21}$

نقوم بحساب قيم معاملات الارتباط r_{jk} بالطريقة التالية:

$$r_{jk}=\frac{1}{n}\frac{\sum(X_{ij}-\bar{X}_j)(X_{ik}-\bar{X}_k)}{\partial_j\partial_k}$$

من معطيات الجدول نجد:

1	2	3	1*1	1*2	1*3	2*2	2*3	3*3
1	1	-1	1	1	-1	1	-1	1
-1	1	0	1	-1	0	1	0	0
2	2	1	4	4	2	4	2	1
-1	-1	1	1	1	-1	1	-1	1
1	-1	0	1	-1	0	1	0	0
-2	-2	-1	4	4	2	4	2	1
0	0	0	12	8	2	12	2	4

حيث : $1= X_{ij}-\overline{X}_J$ ؛ $2= X_{ij}-\overline{X}_J$ ؛ $3= X_{ij}-\overline{X}_j$

مثلا بحساب معاملات الارتباط r_{11} و r_{12} نجد:

$$r_{11} = \frac{1}{6}\frac{\sum(r_{11}-\bar{r})(r_{11}-\bar{r})}{\partial_1\partial_1}=\frac{1}{6}\frac{12}{1.414*1.414}=1$$

$$r_{12} = \frac{1}{6}\frac{\sum(r_{11}-\bar{r})(r_{22}-\bar{r})}{\partial_1\partial_2}=\frac{1}{6}\frac{8}{1.414*1.414}=0.667$$

$$r_{13} = \frac{1}{6}\frac{\sum(r_{11}-\bar{r})(r_{33}-\bar{r})}{\partial_1\partial_3}=\frac{1}{6}\frac{2}{1.414*0.8\ 16}=0.289$$

بنفس الطريقة نحسب باقي القيم لنتحصل على المصفوفة التالية:

$$C = \begin{pmatrix} 1 & 0.667 & 0,289 \\ 0.667 & 1 & 0.289 \\ 0.289 & 0.289 & 1 \end{pmatrix}$$

نلاحظ ان كل القيم محصورة بين 1 و -1

3.2.3. تحديد القيم الذاتية والاشعة الذاتية

سنقوم أولا بالتذكير بخصائص القيم الذاتية للمصفوفة:

- نقول عن λ انها قيمة ذاتية للمصفوفة C اذا وفقط اذا كان $|C - \lambda I_p| = 0$، حيث I_p هو عنصر الوحدة بالنسبة لعملية الجداء
- نقوم بحساب القيم الذاتية لقيم المصفوفات المربعة فقط وترتيبها ترتيبا تنازليا
- مجموع القيم الذاتية = مجموع عناصر القطر للمصفوفة C ونكتب: $\sum_{i=1}^{n}\lambda i = trace\ C = P$

حساب $|C - \lambda I_p| = 0$

$$\left|\begin{pmatrix} 1 & 0.667 & 0.289 \\ 0.667 & 1 & 0.289 \\ 0.289 & 0.289 & 1 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right| = 0$$

$$\left|\begin{pmatrix} 1-\lambda & 0.667 & 0.289 \\ 0.667 & 1-\lambda & 0.289 \\ 0.289 & 0.289 & 1-\lambda \end{pmatrix}\right| = 0$$

بعد الحساب نجد: $\lambda_1=1.86$ ؛ $\lambda_2=0.81$ ؛ $\lambda_3=0.33$

لحساب الاشعة الذاتية نقوم بتعويض كل قيمة من قيم الاشعة الذاتية التي تم حسابها ونجد:

ملاحظة

نقول عن الشعاع μ_n حيث $\mu_n=\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}$ شعاع ذاتي لـ C المرفوقة بالقيم الذاتية λ_n ، اذا وفقط اذا كانت :

$$\left(C-\lambda_n I_p\right)\mu_n=0$$

عند $\lambda_1=1.86$ فان $\left(C-\lambda_1 I_3\right)\mu_1=0$

$$\left[\begin{pmatrix}1 & 0.667 & 0.289\\ 0.667 & 1 & 0.289\\ 0.289 & 0.289 & 1\end{pmatrix}-1.86\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right]\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

$$\left[\begin{pmatrix}1 & 0.667 & 0.289\\ 0.667 & 1 & 0.289\\ 0.289 & 0.289 & 1\end{pmatrix}-\begin{pmatrix}1{,}86 & 0 & 0\\ 0 & 1{,}86 & 0\\ 0 & 0 & 1{,}86\end{pmatrix}\right]\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

$$-0.86\mu_1+0.667\mu_2+0.289\mu_3=0 \quad (1)$$
$$0.667\mu_1-0.86\mu_2+0.289\mu_3=0 \quad (2)$$
$$0.289\mu_1+0.289\mu_2-0.86\mu_3=0 \quad (3)$$

لو نطرح العلاقة 1 من العلاقة 2 نجد : $-1.527\mu_1+1.527\mu_2=0$ معناه $\mu_1=\mu_2$

لو نضرب العلاقة 2 في 0,775 ثم نجمعها مع العلاقة 1 نحصل على : $-0.343\mu_1+0.513\mu_3=0$

معناه $\mu_3=0.67\mu_1$

في الأخير نجد:

$$\mu_1=\begin{pmatrix}0.639\\ 0.639\\ 0.429\end{pmatrix}$$

بنفس الطريقة عند كل من القيمتين الذاتيتين المتبقيتين: 081 و 0.33

نجد على التوالي: $\mu_2=\begin{pmatrix}-0.303\\ -0.303\\ 0.906\end{pmatrix}$ و $\mu_3=\begin{pmatrix}0.707\\ -0.707\\ 0\end{pmatrix}$

ان تحليل المصفوفة C يعطي لنا p قيمة ذاتية مرتبة ترتيبا تنازليا و p اشعة ذاتية متعامدة فيما بينها في الفضاء.

حيث ان المحور الأول F1 هو احسن محور على الاطلاق حيث يحتوي على اكبر نسبة من المعلومات ويعطي احسن تمثيل بياني وشعاع التوجيه لهذا المحور هو μ_1 (شعاع التوجيه للمحور F2 هو μ_2، الخ).

تجدر الاشارة هنا الى انه بعد الحصول على القيم الذاتية و الاشعة الذاتية لمصفوفة الارتباط يتم تحديد المركبات الاساسية. ويتم تحديد عدد المركبات الاساسية كأقصى عدد ممكن والذي يكون مساويا لعدد المتغيرات في مصفوفة الارتباط . كما انه يتم اجراء اختبارات صلاحية عينة المعطيات المدروسة من خلال وجود ارتباطات كافية ما بين المتغيرات للظاهرة المدروسة ومن بين هذه الاختبارات:

الجدول رقم 3-02: أهم اختبارات الصلاحية

الاختبار	الدور
اختبار بارتلت Bartlett	بفحص مصفوفة الارتباط و اختبار فرضية العدم
اختبار كيزر-ماير-أولكن KMO	يبين هذا الاختبار ان الارتباطات الجزئية بين المتغيرات ليست ضعيفة وتكون قيمة هذا المعيار محصورة بين 0.3 و 0.7

أما فيما يخص المعايير المستخدمة في تحديد عدد المركبات الاساسية يتم اختيار هذه المعايير طبقا لنسبة التباين المفسر من قبل العوامل والتي تبين مقدار المحافظة على المعلومات من العوامل الاساسية ويوجد ثلاث معايير أساسية:

- معيار القيم الذاتية (قاعدة guttman-kaiser): تمثل القيم الذاتية كمية المعلومات الملتقطة من طرف العامل بحيث يتم اختيار العوامل التي لها قيمة ذاتية أكبر.

- اختبار الانعطاف (coude de test) يتم اختيار العوامل التي تقع قبل مرحلة الانعطاف الذي يليه انخفاض مستمر من منحنى القيم الذاتية.

- معيار قيمة التباين المشروح: يتم ملاحظة قيمة التباين المشروح من قبل العوامل وهذا من اجل ضمان أن العوامل المستخلصة تشرح بأكبر معنوية ممكنة للتباين تكون محددة مسبقا.

ملاحظة

قبل اجراء تفسير النتائج للمركبات الاساسية عادة ما يتم اجراء **تدوير للمحاور** من أجل التعرف على الارتباطات القوية بين المتغيرات. ويهدف تدوير المحاور الى تقليل المسافة بين المتغيرات و المحاور بحيث أن كل عنصر يكون مرتبط فقط بمركبة أساسية واحدة. مع العلم أن التدوير لا يمكنه تغيير نسبة التباين المفسر بالمركبات الاساسية قبل التدوير ويكون التدوير اما متعامدا أو مائلا بحيث تكون العوامل في التدوير المتعامد غير مرتبطة وهذا ما يقود الى تفسير جيد للنتائج على عكس التدوير المائل و التي يكون فيه ارتباط العوامل مع المتغيرات مما يجعل التفسير أكثر صعوبة. ويكون التحليل الى مركبات أساسية المتعامد بأحد الاشكال التالية: varimax، quartimax

4.2.3. تمثيل الافراد والمتغيرات على المحاور

نسمي Z_λ نسبة التمثيل على المحور حيث:

المحاور	القيم الذاتية	نسب التمثيل	النسب التجميعية
1	1.86	$\frac{1.86}{3}.100 = 62$	62
2	0.81	$\frac{0.81}{3}.100 = 27$	89
3	0.33	$\frac{0.33}{3}.100 = 11$	100
المجموع	3	100	-

من الجدول نقول ان 62% من البيانات ممثلة على المحور الأول كما ان 27% من البيانات ممثلة على المحور الثاني. ويمكننا القول أيضا ان 89% من البيانات (62+27) ممثلة على المستوى الأول وهي نسبة معتبرة يمكن الاعتماد عليها في التحليل والدراسة.

والان سنقوم بحساب احداثيات كل الافراد على المحور بالطريقة التالية:

$$F_n - \hat{X}^* \mu_n \qquad F_1 - \hat{X} \mu_1$$

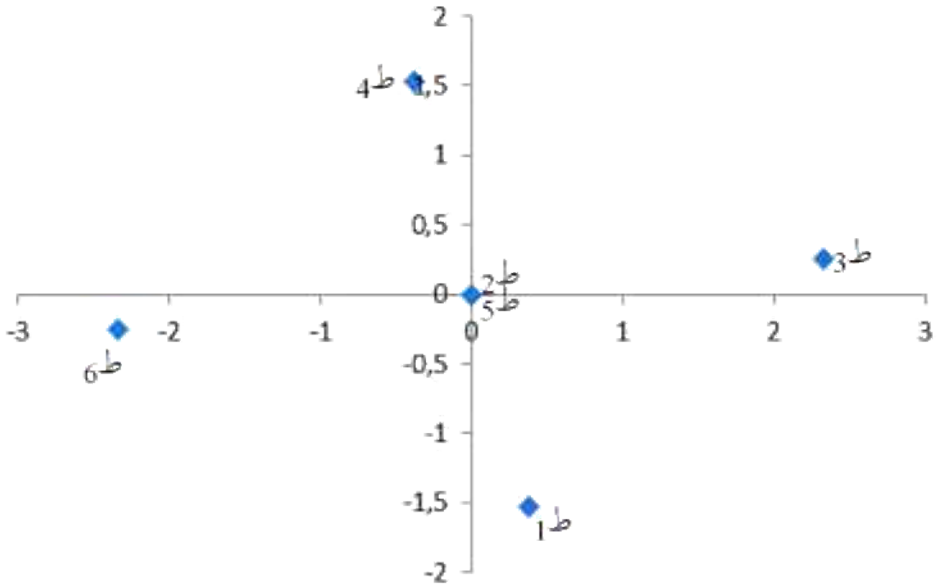
$$F_1 - \begin{pmatrix} 0.707 & 0.707 & -1.22 \\ -0.707 & 0.707 & 0 \\ 1.414 & 1.414 & 1.22 \\ -0.707 & -0.707 & 1.22 \\ +0.707 & -0.707 & 0 \\ -1.414 & 1.414 & -1.22 \end{pmatrix} \begin{pmatrix} 0.639 \\ 0.639 \\ 0.429 \end{pmatrix}$$

$$F_1 - \begin{pmatrix} 0.378 \\ 0 \\ 2.332 \\ -0.378 \\ 0 \\ -2.332 \end{pmatrix}$$

بنفس الطريقة نحسب F2و F3 ونبوب النتائج في الجدول التالي :

F_3	F_2	F_1	الافراد
0	-1.535	0.378	**1**
-1	0	0	**2**
0	0.249	2.332	**3**
0	1.535	-0.378	**4**
1	0	0	**5**
0	-0.249	-2.332	**6**

ومن ثم نقوم بتمثيل الافراد على المستوى الاول: (F_1، F_2)



ملاحظة

كلما كانت النقاط بعيدة عن المركز تتزايد جودة تمثيلها على المحور والعكس. مثلا: نقول ان الفرد 3 و6 ممثلين احسن تمثيل على المحور الأول بينما الفردين 1 و4 ممثلين احسن تمثيل على المحور الثاني

بعد ذلك نقوم بحساب G والتي تعبر عن احداثيات كل المتغيرات على المحور حيث: $G_n = \sqrt{\lambda_n} * \mu_n$

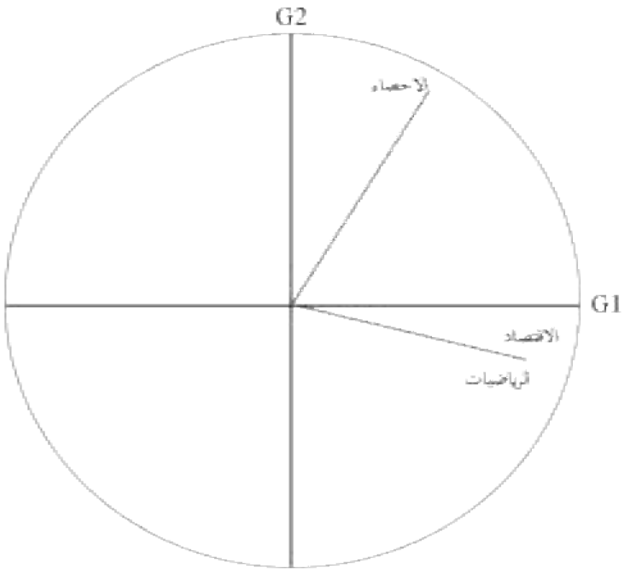
$$G_1 = \sqrt{\lambda_1} * \mu_1 = \sqrt{1.86} * \begin{pmatrix} 0.639 \\ 0.639 \\ 0.429 \end{pmatrix}$$

$$G_1 = \begin{pmatrix} 0.871 \\ 0.871 \\ 0.584 \end{pmatrix}$$

بنفس الطريقة نحسب بقية القيم ونلخص النتائج في جدول:

المتغيرات	G_1	G_2	G_3
الاقتصاد	0.871	-0.272	0.407
الرياضيات	0.871	-0.272	-0.407
الاحصاء	0.584	0.815	0

وللتعليق على هذا الجدول نقول انه مثلا 0.871 هي معامل الارتباط البسيط لمتغيرة الاقتصاد على المحور الأول ونمثل ذلك بيانيا كما يلي:



ملاحظة

-كل المتغيرات تقع على سطح دائرة مركزها g ونصف قطرها 1 و لا يمكن ان تخرج عن هذا السطح.

-كلما كانت المتغيرات بعيدة عن المركز تزداد جودة تمثيلها.

-بالنسبة لهذا المثال نلاحظ ان كل المتغيرات بعيدة عن المركز وبالتالي فهو ذا جدوة عالية ومقبولة في الدراسة

3.3. التحليل الى مركبات اساسية باستخدام برنامج SPSS

فيما يلي نوضح كيفية تنفيذ تحليل PCA باستخدام برنامج SPSS ، بالإضافة إلى الخطوات التي نحتاج إلى اتباعها لتفسير النتائج. ومع ذلك ، قبل أن نقدم على هذا الإجراء ، تحتاج إلى فهم الفرضيات المختلفة التي يجب أن تفي بها البيانات التي سنقوم بدراستها حتى منحنا تحليل PCA نتيجة صحيحة.

حيث عندما نختار تحليل البيانات باستخدام طريقة PCA، فإن هذه العملية تتطلب التحقق من إمكانية تحليل البيانات المراد تحليلها باستخدام هذه الطريقة. لذا لا بد من التحقق من النقاط التالية:

- وجود متغيرات متعددة يجب قياسها، حيث تتضمن أمثلة المتغيرات المستخدمة بشكل شائع في PCA مقاييس ليكرت.

- يجب أن تكون هناك علاقة خطية بين جميع المتغيرات وهذا يرجع الى كون تحليل PCA يعتمد على معاملات ارتباط بيرسون ، وعلى هذا النحو ، يجب أن تكون هناك علاقة خطية بين المتغيرات.
- لكي تنتج PCA نتيجة موثوقة يجب الاعتماد على حجم كاف لعينة الدراسة.
- يجب أن يكون هناك ارتباطات كافية بين المتغيرات من أجل تقليل المتغيرات إلى عدد أقل من المركبات.

بعد التحقق من العناصر السابقة يمكننا تحليل معطياتنا باستخدام برنامج SPSS وبالاعتماد على المثال التالي:

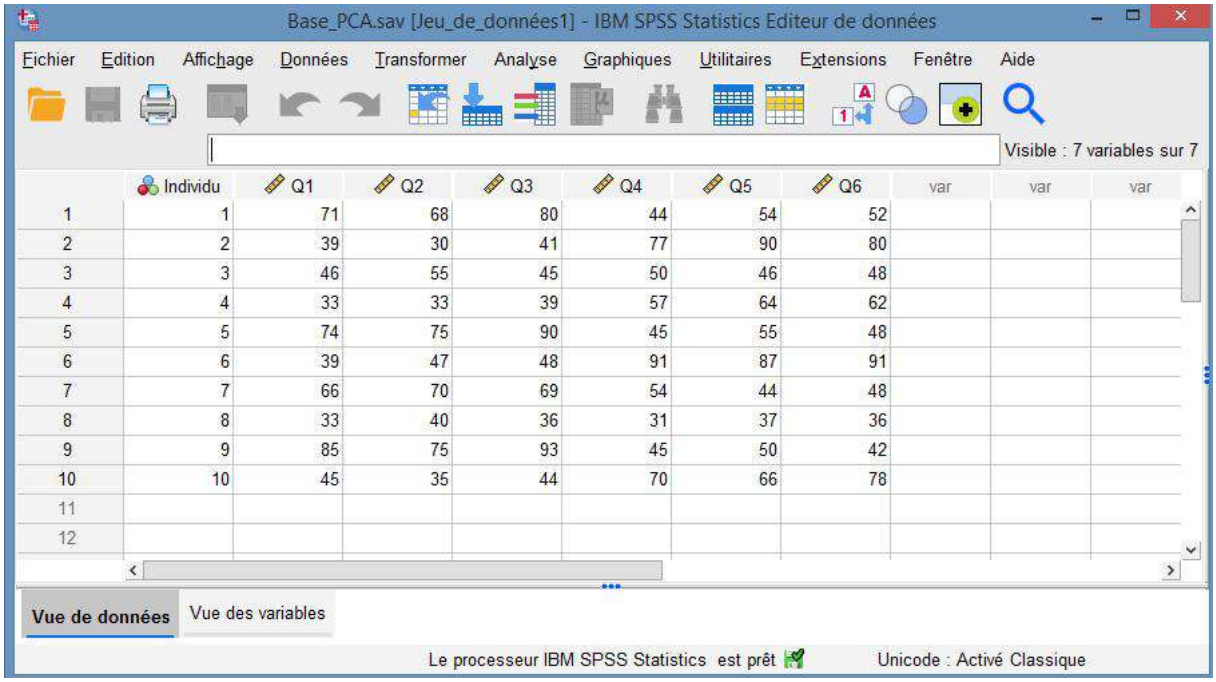
1.3.3. ادراج البيانات في الجدول

أراد مدير مؤسسة معينة تعيين موظف آخر لشركته وكان يبحث عن شخص يظهر مستويات عالية من التحفيز والاعتمادية والحماس والالتزام. من أجل اختيار المرشحين للمقابلة، أعد استبيان يتكون من 5 اسئلة يعتقد أنه قد يجيب عما إذا كان لديه المرشح المناسب. اعطى هذا الاستبيان لـ 10 مرشحا محتملا، وقد تمت صياغة الأسئلة بحيث يتم تمثيل هذه الصفات في الأسئلة مع نطاق تسجيل من 0 إلى 100. وكانت النتائج كما يلي:

الافراد	Q1	Q2	Q3	Q4	Q5	Q6
1	71	68	80	44	54	52
2	39	30	41	77	90	80
3	46	55	45	50	46	48
4	33	33	39	57	64	62
5	74	75	90	45	55	48
6	39	47	48	91	87	91
7	66	70	69	54	44	48
8	33	40	36	31	37	36
9	85	75	93	45	50	42
10	45	35	44	70	66	78

سوف نقوم بإجراء تحليل الى مركبات اساسية لتحديد عدد العوامل المهمة الموجودة في البيانات، بالإضافة الى ذلك سوف نوضح إلى أي مدى يمكن لهذه العوامل أن تشرح الارتباطات بين المتغيرات. كذلك سيتم تدوير العوامل حتى نجعل التحليل والتفسير أكثر وضوحا. وفي الاخير سنحاول تحديد وتسمية هذه المركبات.

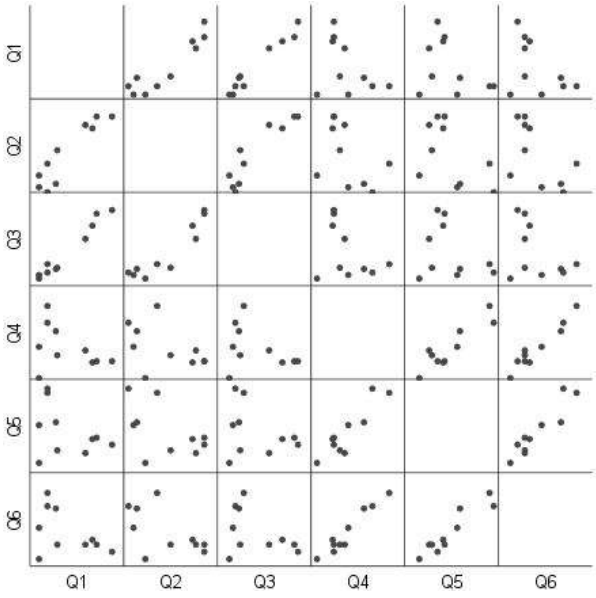
قبل بدأ عملية التحليل سنقوم ادخال البيانات في برنامج SPSS كما في الشكل الموالي:



	Individu	Q1	Q2	Q3	Q4	Q5	Q6	var	var	var
1	1	71	68	80	44	54	52			
2	2	39	30	41	77	90	80			
3	3	46	55	45	50	46	48			
4	4	33	33	39	57	64	62			
5	5	74	75	90	45	55	48			
6	6	39	47	48	91	87	91			
7	7	66	70	69	54	44	48			
8	8	33	40	36	31	37	36			
9	9	85	75	93	45	50	42			
10	10	45	35	44	70	66	78			
11										
12										

باستخدام تحليل الارتباط يمكننا البحث عن الارتباطات الموجودة بين مختلف المتغيرات والغرض هنا هو التحقق من عدم وجود متغيرات مترابطة تماما (والتي قد تحتاج إلى إزالة).

تم انشاء الإخراج التالي بواسطة الأمر Graphiques و Dispersions/Points. كما يمكننا ايضا انشاء مصفوفة الارتباطات للتحقق.



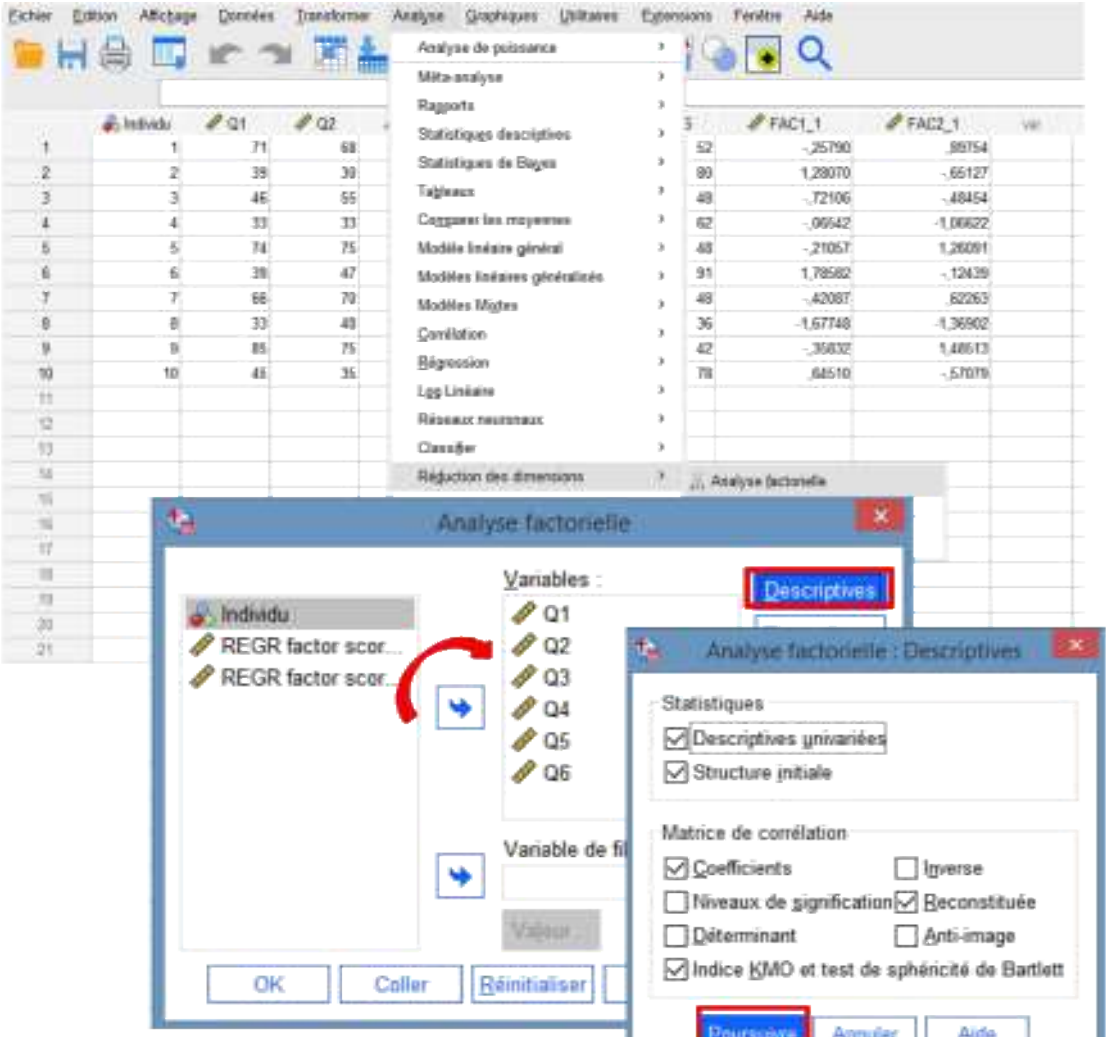
		Q1	Q2	Q3	Q4	Q5	Q6
Corrélation	Q1	1,000					
	Q2	,921	1,000				
	Q3	,979	,921	1,000			
	Q4	-,389	-,461	-,366	1,000		
	Q5	-,365	-,508	-,301	,905	1,000	
	Q6	-,462	-,569	-,425	,967	,934	1,000

ما يمكن ملاحظته هو أن معظم الارتباطات أعلى بكثير من 0.3 وهذا مؤشر جيد يدل على أننا سنحصل على نتيجة. كما يبدو من مصفوفة الارتباطات أن هناك مجموعتين من المتغيرات (باللون الرمادي). المتغيرات Q1، Q2 و Q3 في مجموعة اولى والمتغيرات Q4، Q5 و Q6 في المجموعة الثانية.

وبالتالي يمكننا التعمق اكثر في ذلك من خلال تحليل PCA.

2.3.3. إجراء التحليل الى مركبات اساسية

من اجل ذلك ننقر في برنامج SPSS على Analyse ثم Réduction des dimensions ثم Analyse factorielle لفتح مربع حوار تحليل العوامل.



ثم ننقل المتغيرات الستة إلى مربع المتغيرات

ننقر على Descriptives

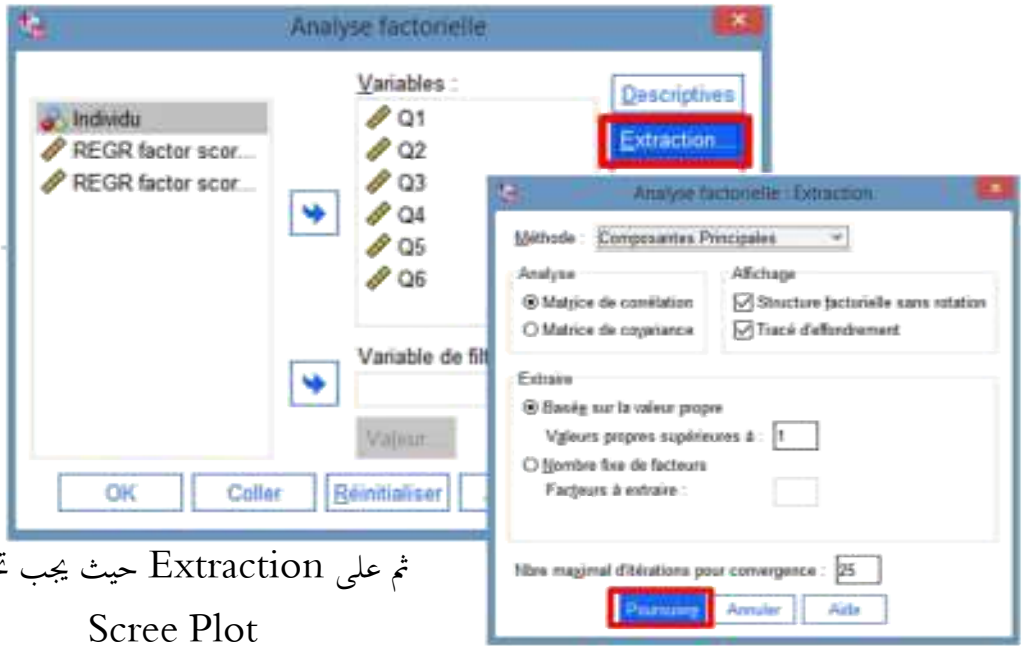
ثم نختار :

Descriptives Univariées

Coefficients, Reconstituée

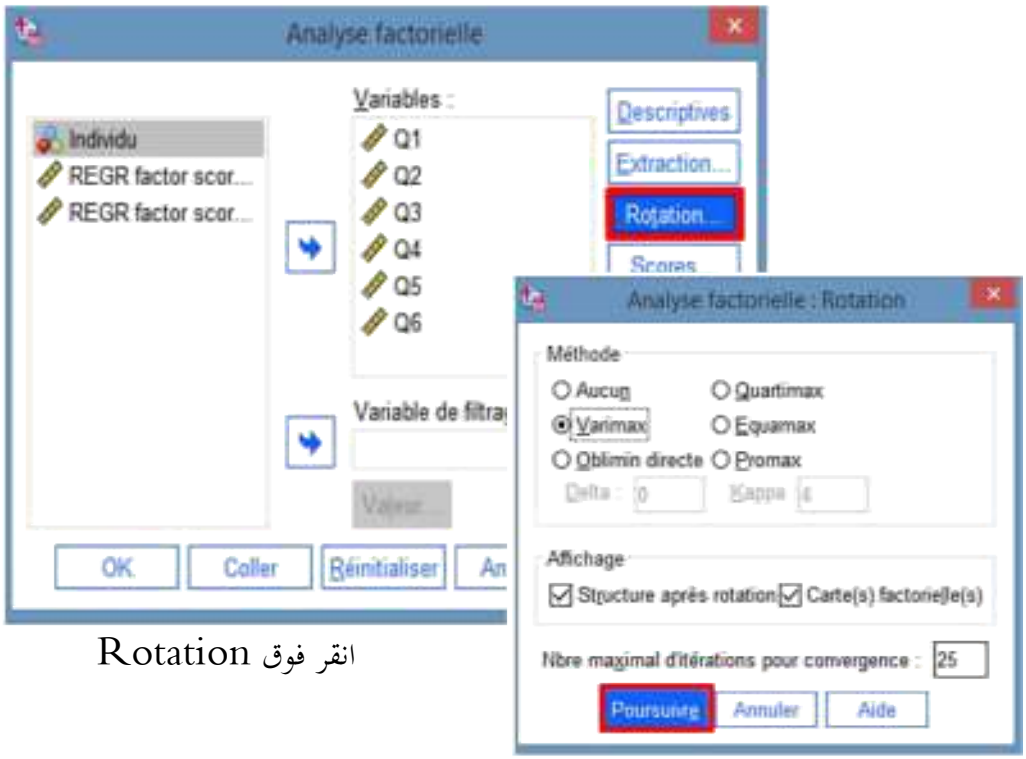
Indice KMO

ثم ننقر علىPoursuivre



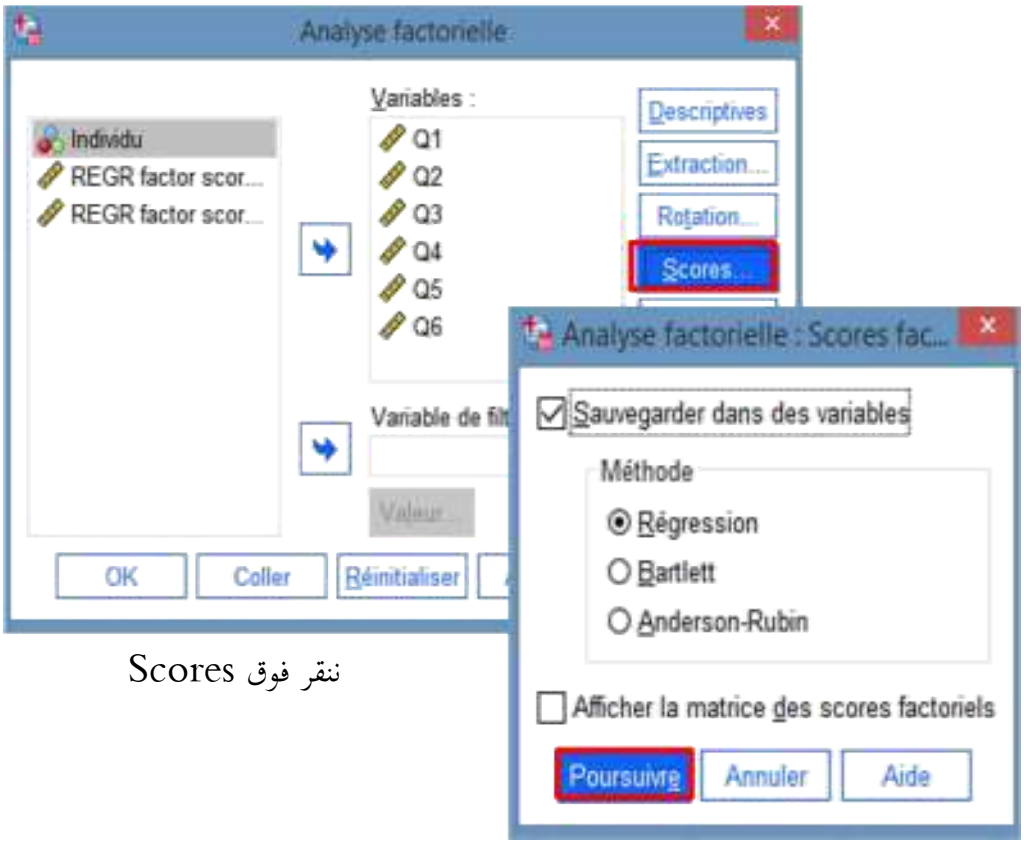
ثم على Extraction حيث يجب تحديد Scree Plot

بعد التأكد من أن الطريقة المختارة هي المكونات الرئيسية وأن التحليل سيتم إجراؤه على مصفوفة الارتباط بالإضافة الى استخراج العوامل ذات القيم الذاتية التي تزيد عن 1



انقر فوق Rotation

ثم يجب تحديد Varimax و Cartes factorielles ثم ننقر على Poursuivre



ننقر فوق Scores

ثم على النتائج لتحديد نوع درجات العوامل التي تريد حفظها في مجموعة البيانات ، وهنا نحدد Régression

ثم ننقر على Poursuivre

يعد اكمال كل هذه العمليات سيقوم برنامج SPSS بإخراج نتائج التحليل

3.3.3. تفسير المخرجات و تحليل النتائج

الجدول الموالي يقدم وصف للمتغيرات من خلال مؤشرات الاحصاء الوصفي والمتمثلة في المتوسطات الحسابية والانحرافات المعيارية وحجم العينة لكل متغير. نلاحظ أن قيم المتوسطات الحسابية لكل المتغيرات متقاربة ولها تشتت متقارب نسبيا.

	المتوسط الحسابي	الانحراف المعياري	N
Q1	53,10	19,041	10
Q2	52,80	18,085	10
Q3	58,50	22,237	10
Q4	56,40	17,989	10
Q5	59,30	17,695	10
Q6	58,50	18,447	10

بالنسبة لمصفوفة معاملات الارتباط فقد سبق أن ادرجناها، وقد ساعدنا قيم معاملات الارتباط في استنتاج بعض العلاقات بين المتغيرات.

يمكننا ايضا التعقيب على قيم KMO و اختبار Bartlett والتي تشير قيمها انه يمكننا اجراء تحليل PCA.

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage		,763
Test de sphéricité de Bartlett	Khi-carré approx.	73,582
	Ddl	15
	Signification	<,001

وتظهر قيم KMO واختبار Bartlett على التوالي ان عناصر العينة كافية لإجراء الدراسة و انه لا يوجد معاملات ارتباط معدومة.

وفيما يخص نوعية التمثيل فنلاحظ أن تمثيل كل المتغيرات اكبر من 0.4 وبالتالي لا نحذف اي متغير وكل المتغيرات ستدخل في عملية التحليل (Jolliffe, 2002).

	Initiales	Extraction
Q1	1,000	,976
Q2	1,000	,942
Q3	1,000	,982
Q4	1,000	,954
Q5	1,000	,942
Q6	1,000	,980
Méthode d'extraction : Analyse en composantes principales.		

بعد ذلك لا بد من تحديد القيم الذاتية لكل متغيرات الدراسة والتي عددها ستة متغيرات.

من بين مخرجات برنامج SPSS نجد جدول يوضح أهمية كل مركب من المركبات الاساسية. الأولين فقط لهما قيم ذاتية تزيد عن 1.00.

من الجدول نلاحظ انه تم اختزال الجدول الى محورين حيث يضم المحور الاول ثلاث متغيرات بحجم معلومات قدره 48.25 % والمحور الثاني يضم ايضا ثلاث متغيرات بحجم معلومات يقدر بـ 48%. ومنه فان المحورين معا اختزلا ما قيمته 96.25% من حجم المعلومات.

الجدول رقم 3-03: جدول تفسير التباين

Composante	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	4,164	69,397	69,397	4,164	69,397	69,397	2,895	48,251	48,251
2	1,612	26,862	96,259	1,612	26,862	96,259	2,881	48,008	96,259
3	,144	2,396	98,655						
4	,052	,867	99,522						
5	,023	,383	99,905						
6	,006	,095	100,000						

Méthode d'extraction : Analyse en composantes principales.

نفس النسبة الاجمالية ولكن القسم مختلف جدا (⇕)

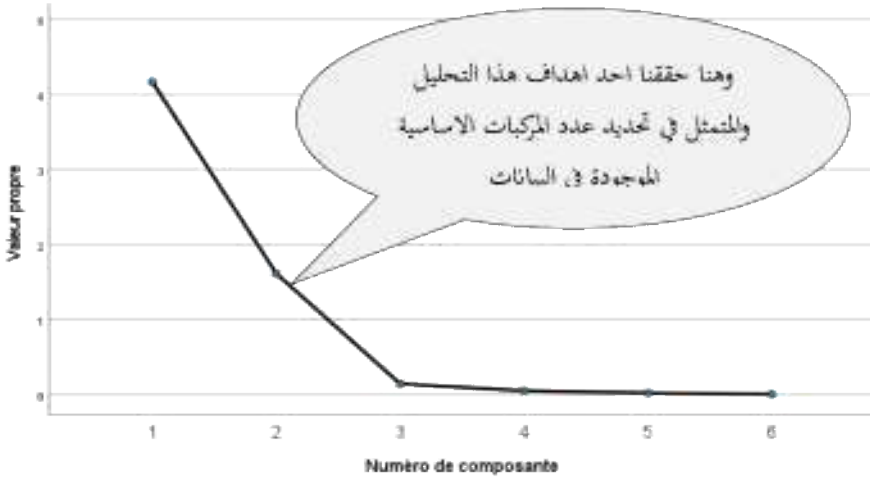
ملاحظة

تجدر الاشارة الى انه في أعمدة قيم eigenvalues الأولية، يتم عرض قيم eigenvalues، ونسبة التباين التي تستطيع العوامل تفسيرها ، ونسبة التباين المفسر في كل عامل. في الأعمدة الثلاثة التالية، تتكرر فقط قيم العوامل المحفوظة في التحليل بعد الاستخراج ويتم حذف قيم العوامل المستبعدة. في العمود الأخير هي القيم الذاتية للعوامل بعد الدوران. يعمل الدوران على تحسين هيكل العامل ونتيجة لذلك تم معادلة الأهمية النسبية للعوامل المتبقية.

يوضح الجزء الأوسط من الجدول قيم eigenvalues ونسبة التباين الموضحة لعاملي الحل الأولي فقط اللذين يعتبران مهمين. من الواضح أن العامل الأول أكثر أهمية من الثاني، ومع ذلك في الجزء الأيمن من الجدول، يتم عرض قيم eigenvalues ونسبة التباين الموضحة للعوامل التي تم تدويرها (Carricano et al., 2010).

وفي حين ان العاملان يفسران معا نفس التباين فإن تقسيم الأهمية بين العاملين مختلف تماما، ويظهر هنا تأثير التدوير في توزيع الأهمية بشكل متساو إلى حد ما بين العاملين. من الجدول نلاحظ أن قيم eigenvalues للعاملين هي على التوالي 2.895 و 2.881 ، مقارنة بـ 4.164 و 1.612 في الحل الأولي. وهذا ما يوضح مدى أهمية استخراج عدد مناسب من العوامل.

يتدعم هذا الاستنتاج من خلال مخطط الانعطاف (coude) والذي يعرض في الواقع نفس البيانات بشكل أكثر وضوحا:



يقدم الجدول التالي معلومات حول مدى دوران العوامل. في هذه الحالة تم تدوير العوامل بمقدار 45 درجة، ويمكن حساب الزاوية بمعالجة معامل الارتباط باعتباره جيب التمام (جيب تمام 45 درجة هو 707.).

Composante	1	2
1	-,709	,705
2	,705	,709
Méthode d'extraction : Analyse en composantes principales. Méthode de rotation : Varimax avec normalisation Kaiser.		

أما والجدول الموالي يوضح احداثيات المتغيرات قبل وبعد التدوير.

الجدول رقم 3-04: احداثيات المتغيرات قبل وبعد التدوير

قبل التدوير				بعد التدوير		
	Composante				Composante	
	1	2			1	2
Q1	,824	,545		Q1	-,200	,967
Q2	,878	,415		Q2	-,330	,913
Q3	,799	,586		Q3	-,154	,979
Q4	-,818	,533		Q4	,956	-,199
Q5	-,804	,544		Q5	,953	-,181
Q6	-,873	,467		Q6	,948	-,284
Méthode d'extraction : Analyse en composantes principales.				Méthode d'extraction : Analyse en composantes principales. Méthode de rotation : Varimax avec normalisation Kaiser.		
a. 2 composantes extraites.				a. Convergence de la rotation dans 3 itérations.		

يوضح الجدول التالي إلى أي مدى يمكن إعادة إنتاج مصفوفة الارتباط الأصلية من عاملين:

الجدول رقم 3-05: نتائج الارتباطات

Corrélations reconstituées							
		Q1	Q2	Q3	Q4	Q5	Q6
Corrélation reconstituée	Q1	,976[a]	,949	,978	-,383	-,365	-,464
	Q2	,949	,942[a]	,944	-,497	-,479	-,572
	Q3	,978	,944	,982[a]	-,341	-,323	-,423
	Q4	-,383	-,497	-,341	,954[a]	,948	,963
	Q5	-,365	-,479	-,323	,948	,942[a]	,956
	Q6	-,464	-,572	-,423	,963	,956	,980[a]
Résidu[b]	Q1		-,028	,002	-,006	,000	,002
	Q2	-,028		-,023	,036	-,028	,003
	Q3	,002	-,023		-,025	,022	-,002
	Q4	-,006	,036	-,025		-,042	,003
	Q5	,000	-,028	,022	-,042		-,022
	Q6	,002	,003	-,002	,003	,022	
Méthode d'extraction : Analyse en composantes principales.							
a. Qualités de représentation reconstituées							
b. Les résidus sont calculés d'après les corrélations observées et reconstituées. Il existe 0 [illegible] ndants avec des valeurs absolues supérieures à 0.05.							

وهذا نكون قد حققنا هدفا اخر لهذا التحليل والمتمثل في دراسة إلى أي مدى يمكن للمركبات تفسير الارتباطات بين المتغيرات

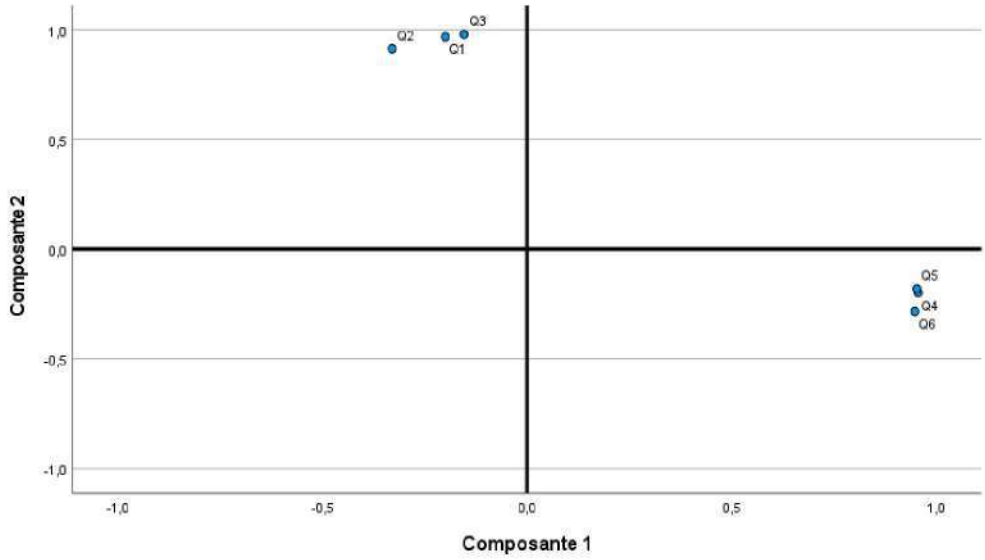
وهذا نكون قد حققنا هدفا اخر لهذا التحليل والمتمثل في دراسة إلى أي مدى يمكن للمركبات تفسير الارتباطات بين المتغيرات

تظهر القيم المتبقية الصغيرة أن هناك اختلاف بسيط جدا بين الارتباطات المستنسخة والارتباطات التي لوحظت بالفعل بين المتغيرات مما يوفر ملخص دقيق للعلاقات في البيانات.

كما ان عملية تحليلي PCA باستخدام برنامج SPSS ينتج عنه مخطط يوضح المتغيرات الستة على المحاور التي تمثل العاملين. ويبدو من المعقول أن نحدد مبدئيا العامل المدور الأول على أنه Fact1 ، حيث يضم المتغيرات الثلاث الاولى. ويبدو ان العوامل المدورة الثانية تمثل العامل Fact2 ;والذي يضم المتغيرات المتبقية.

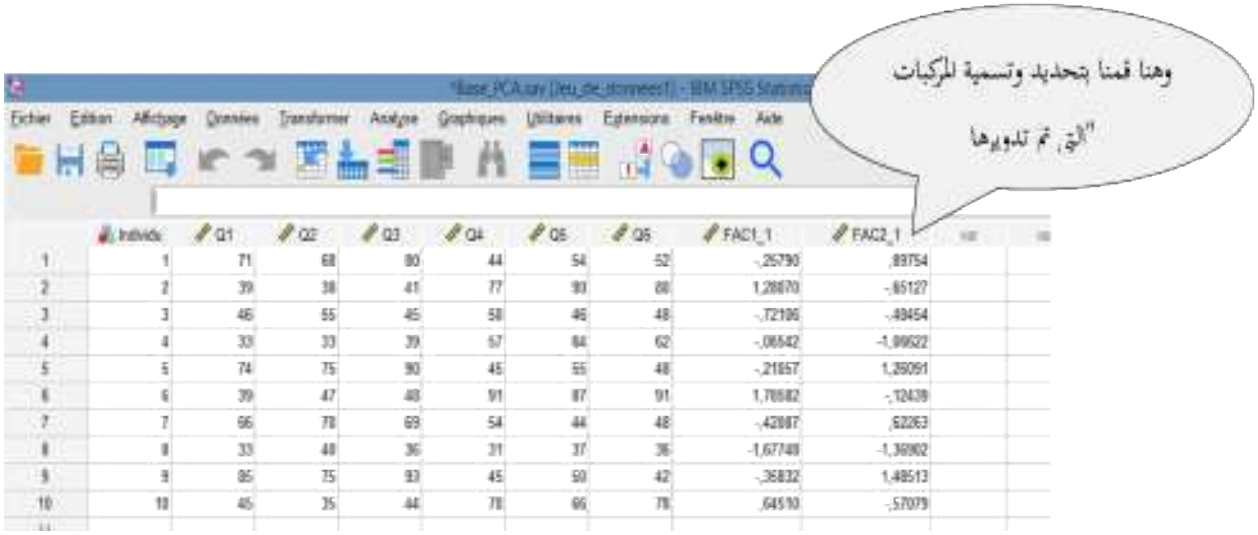
و يتم الاستعانة بهذا المخطط لتسمية المحاور مع مراعاة ان المتغيرات التي نقع في محيط الدائرة يساوي الواحد تعتبر ذات تمثيل جيد يعكس المتغيرات القريبة من مركز الدائرة (Labatte, 2012).

الشكل رقم 3-04: منحنى المكونات بعد التدوير

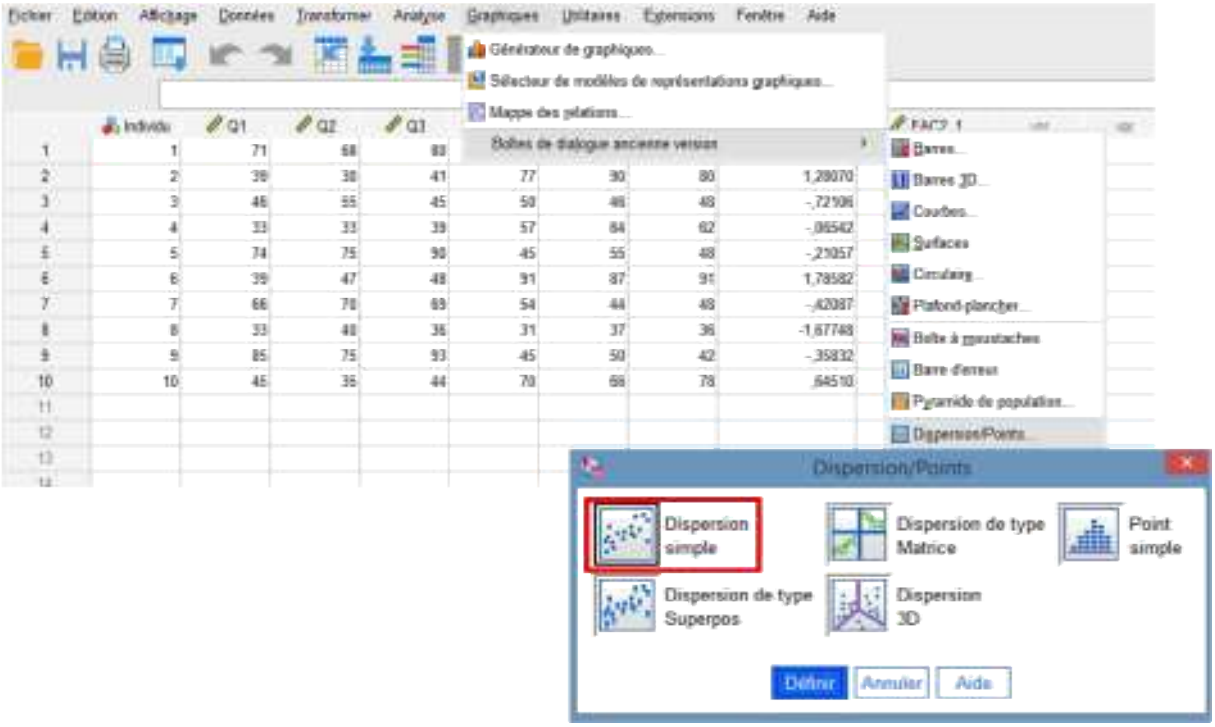


وتمت إضافة درجات العوامل إلى البيانات، كما سنرى في جدول الموالي. هذه هي الدرجات المعيارية التي تم الحصول عليها من خلال تطبيق تحميل عامل الدوران على الدرجة الموحدة لكل مشارك في كل من المتغيرات (تماما مثل إجراء التنبؤ باستخدام معادلة الانحدار).

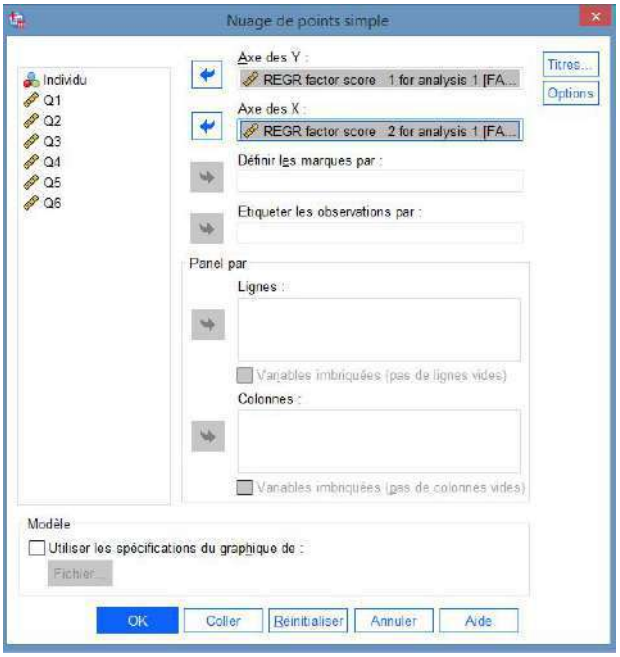
المشترك الثامن (8) لديه درجة معيارية منخفضة على العامل الأول (-1.68) وبالتالي يمكن القول أنهلا يؤثر في العامل الاول (Fact1). نفس المشارك لديه أيضا درجة معيارية منخفضة ايضا على العامل الثاني (-1.37). من ناحية أخرى ، حصل المشارك 6 على درجات عالية (1.79) في العامل الاول ، ولكنه حصل على درجة قريبة من المتوسط (-0.12) في العامل الثاني.



وكخطوة اخيرة، بعد التعامل مع المتغيرات الستة وتكوين المحاور وتسميتها (استخراج المركبات) يتم اسقاط الافراد (المترشحين في هذه الحالة) لدراستها لذا سنستعين بالتمثيل البياني باتباع الخطوات التالية:

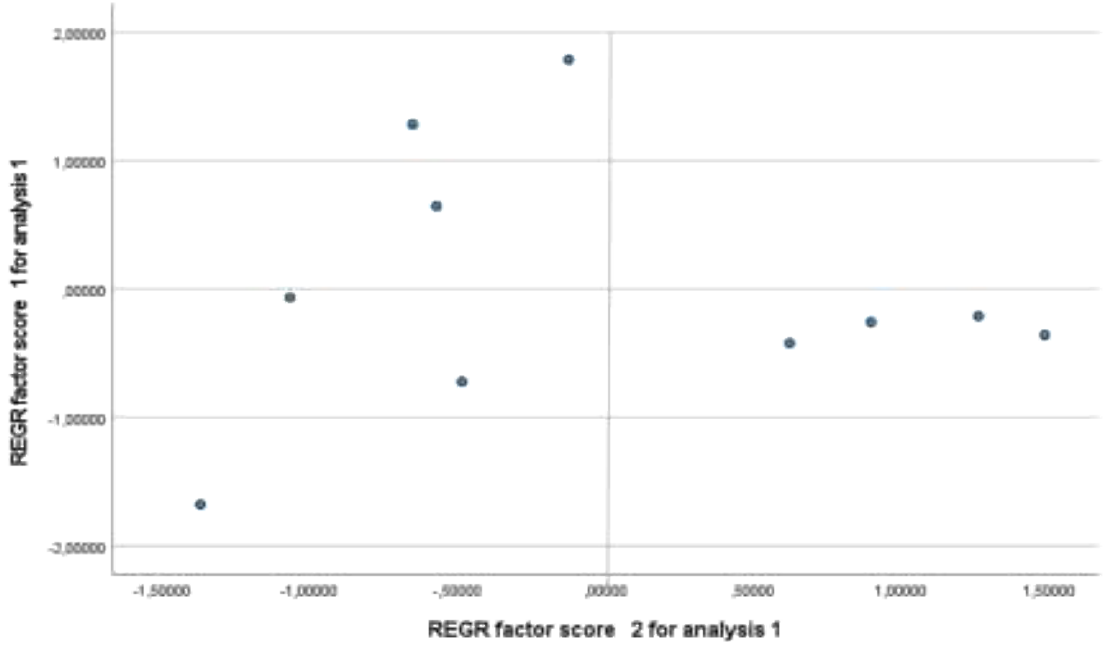


ومن صندوق الحوار نختار الانتشار البسيط، وبعدها نقوم بإدخال المحاور كما يلي:



وبعد الضغط على OK نحصل على المنحنى التالي:

الشكل رقم 3-05: تمثيل الافراد



ومن خلال هذا المنحنى يمكن دراسة ترشحات الاشخاص العشرة لهذا المنصب حسب العامل الاول والعامل الثاني.

وكملخص فان تطبيق طريقة PCA يتطلب إجراء تحليل من المراحل المحددة بوضوح:

1. إعداد البيانات
2. فحص مصفوفة الارتباط الملحوظ
3. إحصاءات لتقييم مدى ملاءمة مجموعة البيانات على أساس KMO و Bartlett والمقاييس المحددة للتحليل الى مركبات اساسية.
4. استخراج العوامل
5. التدوير وتفسير النتائج
6. تسمية العوامل (حسب الخصائص التي يشترك فيها المتغيرات التي تنتمي لكل عامل)

الفصل الرابع
طريقة التحليل بالمعاملات للتوفيقات

مقدمة الفصل

التحليل بالمعاملات للتوفيقات هو امتداد لتحليل PCA الطي تطرقنا اليه في الفصل السابق، وهو اسلوب مناسب لاستكشاف العلاقات بين المتغيرات النوعية أو البيانات الفئوية. لذا نهدف من خلال هذا الفصل الى توضيح تطبيق نوع اخر من التحليل العاملي والتي تعتبر تقنية إحصائية تساعد في تمثيل بنية الجداول المتقاطعة بيانيا لإلقاء الضوء على الآليات الأساسية.

ستعرض هذا الفصل كيفية عمل طريقة التحليل بالمعاملات للتوفيقات وتفسيرها عن طريق امثلة توضيحية مع استخدام برنامج SPSS. بالإضافة إلى ذلك سنعرض كيفية الكشف عن أهم الخصائص التي تشرح المتغيرات في مجموعة البيانات.

1.4. طريقة التحليل بالمعاملات للتوفيقات

تحليل بالمعاملات للتوفيقات، كما تعرف باللغة الانجليزية Correspondence analysis (CA)، هي امتداد لطريقة التحليل الى مركبات الأساسية PCA ، وهي طريقة مصممة للتعامل مع المتغيرات الكيفية (الاسمية والترتيبية). في الأصل ، تم تطوير هذه الطريقة لتحليل الجداول المزدوجة التي يتم فيها وصف عينة من الملاحظات بواسطة متغيرين اسميين، ولكن تم توسيعها بسرعة لتحليل أي مصفوفة بيانات.

يمكن إرجاع أصل كريقة التحليل بالمعاملات للتوفيقات (CA) إلى العمل المبكر لبيرسون و فيشر، لكن النسخة الحديثة من تحليل اCA وتفسيرها الهندسي تأتي من الستينيات في فرنسا وترتبط بالمدرسة الفرنسية لتحليل البيانات.

و تم تطوير هذه الطريقة من طرف Jean-Paul Benzecri كأسلوب ، غالبا ما تم اكتشافه وبالتالي يمكن العثور على أشكال مختلفة من CA تحت عدة أسماء مختلفة مثل : "القياس الأمثل" أو "تحليل التجانس" وغيرها من االتسميات والتي ترتبط اساسا بعدد كبير من الخصائص. حيث تعتبر هذه الطريقة الحل الأمثل للعديد من المشكلات التي تبدو مختلفة على ما هي عليه.

وتتمثل الطريقة رياضيا بتطبيق طريقة تحليلي PCA مع استعمال وحدة قياس كاي تربيع (χ^2) وبصفة عامة تستعمل طريقة CA في جدول مزدوج ذو متغيرات كيفية. حيث انها تسمح بتمثيل نقاط الاسطر والاعمدة لجدول مزدوج ذو متغيرات كيفية (Abdi and Williams, 2010).

1.1.4. النقاط الرئيسية

يحول تحليلي CA جدول البيانات إلى مجموعتين من المتغيرات الجديدة تسمى درجات العوامل، يتم الحصول عليها كمجموعات خطية من الصفوف والأعمدة على التوالي: مجموعة واحدة للصفوف ومجموعة واحدة للأعمدة. تعطي درجات العوامل هذه أفضل تمثيل لهيكل التشابه لصفوف وأعمدة الجدول، بالإضافة إلى ذلك ، يمكن رسم درجات العوامل كخرائط تعرض المعلومات في الجدول الأصلي بالشكل الأمثل.

في هذه الخرائط، يتم تمثيل الصفوف والأعمدة كنقاط تكون إحداثياتها هي درجات العوامل وحيث تسمى الأبعاد أيضا عوامل أو مكونات أو مجرد أبعاد. ومن المثير للاهتمام أن درجات عامل أن درجات عوامل الصفوف والأعمدة لها نفس التباين، وبالتالي يمكن تمثيل الصفوف والأعمدة بشكل ملائم في خريطة واحدة.

في طريقة التحليل بالمعاملات للتوفيقات يتناسب التباين الكلي (غالبا ما يسمى القصور الذاتي) لدرجات العوامل مع إحصائية استقلالية مربع كاي ، وبالتالي فإن درجات العوامل تتحلل إلى مكونات متعامدة.

وتعتبر طريقة CA أسلوب إحصائي يوفر تمثيل بياني للجداول المزدوجة والتي تعرف أيضا باسم علامات التبويب المتقاطعة أو جداول الاقتران (contingency tables). ويعتبر تحليل CA أسلوب يمكن من خلاله إيجاد تمثيل متعدد الأبعاد للتبعيات بين الصفوف والأعمدة في مساحة منخفضة الأبعاد. يسمح ببناء نظام متعامد من

المحاور (يسمى العوامل) حيث يمكن عرض الملاحظات (صفوف الجدول) والمتغيرات (أعمدة الجدول) في وقت واحد ، مما يسهل اكتشاف المعلومات البارزة المضمنة في جدول اقتران معين.

في هذا النظام، يتم تفسير القرب بين الملاحظات أو بين المتغيرات على أنه تشابه قوي. يتم تفسير القرب بين الملاحظات والمتغيرات على أنه علاقة قوية (Bastin et Benzecri, 1980).

وينشأ ذلك عندما يكون من الممكن وضع الأحداث في مجموعتين أو أكثر من مجموعات مختلفة من الفئات، مثل المنتج والموقع للمشتريات في أبحاث السوق أو الأعراض والعلاج في الاختبارات الطبية، الخ. لذلك فمن المنطقي محاولة تحليل أوجه التشابه والاختلاف بين فئات بعض المتغيرات وذلك من خلال مجموعة من المؤشرات كما يوضحها الجدول التالي:

الجدول رقم 4-01: مؤشرات تحليلي الابعاد

الخاصية	الاهمية
ملفات تعريف الصفوف والأعمدة	وهو متجه يحتوي على التكرارات النسبية لكل صف أو عمود (حاصل القسمة بين الخلية وصفها أو عمودها الإجمالي)، ويتم استخدامها لحساب المسافة بين جميع الفئات
المسافة	يتم إجراء العملية للحصول على قياس للبعد أو القرب بين الفئات التي تم تحليلها، ويتم استخدام المسافة الاقليدية
الوزن	يمثل وزن كل صف أو فئة عمود في إجمالي العينة ، أي إجمالي النسب المئوية للفئات
القيمة الذاتية أو القصور الذاتي (inertia)	تدل على وزن كل متغير أصلي في الأبعاد النهائية
المساهمة المطلقة	يشير إلى ازن كل صف أو فئة عمود (بشكل مستقل) في البعد الإجمالي
المساهمة النسبية	تشير إلى وزن كل بعد في كل فئة صف وعمود
مركز الثقل	قياس بعد كل فئة عن المركز في التمثيل البياني

وتوفر طريقة التحليل بالمعاملات للتوفيقات حلا لتلخيص مجموعة البيانات وتصورها في مخططات ثنائية الأبعاد. ويتعلق الامر هنا بالتحليل بالمعاملات للتوفيقات البسيط ، والذي يستخدم لتحليل الترددات المكونة من بيانات فئوية. يوفر درجات العوامل (الإحداثيات) لكل من الصفوف ونقاط العمود في جداول تستخدم هذه الإحداثيات لتصور الارتباط بين عناصر الصف والعمود بيانيا (Hoffman and Franke, 1986). والهدف هو الحصول على نظرة شاملة للبيانات المفيدة للتفسير.

الهدف من هذا الاسلوب من التحليل هو تمثيل الحد الأقصى من القصور الذاتي الكلي على المحور العاملي الأول، والحد الأقصى من القصور الذاتي المتبقي على المحور الثاني ، وهكذا حتى البعد الأخير. وعموما فإن الهدف

الرئيسي هو قراءة المعلومات الواردة في الفضاء متعدد الأبعاد عن طريق تقليل حجم المساحة مع الاحتفاظ بأكبر قدر ممكن من المعلومات الموجودة في المساحة الأصلية.

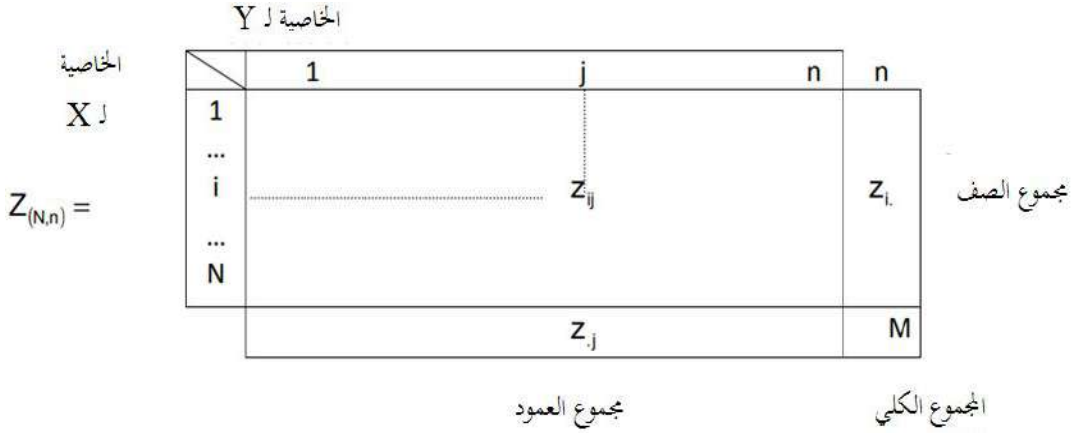
2.4. التحليل بالمعاملات للتوفيقات: النظرية والتطبيق

1.2.4. مراحل التحليل

تطبق طريقة التحليل CA بشكل أساسي على جداول الاقتران (التوافق). إنه جدول يحتوي على تقاطع الصف i والعمود j لعدد Z_{ij} من العناصر.و يتمثل هذا في تقسيم عناصر مجتمع M ما وفقا لـ X صف و Y عمود من الخصائص. ولذلك فهي خصائص كيفية (اسمية أو ترتيبية). وغالبا ما تركز الدراسة التقليدية لمثل هذه الحالات على الاعتماد أو الاستقلال بين عنصرين. يتم إجراؤه عموما باستخدام اختبار χ^2 وبشكل أكثر تحديدا عن طريق تحليل التباين (نسبة الارتباط) والانحدار.

في جدول الاقتران تكون اشكال الخصائص حصرية وشاملة. ويترتب على ذلك أن المجاميع في الصفوف والاعمدة لها معنى. ويأخذ جدول البيانات (N, n) الشكل التالي:

الجدول رقم 4-02: جدول البيانات



مع:

$$Z_{i.} = \sum_{j=1}^{n} Z_{ij}$$

$$Z_{.j} = \sum_{i=1}^{n} Z_{ij}$$

$$M = \sum_{i=1}^{N}\sum_{j=1}^{n} Z_{ij} = \sum_{i=1}^{N} Z_{i.} = \sum_{j=1}^{n} Z_{.j}$$

ومن بين أهداف تحليل CA هو وصف العلاقات القائمة بين متغيرين كيفيين، تم جمعهما في جدول اقتران، على مساحة قليلة الأبعاد، مع وصف العلاقات بين فئات (خصائص) كل متغير في نفس الوقت. حيث تعكس المسافات بين نقاط الفئة على الرسم البياني لكل متغير العلاقات بين الفئات، مع تمثيل الفئات المتشابهة بالقرب من بعضها البعض.

تم تطوير المنهجية بواسطة Benzecri في أوائل الستينيات حيث ان المضمون يتمثل في تقديم نوع خاص من تحليل المركبات الاساسية ولكن يتم إجراؤه على جدول اقتران وباستخدام مسافة إقليدية مرجحة تسمى مربع كاي (Benzecri, 1973). ويعتمد الحد الأقصى لعدد الأبعاد التي يمكن تحديدها في CA على عدد الفئات لكل متغير. على وجه التحديد، إذا كان أحد المتغيرات يحتوي على i من الفئات والآخر يحتوي على j من الفئات ، فإن عدد الأبعاد (أو العوامل) يكون وفقا للقاعدة التالية:

$$Min\,[j-1, i-1]$$

ملاحظة

إن تعميم هذه التقنية على أي عدد من المتغيرات هو ما يعرف بتحليل CA المتعدد حيث الحد الأقصى لعدد الابعاد هو:

$$Min\,[m, N-1]$$

حيث m هو عدد فئات المتغيرات دون فقدان البيانات مطروحا منها عدد المتغيرات المذكورة و N هو حجم العينة.

تسمح نتائج التحليلي بتحديد مدى تمثيل فئات الصفوف والأعمدة في مساحة صغيرة. في هي الحالة ، فإن الأمر هنا يتعلق بفحص الفئات المتشابهة وغير المتشابهة في صف والعمود، بالإضافة الى الفئات الموجودة في الصف المرتبطة بالفئات الموجودة في العمود. أي أن الامر يتعلق باستكشاف العلاقات بين الصفوف والأعمدة.

ان تطبيق طريقة تحليلي CA وتوضيح العلاقات بين البيانات، يتطلب ايجاد مجموعة من الجداول التي تمكننا من توضيح العلاقات هي:

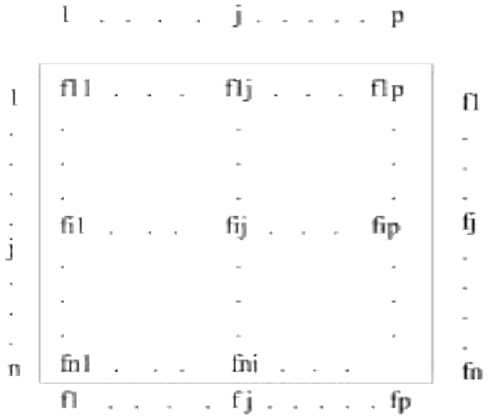
- **جدول التكرارات النسبة:** يتم حساب التكرارات والتكرارت النسبية باستخدام العلاقة التالية:

$$f_{ij} = {Z_{ij}}/{Z_{..}}$$

حيث:

$$f_{.j} = {Z_{.j}}/{Z_{..}} \quad ; \quad f_{i.} = {Z_{i.}}/{Z_{..}}$$

ومنه جدول التكرارات النسبية يعطى كما يلي:



- **جدول التكرارات النسبية للأسطر والاعمدة:** والذي نحتاجه لمقارنة الاسطر فيما بينها والاعمدة فيما بينها، ويتم بناء هذا الجدول من خلال العلاقتين التاليتين:

$$f_j^i = {Z_{ij}}/{Z_{i.}} = {f_{ij}}/{f_{i.}}$$

$$f_j^i = {Z_{ij}}/{Z_{.j}} = {f_{ij}}/{f_{.j}}$$

بعد ذلك نقوم بحساب مؤشر التجاذب/التنافر والذي يرمز له بالرمز d_{ij} و يحسب بالعلاقة التالية:

$$d_{ij} = \frac{f_{ij}}{f_{i.}\, fj_{.}}$$

ويلعب هذا المؤشر دورا هاما في تحليل CA حيث اذا كانت القيمة اكبر من الواحد نقول عن الخاصية i و j تتجاذب فيما بينها والعكس (تنافر). وفي حالة كانت القيمة تساوي الواحد نقول لا يوجد ارتباط بين السطر والعمود.

بعد ذلك نمر الى اختبار مربع كاي، حيث ان عند دراسة اي ظاهرة متعلقة بمتغيرين كيفيين يمكن البحث عن وجود او عدم وجود علاقة بينهما من خلال اختبار χ^2 المعطى بالعلاقة التالية:

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{(z_{ij} - (\frac{z_{i.}z_{.j}}{z_{..}}))^2}{\frac{z_{i.}z_{.j}}{z_{..}}}$$

ملاحظة

إن تحليل المعطيات بطريقة CA يمكن ان يعطي معلومات مهمة حتى وان كان χ^2 عير معنوي

2.2.4. تحويل جدول البيانات (الاسطر / الاعمدة)

في جدول الاقتران لا تحمل خصائص الأفراد والمتغيرات نفس المعنى الموجود في جدول PCA. ففي جدول الاقتران تمثل الصفوف والأعمدة الصفات أو الخصائص المكونة للمتغيرات. للحفاظ على التجانس في عرض التحليلين، يفترض أن الصفوف تمثل الأفراد والأعمدة تمثل المتغيرات.

ويلاحظ أيضا أنه يمكن تمثيل الجدول الأول بطريقة اخرى حيث يوضع X في الصف أو في العمود، ونفس الشيء لـ Y، دون تعديل طبيعة الجدول. وبالتالي تعتبر CA طريقة PCA وبالتالي ، من خلال القياس ، من مصفوفة Z أو تحويلها إلى مصفوفات، يمكننا اعتبار أن المعلومات الواردة في الجدول يمكن تحليلها من مسافتين:

* مساحة R^n للمتغيرات (الاعمدة) حيث يمكننا تمثيل سحابة N نقاط أفراد (الصفوف). كل فرد لديه احداثيات ممثلة بالعلاقة التالية:

$$x_{ij} = \frac{p_{ij}}{p_{i.}}$$

هذا يعني اننا سنهتم هنا بالتقريب النسبي لنقاط الأفراد ومن ثم اختيار هذه المصفوفة.

	1	...	j	...	n
1					
...					
i			$p_{ij}/p_{i.}$		
...					
N					

احداثيات i في R^n

* مساحة R^n للأفراد (الصفوف) حيث يمكننا تمثيل سحابة n نقاط متغيرات (اعمدة). كل متغير لديه احداثيات ممثلة بالعلاقة التالية:

$$y_{ij} = \frac{p_{ij}}{p_{.j}}$$

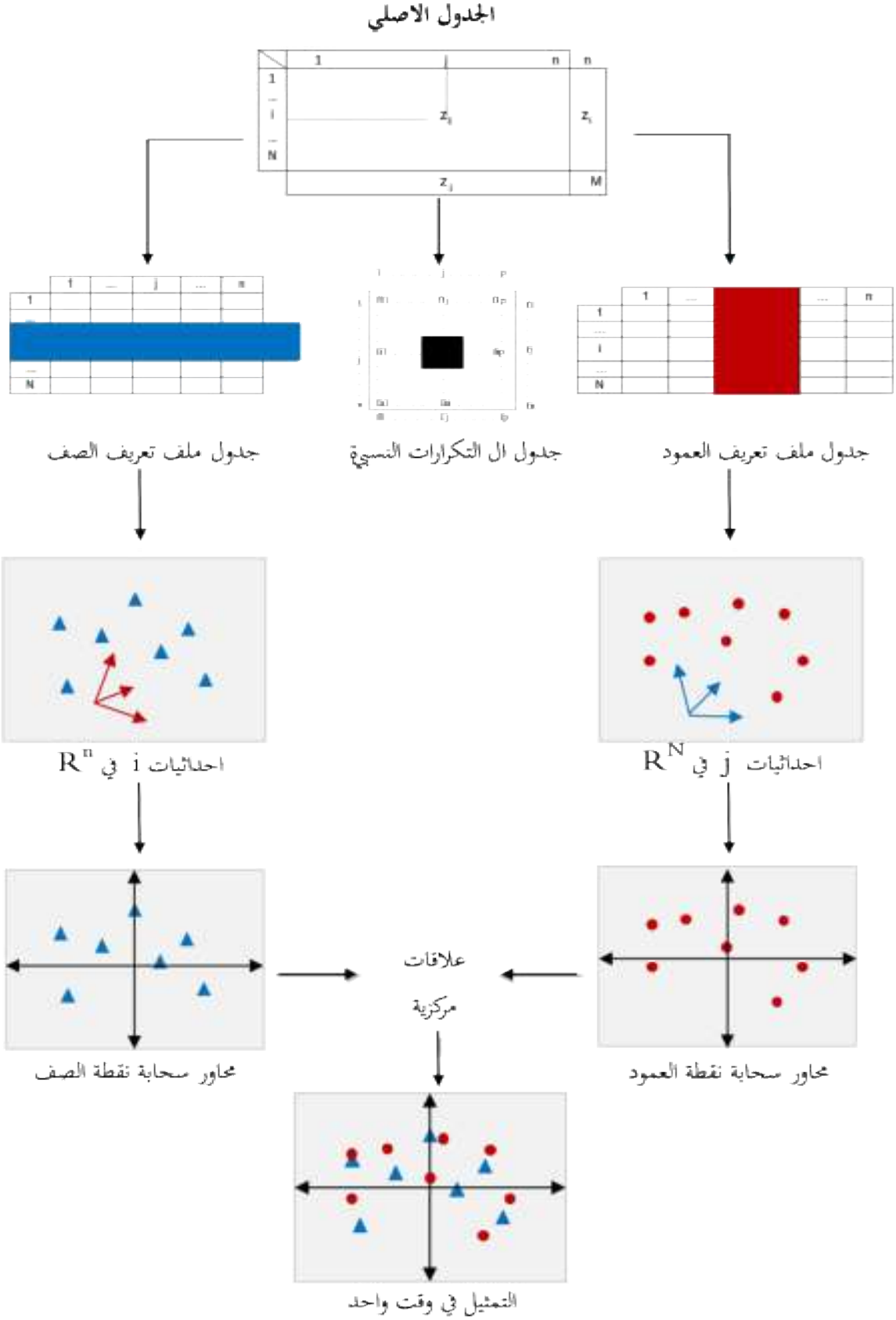
هذا يعني اننا سنهتم هنا بالتقريب النسبي لنقاط المتغيرات ومن ثم اختيار هذه المصفوفة.

	1	...	j	...	n
1					
...					
i			$p_{ij}/p_{.j}$		
...					
N					

احداثيات j في R^N

بالمقارنة مع PCA ، يتم إعطاء المعلومات من خلال المسافة الإقليدية بين نقاط السحب في الفضاءين R^n و R^N. ويمكن توضيح عملية تحويل البيانات من خلال الشكل الموالي:

الشكل رقم 4-01: تحويل جدول البيانات



كما في طريقة PCA التموقع يكون في فضاء المتغيرات، لذلك نستخدم مصفوفة ملفات تعريف الصف. يمكننا حساب خصائص هذه المتغيرات كالمتوسط والتباين المشترك (Covariance).

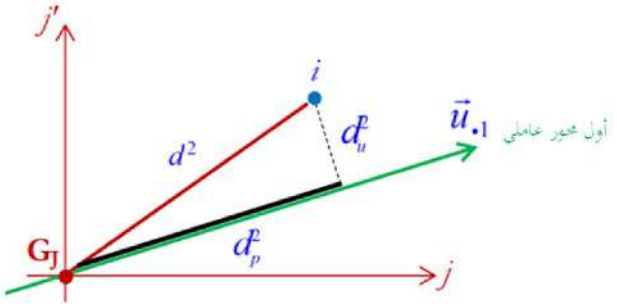
$$\bar{x}_j = \sqrt{p_{.j}} \qquad \mathrm{cov}(x_j, x_{j'}) = V_{jj'} = \sum_i \frac{p_{ij} p_{ij'}}{\sqrt{p_{.j}}\sqrt{p_{.j'}}\, p_{i.}} - \sqrt{p_{.j}}\sqrt{p_{.j'}}$$

من خلال تغيير j و j' من 1 إلى n ، فإننا نبني المصفوفة V(n,n) للتباينات و التباينات المشتركة للمتغيرات. هذه هي مصفوفة المعلومات الخاصة بالمتغيرات وبالقياس مع PCA، فإن الخطوة التالية من CA ستكون قطرية هذه المصفوفة.

تتيح المصفوفة V إمكانية حساب مصفوفة تغيير القاعدة بواسطة القيم الذاتية والمتجهات الذاتية. تجدر الاشارة الى انه في تحليل CA، لا نستخدم دائما المصفوفة V ولكننا نستخدم مصفوفة أبسط تسمى مصفوفة القصور الذاتي.

من مصفوفة التباينات والتباينات المشتركة (Var, Cov) نحاول توليد المتجهات الذاتية والقيم الذاتية. تشير هذه العملية الجبرية إلى أن القصور الذاتي الموضح بواسطة المحاور هو الحد الأقصى، أي أن إسقاط جميع النقاط i على المحور ، والمسافة d_p^2 ، هو الحد الأقصى ، والمسافة d_u^2 من كل نقطة إلى المتجه أو المحور العامل الذي هو الحد الأدنى الذي يتم الحصول عليه.

الشكل رقم 4-02: تحليل تركيب القصور الذاتي



3.2.4. مثال توضيحي

سنقوم بالاعتماد على البيانات الجدول الموالي، باجراء تحليل المعاملات للتوفيقات باتباع الخطوات السابقة.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Y_1	16	17	10	19	16	48	15	19	18
Y_2	8	15	45	17	27	20	0	14	53
Y_3	20	20	23	21	29	20	18	19	56
Y_4	11	18	12	17	20	16	15	10	44
Y_5	27	25	25	22	30	26	24	22	28
Y_6	21	21	20	24	27	34	13	17	24
Y_7	21	21	21	23	26	15	16	18	21

ترتبط النتيجة الأولى التي تم الحصول عليها بعد الحساب بالهدف الرئيسي لطريقة CA، لأنها طريقة ذات طبيعة اختزالية، ومنه فالهدف هو تقليل الأبعاد مع ضمان شرح الحد الأقصى من المعلومات الأصلية في الأبعاد الجديدة.

يعرض الجدول الموالي القيم الذاتية التي تعبر عن قيم إرشادية للقوة التفسيرية لكل من الأبعاد الجديدة التي يتم فيها تنظيم البيانات.

الجدول رقم 4-03: القيم الذاتية ونسبة التباين

	القيم الذاتية	% التباين	%التجميعية
البعد 1	6.251	60.06	60.06
البعد 2	2.170	20.86	80.91
البعد 3	1.667	16.01	96.92
البعد 4	0.020	1.9	98.82
البعد 5	0.077	0.74	99.56
البعد 6	0.045	0.44	100

يظهر العمودان المتبقيان نسبة التباين التي تشرح كل بعد من الأبعاد الستة. التباين الموضح هو نسبة القيمة الذاتية على مجموعة المعلومات الموضحة، ويوضح لنا التباين التجميعي الحد الأقصى لمستوى المعلومات. ويرتبط عدد الأبعاد التي يحسبها التحليل بالحد الأدنى لعدد الصفوف أو الأعمدة الموجودة في جدول الاقتران عن طريق طرح وحدة واحدة.

$$Min\,[j-1, i-1]$$

إذا كانت مصفوفة الإدخال تحتوي على i صف وj عمود، فسيكون عدد الأبعاد في التحليل على الأكثر هو الحد الأدنى: (j-1, i-1).

كما هو الحال في تحليل المركبات الاساسية PCA، هناك العديد من المعايير لتحديد العدد الأمثل للأبعاد، ولكن عادة ما نعمل مع تلك الأبعاد التي تصل على الأقل إلى الوحدة (معيار كايزر-غوتمان Kaiser-Gutmann) في القيمة الذاتية. في مثالنا هذا يمكننا أن نرى كيف أن القيمة الذاتية، وبالتالي نسبة كبيرة من التباين الموضح يقع في البعد الثالث.

لهذا فالجدولين المواليين يظهانر مساهمات كل فئة من فئات الصف والعمود المتضمنة في التحليل، بمعنى اخر يوضح لنا المساهمات المطلقة والنسبية لكل فئة من الفئات في التحليل.

الجدول رقم 4-04: مساهمات الصفوف والاعمدة في تشكيل المحاور

مساهمات الصفوف

المتغيرات	المساهمات المطلقة					
	البعد 1	البعد 2	البعد 3	البعد 4	البعد 5	البعد 6
X_1	7.64	7.08	6.18	0.74	24.48	16.04
X_2	1.3	2.93	0.21	7.61	25.71	0
X_3	22.36	11.49	34.67	6.99	7.3	0.06
X_4	1.26	0.23	0.43	41.17	0.32	21.57
X_5	0.54	1.8	4.42	7.52	1.84	14.38
X_6	18.28	52.18	11.51	0.17	0.42	4.4
X_7	9.37	22.44	1.84	20.82	19.78	0
X_8	2.65	0.1	1.56	17.64	15.52	43.53
X_9	36.61	1.74	39.18	0.34	2.82	0
المجموع	**100**	**100**	**100**	**100**	**100**	**100**

مساهمات الأعمدة

المتغيرات	المساهمات المطلقة					
	البعد 1	البعد 2	البعد 3	البعد 4	البعد 5	البعد 6
Y_1	28.34	29.24	12.56	5.01	0.05	13.21
Y_2	50.56	26.53	5.03	0.58	2.15	0.2
Y_3	5.09	9.93	9.73	7.87	51.12	0.29
Y_4	2.97	11.79	34.92	9.33	28.57	0.11
Y_5	5.57	7.02	12.72	31.08	12.94	18.03
Y_6	5.81	4.69	1.69	36.5	5.18	30.92
Y_7	1.57	10.79	23.35	9.64	0.23	39.35
المجموع	**100**	**100**	**100**	**100**	**100**	**100**

تجدر الاشارة الى ان المساهمة المطلقة لعنصر ما تقاس على اساس تأثير الفئة على تكوين البعد او العامل. وكما نلاحظ من الجدولين فان لكل مجموعة من فئات الصفوف وفئات الأعمدة يكون المجموع يساوي الأحمال 100.

الجدول رقم 4-05: جودة التمثيل

تمثيل الفئات حسب الاعمدة

المتغيرات	المساهمات النسبية						المجموع
	البعد 1	البعد 2	البعد 3	البعد 4	البعد 5	البعد 6	
X_1	62.63	20.14	13.5	0.19	2.48	0.95	**100**
X_2	43.99	34.39	1.91	8.16	11.49	0	**100**
X_3	62.88	11.11	25.75	0.62	0.25	0	**100**

X_4	43	2.76	3.93	44.67	0.14	5.36	**100**
X_5	19.88	23.07	43.45	8.8	0.84	3.85	**100**
X_6	43.27	45.85	7.77	0.01	0.01	0.08	**100**
X_7	50.49	41.99	2.64	3.56	1.32	0	**100**
X_8	65.22	0.87	10.25	11.4	4.71	7.76	**100**
X_9	76.74	1.27	21.89	0.02	0.07	0	**100**

تمثيل الفئات حسب الصفوف

المتغيرات	**المساهمات المطلقة**						
	البعد 1	**البعد 2**	**البعد 3**	**البعد 4**	**البعد 5**	**البعد 6**	**المجموع**
Y_1	67.34	24.12	7.95	0.38	0	0.23	**100**
Y_2	82.7	15.04	2.19	0.03	0.04	0	**100**
Y_3	42.37	28.7	21.61	2.08	5.26	0.02	**100**
Y_4	17.44	24.04	54.68	1.74	2.07	0	**100**
Y_5	44.04	19.27	26.81	7.79	1.26	1.03	**100**
Y_6	62.21	17.43	4.82	12.39	0.68	2.4	**100**
Y_7	12.92	30.82	51.22	2.51	0.02	3.35	**100**

وتتمثل المساهمة النسبية لعنصر ما في التأثير في جودة التمثيل لكل فئة من فئات الصف أو العمود في البعد أو العامل (أي جودة تمثيل الفئة في كل البعد).

يوضح الجدول الموالي ثلاثة مؤشرات هامة تتعلق بالتمثيل البياني (المخطط)، وهي:

- المؤشر الأول: الوزن أو مساهمة كل فئة من فئات الصف أو العمود على إجمالي الارتباطات؛
- المؤشر الثاني: المسافة والتي تمثل قرب أو بعد كل فئة من الصفوف أو العمود من مركز الثقل؛
- المؤشر الثالث: مصفوفة تحتوي على إحداثيات كل فئة من فئات الصفوف أو الأعمدة في خريطة تحديد المواقع التي تم الحصول عليها في كل الأبعاد.

الجدول رقم 4-06: الوزن والمسافة في تحليل CA

المتغيرات	**الوزن**	**المسافة**
X1	0.09	0.08
X2	0.100	0.02
X3	0.114	0.2
X4	0.104	0.02
X5	0.127	0.01
X6	0.128	0.19
X7	0.075	0.15
X8	0.086	0.03
X9	0.176	0.17
Y_1	0.127	0.21
Y_2	0.147	0.26
Y_3	0.164	0.05

Y_4	0.118	0.09
Y_5	0.165	0.05
Y_6	0.146	0.01
Y_7	0.132	0.06

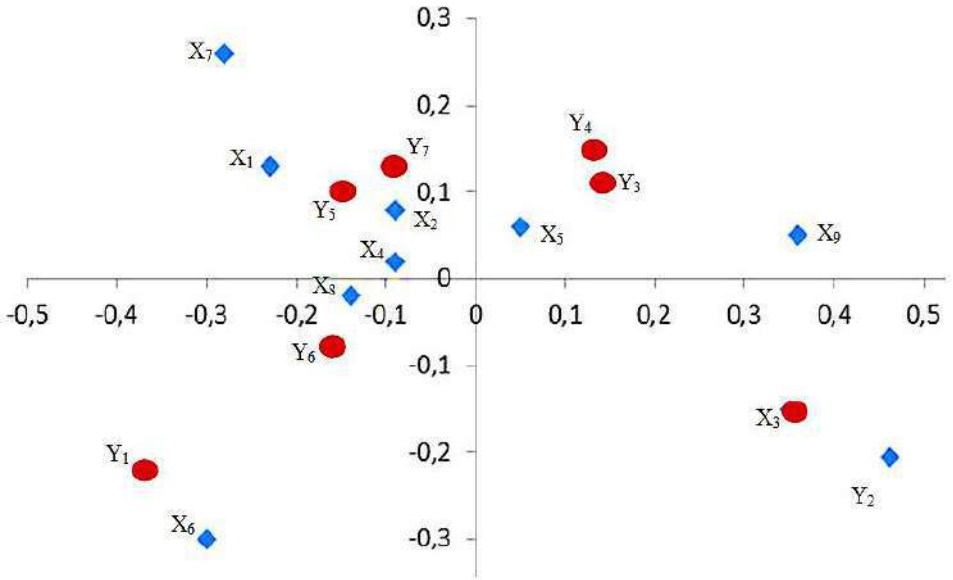
كما يوضح الجدول الموالي احداثيات كل فئة على الابعاد الستة

الجدول رقم 4-07: الإحداثيات

المتغيرات	البعد 1	البعد 2	البعد 3	البعد 4	البعد 5	البعد 6
X1	-0.23	0.13	0.11	0.01	0.05	0.03
X2	-0.09	0.08	0.02	-0.04	-0.05	0
X3	0.35	-0.15	0.23	0.03	-0.02	0
X4	-0.09	0.02	0.03	-0.09	0	-0.03
X5	0.05	0.06	0.08	-0.03	0.01	0.02
X6	-0.3	-0.3	-0.12	0.01	-0.01	0.01
X7	-0.28	0.26	-0.06	0.07	-0.05	0
X8	-0.14	-0.02	0.05	0.06	0.04	-0.05
X9	0.36	0.05	-0.19	0.01	0.01	0
Y_1	-0.37	-0.22	-0.13	0.03	0	-0.02
Y_2	0.46	-0.2	0.08	0.01	-001	0
Y_3	0.14	0.11	-0.1	0.03	0.05	0
Y_4	0.13	0.15	-0.22	-0.04	-0.04	0
Y_5	-0.15	0.1	0.11	0.06	-0.02	0.02
Y_6	-0.16	-0.08	0.04	-0.07	0.02	0.03
Y_7	-0.09	0.13	0.17	-0.04	0	-0.04

بعد ذلك يتم تمثيل الحداثيات على المخطط باستخدام البعدين (المحورين) الرئيسيين:

الشكل رقم 4-03: التمثيل البياني



على عكس ما يحدث في تحليل المركبات الاساسية فانه في تحليل المعاملات للتوفيقات من الممكن إجراء تحليل تقارب بين فئات الصف والعمود على المخطط، حيث:

- القرب بين فئتي صف أو فئتين من الأعمدة يعني وجود خصائص متشابهة يدل هذا على حدوث ارتباطات ذات طبيعة مشتركة أو متشابهة، مما يعني وجود مستوى عال من الارتباط. على سبيل المثال، إذا أردنا التعليق على بيانات المثال، فسنقول ان Y_3 و Y_4 خاصيتان متشابهتان جدا نظرا لعلاقة الارتباط بينهما.
- القرب من المركز يعني أن هذه الفئات لديها قوة وصفية أقل.
- نستنتج من القرب بين فئات الصفوف والأعمدة أن هناك ارتباط قوي بين كل صف وعمود.

3.4. تحليل CA باستخدام برنامج SPSS

سنركز هنا على جملة من العناصر الأساسية في تطبيق تحليل المعاملات للتوفيقات من خلال المؤشرات المختلفة التي تظهر في الجداول والتمثيل البياني. سوف نأخذ مثال على جدول الاقتران الذي يتعلق بمجموعة من الخصائص وسنقدم المخرجات الناتجة عن برنامج SPSS.

من خلال هذا التمرين سنستخدم قاعدة بيانات وهي بيانات وهمية تم الاعتماد عليها فقط لتوضيح تطبيق طريقة تحليل CA باستخدام برنامج SPSS. تحتوي قاعدة البيانات هذه على بيانات دراسة تم اجراها حول سؤال متعلق بالبحث في سلوكيات طلبة كلية العلوم الاقتصادية لجامعة خميس مليانة.

حيث سينظر الى سلوك التي يرجح أن يتعرض لها الطلبة مع تحليل خطر تطويره هذا السلوك كتعاطي المخدرات، التخلي عن الدراسة، العنف، مشكلات الصحة العقلية وهذا بالنظر الى فئتهم العمرية (أي ، 17-19 سنة ، 20- 22 سنة ، 23-25 سنة ، 26-28 سنة).

1.1.4. تحضير وادخال البيانات

يمكننا إدخال البيانات في SPSS باتباع الخطوات التالية:

من خلال إجراء تحليل المعاملات للتوفيقات CA نريد تحديد العلاقات المحتملة بين هذين المتغيرين. ويوضح الجدول الموالي جزء من بيانات التمرين:

الطالب	الفئة العمرية	الخطر
الطالب 1	22-20	العنف
الطالب 2	19-17	التخلي عن الدراسة
الطالب 3	22-20	مشكلات الصحة العقلية
الطالب 4	28-26	تعاطي المخدرات
الطالب 5	22-20	التخلي عن الدراسة
الطالب 6	19-17	العنف
.	.	.
.	.	.
الطالب 4996	25-23	العنف

نفرض اننا نريد تحليل العلاقة بين المتغير الاول (العمر) والمتغير الثاني (الخطر). اذن سنقوم بإدخال البيانات بحيث يتم انشاء ثلاث (3) متغيرات يكون المتغير الاول العمر والمتغير الثاني هو الخطر، مع ضرورة تسمية كل متغير بناء على عدد الفئات داخل المتغير. يضاف الى هذين المتغيرات متغير اخر (ثالث) يتمثل في التكرار.

بالنسبة للعمر هناك أربع فئات عمرية مختلفة في علامة تبويب القيم في SPSS سنقدم أربع قيم مختلفة لمتغير العمر. ستكون القيمة 1 هي الفئة العمرية 17-19 ، والقيمة 2 ستكون الفئة العمرية 20-22 ، وهكذا. يحتوي متغير الخطر أيضا على أربع فئات وسنسميها بنفس الطريقة. القيمة 1 ستكون خطر تعاطي المخدرات، والقيمة 2 ستكون التخلي عن الدراسة، الخ.

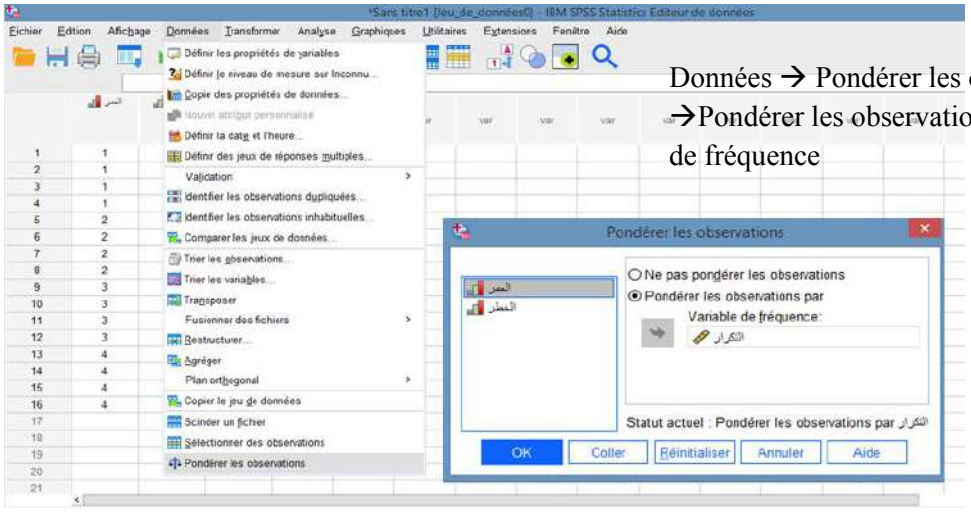
في جدول عرض البيانات لملف SPSS الخاص بهذا المثال سيكون لدينا 16 إدخال، بحيث يتم مطابقة كل فئة عمرية مع كل فئة من فئات الخطر. سيعطي متغير التكرار (frequence) تكرار بيانات العينة لكل فئة عمرية مع كل عامل خطر. على سبيل المثال ، يمكننا أن نرى أن تكرار الفئة العمرية 17-19 والتي قد تتعرض لخطر التخلي عن الدراسة هو 12.

ادراج كل البيانات في برنامج SPSS تشكل لدينا جدول البيانات.

ملاحظة

يمكن إدخال البيانات بإنشاء متغيرات الدراسة بالإضافة الى متغير اخر هو التكرار بالطريقة الموضحة اعلاه، بناء على افتراض أن تكرار كل متغير قد تم حسابه مسبقا. إذا لم يتم حساب تكرار كل متغير وكانت البيانات الأولية فقط متاحة، فمن الممكن استخدام نفس التحليل بدون استخدام هذه الخطوة. ومع ذلك فإن تحليل البيانات الأولية بهذه الطريقة غير شائع بسبب الحجم الكبير للبيانات، ففي مثالنا سيؤدي إدخال البيانات الأولية إلى إنتاج ملف مكون من 4996 سطرا.

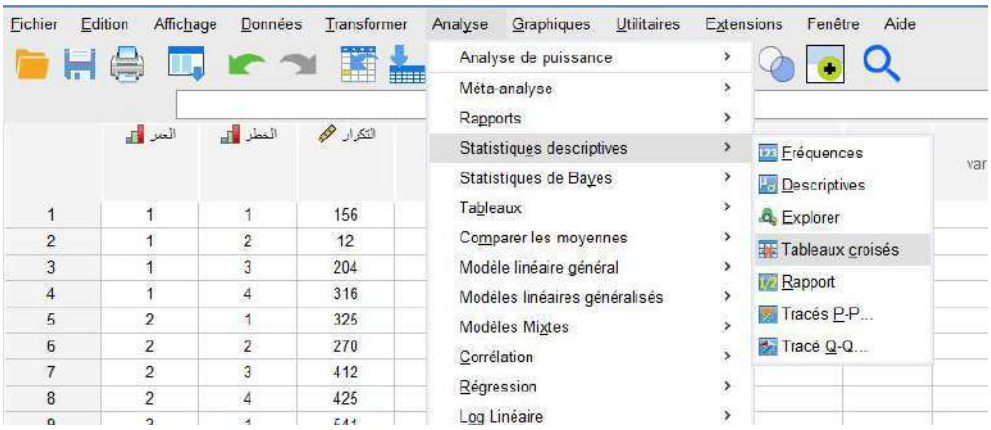
بمجرد إدخال جميع البيانات في SPSS ، فإن الخطوة التالية هي ترجيح الحالات حسب التكرار، وللقيام بذلك نضغط على:



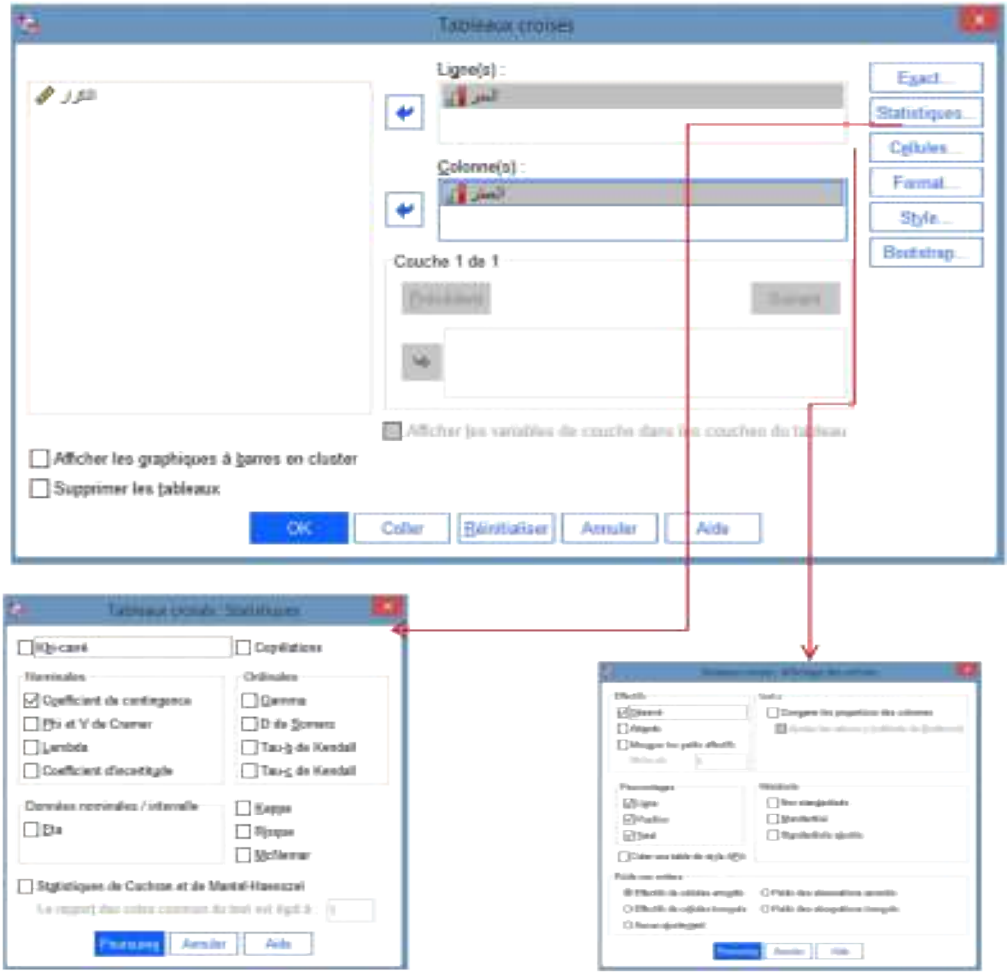
Données → Pondérer les observations
→Pondérer les observations par : variable de fréquence

سنشرع الان في تحليل البيانات باستخدام SPSS، لأجل ذلك سنبدأ بجدول الاقتران (التقاطع) وذلك باتباع الخطوات التالية:

الضغط على Analyse ثم نختار من القائمة Statistiques descriptive ثم نضغط على Tableau croisés كما في الصورة الموالية:



ومن مربع الحوار نقوم بإدراج المتغيرات حتى نستطيع تحليل العلاقة بينها، كما يلي:



و بما ان المتغيرين نوعيين فإن معامل ارتباط بيرسون لا يفي بالغرض لذا سنستعمل اختبار التوافق (Stafford, et Bodson, 2007)، ثم نضغط على statistiques في مربع الحوار ونختار coefficient de contingence.

ويمكننا ايضا استخراج النسب المئوية وذلك بالضغط على Cellules في مربع الحوار، وبالتالي نحصل غبى النتائج التالية:

اختبار التوافق

		Valeur	Signification approximative
Nominal par Nominal	Coefficient de contingence	,481	,000
N d'observations valides		4996	

وبما ان مستوى الدلالة أقل من0.05 اذا نرفض الفرضية الصفرية ونقول ان يوجد ارتباط بين الفئة العمرية ومختلف المخاطر التي قد يتعرض لها الطلبة. ولكن هذه العلاقة قريبة من المتوسط بما ان القيمة بلغت 0.481 (انظر الملحق رقم 2)، علما ان معامل التوافق محصور بين 1 و 0.

أما الجدول الموالي فيوضح النسب وهي نسب بالنسبة للأسطر، بالنسبة للأعمدة، وبالنسبة الى المجموع.

الخطر * العمر Tableau croisé

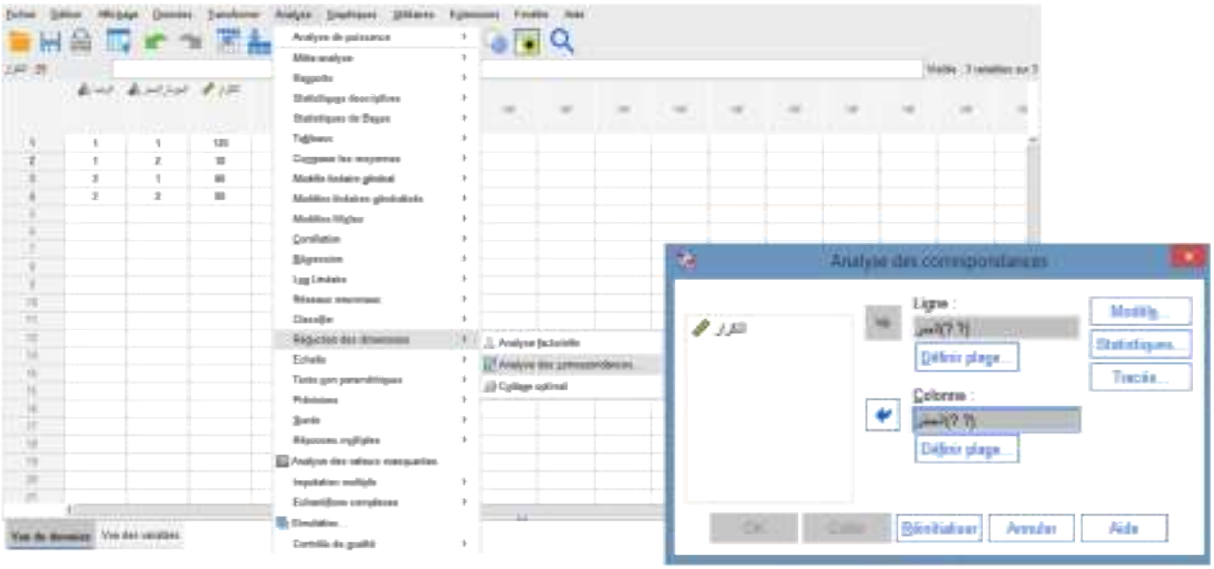
			الخطر				Total
			تعاطي المخدرات	التخلي عن الدراسة	العنف	مشاكل الصحة العقلية	
العمر	17-19	Effectif	156	12	204	316	688
		% dans العمر	22,7%	1,7%	29,7%	45,9%	100,0%
		% dans الخطر	11,9%	1,9%	13,9%	19,9%	13,8%
		% du total	3,1%	0,2%	4,1%	6,3%	13,8%
	20-22	Effectif	325	270	412	425	1432
		% dans العمر	22,7%	18,9%	28,8%	29,7%	100,0%
		% dans الخطر	24,8%	42,7%	28,2%	26,7%	28,7%
		% du total	6,5%	5,4%	8,2%	8,5%	28,7%
	23-25	Effectif	541	197	481	437	1656
		% dans العمر	32,7%	11,9%	29,0%	26,4%	100,0%
		% dans الخطر	41,3%	31,1%	32,9%	27,5%	33,1%
		% du total	10,8%	3,9%	9,6%	8,7%	33,1%
	26-28	Effectif	289	154	366	411	1220
		% dans العمر	23,7%	12,6%	30,0%	33,7%	100,0%
		% dans الخطر	22,0%	24,3%	25,0%	25,9%	24,4%
		% du total	5,8%	3,1%	7,3%	8,2%	24,4%
Total		Effectif	1311	633	1463	1589	4996
		% dans العمر	26,2%	12,7%	29,3%	31,8%	100,0%
		% dans الخطر	100,0%	100,0%	100,0%	100,0%	100,0%
		% du total	26,2%	12,7%	29,3%	31,8%	100,0%

ويمكننا ترجمة نتائج الجدول السابق كما يلي: نقول انه مثلا 156 طالب ممن تتراوح اعمارهم بين 17 و 19 سنة معرضين لخطر تعاطي المخدرات.

فيما يخص نسبة الاسطر فيمكننا القول ان 22.7% من الطلاب اعمارهم تنتمي الى الفئة العمرية 17-19، كما ان 11.7 % من الطلاب معرضين لخطر تعاطي المخدرات. أما النسبة 3.1% فيمثلون النسبة في مجموع العينة الكلية الذين ستركون في خاصيتين هما: العمر بين 17 و 19 سنة وتعاطي المخدرات.

2.1.4. تحليل المعاملات للتوفيقات في SPSS

بمجرد ترجيح البيانات حسب التكرار أصبح من الممكن اجراء التحليل في SPSS. للقيام بذلك في Analyse نختار من القائمة Réduction des dimensions ثم analyse des correspondances. كما يظهر في الجهة اليسرى من الشكل الموالي:



بعد ذلك ، ندرج كل من متغيري العمر والخطر في ملفات تعريف الصف والعمود على التوالي (كما يظهر في الجهة اليمنى من الشكل السابق). لا يهم أي متغير يقع على أي محور عند إجراء التحليل.

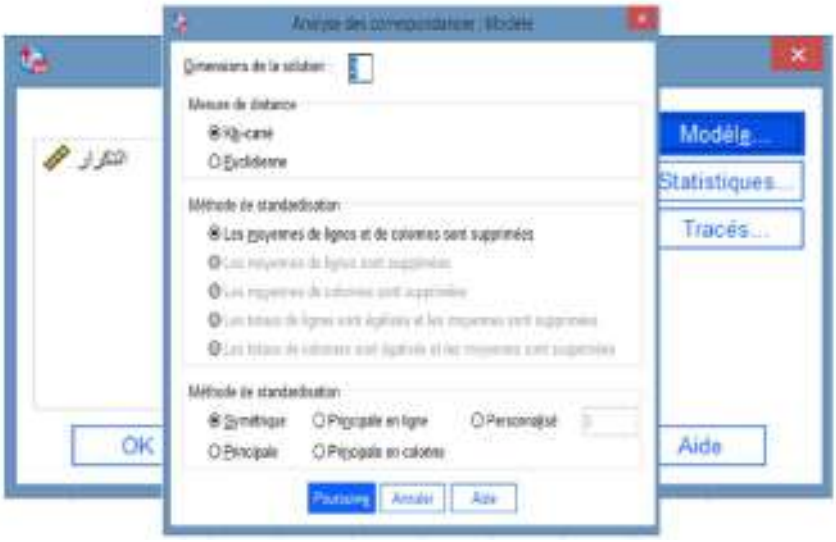
ليكون التحليل كاملا يجب تحديد نطاق الصفوف والأعمدة لكل متغير. ضمن الصف نضغط على تعريف النطاق (Définir plage). في متغير العمر لدينا أربع فئات فقط، وسنقوم بتضمين جميع الفئات الأربع هنا. بجانب ادنى قيمة نضع القيمة 1 ؛ بجانب أكبر قيمة نضع القيمة 4.ونجري نفس العملية للمتغير الثاني (الخطر).

ثم نضغط على Mettre à jour ثم Poursuivre

ملاحظة

باستخدام CA من الممكن إجراء تحليل أولي يتم فيه تحليل جزء فقط من البيانات. مع مجموعات البيانات الكبيرة، حيث يستطيع الباحث تضمين جزء فقط من بياناتهم في التحليل في البداية، ثم ادراج جميع المتغيرات لاحقا. قد يكون هذا مفيدا إذا كان هناك العديد من الفئات داخل كل متغير ويرغب الباحث في التقليل منها والتركيز على ارتباط معين داخل مجموعة البيانات قبل النظر إلى الصورة العامة. في مثالنا هذا سيتم استخدام جميع البيانات للتحليل.

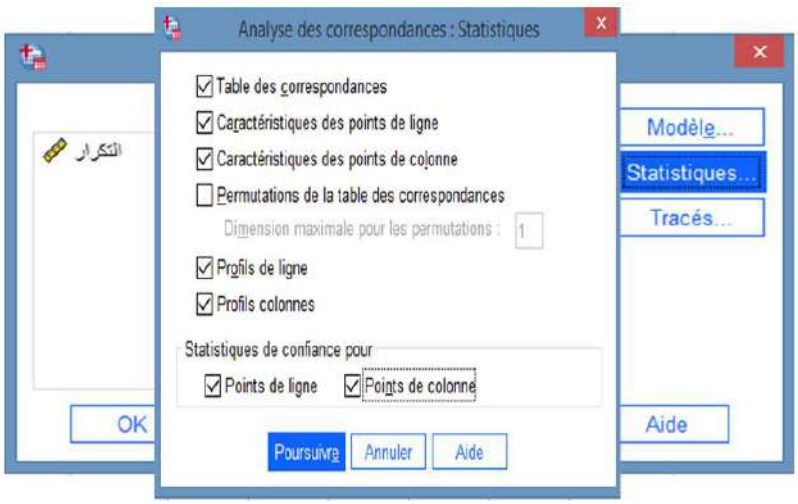
وعند الضغط على Modèle في مربع الحوار يمكن تحديد الكيفية التي ينتج من خلالها SPSS نتائج التحليل. الخيار الأول هو اختيار عدد الأبعاد المراد تضمينها في الحل. في هذا المثال تم تعيين عدد الأبعاد على اثنين (2). هذا هو الإعداد الافتراضي في SPSS عند تشغيل CA.



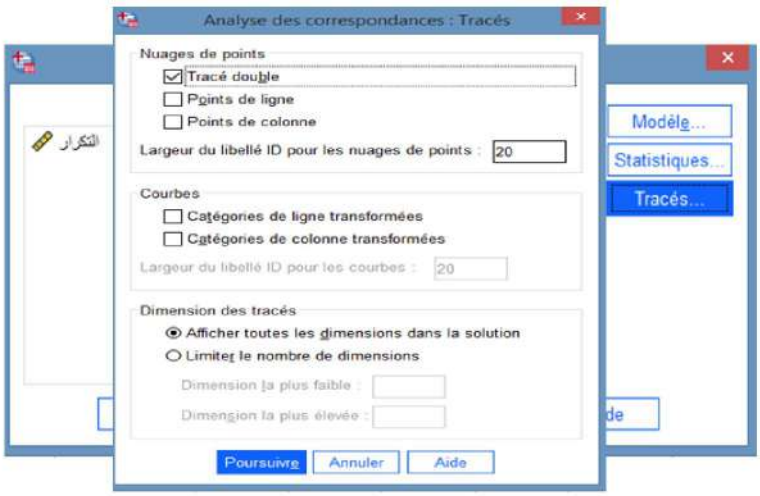
ومع ذلك ، يمكن زيادة الأبعاد وفقا لتقدير الباحث اعتماد طبعا على نوع البحث الذي يقوم بإجرائه. يجب تعيين مقياس المسافة على مربع كاي (Khi-carré) ، ويجب تعيين طريقة التوحيد على " Les moyennes de lignes et de colonnes sont supprimées" ، ويمكن اختيار طريقة التوحيد

حسب الطريقة التي تراد بها تفسير النتائج. في هذا المثال قمنا باختيار Symétrique لكي نتمكن من مقارنة الصفوف بالأعمدة ثم نضغط على Poursuivre.

بعدها نضغط على Statistiques في مربع الحوار وذلك بغرض اختيار جداول الإخراج المراد تضمينها في المخرجات. بشكل تلقائي في SPSS يتم ادراج الخيارات الثلاثة الأولى. في مثالنا سنختار أيضا profils de ligne، profils de colonnes ، بالإضافة إلى إحصائيات الثقة لنقاط الصفوف (points de ligne) ونقاط العمود (Points de colonne). كما في الشكل التالي:



وبالضغط على Tracés يمكن اختيار كيفية عرض التحليل بيانيا، ويعتبر هذا من أهم أجزاء نتائج التحليل.



ضمن المخططات نختار عرض tracé double البياني، بالإضافة إلى نقاط الصف فقط ونقاط العمود فقط في رسم بياني منفصل. تجدر الاشارة الى ان عرض نقاط الصفوف أو نقاط العمود في مخطط التشتت يعد مفيدا عند مقارنة نقاط الصف أو نقاط العمود من أجل تبسيط البيانات التي سيتم إنتاجها في tracé double، ولكن في هذا المثال سنكتفي فقط بالمخطط المختار.

وأخيرا تسمح أبعاد الرسم باختيار تضمين أو عدم تضمين جميع الأبعاد التي طلب من SPSS إنتاجها في التحليل، أو تقييد الأبعاد المدرجة في التمثيل البياني للمعطيات. في هذا المثال ، سنعرض جميع الأبعاد. الان نضغط على Poursuivre.

بعد انتهاء هذه المرحلة نكون قد قمنا بإدراج جميع المعالم المتعلقة بطريقة التحليل CA ويمكن الآن الحصول على نتائج هذا التحليل وذلك بالضغط على OK.

3.1.4. التعليق على النتائج

سيقوم برنامج SPSS بإخراج عدة جداول من بينها الجدول الموالي والذي المعاملات للتوفيقات (Table des correspondances)، حيث تستند البيانات الواردة في هذا الجدول إلى البيانات المتعلقة بمتغيري العمر والمخاطر والتي قمنا بإدراجها في البرنامج.

Table des correspondances					
	الخطر				
العمر	تعاطي المخدرات	التخلي عن الدراسة	العنف	مشاكل الصحة العقلية	Marge active
17-19	156	12	204	316	688
20-22	325	270	412	425	1432
23-25	541	197	481	437	1656
26-28	289	154	366	411	1220
Marge active	1311	633	1463	1589	4996

يتضمن هذا الجدول التكرارات لكل فئة من فئات الصفوف والأعمدة بالإضافة الى التجميعي لكل فئة من فئات الصفوف وفئات الأعمدة تسمى الهامش النشط (Marge active). على سبيل المثال ، يمكن أن نلاحظ أن مجموع التكرارات للفئة العمرية 17-19 عبر كل فئة من فئات الخطر يصل إلى 688. وبالمثل ، فإن جميع تكرارات خطر العنف عبر كل فئة من الفئات العمرية تصل إلى 1463.

ينتج ايضا من عملية التحليل جدول ملفات تعريف الصف وملفات تعريف العمود (Profils lignes/ Profils colonnes) كما هو موضح في الجدول التالي.

يعطي هذا الجدول التكرار الموزون لكل نقطة من نقاط الصف / العمود ، بحيث يتم جمع إجمالي الصف / العمود بأكمله إلى 1. يتم حساب ملفات تعريف الصف / العمود بأخذ كل نقطة صف / العمود وتقسيمها على الهامش النشط الخاص بذلك الصف / العمود.

على سبيل المثال ، بالنسبة للفئة العمرية 17-19 وعامل الخطر تعاطي المخدرات، فإن التكرار (على النحو الوارد في جدول العاملات للتوفيقات) هو 156. الهامش النشط لهذا الصف هو 688. لذلك ، 688/156 = 0.227، بهذه الطريقة يتم حساب القيم الاخرى في الجدول.

Profils lignes					
	الخطر				
العمر	تعاطي المخدرات	التخلي عن الدراسة	العنف	مشاكل الصحة العقلية	Marge active
17-19	,227	,017	,297	,459	1,000
20-22	,227	,189	,288	,297	1,000
23-25	,327	,119	,290	,264	1,000
26-28	,237	,126	,300	,337	1,000
Masse	,262	,127	,293	,318	

Profils colonnes					
العمر	الخطر				
	تعاطي المخدرات	التخلي عن الدراسة	العنف	مشاكل الصحة العقلية	Masse
17-19	,119	,019	,139	,199	,138
20-22	,248	,427	,282	,267	,287
23-25	,413	,311	,329	,275	,331
26-28	,220	,243	,250	,259	,244
Marge active	1,000	1,000	1,000	1,000	

ويعتبر جدول الملخص (Récapitulatif) أهم جدول ينتج عن تحليل CA في برنامج SPSS. حيث يستخدم تحليل المعاملات للتوفيقات إحصائية مربع كاي لاختبار التباين الكلي مع الاحتمال المرتبط به. تكون إحصائية مربع كاي عالية عندما يكون هناك تطابق كبير بين صفوف وأعمدة الجدول (Fellenberg et al ، 2001).

Récapitulatif								
					Proportion d'inertie		Valeur singulière de confiance	
Dimension	Valeur singulière	Inertie	Khi-deux	Sig.	Représentation	Cumulé	Ecart type	Corrélation 2
1	,173	,030			,713	,713	,011	,102
2	,110	,012			,285	,998	,014	
3	,008	,000			,002	1,000		
Total		,042	210,373	,000[a]	1,000	1,000		
a. 9 degrés de liberté								

أول شيء يجب ملاحظته في الجدول السابق هو ما إذا كان النموذج مهما أم لا. في هذا المثال يعتبر النموذج مهم عند المستوى 0.05 ، حيث مستوى المعنوية بلغ 0.000 وبلغت قيمة مربع كاي 210.373. نلاحظ أيضا أن تحليل CA قد نتج عنه ثلاثة عوامل.

ملاحظة

تجدر الاشارة هنا ان برنامج SPSS ينتج فقط العوامل التي يمكن تفسيرها، بدلا من تضمين جميع العوامل. لهذا السبب لا يصل مجموع القصور الذاتي دائما إلى 100%.

كما ان من الممكن أيضا حساب متوسط القيمة الذاتية الذي يجب أن يظل المحور فوقه في النتيجة. يمكن اعتبار النقطة التي يظهر عندها الرسم البياني للقيم الذاتية منعطفا (coude) على أنها تشير إلى العدد الأمثل من المحاور الرئيسية التي يجب تذكرها. إذا كانت البيانات عشوائية، فإن القيمة المتوقعة للقيمة الذاتية لكل محور ستكون: 1/(n-1). ووفقا لـ (Bendixen, 1995) يجب اعتبار أي محور بمساهمة أكبر من الحد الأقصى لهاتين النسبتين هاما وإدراجه في حل تفسير البيانات.

ويوضح عمود القصور الذاتي (Inertie) في الجدول، التباين الكلي الموضح في كل عامل. في هذا المثال نلاحظ ان إجمالي القصور الذاتي (شرح التباين الكلي) هو 4.2%، مما يشير إلى أنه بالنسبة لبيانات المثال فإن معرفة شيء ما عن العمر يفسر حوالي 4% مما يتعلق بالخطر والعكس صحيح. هذا الارتباط ضعيف ، لكنه لا يزال مهما كما يتضح من قيمة مربع كاي.

ويتم تشكيل كل عامل وفقا لمقدار التباين المشروح، حيث يتضمن العامل الاول (1) دائما معظم التباين، متبوعا بالعامل الثاني (2)، ...الخ.

يمكننا ان نقول مثلا، من خلال نتائج الجدول السابق ان العامل الاول يفسر 3% من إجمالي 4.2% من إجمالي التباين. علاوة على ذلك يفسر العامل الثاني ما قيمته 1.22% من إجمالي التباين الذي تم حسابه (4.2%). أما العامل الثالث والاخير فيمثل 0% من إجمالي التباين، وبالتالي سيتم استبعاده من التحليل الإضافي.

بينما يوضح عمود Valeur singulière الجذور التربيعية للقيم الذاتية (eigenvalues)، والتي تصف الحد الأقصى من الارتباط بين فئات المتغيرات في التحليل لعامل معين (Garson، 2008). في تحليل CA، تعتبر قيم eigenvalues والقصور الذاتي مترادفتين، بمعنى ان كل محور له قيمة ذاتية يساوي مجموعها قصور سحابة النقاط (Benzecri، 1992).

تعطي القيم الموجودة في عمود نسبة القصور الذاتي (Proportion d'inertie) نسبة التباين التي يشرحها كل عامل من إجمالي التباين. في هذا المثال يفسر العامل الاول 71.3% من إجمالي 4.2% من التباين، والعامل الثاني يفسر حوالي 29% من التباين، بينما العامل الثالث يفسر 0.2 % فقط.

كلما زادت نسبة التفسير زادت الدقة الموجودة في البيانات الأصلية في تحليل CA منخفض الأبعاد (Bendixen, 2003).

ملاحظة

لا توجد قاعدة عامة أو معايير للاحتفاظ بالعوامل أو رفضها في التحليل على أساس نسبة القصور الذاتي، حيث يعتمد ذلك على سؤال البحث وقرار الباحث حول ما هو مهم بالموازاة مع الدلالة الإحصائية لأي حالة معينة. ففي هذا المثال نقول انه يوجد عاملين يفسران التباين بين فئات الخطر والفئة العمرية، وقد تكشف بعض الأسئلة البحثية الاخرى أو بعض الاهداف التي يرغب الباحث في تحقيقيها أن الأبعاد الثلاثة ضرورية لشرح معظم التباين.

ويوجز الجدولين المواليين معلومات حول كيفية رسم كل نقطة من نقاط الصف او العمود في التمثيل البياني. يشير عمود الكتلة (Masse) في هذا الجدول إلى نسبة كل فئة لمتغير العمر وكل كل فئة لمتغير الخطر فيما يتعلق بجميع الفئات العمرية وفئات المخاطر على التوالي في التحليل.

يشير العمود Score de la dimension إلى الإحداثيات في كل عامل (1 و 2) حيث سيتم تمثيل كل فئة صف وكل فئة عمود بيانيا. يوضح عمود Contribution مدى جودة تحميل كل نقطة على كل عامل، بالإضافة إلى مدى جودة تفسير استخراج العوامل لكل نقطة.

Présentation des points de ligne[a]

العمر	Masse	Score de la dimension		Inertie	Contribution				
					Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point		
		1	2		1	2	1	2	Total
17-19	,138	-,962	,111	,022	,735	,015	,991	,008	1,000
20-22	,287	,373	,336	,010	,230	,295	,660	,339	,999
23-25	,331	,124	-,456	,008	,029	,630	,104	,896	1,000
26-28	,244	-,063	,163	,001	,006	,059	,181	,767	,947
Total actif	1,000			,042	1,000	1,000			

a. Normalisation symétrique

Présentation des points de colonne[a]

الخطر	Masse	Score de la dimension		Inertie	Contribution				
					Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point		
		1	2		1	2	1	2	Total
تعاطي المخدرات	,262	,087	-,511	,008	,011	,626	,044	,955	,999
التخلي عن الدراسة	,127	,946	,392	,022	,654	,178	,902	,098	1,000
العنف	,293	-,024	,006	,000	,001	,000	,374	,016	,390
مشاكل الصحة العقلية	,318	-,426	,260	,012	,333	,196	,809	,190	,999
Total actif	1,000			,042	1,000	1,000			

a. Normalisation symétrique

نلاحظ أن الفئة العمرية 17-19 مثلا، تمثل 73.5% في الى العامل الاول وحوالي 2% (1.5%) فقط في العامل الثاني. يمكن أيضا أن نقول ان العامل الاول يفسر 99% من تباين الفئة العمرية 17-19 عبر متغير الخطر، في حين أن العامل الثاني يفسر فقط حوالي 1% من تباين الفئة العمرية 17-19.

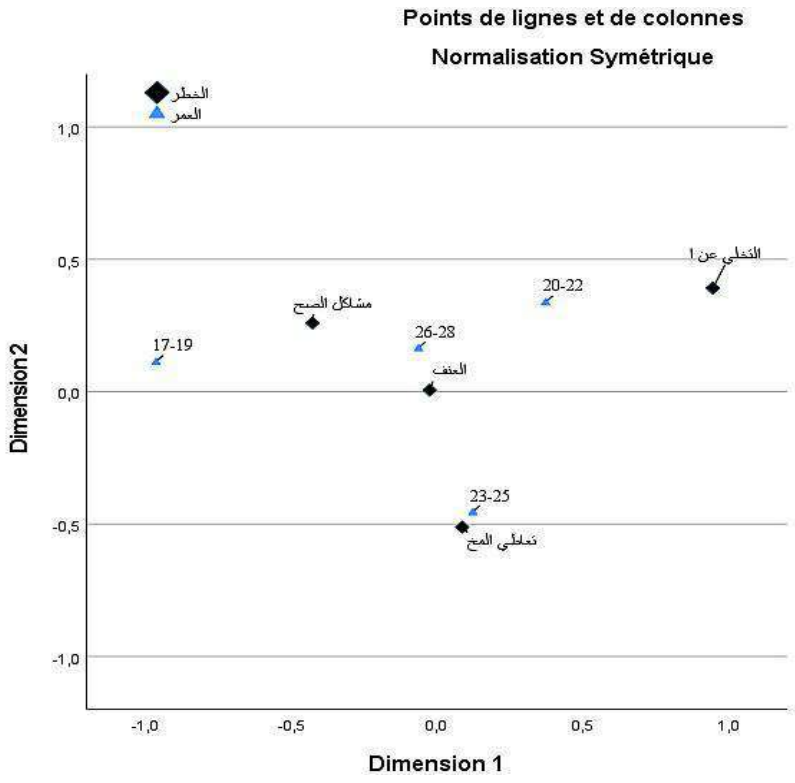
ايضا فئة خطر التخلي عن الدراسة ممثل بشكل جيد في العامل الاول (65%) بينما في العامل الثاني 18%. علاوة على ذلك ، يفسر العامل الاول حوالي 90% من تباين التخلي عن الدراسة عبر العمر، ويفسر العامل الثاني حوالي 10% من هذه الفئة.

يوفر الجدولين المواليين الانحرافات المعيارية لدرجات الصفوف والأعمدة في كل عامل، والتي تستخدم لتقييم دقة تقديرات النقاط على محاورها، تماما مثل استخدام مجالات الثقة في التحليل الاحصائي.

Points de ligne de confiance			
	Ecart type de la dimension		Corrélation
العمر	1	2	1-2
17-19	,046	,132	,540
20-22	,079	,061	-,662
23-25	,094	,036	,473
26-28	,065	,072	,016

Points de colonne de confiance			
الخطر	Ecart type de la dimension		Corrélation
	1	2	1-2
تعاطي المخدرات	,105	,045	,225
التخلي عن الدراسة	,078	,122	-,864
العنف	,054	,067	-,008
مشاكل الصحة العقلية	,063	,065	,741

كما ينتج عن تحليل CA باستخدام برنامج SPSS التمثيل البياني الذي يوفر عرض بياني يوضح كل القيم التي تنتمي لمجموعة البيانات، وهذا ما يوفر هذا رؤية شاملة للاتجاهات داخل البيانات. في هذا المثال ولان التحليل نتج عنه عاملين فقط يمكن عرض النتائج في SPSS من خلال الشكل الموالي:



ملاحظة

تجدر الاشارة الى ما يلي:

-في حالة استخدام ثلاثة أبعاد سيتم تمثيل النقاط برسم بياني ثلاثي الأبعاد لتمثيل كل محور.

-تكشف إحصائية chi-square عن قوة الاتجاهات داخل البيانات والتي تستند إلى مسافات النقاط للفئات.

-تعطي المسافة بين أي نقاط صفوف أو نقاط عمود مقياسا للتشابه (أو الاختلاف).

-النقاط التي يتم تعيينها بالقرب من بعضها البعض لها ملفات تعريف متشابهة، في حين أن النقاط التي يتم تعيينها بعيد عن بعضها البعض لها ملفات تعريف مختلفة تماما.

-يتم تفسير المسافات بين نقاط الصف والعمود بشكل مختلف.

وهكذا قمنا بتمثيل جميع المعطيات بيانيا، ويمكن ملاحظة بعض الاتجاهات العامة. فمثلا نلاحظ أن الطلبة الذين تتراوح اعمارهم بين 23 و 25 سنة معرضون بشكل خاص للمشاكل المتعلقة بتعاطي المخدرات، كما أن الفئة العمرية 17-19 لا تعاني كثيرا من خطر التخلي عن الدراسة، ولكنها أكثر عرضة لمشاكل الصحة العقلية. يبدو أن الفئة العمرية 26-28 أكثر عرضة لعوامل الخطر المتعلقة بالعنف والصحة العقلية.

وهذه النتائج تعتبر ذات اهمية بالغة، حيث يمكن إجراء بحوث إضافية لتحليل واستكشاف ما يجعل الطلبة بمختلف الفئات العمرية تتخلى عن الدراسة او تتعاطى المخدرات/ وما الى ذلك. وبشكل عام ، يوفر تحليل CA صورة شاملة عن الارتباطات بين مختلف المتغيرات

الفصل الخامس

طريقة التصنيف التسلسلي

مقدمة الفصل

لقد رأينا من خلال الفصول السابقة أنه يمكننا استخدام مختلف طرق التحليل العاملي لتجميع المتغيرات وفقا للتباين المشترك. حيث نأخذ في التحليل العاملي عدة متغيرات ونفحص مقدار التباين الذي تشاركه هذه المتغيرات، ثم نقوم بتجميع المتغيرات التي تشترك في نفس العامل. كما حددنا عدة طرق مختلفة لقياس المسافة (أو الاختلاف) بين الصفوف أو بين أعمدة مصفوفة البيانات.

ولكن عوض دمج المتغيرات في عامل واحد يمكننا تصنيفها في مجموعات متجانسة بناءا على استجابة كل عنصر من عناصر المجموعة لخصائص كل مجموعة. وهذا ما سنهتم به في هذا الفصل وذلك بعرض طريقة التصنيف التسلسلي التي تساعد على تجميع البيانات بناء على تشابه او استجابتها مع العديد من المعايير والمتغيرات. حيث تعمل هذه الطريقة بأسلوب تسلسلي او هرمي، حيث يتم ربط العناصر معا بطريقة تسلسلية من الأقرب (الأكثر تشابها) إلى الأبعد.

وفي هذا الفصل نوضح طريقة التصنيف التسلسلي من خلال مثال توضيحي مع عرض الخطوات المختلفة للمتبعة في التحليلي، بالإضافة تطبيق عملي باستخدام برنامج SPSS.

1.5. المبدأ الاساسي للتصنيف التسلسلي

تحليل التصنيف التسلسلي او بالإنجليزية Hierarchical cluster analysis (HCA) هو طريقة تحليل استكشافية مصممة للكشف عن المجموعات المتشابهة داخل مجموعة البيانات، والتي لم تكن واضحة ولم يكن ممكنا ملاحظتها.

تعتبر عملية التصنيف التسلسلي للبيانات من أهم مراحل بناء النظريات العلمية، ويعتبر التحليل التسلسلي (Cluster Analysis) أسلوبا مناسبا لتصنيف الظواهر. ويهتم هذا النوع من التحليل بتقسيم وتصنيف عناصر بيانات المتغيرات إلى عدة مجموعات جزئية تكون متجانسة داخل المجموعة الواحدة، وفي نفس الوقت تكون كل مجموعة متباينة ومختلفة بالنسبة إلى المجموعات الأخرى.

و تقوم التصنيف التسلسلي بتجميع المفردات المتشابهة في مجموعات تسمى Clusters، حيث تكون نقطة النهاية عبارة عن مجموعة من المجموعات. وتختلف كل مجموعة عن المجموعة الاخرى، وتكون المفردات داخل كل مجموعة متشابهة إلى حد كبير مع بعضها البعض.

يسمح التصنيف التسلسلي بدمج عمليات التصنيف والترتيب البسيطة، حيث يحتوي التسلسل على هيكل شجرة منظم حسب المستويات والفروع، وفي كل مستوى توجد الفئات المقابلة لنفس المتغير وفي كل فرع يتم توزيع عناصر التصنيف التسلسلي من العام إلى الخاص أو العكس. في حياتنا اليومية مثلا نقوم بتصنيف وترتيب عدة أمور بشكل تسلسلي، فنقوم بتحديد الأنشطة التي يجب القيام بها أولا خلال اليوم أو نحدد الاشياء التي يجب اقتنائها خاصة تلك التي يعتبر استهلاكها من الضروريات.

1.1.5. الفكرة الاساسية

التصنيف هو تجميع عناصر او مفردات متشابهة فيما بينها وفقا لمعيار معين. ومن الناحية العملية لا يمكن أو من الصعب فحص جميع الاحتمالات التي تساعدنا في تجميع العناصر المتشابهة، حتى مع أسرع أجهزة الكمبيوتر. لذلك تم تطوير طرق تحليلية تساعد على تصنيف العناصر وتجميعها في مجموعات.

ومن الناحية العملية لا يمكن أو من الصعب فحص جميع الاحتمالات التي تساعدنا في تجميع العناصر المتشابهة، حتى مع أسرع أجهزة الكمبيوتر. يعتبر التصنيف التسلسلي أسلوب تتمثل فكرته الأساسية في تجميع مجموعة من الملاحظات في عدد معين من المجموعات أو Clusters. يعتمد هذا التجميع او التصنيف على فكرة المسافة أو التشابه بين الملاحظات التي تعتمد اساسا على القياس متعدد الأبعاد. بمعنى ان الحصول على هذه المجموعات يعتمد على المعيار المسافة ويتوقف على ما نعتبره متشابها (Sedkaoui and Khelfaoui, 2020).

يتشابه المبدأ الاساسي لكل طريقة من كرق التصنيف من حيث أنه يبدأ بمعالجة جميع الحالات كمجموعة واحدة في حد ذاتها. يتم بعد ذلك دمج المجموعات بناء على معيار خاص بالطريقة المختارة. لذلك ففي جميع انواع التصنيف نبدأ بالمجموعات (Clusters) بقدر عدد الحالات وننتهي بمجموعة واحدة فقط تحتوي على جميع الحالات. ومن خلال فحص تقدم الدمج التسلسلي يمكن عزل مجموعات الحالات ذات التشابه العالي.

والتصنيف التسلسلي هو تحليل استكشافي يحاول تحديد الهياكل داخل البيانات، ويسمى أيضا تحليل التجزئة (segmentation analysis) أو التحليل العنقودي. نظرا لأن هذا النوع من التحليلي استكشافي فإنه لا يميز بين المتغيرات التابعة والمستقلة (Sedkaoui, 2018).

وغالبا ما يستخدم التصنيف التسلسلي بالاقتران مع التحليلات الأخرى، مثل التحليل التمييزي (discriminant analysis).

ملاحظة

يقوم التحليل التمييزي (discriminant analysis) بإجراء اختبار متعدد المتغيرات للاختلافات بين المجموعات. بالإضافة إلى ذلك يتم استخدام التحليل التمييزي لتحديد الحد الأدنى لعدد الأبعاد اللازمة لوصف هذه الاختلافات. يتم التمييز أحيانا بين التحليل الوصفي التمييزي والتحليل التمييزي التنبئي. وتجدر الاشارة الى اننا لن نقوم بتفضيل الطريقة في هذه المطبوعة. التحليل التمييزي هو طريقة إحصائية متعددة الاستخدامات غالبا ما تستخدم لتصنيف الملاحظات إلى مجموعتين أو أكثر من المجموعات أو الفئات. بمعنى آخر، يتم استخدام هذا النوع من التحليل لتعيين كائنات لمجموعة واحدة من بين عدد من المجموعات المعروفة.

يجب أن يكون الباحث قادرا على تفسير التصنيف التسلسلي بناء على اتعابه لبيانات الدراسة حتى يتمكن من تحديد ما إذا كانت النتائج التي سيحصل عليها من التحليل ذات مغزى. ومن بين الأسئلة البحثية النموذجية التي يجيب عليها التصنيف التسلسلي نجد:

في التسويق مثلا يعمل هذا النوع من التحليل على تصنيف شرائح العملاء. فاذا قام الباحث باجراء مسح يغطي احتياجات العملاء ، والمواقف ، والتركيبة السكانية ، وسلوك العملاء، فقد يستخدم التصنيف التسلسلي لتحديد مجموعات متجانسة من العملاء لديهم احتياجات ومواقف متشابهة.

اما في التعليم فقد يستخدم التصنيف التسلسلي لتقسيم الطلاب الى مجموعات بغرض تحديد مجموعات الطلاب التي تحتاج إلى اهتمام خاص ومتابعة مثلا. فقد يحدد هذا التحليل المجموعات المتجانسة الموجودة بين الطلاب المتفوقون في جميع المقاييس، أو الطلاب الذين يتفوقون في مقاييس معينة ولكنهم يفشلون في مقاييس أخرى.

وهذا التحليل عبارة عن عملية تجميع العناصر في مجموعات فرعية لها معنى في سياق مشكلة معينة. وبالتالي يتم تنظيم هذه العناصر في تمثيل فعال يميز المجتمع الذي يتم أخذ عينات منه.

ويمكن القول أن الأفراد متشابهون إذا كانوا ينتمون إلى نفس الفئة أو المجموعة بطريقة تجعل جميع الأفراد في نفس المجموعة يشبهون بعضهم البعض، ويختلفون عن أفراد مجموعة أخرى. من خلال تحديد المجموعات من الممكن العمل بطريقة مجزأة دون الحاجة إلى مواصلة العمل مع العينة بأكملها.

وعند استخدام التحليل التصنيفي لا بد من مراعاة نقاط مهمي تتمثل في:

- تحديد الأفراد قيد الدراسة. يجب إيلاء اهتمام خاص للقيم المتطرفة (outliers) التي يمكن أن تعيق تجميع المجموعات المتجانسة.
- تحديد المتغيرات أو البيانات التي تصف وتميز العينة، ويجب ان يكون الاختيار متناسقا مع اهداف البحث.
- تحديد مقياس المسافة بين الافراد والذي يساعد في تفسير التجمعات الناتجة وتحديد أنسب نقطة فاصلة.
- تحديد طرق التصنيف وذلك بناء على الهدف المراد تحقيقه (تصنيف تسلسلي، عير تسلسلي)
- تلخيص المجموعات الناتجة. عن طريق التمثيل البياني
- التعليق على النتائج وتفسيرها

2.1.5. أنواع التحليل التصنيفي

ومن خلال ما سبق يمكننا تعريف التصنيف التسلسلي كما يلي:

" هو احد الاساليب الاحصائية الرياضية لتقسيم عناصر العينة المدروسة الى عدة مجموعات متجانسة داخليا (متشابهة) ومتباينة خارجيا عن بعضها البعض".

التحليل التصنيفي هو أسلوب متعدد المتغيرات يهدف إلى تصنيف العناصر بناء على طابعها المتشابه. تهدف طرق التصنيف المختلفة إلى توزيع n عدد من الأفراد، مع وجود p متغير $(X_1, X_2, \ldots X_p)$ في عدد m من المجموعات الفرعية التي تكون متجانسة قدر الإمكان.

بشكل عام تستخدم قياس المسافة بالطريقة الإقليدية، ويهدف هذا التحليل الى جعل تباين العناصر داخل كل مجموعة اصغر ما يمكن، والتباين بين مراكز المجموعات اكبر ما يمكن (Sedkaoui, 2018). وبضورة عامة يتفرع هذا التحليل الى نوعين اساسيين هما: التصنيف التسلسلي ، والتصنيف غير التسلسلي.

- **التصنيف التسلسلي**: وهي طريقة تصنيف تقوم بتجميع n عنصر في مجموعات، حيث يكون عدد المجموعات التي ستظهر غير معروف لذلك نحتاج إلى إيجاد العدد الأمثل للمجموعات. التصنيفات التسلسلية هي في الأساس طريقة تصنيف غير خاضعة للإشراف. في هذا النوع من التصنيف يتم تضمين

المجموعات المتشابهة معا وترتيبها بطريقة هرمية. يعمل التصنيف التسلسلي بشكل متكرر على ربط أزواج المجموعات حتى يتم تضمين كل عنصر من عناصر البيانات في التسلسل الهرمي.

- **التصنيف غير التسلسلي**: أما بالنسبة لطريقة التصنيف غير التسلسلي فهي طريقة تجمع العناصر المعروفة أو المحددة مسبقا بحيث يمكن في هذه الطريقة تعديل عدد المجموعات وفقا لغرض البحث. والتصنيفات غير التسلسلية تؤدي إلى تحلل مجموعة الأفراد إلى m مجموعات منفصلة أو فئات متكافئة، حيث يحدد العدد m من الفئات مقدما. والنتيجة التي يتم الحصول عليها هي تقسيم مجموعة الأفراد أو مجموعة من الأجزاء أو فئات المجموعة I من الأفراد. يتضمن التصنيف غير التسلسلي تكوين مجموعات جديدة من خلال دمج أو تقسيم المجموعات، ولا يتبع هيكل شجرة مثل التصنيف التسلسلي. ومن بين اهم الطرق نجد كريقة K-means حيث تكون العلاقة بين المجموعات غير محددة.

في كلتا الطريقتين تتمثل إحدى المشكلات المهمة في كيفية تحديد التشابه بين عدة عناصر بحيث يمكن تكوين مجموعات من العناصر ذات تشابه كبير مع بعضها البعض.

ويعتبر التصنيف التسلسلي من الاساليب المفضلة في التصنيف، لأنه يعتمد على اسس بسيطة، ويعمل على تشكيل مجموعات متكونة من افراد او عناصر العينة (n عنصر). وبشكل متتالي، ضمن m مجموعة (او عنقود) عن كريق دمج العناصر المتقاربة ضمن مجموعات بحيث تكون المجموعة الاولى ابسط المجموعات. وتتألف كل مجموعة (او عنقود) من عدة مجموعات متقاربة ومرتبطة مع بعضها بواسطة علاقات تحقق شروط التقارب.

في الأساس هناك نوعان من الاساليب لتحليل التصنيف التسلسلي (Sedkaoui and Khelfaoui, 2020):

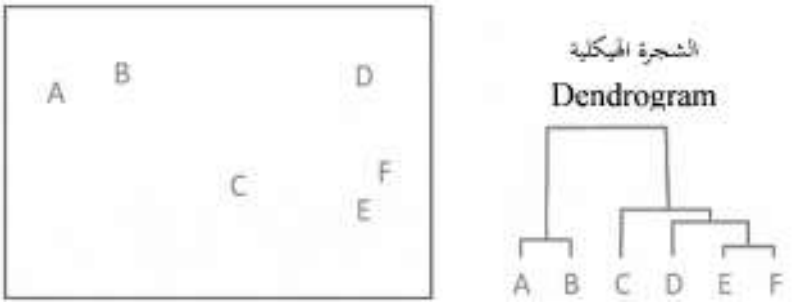
- **اسلوب التجميع (hierarchical agglomerative clustering)**: يعرف أيضا بالنهج التصاعدي. يتم التعامل مع البيانات من الأسفل إلى الأعلى، اي يتم التعامل معها على أنها كتلة مفردة في البداية ثم تتشكل مجموعات على التوالي حتى يتم دمج جميع المجموعات في مجموعة واحدة تحتوي على جميع البيانات. ويعتمد هذا الاسلوب على مصفوفة التقارب بين مفردات العينة حسب المسافات المحسوبة.
- **اسلوب التقسيم (hierarchical divisive clustering)**: لا يتطلب هذا الاسلوب التحديد المسبق لعدد المجموعات، ويتطلب التجميع من أعلى إلى أسفل طريقة لتقسيم المجموعة التي تحتوي على البيانات بأكملها وتستمر عن طريق تقسيم المجموعات بشكل متكرر حتى يتم تقسيم البيانات الفردية إلى مجموعة مفردة.

وتجدر الاشارة الى ان هذين الاسلوبين يعتمدان على بيانات العينة المدروسة وعلى طبيعة المتغيرات المستقلة $X_1, X_2, \ldots . X_p$ المستخدمة في عملية العنقدة (clustering).

ملاحظة

اذا كانت المتغيرات X1, X2, ... Xp كمية، فاننا نقوم بحساب عناصر مصفوفة التباعد (Dissimilarity) والتي تعبر عن المسافات التي تفصل بين افراد العينة. اما اذا كامن المتغيرات نوعية أو مختلطة فاننا نقوم بحساب عناصر مصفوفة اخرى تعرف بمصفوفة التشابه (Similarity) وهي عبارة عن اوزان التكرارات التي تقابل العناصر المتشابهة.

ويمكن تمثيل نتائج هذين الاسلوبين بيانيا غلى شكل عناقيد مترابطة يعرف بالشجرة الهيكلية (Dendrogram) والذي يظهر العلاقة الهرمية بين المجموعات:



وتجدر الاشارة الى التصنيف التسلسلي يعتمد على عدة طرق تعرف بطرق الربط وهي (Cordan, 1999): طريقة الربط المنفرد (Single Linkage)، طريقة الربط التام (Complete Linkage)، طريقة الربط المتوسط (Average linkage) بالإضافة الى طريقتين هما: طريقة Ward وطريقة التمركز (Centroid). والجدول الموالي يوضح هذه الطرق:

الجدول رقم 5-01: وصف طرق حساب المسافة بين الافراد والمجموعات

الطريقة	الوصف
طريقة الربط المنفرد	يتم ضم المجموعات مع الأخذ في الاعتبار أصغر المسافات الموجودة بين أقرب عناصر المجموعات المختلفة.
طريقة الربط التام	يتم ضم المجموعات من خلال النظر في أصغر المسافات الموجودة بين العناصر الأبعد في المجموعات المختلفة.
طريقة الربط المتوسط	يتم حساب المسافة بين المجموعات من خلال حساب متوسط المسافة بين جميع أزواج الملاحظات بغض النظر عما إذا كانت متقاربة أو متباعدة.
طريقة Ward	يتم حساب المسافة بين مجموعتين كمجموع المربعات بين المجموعات في ANOVA. الهدف هو تقليل التباين داخل المجموعة وتعظيم التجانس داخل المجموعات.
طريقة التمركز	يتم حساب المسافة بين مجموعتين على أنها المسافة بين مركزهما.

علما ان كل طريقة من هه الطرق تعتمد عند دمج الافراد او العناقيد (المتكونة من الافراد) على معيار معين لقياس التشابه بين الجانبين (Zhang et al., 2017).

كما انه يتم تنظيم بيانات العينة المتكونة من n عنصر او مشاهدة حسب المتغيرات النظامية (المعيارية او الثنائية) المعتمدة في عملية التحليل وتوضع في جدول كما يلي:

الجدول رقم 5-02: نموذج جدول بيانات التصنيف التسلسلي

المتغيرات / الافراد	X_1	X_2	... X_i ...	X_p
1	x_{11}	x_{12}	... x_{1i}...	x_{1p}
2	x_{21}	x_{22}	... x_{2i}...	x_{2p}
.	...	...	...	...
j	x_{j1}	x_{j2}	... x_{ji}...	x_{jp}
.	...	...	...	...
N	x_{n1}	x_{n2}	... x_{ni}...	x_{np}

وعموما ينتج عن طريقة التصنيف التسلسلي سلسلة من التكتلات التي يتم فيها دمج مجموعات صغيرة تحتوي في البداية على جزيئات فردية لتشكيل مجموعات أكبر تدريجيا.

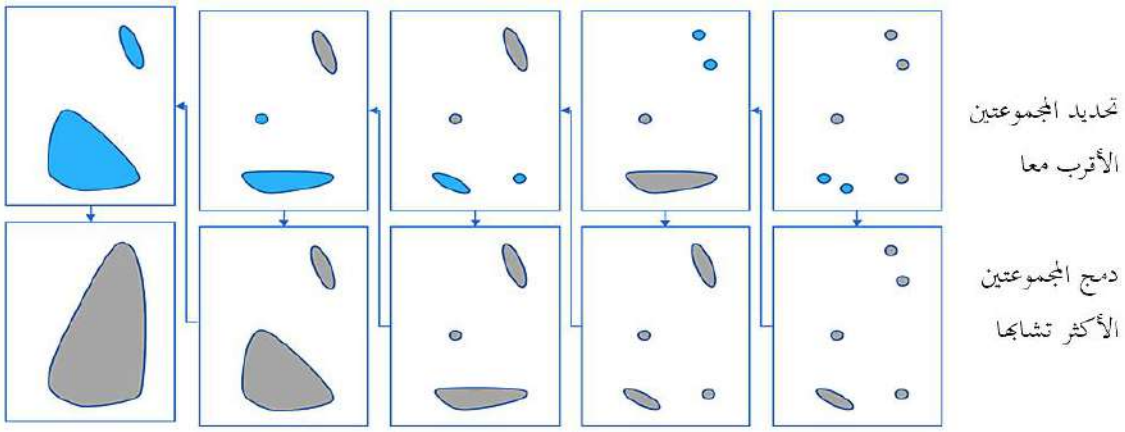
3.1.5. طريقة عمل التصنيف التسلسلي

يبدأ التصنيف التسلسلي بمعالجة كل ملاحظة على أنها مجموعة منفصلة، بعد ذلك ينفذ بشكل متكرر الخطوتين التاليتين:

- تحديد المجموعتين الأقرب معا؛
- دمج المجموعتين الأكثر تشابها.

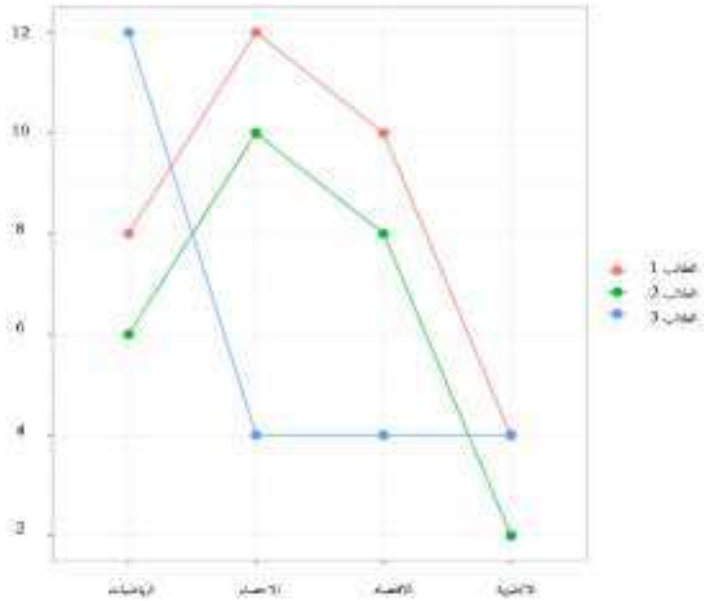
وتستمر هذه العملية بصفة متكررة حتى يتم دمج كل المجموعات، ويمكن توضيح ذلك من خلال الشكل التالي:

الشكل رقم 5-01: طريقة عمل التصنيف التسلسلي



وحتى نبسط العملية اكثر نفرض اننا اردنا قياس علامات ثلاث طلاب في اربع مقاييس هي: الرياضيات، الاحصاء، الاقتصاد، والانجليزية. ومن خلال التحليل العاملي نحصل على العوامل لكل من هذه المكونات الأربعة كمقياس بسيط للتشابه بين درجاتهم. نفرض ان نتائج التحليل العاملي كانت كما في الشكل الموالي:

الشكل رقم 5-02: العلاقة بين علامات الطلبة



من خلال الشكل السابق يتضح أن لدى الطالب الاول والطالب الثاني نمط ردود متشابه جد في المقاييس الاربعة، وهذا مت تظهره الخطوط المتوازية نسبيا. وهذا ما يشير الى أن الاختلاف النسبي في علاماتهم في المقاييس الاربعة هو نفسه.

اما الطالب الثالث فمن الواضح ان مجموعته مختلفة جدا مقارنة بالطالبين الاخرين، لكن يمكننا ملاحظة ان علامته في الانجليزية مشابهة للطالب الاول وقريبة للعلامة الطالب الثاني، ولكن بالنسبة للمقاييس الثلاثة الاخرى فان علاماته مختلفة كثيرا عن الطالبين الاول والثاني. لذلك يمكننا تجميع الطالبين الاول والثاني معا بناء على حقيقة أن علاماتهم متشابهة.

قد يكون هذا الاستنتاج ممكنا في حالة لكن اذا اردنا ان نقارن كل طلبة السنة الثالة علوم التسيير في كل المقاييس المدرسة فانه سيكون من الصعب جدا تفسير واستخراج نتائج من التمثيلات البيانية. لذلك نحتاج إلى طريقة موضوعية لقياس درجة التشابه بين علامات كل الطلبة في جميع المقاييس. وعادة ما نستخدم نوعان من القياسات: معاملات التشابه ومعاملات الاختلاف (Coquillard, 2018).

فمثلا عادة ما نعتمد في الاحصاء على معامل الارتباط كمقياس لتفسير التشابه بين متغيرين لانه يشير الى ما إذا كانت التغيرات التي تحدث في متغير واحد تتبع الى حد ما نفس النمط مقارنة بمتغير اخر. ويعد معامل الارتباط مقياسا موحدا لانه لا يتأثر باختلافات التشتت بين المتغيرات. ومع ذلك هناك مشكلة في استخدام معامل ارتباط فهو يتجاهل المعلومات حول ارتفاع العلامات مثلا. لذلك على الرغم من كونه يشير الى نمط الاستجابات بين علامات الطلبة إلا أنه لا يترجم المسافة بين مختلف العلامات. على هذا النحو فإن معامل الارتباط يفتقد إلى معلومات مهمة.

في هذه الحالية نلجأ الى مقياس بديل يتمثل في المسافة الإقليدية (Euclidean Distance)، والتي يرمز لها بالرمز d وتمثل المسافة الهندسية بين عنصرين (أو حالاتين). لذلك ، إذا كنا سنسمي الطالبين الاول والثاني بالعنصرين الاول والثاني على التوالي فيمكننا التعبير عن المسافة الإقليدية الخاصة بهم من خلال العلاقة التالية:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$$

تعني هذه المعادلة ببساطة أنه يمكننا اكتشاف المسافة بين الطالب الاول و الطالب الثاني من خلال أخذ علاماته على اساس متغير k وحساب الفرق.

فبالنسبة لبعض المقاييس (المتغيرات) ستكون علامات الطالب الاول اعلى من علامات الطالب الثاني والعكس. لذلك ستكون بعض الاختلافات (الفرق) إيجابية وبعضها الآخر سلبية. في النهاية ، نريد جمع الاختلافات عبر عدد من المتغيرات وبالتالي إذا كان لدينا فرق موجب وسالب فقد يتم حذفهما، لتجنب هذه المشكلة نقوم ببساطة بترتيب كل فرق قبل جمعها مع تربيع كل فرق.

هذه العملية تعني اننا انتقلنا إلى المتغير (المقياس) التالي، وفي كل مرة نكرر العملية فاننا نضيف جميع الاختلافات، وهذه العملية تشبه حساب التباين.

ملاحظة

وعند حساب المسافة الإقليدية فانه كلما كانت المسافة أصغر كانت الحالات أكثر تشابها. ومع ذلك ، يتأثر هذا المقياس بشدة بالمتغيرات ذات الحجم الكبير أو اختلافات التشتت. لذلك ، إذا تمت مقارنة الحالات عبر متغيرات لها تباينات مختلفة جدا، فإن المسافات الإقليدية ستكون غير دقيقة. على هذا النحو من المهم توحيد الدرجات قبل متابعة التحليل. يعد توحيد الدرجات أمرا مهما بشكل خاص إذا تم قياس المتغيرات على مستويات مختلفة.

تجدرالاشارة الى انه يمكن استخدام مقاييس اخرى لقياس المسافة كمقياس تشيبيشف (Tchebychev) والذي يحسب القيم المطلقة الاعظمية لقيم المتغيرات المعينة في التصنيف.

بعد تحديد مقياس المسافة من الضروري تحديد مكان حساب المسافة. على سبيل المثال يمكن حسابها بين أكثر جزأين تشابها في المجموعة (طريقة الربط المنفرد)، أو بين العنصرين الأقل تشابها (طريقة الربط التام) ، أو مركز المجموعات (طريقة الربط المتوسط) .كما هو الحال مع مقاييس المسافة، يجب أن يتم اختيار معايير الربط بناء على الاعتبارات النظرية من مجال التطبيق. القضية النظرية الرئيسية هي ما الذي يسبب الاختلاف.

في حالة عدم وجود مبررات نظرية واضحة لاختيار معايير الربط فإن طريقة Ward هي الطريقة الافضل، لانها تعمل على تحديد الملاحظات التي يتم تجميعها بناء على تقليل مجموع المسافات المربعة لكل ملاحظة من متوسط الملاحظة في مجموعة (Cordan, 1999).

وبمجرد قياس التشابه بين العناصر يمكننا البدا في عملية تصنيف او تجميع العناصر بناء علىدرجة التشابه. وهنا يعمل التصنيف التسلسلي بطريقة هرمية حيث يبدأ بمعالجة جميع الحالات كمجموعة في حد ذاتها. يتم بعد ذلك دمج المجموعات، ويندأ بالعناقيد بقدر عدد الحالات وينتهي بمجموعة واحدة فقط تحتوي على جميع الحالات.

2.5. خطوات التصنيف التسلسلي

ان إجراء التحليل باستخدام التصنيف التسلسلي يتطلب اتباع بعض الخطوات التي تبدأ من الملاحظة وصولا الى التمثيل البياني وتفسير النتائج. ولإبراز خطوات التصنيف التسلسلي نعتمد على بيانات المصفوفة التالية:

A	18	0
B	22	0
C	43	0
D	42	0
E	27	0
F	25	0

1.2.5. تشكيل المجموعات (العناقيد)

تتمثل الخطوة الأولى في اختيار مقياس الاختلاف أو المسافة بين العناصر.

يتيح استخدام المسافة الإقليدية حساب مصفوفة التباعد كما يلي:

الجدول رقم 5-03: مصفوفة التباعد

	A	B	C	D	E	F
A	0	4	25	24	9	7
B	4	0	21	20	5	3
C	25	21	0	1	16	18
D	24	20	1	0	15	17
E	9	5	16	15	0	2
F	7	3	18	17	2	0

مثلا نحسب المسافة الاقليدية بين A و B كما يلي:

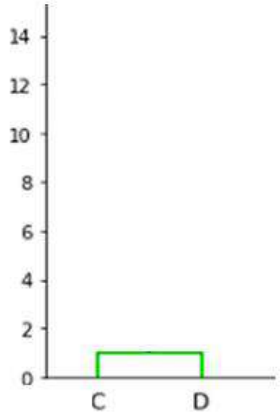
$$d_{AB} = \sqrt{(18-22)^2 + (0-0)^2}$$

$$d_{AB} = \sqrt{(-4)^2 + (0)^2} = \sqrt{16} = 4$$

وتتمثل الخطوة الموالية في عملية التصنيف التسلسلي في البحث عن زوج من العناصر الأكثر تشابها. بمعنى وجود أقل اختلاف بينهما. الحد الأدنى غير الصفري في هذا الجدول هو 1 ويتوافق مع المسافة بين العنصرين C و D، لذلك سيتم تشكيل المجموعة الأولى من خلال الجمع بين هذين العنصرين.

	A	B	C	D	E	F
A	0					
B	4	0				
C	25	21	0			
D	24	20	1	0		
E	9	5	16	15	0	
F	7	3	18	17	2	0

بما ان أقصر مسافة في المصفوفة هي 1 والعناصر المرتبطة بها هي C و D فإن المجموعة (First Cluster) الأولى هي C-D. وننقوم بضم هذين العنصرين عند المستوى 1 من مخطط الشجرة الهيكلية، كما يلي:



النقطة التي يلتقون عندها تسمى عقدة.

سنواصل بشكل أساسي تكرار هذه الخطوة لكن المشكلة الوحيدة هي كيفية حساب الاختلاف بين المجموعة الاولى CD والعناصر الأخرى. يحدد هذا القرار نوع التصنيف التسلسلي الذي نعتزم القيام به، وهناك العديد من الخيارات. في الوقت الحالي ، نختار واحدة من أكثر الطرق شيوعا زهي طريقة الربط المنفرد سيكون الاختلاف بين الزوج المدمجفي المجموعة والعناصر الاخرى هو الحد الادنى لزوج الاختلافات في كل حالة.

الان نقوم بحساب المسافة بين العناصر الاخرى والمجموعة CD كما يلي:

المسافة بين A و CD:

$$\text{Min } (A \rightarrow C, A \rightarrow D) = \text{Min } (25, 24) = 24$$

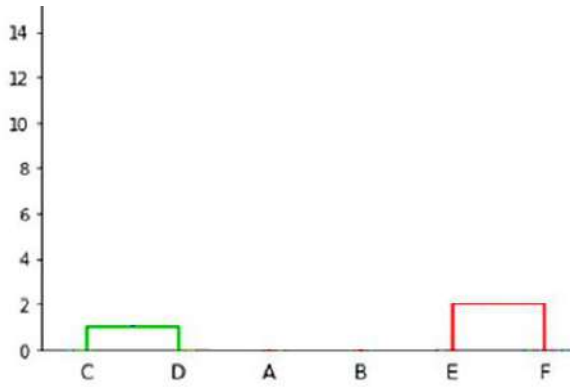
المسافة بين B و CD:

$$\text{Min } (B \rightarrow C, B \rightarrow D) = \text{Min } (21, 20) = 20$$

على سبيل المثال ، الاختلاف بين A و C هو 25، بينما الاختلاف بين A و D هو 24، ومن ثم نختار اقل قيمة من بين القيمتيم وهي 24 لتقدير الاختلاف بين A والمجموعى الاولى CD. وبالمثل نحسب المسافة بين كل من E و F والمجموعة CD وندرج النتائج في الجدول المالي:باستمرار بهذه الطريقة نحصل على مصفوفة جديدة:

	A	**B**	**CD**	**E**	**F**
A	0				
B	4	0			
CD	24	20	0		
E	9	5	15	0	
F	7	3	17	2	0

تتكرر العملية الآن حيث نبحث عن أصغر اختلاف في الجدول وهي 2 وتربط العنصرين E و F، ثم نقوم بتمثيلها عند مستوى 2، كما هو موضح في الشكل الموالي:



بعد البحث عن ادنى مسافة والعناصر المرتبطة بها وتشكيل المجموعة الثانية والمرتبطة بالعناصر E و F، نقوم بحساب المسافة بين هذه المجموعة والعناصر المتبقية كما يلي:

المسافة بين A و EF:

$$\text{Min}(A \rightarrow E, A \rightarrow E) = \text{Min}(9, 7) = 7$$

المسافة بين B و EF:

$$\text{Min}(B \rightarrow E, B \rightarrow E) = \text{Min}(5, 3) = 3$$

المسافة بين CD و EF:

$$\text{Min}(CD \rightarrow E, CD \rightarrow E) = \text{Min}(15, 17) = 15$$

ونجد:

	A	**B**	**CD**	**EF**
A	0			
B	4	0		
CD	24	20	0	
EF	7	3	15	0

نلاحظ ان ادنى قيمة هي 3 وتربط العنصرين B و EF وهي المجموعة الثالثة، نقوم بحساب المسافة بينها وبين العناصر الاخرى:

المسافة بين A و B-EF:

$$\text{Min}(A \rightarrow B, A \rightarrow EF) = \text{Min}(4, 7) = 4$$

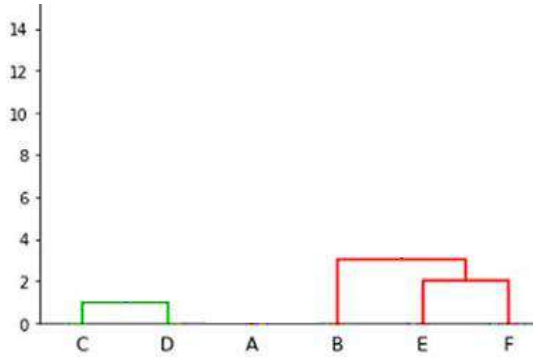
المسافة بين CD و BEF:

$$\text{Min}(CD \rightarrow B, CD \rightarrow EF) = \text{Min}(20, 15) = 15$$

ونكتب:

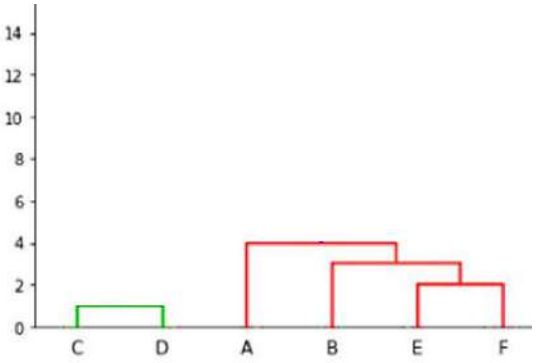
	A	BEF	CD
A	0		
BEF	4	0	
CD	24	15	0

بعد حساب المسافة (الاختلاف) بين المجموعة BEF وباقي العناصر نحصل على الشكل التوضيحي التالي (عند المستوى 3):



كما هو موضح في الرسم البياني السابق تم عرض المجموعتين EF و BEF مدمجتين، وهذا يطلق عليه اسم الشجرة الثنائية لأنه في كل خطوة يتم دمج مجموعات.

ومن الجدول السابق نلاحظ ان القيمة 4 هي ادنى قيمة (مسافة) وهي تربط العنصرين A و BEF وبالتالي فالمجموعة الرابعة هي A-BEF. الان نقوم بحساب المسافة بين هذه المجموعة وبين العنصر الاخرCD . يتم دمج هذه المجموعة عند المستوى 4 كما في الشكل الموالي:



ونكرر العملية ونقوم بحساب الاختلاف أو المسافة بين المجموعة CD و المجموعة ABEF:

$$\text{Min (CD}\rightarrow\text{A, CD}\rightarrow\text{BEF)= Min (24, 15) = 15}$$

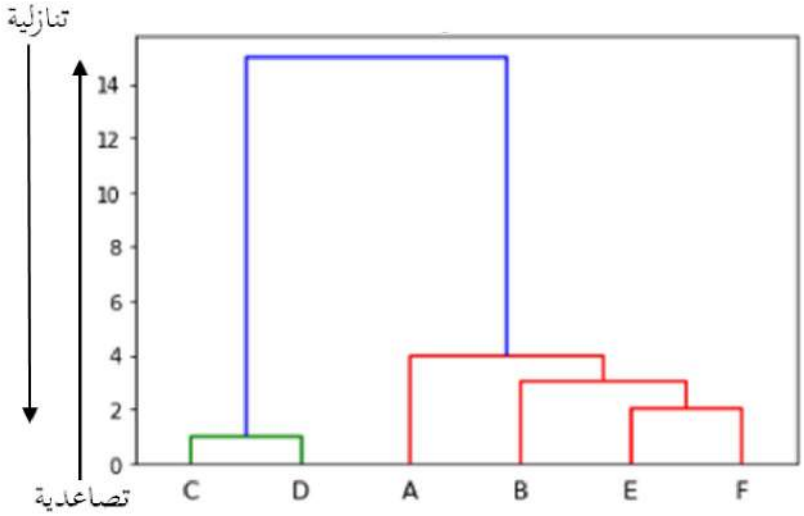
وبالتالي:

	A	CD
ABEF	0	
CD	15	0

والمجموعة الاخير هي المجموعة CD-ABEF

2.2.5. التمثيل البياني

والان يمكننا تمثيل هذه المجموعات بيانيا (الشجرة الهيكلية) لمجموعة الربط المنفرد.



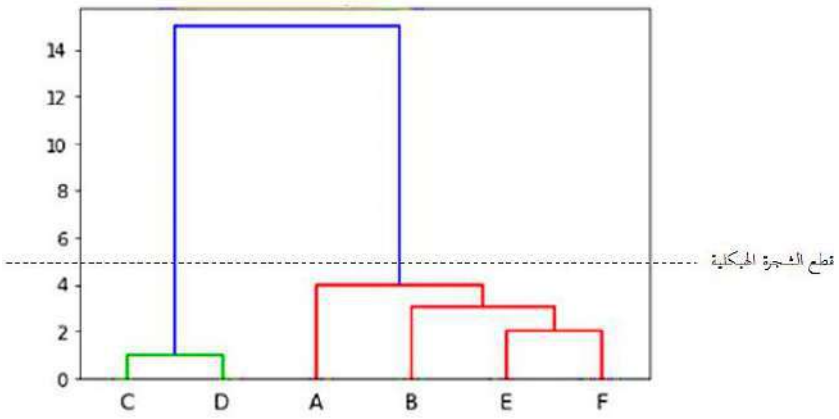
يعطينا مخطط dendrogram تكوين الفئات المختلفة وكذا الترتيب الذي تشكلت فيها. كما يشير المحور العمودي الى المستوى بين الفئتين اللتين تم تجميعهما في مرحلة معينة. مخطط الشجرة الهيكلية النهائي هو تصور لمصفوفة التباعد بناء على بيانات العناصر الستة الموضحة سابقا. وبواسطة هذه الشجرة أصبح تفسير بنية البيانات أسهل بكثير مما كان عليه قبل ادراء التحليل التصنيفي التسلسلي.

فيمكننا أن نرى أن هناك اربع أزواج من بيانات العينة قريبة إلى حد ما، ثلاثة منها (EF, BEF, ABEF) قريبة من بعضها البعض. في حين ان المجموعة الرابعة بعيدة بعض الشيء عن المجموعات الثلاث الاخرى.

ولأننا استخدمنا طريقة الحد الادنى، فإن جميع المجموعات المصنفة تحت مستوى معين من الاختلاف سيكون لها اختلافات بين المجموعات اعلى من هذا المستوى. على سبيل المثال 5 هو المستوى الذي تكون عنده المجموعات

متشابهة تماما مع بعضها البعض، وإذا نظرنا إلى المجموعات التي تقع تحت المستوى 5 أي (CD, EF, BEF) و (ABEF)، فبالتالي داخل كل مجموعة يكون للعناصر تشابه أكثر من 50% .

يحدث أيضا أن يتزامن المستوى 5 في مخطط الشجرة الهيكلية مع قفزة كبيرة في مستويات التصنيف، حيث تم دمج المجموعات الاربعة CD ، EF، BEF، و ABEF بينما العنقود الاخر (المجموعة الاخيرة) CDABEF تم دمجها عن المستوى 15، وبالتالي نقول ان المستوى 5 يعتبر مستوى مناسب جدا لقطع الشجرة الهيكلية.



إذا تم قطع الفروع عند المستوى 5 فسيتبقى لدينا المجموعات الاربع لعناصر البيانات: CD ، EF ، BEF، وABEF والتي يمكن تسميتها بالأنواع 1 ، 2 ، 3، و 4 على التوالي. بمعنى آخر قمنا بإنشاء متغير فئوي بأربع فئات ويتم تصنيف المجموعات على النحو التالي:

A	B	C	D	E	F
4	3	1	1	2	2

إذا تم إجراء تحليل التصنيف على مصفوفة بيانات فيمكن دائما الحصول على مجموعة من العناقيد، حتى لو لم يكن هناك تجميع فعلي للعناصر. إذن كيف يمكننا تقييم ما إذا كانت المجموعات الاربعة في هذا المثال ليست مجرد مجموعات قديمة كنا سنحصل عليها بناء على بيانات عشوائية بدون بنية.

هناك عدد كبير من الباحثين الذين اهتموا بتقييم نتائج التصنيف التسلسلي ومدى صلاحية صحة العناقيد وهنا سنشرح نهج واحدا يعتمد على التبديل. في مثالنا تم تشكيل المجموعات الاربع بحيث تكون الاختلافات بين عناصر العينة أقل من 5 داخليا في كل مجموعة مكونة من أكثر من عنصر واحد.

في الواقع إذا نظرنا إلى النتيجة في الشجرة الهيكلية يمكن خفض نقطة القطع لثلاث مجموعات إلى المستوى 4 المجموعتين BEF و ABEF متصلين معا. كما هو الحال في جميع الاعتبارات الإحصائية ذات الأهمية نسأل عما إذا كانت هذه نتيجة غير عادية أم أنها نشأت عن طريق الصدفة فقط.

للإجابة على هذا السؤال نحتاج إلى فكرة عما يمكن أن يحدث حتى نتمكن من الحكم على نتائجنا الفعلية. يمكن إنشاء ذلك من خلال ادراء تبديلات في البيانات بطريقة معقولة وتقييم النتيجة، مع ضرورة القيام بهذه العملية عدة مرات للتأكد من النتائج.

3.5. التصنيف التسلسلي باستخدام برنامج SPSS

يمكن أن تتعامل طرق تحليل التصنيف التسلسلي التي يقدمها SPSS مع البيانات الاسمية والترتيبية والكمية. من خلال استخدام SPSS نحصل على نوعين من المعلومات، معلومات على شكل ارقام والتي تحدد المسافات بين المجموعات ومختلف المعلومات الاخرى والرسم البياني لمخطط الشجرة الهيكلية والطي آخر يسمح بصريا بتحديد المجموعات أو العناقيد التي تم تشكيلها.

لأجل هذا قمنا بإعداد بيانات وهمية ولكنها مناسبة لتحليل التصنيف التسلسلي. ، ثم انشاء هذه البيانات حيث كانت الافراد أو الحالات متمثلة في 44 استاذ كلية العلوم الاقتصادية لجامعة خميس مليانة. أما بالنسبة للمتغيرات فيهي تتضمن ما يلي:

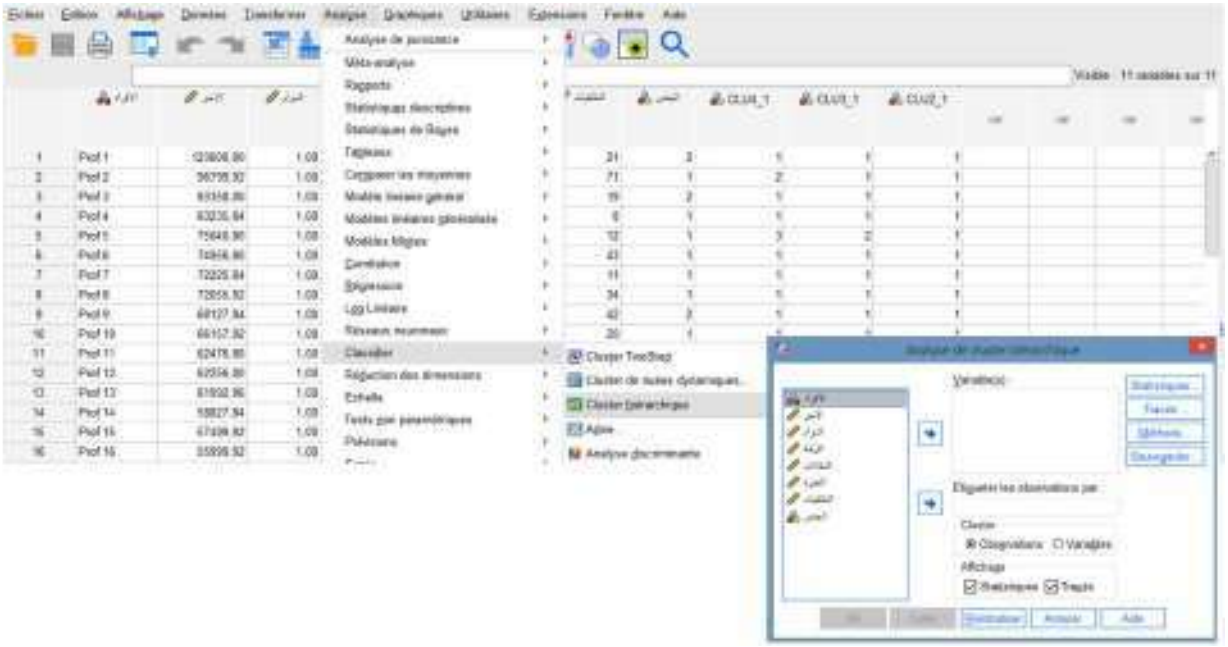
- الافراد: والذي يتضمن رقم الاساتذة من Prof 1 الى Prof 44.
- الاجر: على الرغم من أن رواتب الاساتذة هي معلومات عامة إلا أننا قمنا بإدراج اجر كل استاذ بالدينار الجزائري.
- الدوام: العمل المكافئ لدوام كامل
- الرتبة: حيث 1 تمثل استاذ مساعد ب ، 2 تمثل استاذ مساعد أ ، 3 تمثل استاذ محاضر ب، 4 تمثل استاذ محاضر أ، و 5 تمثل استاذ التعليم العالي
- المقالات: اي عدد المقالات العلمية المنشورة
- الخبرة: عدد سنوات العمل كأستاذ بدوام كامل في الكلية.
- الملتقيات: عدد الملتقيات الدولية التي شارك فيها الاستاذ.
- الجنس: 1 يمثل ذكر، و 2 يمثل انثى.

1.3.5. مراحل التصنيف التسلسلي في SPSS

بعد ادراج كل اجابات الاساتذة في برنامج SPSS سنبدأ الان بتحليل البيانات باستخدام التصنيف التسلسلي. من اجل ذلك نضغط على:

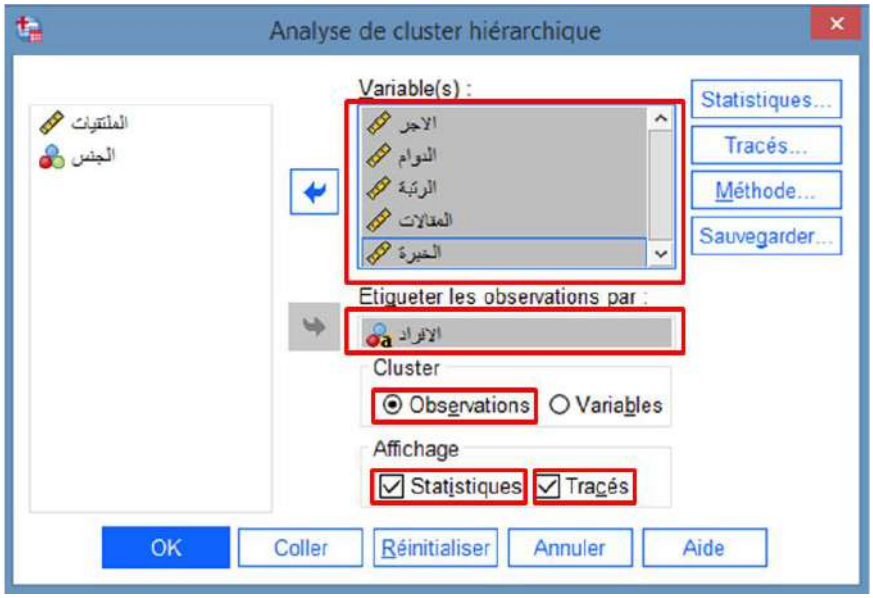
Analyse → Classifier →Cluster Hiérarchique

ونحصل على مربع الحوار المبين على يمين الصورة الموالية

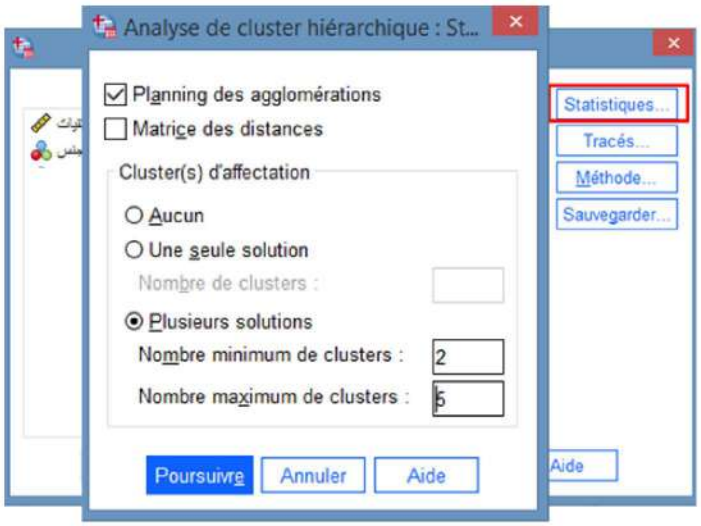


نقوم بتحديد الافراد على أنه المتغير الذي يمكن من خلاله تسمية الحالات كما نقوم بإدراج المتغيرات التالية: الافراد، الاجر، الدوام، الرتبة، المقالات، الخبرة كمتغيرات.

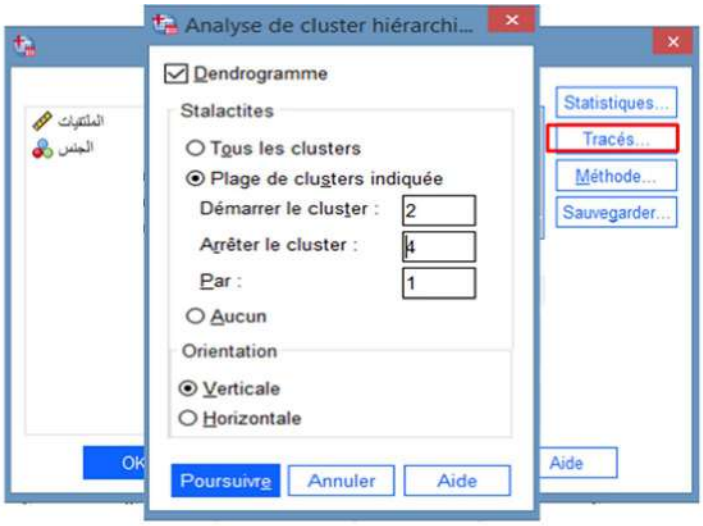
يجب ان نشير في مربع الحوار الى اننا نريد تصنيف الحالات او المشاهدات (observations) بدلا من المتغيرات (variables) كما يجب الاشارة الى عرض اكل من الإحصاءات الممكنة (Statistiques)والتمثيلات البيانية (Tracés).



في نفس مربع الحوار نقوم بالنقر على Statistiques ونقوم بالإشارة على اننا نريد رؤية جدول تصنيف Planning des agglomérations بالإضافة الى Plusieurs solutions لرؤية حلول المجموعة 2، 3، 4، و5. ثن نضغط على Poursuivre.

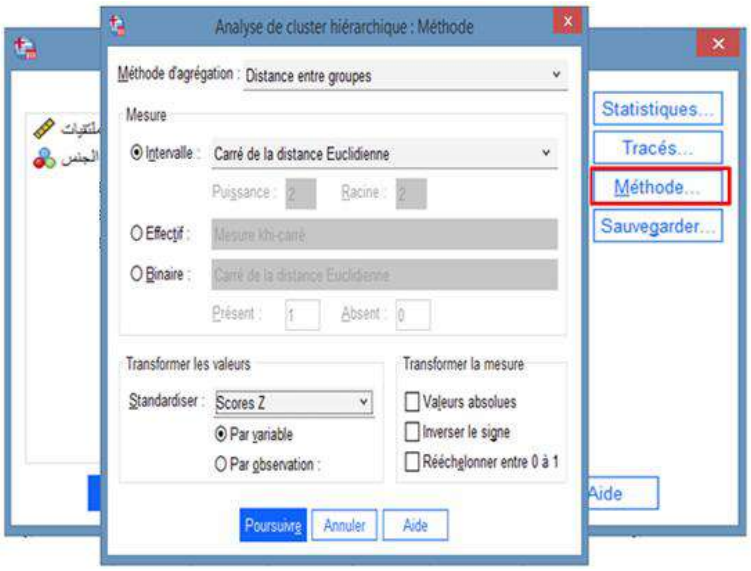


في مربع الحوار نضغط على Tracés ونختار وقم بالإشارة إلى أننا نريد مخطط الشجرة الهيكلية Dendrogram ومخطط Stalactites عمودي مع حلول عنقودية هي 2 و 3 و 4، ثم نضغط على Poursuivre.

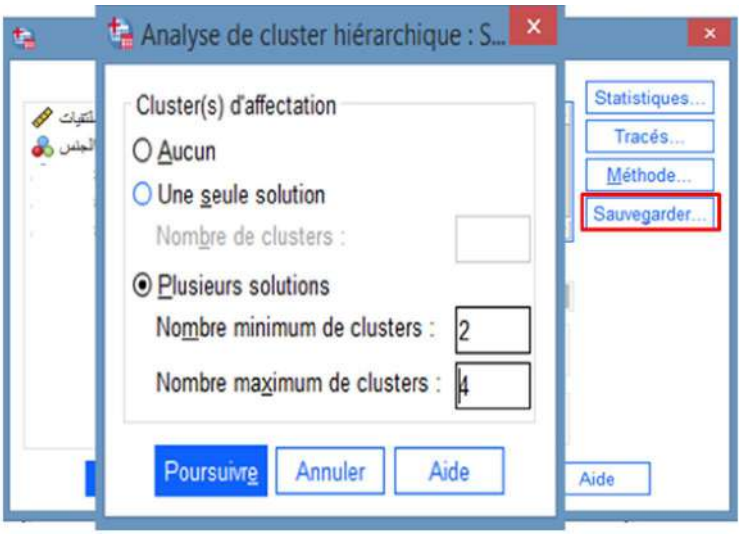


ثم نضغط على Méthode للإشارة الى اننا نريد استخدام طريقة الربط بين المجموعات للتصنيف (Distance entre groupes) ، ونختار المسافات الإقليدية التربيعية (Carré de la distance Euclidienne)،

والمتغيرات الموحدة إلى درجات z بحيث يساهم كل متغير بالتساوي (Z Scores)، ثم نضغط على Poursuivre.



ثم نضغط على sauvegarder حتى نتكمن من تسجيل المجموعة التي تم تصمين الفرد فيها من أجل حلول المجموعة 2 و 3 و 4 لكل حالة، ثم نضغط على Poursuivre.



بعد ذلك نضغط على OK حتى يبدا برنامج SPSS في اجراء التحليل، حيث يبدأ SPSS بتوحيد جميع المتغيرات المتوسط 0 ، التباين 1. وهذا يجعل جميع المتغيرات على نفس المقياس ويتم ترجيحها بالتساوي.

في الخطوة الأولى ،يقوم SPSS بحساب لكل زوج من الأفراد المسافة الإقليدية المربعة كما يلي:

$$\sum_{i=1}^{p}(x_i - y_i)^2$$

2.3.5. تفسير مخرجات SPSS

يتم تحديد الحالتين اللتين تفصلهما أصغر مسافة إقليدية ثم يتم تصنيفهما معا في المجموعة الأولى. ويقوم برنامج SPSS بإعادة حساب المسافات الإقليدية المربعة بين كل مشاهدة أو مجموعة، ومشاهدة اخرى او مجموعة اخرى (كما راينا سابقا من خلال المثال) حتى يتم تجميع جميع الحالات في مجموعة كبيرة واحدة.

في الجدول الموالي، في الخطوة الأولى ، المشاهدة (الفرد) 32 تم تصنيفها مع المشاهدة (الفرد) 33 ، والمسافة الإقليدية المربعة بين هاتين الحالتين هي 0.000. في المراحل من 2 إلى 4 ، انشأ SPSS ثلاث مجموعات أخرى، تحتوي كل منها على مشاهدتين. في المرحلة الخامسة ، يضيف SPSS الفرد (المشاهدة) 39 إلى المجموعة التي تحتوي على الحالتين 37 و 38. بحلول المرحلة 43 تم تجميع أو تصنيف جميع المشاهدات في مجموعة واحدة

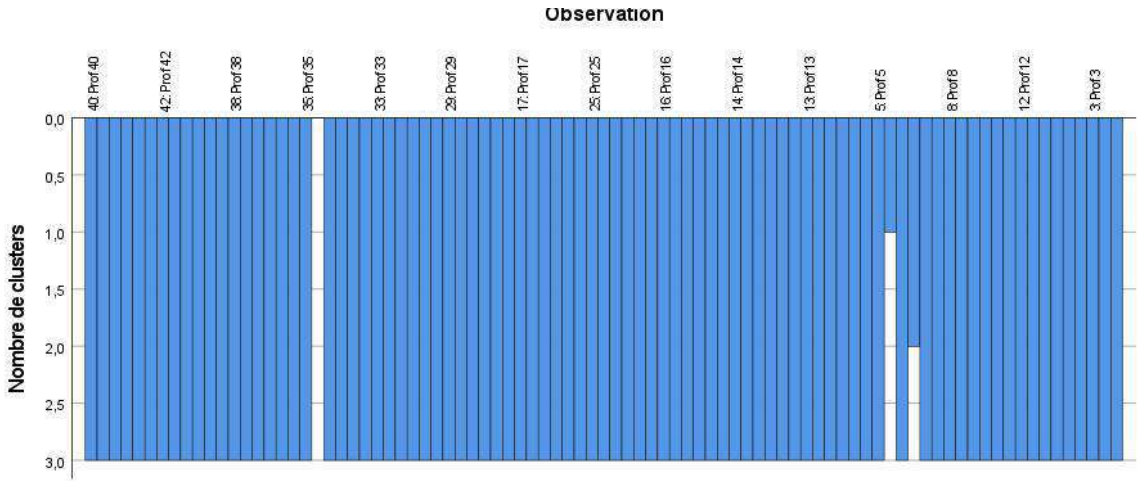
الجدول رقم 5-04: تصنيف المجموعات في SPSS

Etape	Cluster combiné		Coefficients	Etape de première apparition du cluster		Etape suivante
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	32	33	,000	0	0	9
2	41	42	,000	0	0	6
3	43	44	,000	0	0	6
4	37	38	,000	0	0	5
5	37	39	,001	4	0	7
6	41	43	,002	2	3	27
7	36	37	,003	0	5	27
8	20	22	,007	0	0	11
9	30	32	,012	0	1	13
10	21	26	,012	0	0	14
11	20	25	,031	8	0	12
12	16	20	,055	0	11	14
13	29	30	,065	0	9	26
14	16	21	,085	12	10	20
15	11	18	,093	0	0	22
16	8	9	,143	0	0	25
17	17	24	,144	0	0	20

18	13	23	,167	0	0	22
19	14	15	,232	0	0	32
20	16	17	,239	14	17	23
21	7	12	,279	0	0	28
22	11	13	,441	15	18	29
23	16	27	,451	20	0	26
24	3	10	,572	0	0	28
25	6	8	,702	0	16	36
26	16	29	,768	23	13	35
27	36	41	,858	7	6	33
28	3	7	,904	24	21	31
29	11	28	,993	22	0	30
30	5	11	1,414	0	29	34
31	3	4	1,725	28	0	36
32	14	31	1,928	19	0	34
33	36	40	2,168	27	0	40
34	5	14	2,621	30	32	35
35	5	16	2,886	34	26	37
36	3	6	3,089	31	25	38
37	5	19	4,350	35	0	39
38	1	3	4,763	0	36	41
39	5	34	5,593	37	0	42
40	35	36	8,389	0	33	43
41	1	2	8,961	38	0	42
42	1	5	11,055	41	39	43
43	1	35	17,237	42	40	0

من خلال الشكل الموالي وبالنسبة للحل العنقودي يمكننا أن نلاحظ ان المجموعة الواحدة تتكون من عشر حالات، (من الفرد 44 الى 36) متبوعا بعمود أبيض.

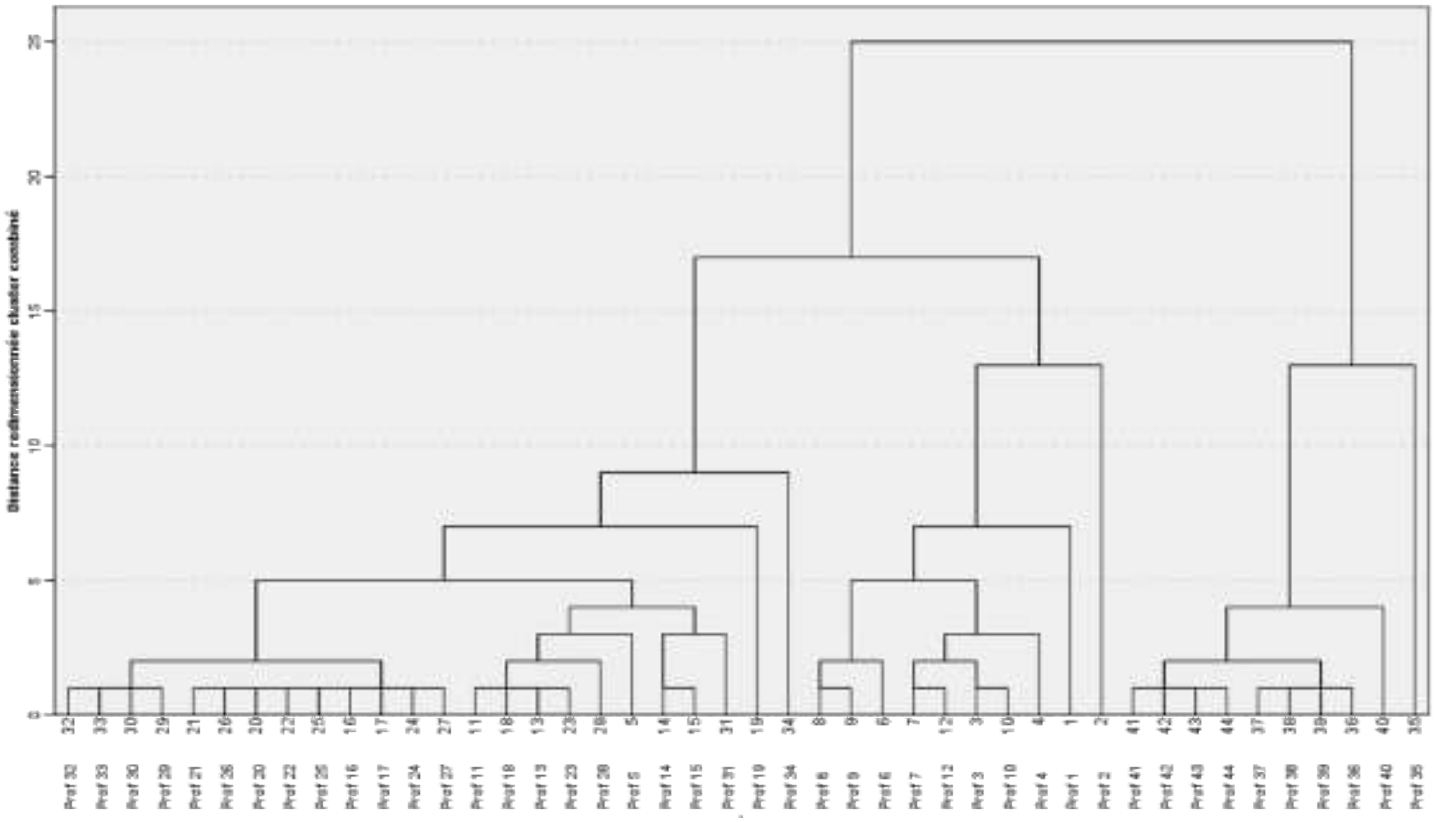
ويمثل هؤلاء الاساتذة أعضاء هيئة التدريس بدوام جزئي (باستثناء واحد) وتتألف المجموعة الثانية من أي استاذ آخر.



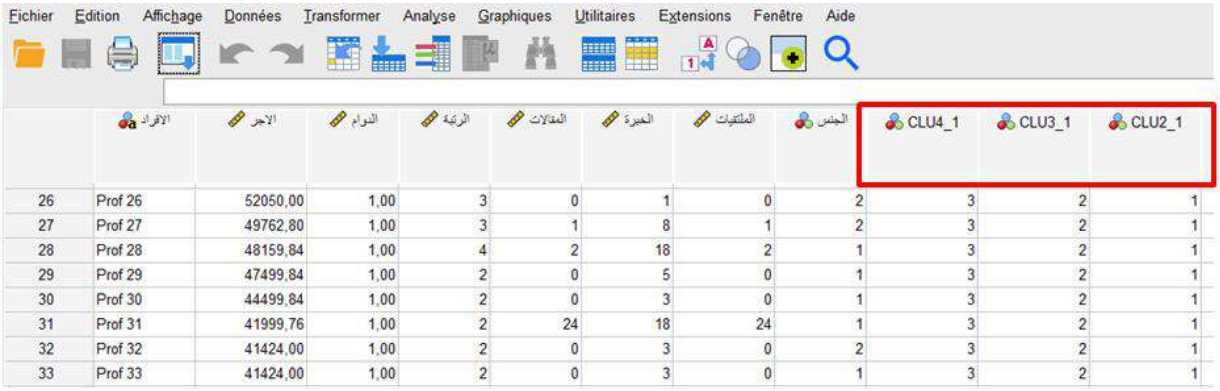
بالنسبة إلى المجموعات الثلاثة ، يمكننا ملاحظة أن مجموعة الاساتذة المساعدين ظلت كما هي ولكن المجموعة الأخرى تنقسم إلى قسمين. من الاستاذ رقم Prof 35 الى الاستاذ رقم Prof 5 يمثلون الاساتذة المحاضرين والمساعدين أ. أما الاساتذة المتبقين فهم اساتذة التعليم العالي. بالنسبة إلى المجموعات الأربعة نلاحظ ان أن حالة واحدة (Prof 35) تشكل مجموعة خاصة بها.

ومن خلال مخطط الشجرة الهيكلية التالي يمكننا استخلاص نفس النتائج الموجودة في المذكورة سابقا.

الشكل رقم 5-03: مخطط الشجرة الهيكلية



اذا قمنا بإلقاء نظرة على قاعدة البيانات في SPSS سنجد ثلاثة متغيرات جديدة: CLU2_1 عبارة عن عنقود يشمل المجموعات المكونة من مجموعتين، CLU3_1 لحل الثلاث مجموعات ناتجة عن التحليلي، و CLU4_1 كل المجموعات الأربعة.



	الافراد	الاجر	الدوام	الرتبة	المقالات	الخبرة	الملتقيات	الجنس	CLU4_1	CLU3_1	CLU2_1
26	Prof 26	52050,00	1,00	3	0	1	0	2	3	2	1
27	Prof 27	49762,80	1,00	3	1	8	1	2	3	2	1
28	Prof 28	48159,84	1,00	4	2	18	2	1	3	2	1
29	Prof 29	47499,84	1,00	2	0	5	0	1	3	2	1
30	Prof 30	44499,84	1,00	2	0	3	0	1	3	2	1
31	Prof 31	41999,76	1,00	2	24	18	24	1	3	2	1
32	Prof 32	41424,00	1,00	2	0	3	0	2	3	2	1
33	Prof 33	41424,00	1,00	2	0	3	0	1	3	2	1

تجدر الاشارة الى ان هناك العديد من الأمور التي يجب أن يكون الباحث على دراية بها عند إجراء التحليل التصنيفي:

- عادة ما تعطي الطرق المختلفة للتصنيف نتائج مختلفة جدا، يحدث هذا بسبب المعيار المختلف لدمج المجموعات او العناقيد.
- باستثناء الربط المنفرد ستتأثر النتائج بطريقة ترتيب المشاهدات والمتغيرات.
- التحليل التصنيفي هو طريقة استكشافية وجودة حل المجموعة الخاصة بكل حالة تكون ذاتية بطبيعتها.

كما انه من المهم اتخاذ قرار بشأن المجموعات المنطقية والقابلة للتنفيذ. من الطرق الجيدة لاختبار هذه المعايير أن تسأل هل هذه العينة ذات حجم كاف لاجراء التحليل، وهل يمكننا وصف ما المعلومات الناتجة عن تحليل بيانات العينة.

الفصـل الـسادس

تمـاريـن محـــلولـة وتمـاريـن مقـترحـة

مقدمة الفصل

في هذا الفصل سنقد مجموعة من التمارين مع اقتراح حلول لبعض منها مع تطبيق برنامج SPSS.كما ستضمن هذا الفصل ايضا مجموعة من التمارين المقترحة والتي يمكن للطلبة حلها وذلك بهدف التدرب على تحليل المعطيات باستخدام الطرق المبينة في هذه المطبوعة وتفسير مختلف النتائج المتحصل عليها.

1.6. تمارين محلولة

فيما يلي عرض لمجموعة التمارين الخاصة بالفصول الخمسة لهذه المطبوعة مع حلول مقترحة لها. تجدر الاشارة الى اننا اعتمدنا في بعض التمارين على برنامج SPSS.

التمرين 01

تهتم السلطات الصحية في بلدية خميس مليانة بتقييم جودة مياه الشروب من حيث تحليل وجود البكتيريا في طبقة المياه الجوفية بالقرب من المدينة. وقد قامت هذه السلطات بالنظر في منطقتين مختلفتين من الخزان الجوفي وحصلت على النتائج التالية (عدد البكتيريا لكل 1000 مم):

المنطقة1	194	199	191	202	215	214	197	204	199	202	230	193	194	209
المنطقة2	158	161	143	174	220	156	156	156	198	161	188	139	147	116

المطلوب دراسة مقارنة لنوعية المياه في المنطقتين باستخدام المقاييس الإحصائية والتمثيلات البيانية.

الحل

للقيام بذلك سوف نعتمد على برنامج SPSS مع ادراج نتائج التحليل فقط لان خطوات التحليلي في البرنامج تم التطرق اليها في الفصل الاول والملحق رقم 4. ومن خلال برنامج SPSS سنقوم بحساب المتوسط، التباين، الوسيط، الحد الادنى والحد الاقصى لكلا المنطقتين. ثم سنقوم بتمثيل المعطيات بيانيا.

Statistiques

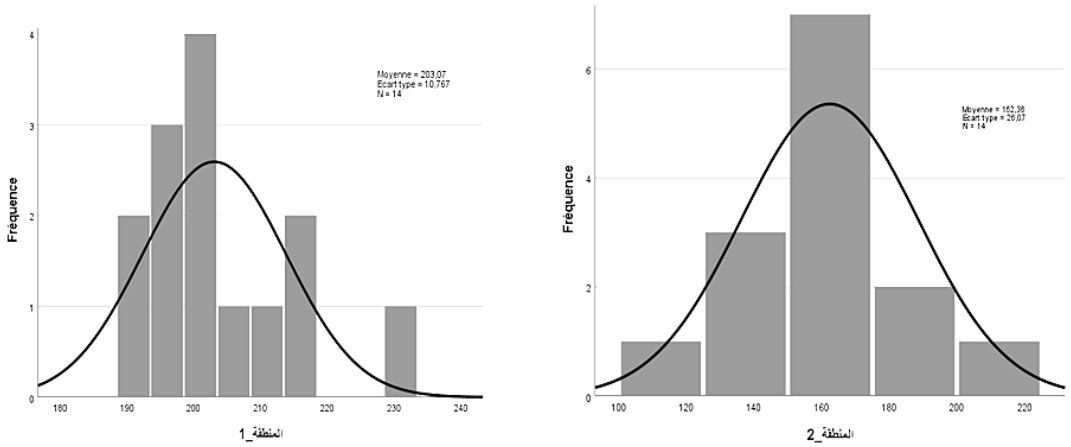
		المنطقة 2	المنطقة 1
N	Valide	14	14
	Manquant	0	0
Moyenne		203,07	162,36
Médiane		200,50	157,00
Mode		194[a]	156
Ecart type		10,767	26,070
Variance		115,918	679,632
Asymétrie		1,302	,664
Erreur standard d'asymétrie		,597	,597
Kurtosis		1,702	,987
Erreur standard d'aplatissement		1,154	1,154
Minimum		191	116
Maximum		230	220
Percentiles	25	194,00	146,00
	50	200,50	157,00
	75	210,25	177,50

a. Présence de plusieurs modes. La plus petite valeur est affichée.

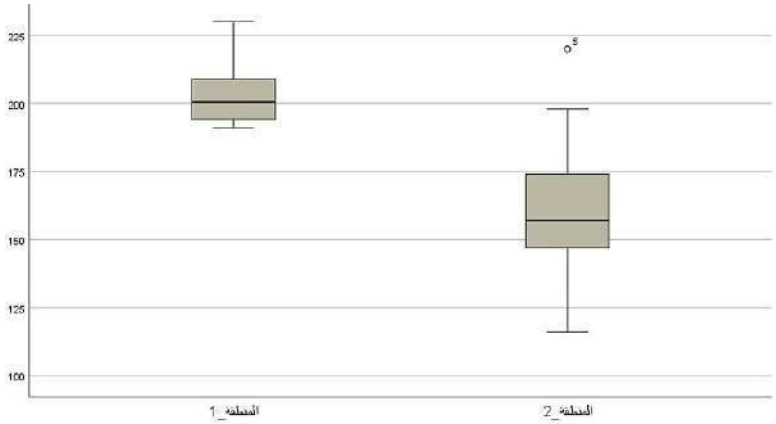
نلاحظ أن قيم المتوسط والوسيط لعدد البكتيريا في المنطقة 2 أقل مقارنة بالمنطقة 1، وهذا ما يدفعنا للقل ان جودة المياه في المنطقة 2 أعلى بشكل واضح. فيما يتعلق بقيم التباين التي تم الحصول عليها يمكننا أن نلاحظ ان التباين أكبر في المنطقة 2.

أما بيانيا فيمكننا ان نلاحظ ان الاختلافات بين قيم المتغيرات واضحة، حيث في حالة عدد البكتيريا في المنطقة 1، فإن الجزء الأكبر من الملاحظات يقع بين القيم 190 و 210 ، وبالنسبة للمنطقة 2 فاذ ذلك يقع بين القيم

140 و 180. مع النتائج التي تم الحصول عليها سابقا، تكون قيم المتوسط الحسابي والوسيط والمنوال لعدد البكتيريا في المنطقة 2 أقل من مقاييس المنطقة 1.



اذا قمنا بتمثيل البيانات باستخدام مخطط الصندوق (حيث البيانات في الرسم البياني هي ملخصات لمتغيرات مختلفة)، فاننا نحصل على الشكل الموالي:



يمثل العمودي قيم المتغيرات قيد الدراسة (عدد البكتيريا). ويسلط هذا الرسم البياني الضوء مرة أخرى على الاختلافات بين قيم كلا المنطقتين، مما يوفر المزيد من الأدلة لتأكيد وجود جودة مياه أعلى في المنطقة 2.

بالنسبة للوسيط يمكن ملاحظة أن القيمة المتوسطة أقل بالنسبة لعدد المستعمرات في المنطقة 2. بالنسبة للربيع الثاني فان ما يمثل 50% من القيم التي يحددها الصندوق لعدد البكتيريا في المنطقة 1 أعلى من قيم عدد البكتيريا في المنطقة 2. بالنسبة للقيمة خارج الحدود: عدد البكتيريا في المنطقة 2 له قيمة خارج الحدود، وهو ما يتوافق مع الحالة 5.

التمرين 02

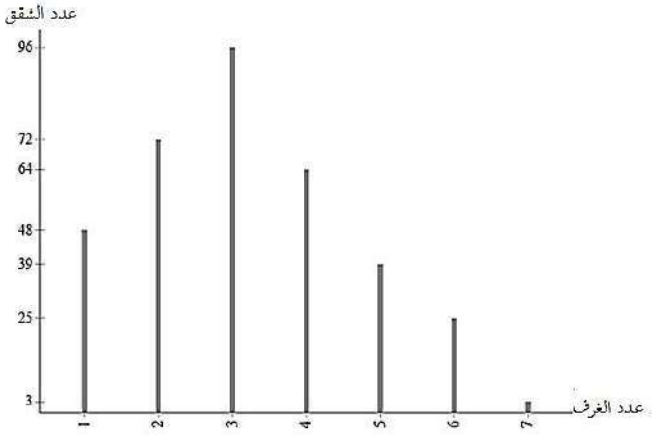
في بلدة صغيرة ،كان هناك عدد من الغرف لكل شقة، باعتبار عدد الغرف لكل شقة متغير عشوائي منفصل بقيم عددية، ومن خلال بيانات الجدول الموالي:

عدد الغرف	1	2	3	4	5	6	7
عدد الشقق	48	72	96	64	39	25	3

مثل بيانيا معطيات الجدول، والتكرارات المتجمعة الصاعدة، ثم احسب كل من المنوال، المتوسط الحسابي، المدى، التباين، والانحراف المعياري.

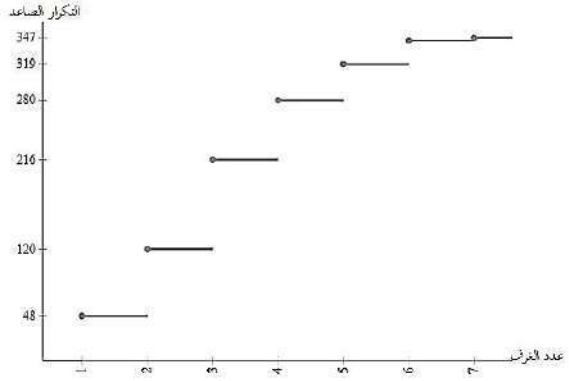
الحل

يمكن تمثيل بيانات الجدول باستخدام الاعمدة البيانية كما يلي:



نحسب التكرارات الصاعدة ونمثلها بيانيا

عدد الغرف	1	2	3	4	5	6	7
عدد الشقق	48	72	96	64	39	25	3
التجميعي الصاعد	48	120	216	280	319	344	347



حساب مقاييس النزعة المركزية ومقاييس التشتت:

المنوال: 3 ؛ المتوسط الحسابي: 3.17579250

المدى: 6 ؛ التباين: 2.15065319 ؛ الانحراف المعياري: 1.46651054

التمرين 03

لتكن المعطيات المبينة في الجدول التالي:

المتغيرات / الأفراد	x_1	x_2
1	4	5
2	6	7
3	8	0

احسب: $\bar{x}_1$ و $\bar{x}_2$ ؟ احسب: σ_{x_1} و σ_{x_2} ؟

لتكن المصفوفة Z المعرفة كما يلي:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_i}}$$

ماذا تمثل هذه المصفوفة؟ احسب المصفوفة Z؟ تحقق ان: $\bar{Z}_1 = \bar{Z}_2 = 0$ و $\sigma_{Z_1} = \sigma_{Z_2} = 1$

لتكن المصفوفة V حيث : $V = \frac{1}{n} Z'Z$ ، احسب المصفوفة V ؟ ماذا تمثل هذه المصفوفة؟

الحل

$$x_1 = \begin{pmatrix} 4 \\ 6 \\ 8 \end{pmatrix} \quad ; \quad x_2 = \begin{pmatrix} 5 \\ 7 \\ 0 \end{pmatrix}$$

$$\bar{x}_1 = \frac{8+6+4}{3} = 6 \quad ; \quad \bar{x}_2 = \frac{5+7+0}{3} = 4$$

الان نحسب القيم $(x_{ij} - \bar{x}_j)$

N	x_1	x_2	$x_1 - \bar{x}_1$	$x_2 - \bar{x}_2$
1	4	5	4-6=-2	5-4=1
2	6	7	6-6=0	7-4=3
3	8	0	8-6=2	0-4=-4

حساب الانحرافات المعيارية:

$$\sigma_{x_1} = \sqrt{\frac{1}{3}\left[(-2^2) + (2^2)\right]} \rightarrow \sigma_{x_1} = 2\sqrt{\frac{2}{3}}$$

$$\sigma_{x_2} = \sqrt{\frac{1}{3}\left[(1^2) + (-4^2)\right]} \rightarrow \sigma_{x_2} = \sqrt{\frac{26}{3}}$$

المصفوفة Z_{ij} تمثل مصفوفة معاملات الارتباط.

ويتم حسابها كما يلي:

$$\Rightarrow \bar{Z}_1 = \bar{C}_1^* = 0 \quad \text{و} \quad \bar{Z}_2 = \bar{C}_2^* = 0$$

$$\sigma_{yj} = 1; \quad j = 1,2$$

بالإضافة الى ان معامل الارتباط بين متغيرين يساوي:

$$r = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$

وبتعويض القيم نجد:

$$Z = \begin{pmatrix} \frac{-2\sqrt{3}}{2\sqrt{2}} & \sqrt{\frac{3}{26}} \\ 0 & \frac{3\sqrt{3}}{\sqrt{26}} \\ \sqrt{\frac{3}{2}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{pmatrix}$$

حساب $V = \frac{1}{n} Z'Z$ كما يلي:

$$\frac{1}{3}\begin{pmatrix} -\sqrt{\frac{3}{2}} & 0 & \sqrt{\frac{3}{2}} \\ \sqrt{\frac{3}{26}} & \frac{3\sqrt{3}}{\sqrt{26}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{pmatrix}\begin{pmatrix} -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{26}} \\ 0 & \frac{3\sqrt{3}}{\sqrt{26}} \\ \sqrt{\frac{3}{2}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{pmatrix}=\frac{1}{3}\begin{pmatrix} 3 & \frac{-15}{2\sqrt{13}} \\ \frac{-15}{2\sqrt{13}} & 3 \end{pmatrix}=\begin{pmatrix} 1 & \frac{-5}{2\sqrt{13}} \\ \frac{-5}{2\sqrt{13}} & 1 \end{pmatrix}=\begin{pmatrix} 1 & -0,69 \\ -0,69 & 1 \end{pmatrix}$$

وتعرف هذه المصفوفة بمصفوفة احداثيات الافراد.

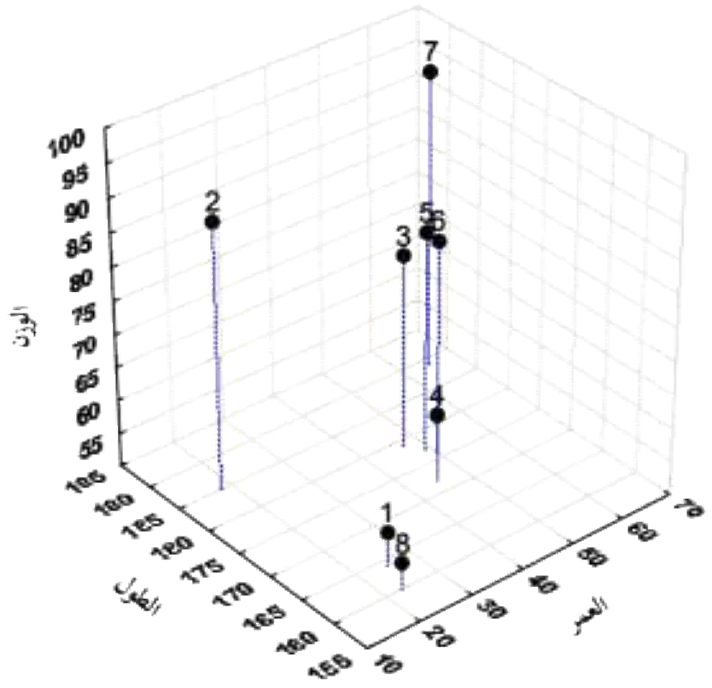
التمرين 04

مثل بيانات الجدول الموالي، والتي تشمل بيانات 8 اشخاص فيما يتعلق بثلاث خصائص هي: X_1 الوزن (كغ)، X_2 الطول (سم)، و X_3 العمر، بيانيا وعلق عليه. قم بتحليل البيانات باستخدام طريقة PCA.

المتغيرات / الافراد	X_1	X_2	X_3
1	55	164	25
2	90	185	18
3	79	179	47
4	60	172	45
5	83	177	49
6	82	176	50
7	95	189	65
8	54	160	23

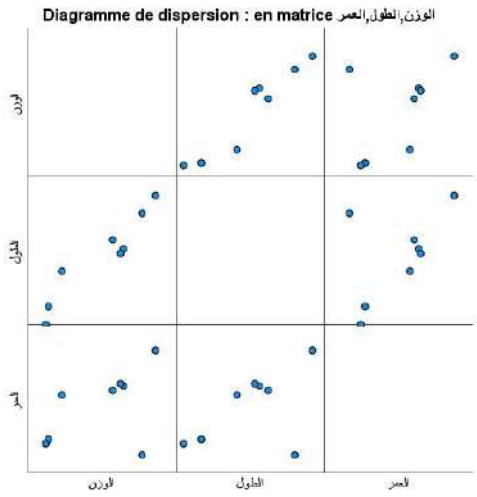
الحل

بالنظر إلى كل صف من مصفوفة البيانات كنقطة إحداثيات في المتغيرات المدروسة (الوزن، الطول، والعمر) يمكن استخدام الفضاء ثلاثي الأبعاد لتمثيل هذه المصفوفة، كما هو مبين في الشكل الموالي:



نلاحظ ان سحابة النقاط تتكون من 8 ملاحظات في الفضاء ثلاثي الأبعاد، وتشير النقاط التي تظهر بالقرب من بعضها البعض إلى الأشخاص لديهم خصائص متشابهة فيما يتعلق بالمتغيرات الثلاثة. فمثلا فإن الأفراد 3، 5، و6 لديهم قياسات متشابهة، ونفس الشيء مع الشخصين 1 و 8. وكلما كان الأشخاص متباعدين كلما اشار ذلك إلى أن اختلاف قياساتهم عن بعضها البعض (في متغير واحد على الأقل) كما هو الحال بالنسبة للشخص 7.

الان يمكننا اجراء تحليل PCA باستخدام برنامج SPSS، وسنحصل على النتائج التالية:



Matrice de corrélation

		الوزن	الطول	العمر
Corrélation	الوزن	1,000		
	الطول	,949	1,000	
	العمر	,499	,526	1,000

ما يمكن ملاحظته هو ان معاملات الارتباط كلها اكبر من 0.4 وهذا مؤشر جيد يدل على ان هناك ارتباط بين المتغيرات وبالتالي يمكننا ادراء تحليل PCA. كما يلخص الجدول الموالي قيم المتوسط الحسابي والانحراف المعياري للمتغيرات الثلاثة.

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
الوزن	74,75	16,122	8
الطول	175,25	9,794	8
العمر	40,25	16,377	8

يمكننا ايضا التعقيب على قيم KMO و اختبار Bartlett والتي تشير قيمها المبينة في الجدول الموالي الى انه يمكننا اجراء تحليل PCA.

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,712
Test de sphéricité de Bartlett	Khi-carré approx.	13,648
	ddl	3
	Signification	,003

حيث يظهر يختبر اختبار Bartlett ان هناك بعض العلاقات بين المتغيرات التي من المتوقع تضمينها في التحليل (p أقل من 0.05). يشير KMO إلى مدى كفاية حجم العينة ، لذلك يعد إجراء التحليل الى مركبات اساسية مناسبا.

وفيما يخص نوعية التمثيل فنلاحظ أن كل تمثيل كل المتغيرات اكبر من 0.4 وبالتالي لا نحذف اي متغير وكل المتغيرات ستدخل في عملية التحليل.

Qualités de représentation

	Initiales	Extraction
الوزن	1,000	,898
الطول	1,000	,914
العمر	1,000	,529

Méthode d'extraction : Analyse en composantes principales.

بعد ذلك يظهر SPSS جدول التباين الكلي الموضح مع القيم الذاتية (eigenvalues) المقابلة لكل عامل.

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	1,710	56,998	56,998	1,710	56,998	56,998	1,710	56,995	56,995
2	1,164	38,815	95,813	1,164	38,815	95,813	1,165	38,817	95,813
3	,126	4,187	100,000						

Méthode d'extraction : Analyse en composantes principales.

من الجدول نلاحظ انه تم اختزال البيانات الى محورين حيث يضم المحور الاول ثلاث متغيرين بحجم معلومات قدره 56.99 % والمحور الثاني يضم متغير واحد بحجم معلومات يقدر بـ 38.81%. ومنه فان المحورين معا اختزلا ما قيمته 95.81% من حجم المعلومات.

Matrice des composantes[a]

	Composante 1	Composante 2
الوزن	,927	-,296
الطول	-,013	,990
العمر	,922	,311

Méthode d'extraction : Analyse en composantes principales.
a. 2 composantes extraites.

Rotation de la matrice des composantes[a]

	Composante 1	Composante 2
الوزن	,923	-,307
الطول	-,001	,990
العمر	,926	,300

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.
a. Convergence de la rotation dans 3 itérations.

تحتوي مصفوفة النمط على الأحمال المتغيرة للعوامل التي تم تدويرها ويمكن مقارنتها بمصفوفة العوامل التي لوحظت في التدويرات المتعامدة. وتأخذ مصفوفة في عين الاعتبار العلاقات بين العوامل وهي في الواقع نتاج مصفوفة الأنماط مع مصفوفة معاملات الارتباط بين العوامل.

ويوضح الجدول التالي إلى أي مدى يمكن إعادة إنتاج مصفوفة الارتباط الأصلية من عاملين:

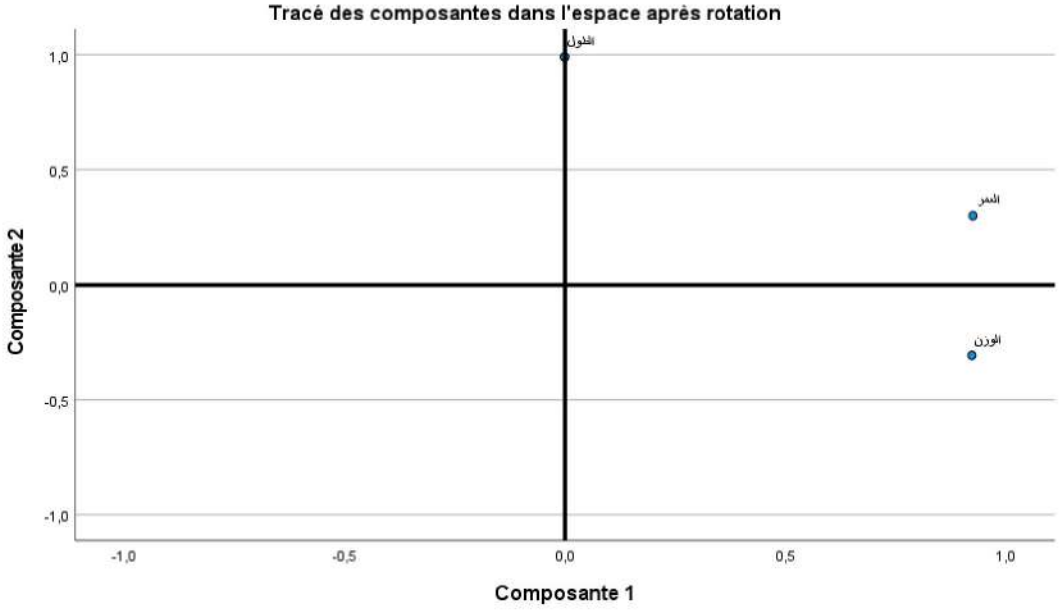
Corrélations reconstituées

		الوزن	الطول	العمر
Corrélation reconstituée	الوزن	,947[a]	-,305	,763
	الطول	-,305	,980[a]	,296
	العمر	,763	,296	,947[a]
Résidu[b]	الوزن		,032	-,053
	الطول	,032		-,032
	العمر	-,053	-,032	

Méthode d'extraction : Analyse en composantes principales.
a. Qualités de représentation reconstituées
b. Les résidus sont calculés d'après les corrélations observées et reconstituées. Il existe 1 (33,0%) résidus non redondants avec des valeurs absolues supérieures à 0.05.

وتسمح مراقبة مصفوفة المكونات بتصور كل متغير للمكونات (العوامل) المستخرجة قبل الدوران. بمعنى آخر هي معاملات الارتباط بين المتغيرات والمكونات غير المستديرة (العوامل). وينتج عن هذا التحليل تمثيل بياني يظهر المتغيرات الثلاثة، ويتم الاستعانة به لتسمية المحاور.

في الرسم البياني الموالي تم وضع متغيرات الدراسة التي تم تحليلها وفقا لتضمين عاملها في المكونين المستخلصين بعد دوران varimax.



التمرين 05

لتكن بيانات الجدول الموالي والتي تبين 13 مهمة منزلية وكيفية توزيعها على الزوجين:

المهام	الزوجة	الزوج	بالتناوب	بشكل مشترك
غسل ملابس	156	2	14	4
اعداد وجبة الغداء	124	5	20	4
وجبة العشاء	77	7	11	13
فطور الصباح	82	15	36	7
الترتيب	53	1	11	57
غسل الاطباق	32	4	24	53
التسوق	33	9	23	55
اصطحاب الاطفال الى المدرسة	12	23	46	15
القيادة	10	75	51	3
دفع الفواتير	13	21	13	66
التامين	8	53	1	77
الاصلاحات	0	160	3	2
التخطيط للعطل	0	6	1	153

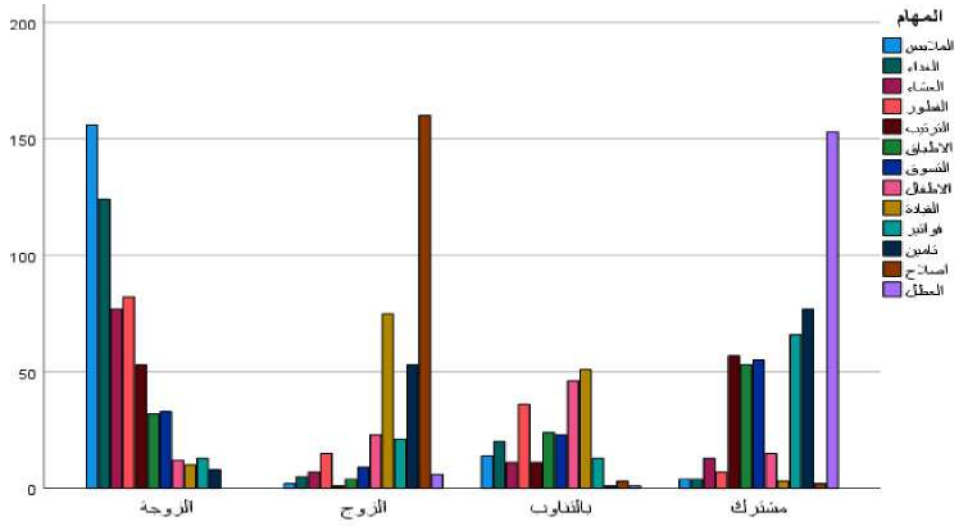
علق على بيانات الجدول؟ ما هو نوع التحليل العاملي المناسب لهذه البيانات؟ قم بتحليل البيانات؟

الحل

جدول أعلاه لا يحتوي بيانات كبيرة جدا، لذلك من السهل فحص وتفسير ملفات تعريف الصفوف والأعمدة، حيث يمكننا ملاحظة ما يلي:

- من الواضح أن الأعمال المنزلية: الغسيل، تحضير الوجبات،... الخ تقوم بها الزوجة بشكل متكرر.
- كثيرا ما ترتبط مهام الإصلاح والقيادة بـ الزوج
- كثيرا ما ترتبط التحضير للعطل بقرارات مشتركة بين الزوجين

وهذا ما يظهر من الشكل البياني التالي:



من الواضح ايضا ان بيانات الجدول عبارة عن تكرارات لفئات من المتغيرات ذان الطابع النوعي وبالتالي فان نوع التحليل العاملي المناسب هو تحليل المعاملات للتوفيقات CA.

و بما ان المتغيرين نوعيين فإنا سنستعمل اختبار التوافق، وجاءت النتائج كما يلي:

Mesures symétriques

		Valeur	Signification approximative
Nominal par Nominal	Coefficient de contingence	,726	,000
N d'observations valides		1744	

بما ان مستوى الدلالة أقل من0.05 اذا نرفض الفرضية الصفرية ونقول ان يوجد ارتباط بين المهمة المنجزة وكيفية انجازها. كما ان هذه الطريقة كانت هذه العلاقة قوية بما ان القيمة بلغت 0.726 (انظر الملحق رقم 2).

ومن خلال الجدول الموالي يمكننا ملاحظة ان مثلا 156 زوجة صرحت انها تقوم بالأعمال المنزلية المتعلقة بغسل الملابس وهو ما يمثل نسبة 26% من عدد الزوجات. كما ان 88.6 % من غسل الملابس تنجزها الزوجة. أما النسبة 8.9% فتمثل النسبة في مجموع العينة الكلية اللاتي تنجون المهمة الاولى.

Tableau croisé الفئات * المهام

			المهام													
			الملابس	الغداء	العشاء	الفطور	الترتيب	الاطباق	التسوق	الاطفال	القيادة	فواتير	تامين	اصلاح	العطل	Total
الفئات	الزوجة	Effectif	156	124	77	82	53	32	33	12	10	13	8	0	0	600
		% dans الفئات	26,0%	20,7%	12,8%	13,7%	8,8%	5,3%	5,5%	2,0%	1,7%	2,2%	1,3%	0,0%	0,0%	100,0%
		% dans المهام	88,6%	81,0%	71,3%	58,6%	43,4%	28,3%	27,5%	12,5%	7,2%	11,5%	5,8%	0,0%	0,0%	34,4%
		% du total	8,9%	7,1%	4,4%	4,7%	3,0%	1,8%	1,9%	0,7%	0,6%	0,7%	0,5%	0,0%	0,0%	34,4%
	الزوج	Effectif	2	5	7	15	1	4	9	23	75	21	53	160	6	381
		% dans الفئات	0,5%	1,3%	1,8%	3,9%	0,3%	1,0%	2,4%	6,0%	19,7%	5,5%	13,9%	42,0%	1,6%	100,0%
		% dans المهام	1,1%	3,3%	6,5%	10,7%	0,8%	3,5%	7,5%	24,0%	54,0%	18,6%	38,1%	97,0%	3,8%	21,8%
		% du total	0,1%	0,3%	0,4%	0,9%	0,1%	0,2%	0,5%	1,3%	4,3%	1,2%	3,0%	9,2%	0,3%	21,8%
	بالتناوب	Effectif	14	20	11	36	11	24	23	46	51	13	1	3	1	254
		% dans الفئات	5,5%	7,9%	4,3%	14,2%	4,3%	9,4%	9,1%	18,1%	20,1%	5,1%	0,4%	1,2%	0,4%	100,0%
		% dans المهام	8,0%	13,1%	10,2%	25,7%	9,0%	21,2%	19,2%	47,9%	36,7%	11,5%	0,7%	1,8%	0,6%	14,6%
		% du total	0,8%	1,1%	0,6%	2,1%	0,6%	1,4%	1,3%	2,6%	2,9%	0,7%	0,1%	0,2%	0,1%	14,6%
	مشترك	Effectif	4	4	13	7	57	53	55	15	3	66	77	2	153	509
		% dans الفئات	0,8%	0,8%	2,6%	1,4%	11,2%	10,4%	10,8%	2,9%	0,6%	13,0%	15,1%	0,4%	30,1%	100,0%
		% dans المهام	2,3%	2,6%	12,0%	5,0%	46,7%	46,9%	45,8%	15,6%	2,2%	58,4%	55,4%	1,2%	95,6%	29,2%
		% du total	0,2%	0,2%	0,7%	0,4%	3,3%	3,0%	3,2%	0,9%	0,2%	3,8%	4,4%	0,1%	8,8%	29,2%
Total		Effectif	176	153	108	140	122	113	120	96	139	113	139	165	160	1744
		% dans الفئات	10,1%	8,8%	6,2%	8,0%	7,0%	6,5%	6,9%	5,5%	8,0%	6,5%	8,0%	9,5%	9,2%	100,0%
		% dans المهام	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% du total	10,1%	8,8%	6,2%	8,0%	7,0%	6,5%	6,9%	5,5%	8,0%	6,5%	8,0%	9,5%	9,2%	100,0%

وسنقوم فيما يلي باجراء التحليل بالاعتماد على برنامج SPSS.

Table des correspondances

	المهام													
الفئات	الملابس	الغداء	العشاء	الفطور	الترتيب	الاطباق	التسوق	الاطفال	القيادة	فواتير	تامين	اصلاح	العطل	Marge active
الزوجة	156	124	77	82	53	32	33	12	10	13	8	0	0	600
الزوج	2	5	7	15	1	4	9	23	75	21	53	160	6	381
بالتناوب	14	20	11	36	11	24	23	46	51	13	1	3	1	254
مشترك	4	4	13	7	57	53	55	15	3	66	77	2	153	509
Marge active	176	153	108	140	122	113	120	96	139	113	139	165	160	1744

يمكن أن نلاحظ أن مجموع التكرارات للفئة الاولى (الزوجة) والتي تقوم بكل انواع المهام 13، هو 600، كم ان العدد الكلي لأداء المهمة الاولى (غسل الملابس(يصل الى 176. وبالمثل فإن جميع مشاركات الزوج في اداء هذه المهام هو 381.

وفيما يلي جداول ملفات تعريف الصف والعمود:

Profils lignes

	الملابس	الغداء	العشاء	الفطور	الترتيب	الاطباق	التسوق	الاطفال	القيادة	فواتير	تامين	اصلاح	العطل	Marge active
الزوجة	,260	,207	,128	,137	,088	,053	,055	,020	,017	,022	,013	,000	,000	1,000
الزوج	,005	,013	,018	,039	,003	,010	,024	,060	,197	,055	,139	,420	,016	1,000
بالتناوب	,055	,079	,043	,142	,043	,094	,091	,181	,201	,051	,004	,012	,004	1,000
مشترك	,008	,008	,026	,014	,112	,104	,108	,029	,006	,130	,151	,004	,301	1,000
Masse	,101	,088	,062	,080	,070	,065	,069	,055	,080	,065	,080	,095	,092	

Profils colonnes

	الملابس	الغداء	العشاء	الفطور	الترتيب	الاطباق	التسوق	الاطفال	القيادة	فواتير	تامين	اصلاح	العطل	Masse
الزوجة	,886	,810	,713	,586	,434	,283	,275	,125	,072	,115	,058	,000	,000	,344
الزوج	,011	,033	,065	,107	,008	,035	,075	,240	,540	,186	,381	,970	,038	,218
بالتناوب	,080	,131	,102	,257	,090	,212	,192	,479	,367	,115	,007	,018	,006	,146
مشترك	,023	,026	,120	,050	,467	,469	,458	,156	,022	,584	,554	,012	,956	,292
Marge active	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	

نلاحظ من الجدول الموالي والذي يلخص مجموعة من القياسات والتي تسمح بمعرفة ما إذا كان النموذج مهما أم لا. في هذا الخصوص يعتبر التحليل مهم عند المستوى 0.05، حيث مستوى المعنوية بلغ 0.000 وبلغت قيمة مربع كاي 1944,456، كما ان هذا الجدول يبين ان تحليل المعاملات للتوفيقات الذي قمنا بانجازه قد نتج عنه ثلاثة عوامل.

Récapitulatif

Dimension	Valeur singulière	Inertie	Khi-deux	Sig.	Proportion d'inertie		Valeur singulière de confiance	
					Représentation	Cumulé	Ecart type	Corrélation 2
1	,737	,543			,487	,487	,013	,245
2	,667	,445			,399	,886	,015	
3	,356	,127			,114	1,000		
Total		1,115	1944,456	,000[a]	1,000	1,000		

a. 36 degrés de liberté

يسمح فحص القيم الذاتية بتحديد عدد المحاور الرئيسية التي يجب مراعاتها. تتوافق هذه القيم مع كمية المعلومات التي يحتفظ بها كل محور، فهي كبيرة بالنسبة للمحور الأول وصغيرة بالنسبة للمحور التالي.

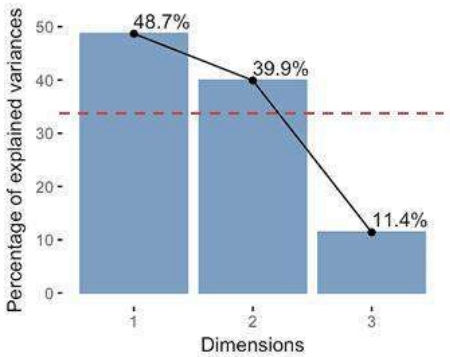
من خلال الجدول نلاحظ ان أول محورين يفسران ما نسبته 88.6% من التباين الكلي، وتعتبر هذه نسبة مقبولة.

هناك طريقة أخرى لتحديد عدد الأبعاد وهي النظر إلى الرسم البياني للقيم الذاتية، مرتبة من الأكبر إلى الأصغر قيمة. يتم تحديد عدد المحاور عند النقطة التي بعدها تكون جميع القيم الذاتية المتبقية صغيرة نسبيا وذات حجم مماثل.

تحتوي بياناتنا على 13 صف و 4 أعمدة. إذا كانت البيانات عشوائية، فإن القيمة المتوقعة للقيمة الذاتية لكل محور ستكون: $\frac{1}{n-1}$

من حيث الصفوف: $\frac{1}{n-1} = \frac{1}{12} = 8.33\%$

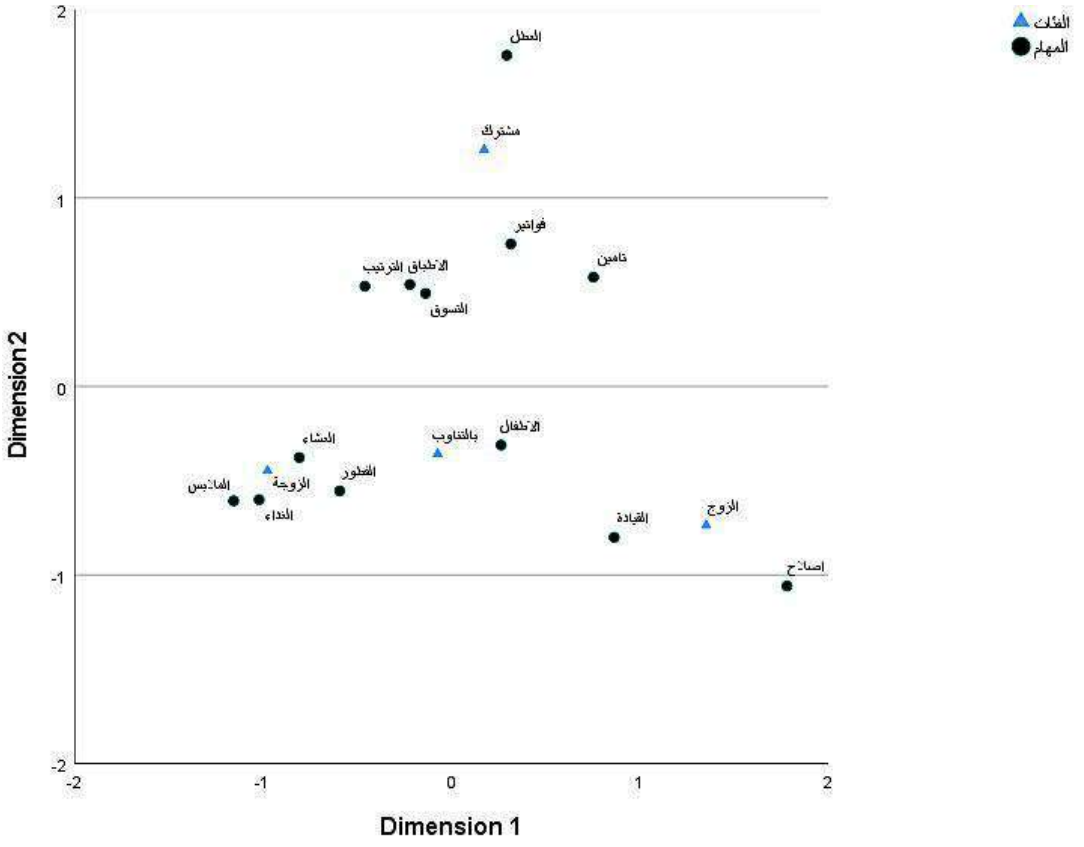
وبالمثل، يجب أن يمثل المحور المتوسط من حيث الاعمدة: $\frac{1}{n-1} = \frac{1}{3} = 33.33\%$



ومنه فانطلاقا من الرسم البياني السابق يجب مراعاة الأبعاد 1 و 2 فقط لتفسير البيانات، لان البعد الثالث يفسر ما قيمته 11.43% فقط من إجمالي القصور الذاتي، وهو أقل من متوسط قيمة المحاور (33.33%) وصغير جدا بحيث لا يمكن الاحتفاظ به.

يمثل البعدان 1 و 2 ما نسبته 48.7% و 39.9% من إجمالي القصور الذاتي على التوالي. هذا يتوافق مع إجمالي تجميعي قدره 88.6 % من إجمالي القصور الذاتي الذي يحتفظ به البعدين.

في التمثيل البياني الموالي يتم تمثيل الصفوف بدوائر سوداء والأعمدة بمثلثات زرقاء. تعطي المسافة بين نقاط الصف أو بين نقاط العمود مقياسا للتشابه (أو الاختلاف). نقاط الخط التي لها ملف تعريف مشابه قريبة من بعضها على الرسم البياني. الشيء نفسه ينطبق على نقاط العمود.



ويظهر من التمثيل البياني ان صفوف اعداد العشاء، وفطور الصباح، والغسيل ترتبط بشكل اكبر مع عمود الزوجة، بينما يرتبط صفا القيادة والإصلاحات بعمود الزوج.

Présentation des points de ligne[a]

		Score de la dimension			Contribution				
					Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point		
الفئات	Masse	1	2	Inertie	1	2	1	2	Total
الزوجة	,344	-,976	-,447	,301	,445	,103	,802	,152	,954
الزوج	,218	1,352	-,737	,381	,542	,178	,772	,208	,980
بالتناوب	,146	-,072	-,357	,118	,001	,028	,005	,105	,110
مشترك	,292	,174	1,257	,315	,012	,691	,021	,977	,998
Total actif	1,000			1,115	1,000	1,000			

Présentation des points de colonne[a]

المهام	Masse	Score de la dimension 1	Score de la dimension 2	Inertie	Contribution Du point vers l'inertie de la dimension 1	Contribution Du point vers l'inertie de la dimension 2	Contribution De la dimension vers l'inertie du point 1	Contribution De la dimension vers l'inertie du point 2	Contribution Total
الملابس	,101	-1,155	-,606	,134	,183	,056	,740	,185	,925
الغداء	,088	-1,020	-,600	,091	,124	,047	,742	,232	,974
العشاء	,062	-,807	-,377	,038	,055	,013	,777	,154	,930
الفطور	,080	-,593	-,554	,041	,038	,037	,505	,400	,905
الترتيب	,070	-,459	,532	,025	,020	,030	,440	,535	,975
الاطباق	,065	-,220	,541	,020	,004	,028	,118	,646	,764
التسوق	,069	-,137	,494	,015	,002	,025	,064	,748	,811
الاطفال	,055	,264	-,311	,053	,005	,008	,053	,066	,119
القيادة	,080	,864	-,800	,102	,081	,076	,432	,335	,767
فواتير	,065	,315	,756	,030	,009	,056	,161	,837	,997
تامين	,080	,754	,580	,058	,061	,040	,576	,309	,885
اصلاح	,095	1,781	-1,058	,313	,407	,159	,707	,226	,933
العطل	,092	,294	1,757	,196	,011	,425	,030	,962	,992
Total actif	1,000			1,115	1,000	1,000			

a. Normalisation symétrique

التمرين 06

أعدت الشركة استبيانا متكون من ثمانية أسئلة حول المشتريات واختارت عينة عشوائية من 20 عميلا للإجابة عليها. يتم قياس الأسئلة باستخدام مقياس ليكرت الخماسي(1: أدنى درجة ، 5: أعلى درجة(، كما يلي:

1. لا أشتري الملابس إلا عندما أحتاجها حقا	1	2	3	4	5
2. اشتري الملابس التي تتماشى مع الموضة	1	2	3	4	5
3. أشتري ملابس فاخرة حتى لو لم أرتديها	1	2	3	4	5
4. عندما أشتري الملابس أختار أيضا قطعا وإكسسوارات أخرى تتناسب معها	1	2	3	4	5
6. عندما يعجبني لا يهمني سعر القطعة	1	2	3	4	5
7. أحاول دائما شراء قطع مميزة	1	2	3	4	5
8. أنا أشتري الملابس فقط إذا كانت من ماركات معروفة	1	2	3	4	5

وجاءت اراد العملاء كما في الجدول الموالي:

العميل	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Q1	2	1	2	1	1	1	3	3	3	3	3	3	5	4	5	4	5	4	1	5
Q2	2	1	2	3	2	2	4	3	2	3	4	3	4	5	4	5	5	5	3	4
Q3	1	2	3	1	2	3	2	3	4	2	3	4	4	5	4	5	4	5	2	5
Q4	1	2	1	2	1	2	2	3	3	2	3	2	4	5	4	5	4	5	2	5
Q5	1	2	2	2	1	1	3	2	3	2	4	3	3	4	5	4	5	3	2	4
Q6	1	1	1	2	1	2	2	3	2	1	3	2	4	5	4	5	4	5	2	4
Q7	2	3	3	2	1	3	3	2	3	2	1	3	4	5	4	4	4	5	2	4
Q8	1	1	1	2	2	2	3	2	2	3	3	2	4	4	5	4	4	5	1	5

بالاعتماد على طريقة التصنيف التسلسلي قم بتحليل بيانات الجدول؟

الحل

باستخدام برنامج SPSS وباتباع الخطوات اللازمة تحصلنا على النتائج التالية:

في الجدول الموالي، في الخطوة الأولى ، المشاهدة (العميل) 14 تم تصنيفها مع المشاهدة (العميل) 16 ، والمسافة الإقليدية المربعة بين هاتين الحالتين هي 0.731، ثم تم النظر في 19 مجموعة والتي يمكن رؤيتها في السطر الاخير من الجدول ثم أقرب حالتين (وفقا لمقياس المسافة المحدد)، وهما العميل 4 والعميل 19 ، وتشكيل 18 مجموعة (السطر 18) ، ثم يتم تجميع العميل 12 والعميل 9 ، وتشكيل 17 مجموعة وما إلى ذلك حتى نحصل على 3 مجموعات.

Planning des agglomérations

	Cluster combiné			Etape de première apparition du cluster		
Etape	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Etape suivante
1	14	16	,731	0	0	5
2	4	19	1,083	0	0	13
3	9	12	1,176	0	0	12
4	15	17	1,185	0	0	11
5	14	18	1,527	1	0	14
6	2	3	2,213	0	0	10
7	13	20	2,236	0	0	11
8	1	5	2,278	0	0	13
9	7	10	2,528	0	0	16
10	2	6	2,740	6	0	15
11	13	15	2,756	7	4	14
12	8	9	3,015	0	3	16
13	1	4	3,439	8	2	15
14	13	14	3,804	11	5	19
15	1	2	4,376	13	10	18
16	7	8	4,591	9	12	17
17	7	11	6,148	16	0	18
18	1	7	8,424	15	17	19
19	1	13	27,219	18	14	0

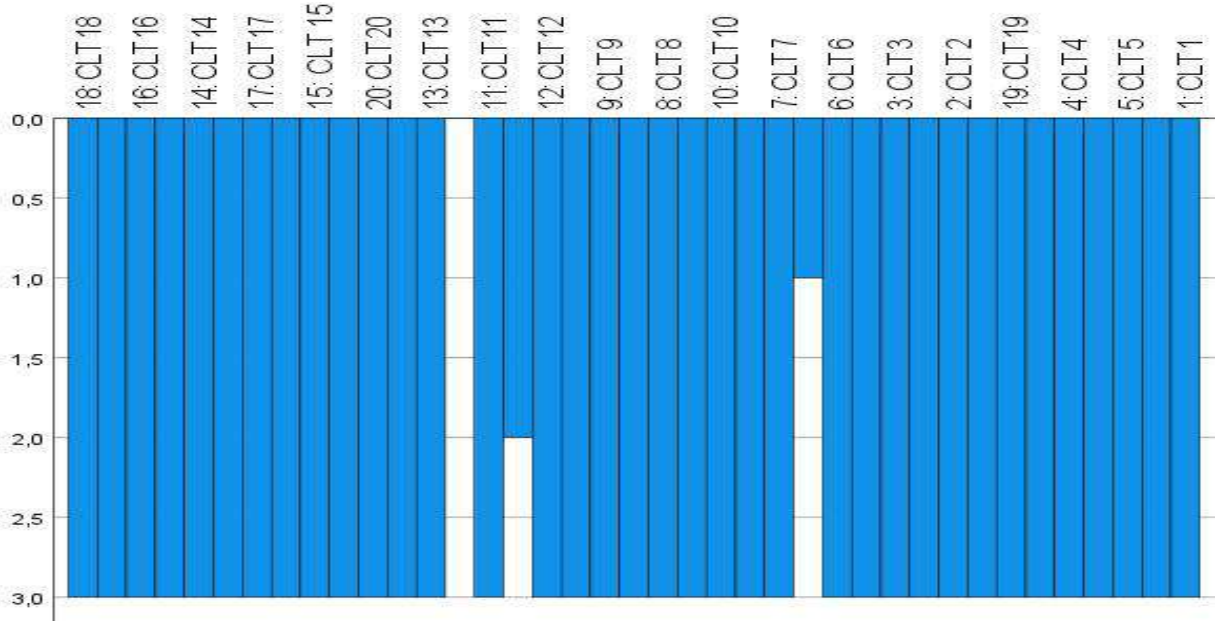
يوضح الجدول السابق كيفية إجراء التصنيفات وفقا للطريقة المختارة، حيث في البداية تم تجميع الحالة 14 و 16 ، بمسافة معتبرة تبلغ 0.731. لاحظ أن كلتا الحالتين 14 و 16 لم يتم تصنيفهما في أي مجموعة من قبل، لذلك يظهر الرقم 0 في عمودي "Cluster 1" و "Cluster 2".

نلاحظ أن العميل رقم 5 يظهر في عمود "Etape suivante" ، لأنه في السطر 5 ، تظهر الحالة 14 مرة أخرى ، عندما يتم تصنيفها مع العميل 18. لاحظ أنه في نفس الصف ، في العمود "Cluster 1" يظهر الرقم 1، لأنه يشير إلى السطر حيث تم بالفعل تجميع المشاهدة 14 مسبقا مع العميل16.

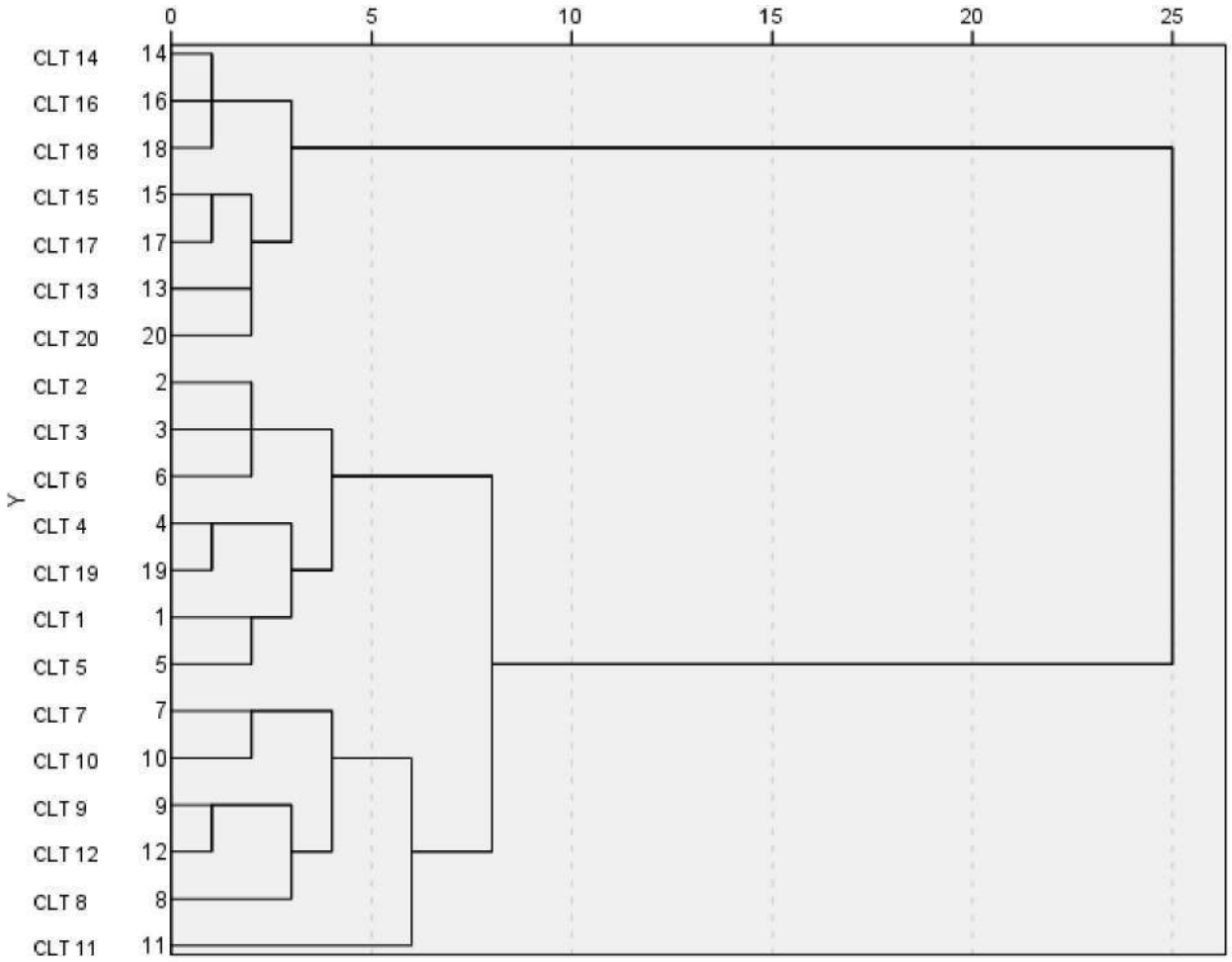
يوضح الجدول الموالي النتيجة النهائية للتصنيف أي العناقيد في كل حالة. على سبيل المثال العميل 1 ثم تصنيفها في المجموعة 1 ؛ العميل 8 في المجموعة 2 ؛ العميل 20 في المجموعة 3، ... الخ.

Observation	3 Clusters 5
1:CLT 1	1
2:CLT 2	1
3:CLT 3	1
4:CLT 4	1
5:CLT 5	1
6:CLT 6	1
7:CLT 7	2
8:CLT 8	2
9:CLT 9	2
10:CLT 10	2
11:CLT 11	2
12:CLT 12	2
13:CLT 13	3
14:CLT 14	3
15:CLT 15	3
16:CLT 16	3
17:CLT 17	3
18:CLT 18	3
19:CLT 19	1
20:CLT 20	3

حيث المجموعة 3 بها المشاهدات (العملاء): 13 ، 14 ، 15 ، 16 ، 17 ، 18 ، 20 ؛ المجموعة 2 لديها: 7 ، 8 ، 9 ، 10 ، 11 ، 12 ؛ المجموعة 1 العملاء: 1 ، 2 ، 3 ، 4 ، 5 ، 6 ، 19. وهذا ما يوضحه الشكل الموالي:

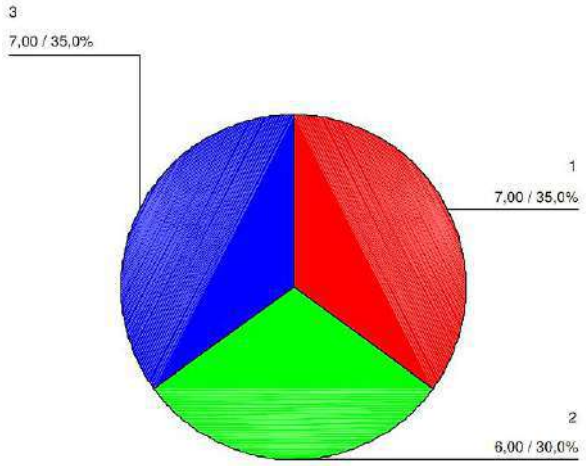


التمثيل البياني الموالي يمثل مخطط الشجرة الهيكلية، ويشير المقياس الرأسي (الموجود على يسار الرسم البياني) إلى مستوى التشابه بين الحالات، ويظهر المقياس الأفقي الحالات خارج الترتيب، بترتيب يسهل تصور التجمعات. يتم إعادة قياس معاملات المسافة إلى قيم بين 0 و 25.



أظهرت لنا مخرجات البرنامج كيف تم تصنيف المجموعات وفي أي مجموعة توجد كل مشاهدة، ولكن ماذا يعني هذا؟

الشكل الدائري الموالي يبين ان 35% من العملاء ينتمون للمجموعة 1؛ 30% للمجموعة 2، و 35% تم تصنيفهم في المجموعة 3 ، كما اننا نلاحظ ان المجموعات الثلاث بنفس الحجم تقريبا. لكن ما هي خصائص كل عنقود؟:



بمقارنة الخصائص الوصفية للأسئلة الثمانية نلاحظ ان إحصائيات المجموعة 1، او العملاء في هذه المجموعة أعطوا درجات منخفضة لجميع الأسئلة ، لذلك يمكننا تصنيفهم على أنهم عملاء يشترون الأجزاء الأساسية وعند الضرورة فقط.

المجموعة 1

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
N	Valide	7	7	7	7	7	7	7	7
	Manquant	0	0	0	0	0	0	0	0
Moyenne		1,2857	2,1429	2,0000	1,5714	1,5714	1,4286	2,2857	1,4286
Ecart type		,4880	,6901	,8165	,5346	,5345	,5345	,7559	,5345
Minimum		1	1	1	1	1	1	1	1
Maximum		2	2	2	2	2	2	2	2

تذكر أنه بالإضافة إلى التحقق من المتوسط، من المهم تحليل مدى تجانس أو عدم تجانس هذا الرأي بالنسبة للمستجوبين. بمعنى آخر يجب أن نتحقق مما إذا كان هناك تباين كبير في البيانات بالنسبة إلى المتوسط. لهذا، يمكننا استخدام معامل الاختلاف (الانحراف المعياري / المتوسط):

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Moyenne	1,2857	2,1429	2,0000	1,5714	1,5714	1,4286	2,2857	1,4286
Ecart type	,4880	,6901	,8165	,5346	,5345	,5345	,7559	,5345
Coef Variation	38%	32%	41%	34%	34%	37%	33%	37%

بشكل عام أن هذه المجموعة متجانسة مع تباين متوسط في البيانات.

المجموعة 2

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
N	Valide	6	6	6	6	6	6	6	6
	Manquant	0	0	0	0	0	0	0	0
Moyenne		3,0000	3,1667	3,0000	2,5000	2,8333	2,1667	2,333	2,5000
Ecart type		,0000	,7028	,8944	,5477	,7528	,7528	,8165	,5477
Minimum		3	2	2	2	2	1	1	2
Maximum		3	4	4	3	4	3	3	3

يوضح جدول الخصائص الوصفية للمجموعة 2، ان العملاء قد اجابوا تقريبا بالعلامة 3 لجميع الأسئلة، وبالتالي أظهروا اللامبالاة ، ولم يتبنوا سلوكا ملائما أو مخالفا للمسألة. لدينا للمجموعة 2:

المجموعة 2

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Moyenne	3,0000	3,1667	3,0000	2,5000	2,8333	2,1667	2,333	2,5000
Ecart type	,0000	,7028	,8944	,5477	,7528	,7528	,8165	,5477
Coef Variation	0%	24%	30%	22%	27%	35%	35%	22%

بشكل عام نلاحظ ايضا أن هذه المجموعة متجانسة مع تباين متوسط في البيانات.

المجموعة 3

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
N	Valide	7	7	7	7	7	7	7	7
	Manquant	0	0	0	0	0	0	0	0
Moyenne		4,5714	4,5714	4,5714	4,5714	4,0000	4,4286	4,2857	4,4286
Ecart type		,5345	,5345	,5345	,5345	,8165	,5345	,4880	,5345
Minimum		4	4	4	4	3	4	4	4
Maximum		5	5	5	5	5	5	5	5

من الخصائص الاحصائية للمجموعة 3 نلاحظ ان العملاء قد أعطوا درجات مرتفعة لجميع الاسئلة، ويمكننا بعد ذلك تصنيفهم على أنهم ممن يتبعون الموضة، ومنه:

المجموعة 3

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Moyenne	4,5714	4,5714	4,5714	4,5714	4,0000	4,4286	4,2857	4,4286
Ecart type	,5345	,5345	,5345	,5345	,8165	,5345	,4880	,5345
Coef Variation	12%	12%	12%	12%	20%	12%	11%	12%

في هذه المجموعة ، لاحظنا أكبر تجانس في الآراء بين المستجوبين، حيث يوجد تباين ضئيل في البيانات فيما يتعلق بالمتوسط.

2.6. تمارين مقترحة

التمرين 01

في مزرعة ما قام مزارع في تاريخ محدد بوزن البيض المنتج (بالجرام). مثل بيانيا المعطيات التي تحصل عليها هذا المزارع، والمبينة في الجدول الموالي، ثم اوجد كل من التكرارات المتجمعة الصاعدة والنازلة، الفئة الوسيطية، المتوسط الحسابي، المنوال، التباين، والانحراف المعياري.

الوزن	28-37	47-38	52-48	57-53	62-58	72-63	52-73
العدد	3	21	74	112	92	62	6

التمرين 02

في عينة من الطلاب تم قياس وزن كل منهم، ثم تم تمثيل النتائج في الجدول التالي:

الوزن (كغ)	45-54	55-59	60-64	65-69	70-74	75-79	80-89
عدد الطلبة	5	14	33	47	26	13	2

أوجد كل من التكرارات التجميعية الصاعدة والنازلة ومثلها بيانيا. احسب المتوسط الحسابي، الوسيط، المدى، التباين، الانحراف المعياري، الربع الأول، والربيع الثالث.

التمرين 03

ليكن الجدول التالي والذي يمثل علامات 6 طلاب في ثلاث مقاييس:

المقياس / الطالب	الاقتصاد	الرياضيات	الاحصاء
1	12	11	11
2	10	11	12
3	13	12	13
4	10	09	13
5	12	09	12
6	09	08	11

باتباع الخطوات المشروحة سابقا قم باجراء التحليل العاملي على معطيات الجدول، علما ان القيم الذاتي والاشعة الذاتية المرافقة لها هي كما يلي:

$$\lambda_1 = 1.86 \quad ، \quad \lambda_2 = 0.81 \quad ، \quad \lambda_3 = 0.33$$

$$U_3 = \begin{pmatrix} 0.707 \\ -0.707 \\ 0 \end{pmatrix} \; ; \; U_2 = \begin{pmatrix} -0.303 \\ -0.303 \\ 0.906 \end{pmatrix} \; ; \; U_1 = \begin{pmatrix} 0.639 \\ 0.639 \\ 0.429 \end{pmatrix}$$

التمرين 04

قامت هيئة مختصة في وزارة السياحة بإجراء عملية تقييم لأربعة فنادق في العاصمة حسب ثلاث متغيرات هي: الخدمة، االتموقع والسعر. وقد قامت مجموعة من الخبراء التابعين لهذه الهيئة بتصنيف هذه الفنادق بالاعتماد على قديرات تتراوح بين: (-3) و $(+3)$. وقد جاءت نتائج هذا التصنيف كما يلي:

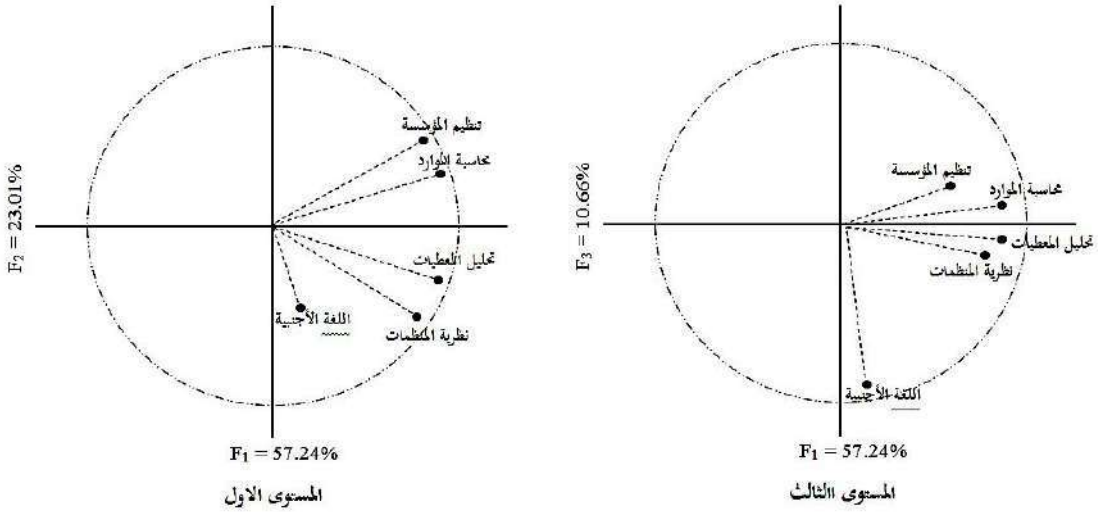
المتغيرات / الفنادق	الخدمة (X_1)	االتموقع (X_2)	السعر (X_3)
الفندق الاول	-2	+3	-1
الفندق الثاني	-1	+1	0
الفندق الثالث	+2	-1	-1
الفندق الرابع	+1	-3	+2

حسب اعتقادك هل بإمكاننا تحليل نتائج هذه الدراسة باستخدام طريقة التحليل الى مركبات اساسية (PCA)؟ علل اجابتك.

باستخدام برنامج SPSS حلل المعطيات الى مركبات اساسية وعلق على النتائج.

التمرين 05

بالاعتماد على طريقة التحليل الى مركبات اساسية، اراد رئيس قسم علوم التسيير بكلية العلوم الاقتصادية لجامعة خميس مليانة، تحليل نتائج طلبة السنة الثالثة (200 طالب) في 5 مقاييس هي: تنظيم المؤسسة، محاسبة الموارد، تحليل المعطيات، نظرية المنظمات واللغة الأجنبية. اذا كان امامه شكلين يمثل كل واحد منهما المقاييس (المتغيرات) في المستوى، كما يلي:



ما هو الشكل الذي يساعد رئيس القسم في تحليل نتائج الطلبة بطريقة أفضل؟ ولماذا؟

التمرين 06

تمثل بيانات الجدول الموالي معطيات عن 6 افراد واربعة متغيرات:

الافراد	X_1	X_2	X_3	X_4
A	32	64	65	67
B	61	37	62	65
C	59	40	45	43
D	36	62	34	35
E	62	46	43	40

قم بتحليل البيانات باستخدام طريقة PCA.

التمرين 07

تم إنشاء قاعدة بيانات اجري على القطاع الصحي ويتعلق باستطلاع رأي المستجوبين لعن وضعهم الصحي والذي يمكن الحكم عليها بأن "جيد جدا"، أو " جيد "، أو "عادي"، أو "سيئ"، أو "سيئ جدا". يوضح الجدول الموالي هذه الردود مع الفئات العمرية للمستجوبين.

العمر	جيد جدا	جيد	عادي	سيئ	سيئ جدا
16-24	243	789	167	18	6
25-34	220	809	164	35	6
35-44	147	658	181	41	8
45-54	90	469	236	50	16
55-64	53	414	306	106	30
65-74	44	267	284	98	20
75+	20	136	157	66	17

اعط تفسيرا مفيدا لكيفية رؤية المرضى لوضعهم الصحي، واوضح كيف يتغير ذلك مع التقدم في السن باستخدام تحليل CA؟

التمرين 08

عينة متكونة من 700 فرد تم تقسيمها حسب المتاجر المختلفة التي يقتنون منها المواد الغذائية. تم ترتيب هؤلاء الافراد وفقا للمتجر والفئة العمرية بحيث تم الحصول على الجدول الموالي"

المتجر	الفئة العمرية			
	16-24	**25-34**	**35-49**	**50+**
A	37	39	45	64
B	13	23	33	38
C	33	69	67	56
D	16	31	34	22
E	8	16	21	35

ما هي طريقة التحليل العاملي الانسب لتفسير هذه المعطيات؟ قم بتحليل البيانات حسب اجابتك؟

التمرين 09

من جدول البيانات التالي:

	X_1	X_2
1	8	3
2	5	3
3	6	4
4	1	6
5	2	8

اتبع الخطوات الموضحة في الفصل الخامس ثم انشئ مخطط الشجرة الهيكلية الذي يلخص هذا التصنيف.

التمرين 10

من خلال بيانات الجدول الموالي والتي توضح 5 مزايا لـ 28 نوع من انواع السيارات، حدد المجموعات الطبيعية للبيانات باستخدام كريقة التصنيف التسلسلي حتى تسهل عملية فهم طبيعة هذه البيانات.

النوع	السعر	الاسطوانة	القوة	الوزن	استهلاك البنزين
Daihatsu Cuore	11600	846	32	650	5,7
Suzuki Swift 1.0 GLS	12490	993	39	790	5,8
Fiat Panda Mambo L	10450	899	29	730	6,1
VW Polo 1.4 60	17140	1390	44	955	6,5
Opel Corsa 1.2i Eco	14825	1195	33	895	6,8
Subaru Vivio 4WD	13730	658	32	740	6,8
Toyota Corolla	19490	1331	55	1010	7,1
Opel Astra 1.6i 16V	25000	1597	74	1080	7,4
Peugeot 306 XS 108	22350	1761	74	1100	9
Renault Safrane 2.2. V	36600	2165	101	1500	11,7
Seat Ibiza 2.0 GTI	22500	1983	85	1075	9,5
VW Golt 2.0 GTI	31580	1984	85	1155	9,5
Citroen ZX Volcane	28750	1998	89	1140	8,8
Fiat Tempra 1.6 Liberty	22600	1580	65	1080	9,3
Fort Escort 1.4i PT	20300	1390	54	1110	8,6
Honda Civic J oker 1.4	19900	1396	66	1140	7,7
Volvo 850 2.5	39800	2435	106	1370	10,8
Ford Fiesta 1.2 Zetec	19740	1242	55	940	6,6
Hyundai Sonata 3000	38990	2972	107	1400	11,7
Lancia K 3.0 LS	50800	2958	150	1550	11,9
Mazda Hachtback V	36200	2497	122	1330	10,8
Mitsubishi Galant	31990	1998	66	1300	7,6
Opel Omega 2.5i V6	47700	2496	125	1670	11,3
Peugeot 806 2.0	36950	1998	89	1560	10,8
Nissan Primera 2.0	26950	1997	92	1240	9,2
Seat Alhambra 2.0	36400	1984	85	1635	11,6
Toyota Previa salon	50900	2438	97	1800	12,8
Volvo 960 Kombi aut	49300	2473	125	1570	12,7

قائمة المراجع

المراجع باللغة العربية

- بن فرحات ساعد، و قطوش عبد الحميد، **محاضرات في الاحصاء 1 مدعمة بتمارين وامتحانات محلولة**، مطبوعة موجهة لطلبة السنة الاولى علوم اقتصادية وتجاريو وعلوم التسيير، جامعة فرحات عباس، سكيف 2013-2014.
- جلاطو جيلالي، **الإحصاء مع تمارين ومسائل محلولة**، ديوان المطبوعات الجامعية، 2002.
- راتول محمد ، **الإحصاء الوصفي**، ديوان المطبوعات الجامعية، الجزائر، 2005.
- رويسات عبد الناصر، **الاحصاء الوصفي ومدخل الاحتمالات "دروس وتمارين"**، ديوان المطبوعات الجامعية، وهران، 2006.
- صواليلي صدر الدين، تحليل المعطيات، دار هومة، الجزائر، 2011.
- معتوق امحمد، **الاحصاء الرياضي والنماذج الاحصائية**، ديوان المطبوعات الجامعية، الجزائر، 2007.
- مواري د. شبيجل، **الإحصاء، سلسلة ملخصات شوم**، الدار الدولية للنشر والتوزيع، 1991.

المراجع باللغة الاجنبية

- Abdi, H, and Williams. L.J. (2010). *Principal Component Analysis*. John Wiley and Sons, Inc. WIREs Comp Stat 2: 433–59.
- Bastin C., et Benzecri J.P. (1980). *Pratique de l'analyse des données : études de cas*, Dunod, Paris, France.
- Bendixen, M.T (2003). A Practical Guide to the Use of Correspondence Analysis in Marketing Research. *Marketing Bulletin* 14. http://marketing-bulletin.massey.ac.nz/V14/MB_V14_T2_Bendixen.pdf
- Bendixen, M.T. (1995). Compositional Perceptual Mapping Using Chi-squared Trees Analysis and Correspondence Analysis. *Journal of Marketing Management* 11 (6): 571–81.
- Benzecri, J.P. (1992). Correspondence Analysis Handbook. New York: Marcel Decker.
- Benzecri, J.P. (1973), L'Analyse des données, Dunod, Paris, France
- Bourbonnais, R. (2004). *Économétrie, Manuel et exercices corriges,* Dunod, Paris, France.
- Bouroche J-M., et Saporta G., (1994), « L'analyse de données », Collection « Que Sais-je ? », Presses Universitaires de France (PUF), 4ème édition.
- Carricano, M., Poujol, F., et Bertrandias, L. (2010). *Analyse de données avec SPSS*, Pearson, Paris, France.
- Casin, P. (2013). *Exercices d'économétrie et d'analyse de données*, Technip, Paris, France.
- Cattell, R. B. (1952). Factor analysis: an introduction and manual for the psychologist and social scientist. Harper.
- Coquillard. P. (2018) Méthodes de classification. Master SV. Cours VIII.
- Cordan, A.D. (1999). *Classification*, Chapman and Hall, London, UK.

- Escofier B. et Pagès J. (1998). Analyses factorielles simples et multiples. 3e ed. Dunod.
- Fellenberg, K., Hauser, N. C., Brors, B., Neutzner, A., Hoheisel, J. D., and Vingron, M. (2001). Correspondence analysis applied to microarray data. Preceedings of the National Academy of Sciences, 98, 10781-10786.
- Hoffman, D. L., and Franke, G. R. (1986). Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. Journal of Marketing Research, 23, 213-227.
- Husson, F., Le, A., and Pagès. J. (2017). *Exploratory Multivariate Analysis by Example Using R*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC.
- Garson, D. (2008). Correspondence Analysis, from Statnotes :Topics in Multivariate Analysis.
- Jolliffe, I. (2002), *Principal Component Analysis*, 2nd edition, Springer-Verlag, Germany.
- Labatte, J-M, (2012), Biostatistiques, université d'Angers, disponible sur : https://math.univ-angers.fr/~labatte/enseignement%20UFR/MTVPS.html, Consulté le 07-15-2021
- Lebart L., Morineau A., Piron M. (2000). *Statistique Exploratoire Multidimensionnelle*, 3ème édition, Dunod, Paris, France.
- Saporta, G. (2011). *Probabilités, Analyse des données et statistique*, Technip, Paris, France.
- Sedkaoui, S. and Khelfaoui, M. (2020). *Sharing economy and big data analytics*, London: ISTE-Wiley.
- Sedkaoui, S. (2018), Data Analytics and Big Data, ISTE-Wiley, London.
- Stafford, J., et Bodson, P. (2007). L'analyse multivariée, presses de l'université du Québec, Canada.
- Zhang, Z., Murtagh, F., Van Poucke, S., Lin S., and Lan, P. (2017). Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. *Ann Transl Med* ;5(4):75

الملاحق

ملحق رقم 01: اسس اعداد الاستبيان

الاستبيان (Questionnaire) هو أحد أدوات البحث، وهي تتكون من مجموعة من الأسئلة بغرض جمع المعلومات من الأفراد موضع البحث، و تكون في الغالب مصممة من أجل التحليل الإحصائي للإجابات. الاستبيان عبارة عن سلسلة من الأسئلة التي يتم طرحها بشكل منهجي من أجل تحديد حالة أو موقف أو تحليل ظاهرة معينة ويشمل عدد كبير من الاسئلة والاحتمالات.

القواعد الأساسية التي يجب اتباعها

يجب أن يتبع تحليل البيانات المسترجعة من الاستبيان قواعد معينة، يتم تقسيمها إلى عدة مراحل. ويسمح احترامها بالحصول على المادة اللازمة القادرة على معالجة النتائج وتحليلها. وتشمل هذه المراحل ما يلي:

- تحديد الهدف من الدراسة: قبل الشروع في تنظيم الاستبيان يوصى بتحديد ما تريد معرفته والمعلومات التي يسعى الباحث الحصول عليها.
- هيكل الاستطلاع : هناك عدة طرق لبناء الدراسة، حيث يتضمن الاستبيان عادة اسئلة عامة في بالداية ثم مطالبة المستجوبين تدريجيا بملء معلومات أكثر تحديدا.
- بناء الاستبيان: عند تطوير الاستبيان يوصى بإدراج نص تمهيدي لتقديم الموضوع وحث المستجوبين على المشاركة. كما يفضل تحديد ما إذا كانت المشاركة مجهولة أم لا. بالإضافة إلى ذلك من الضروري تقديم تفاصيل الاتصال بالشخص المسؤول عن معالجة النتائج. فعلى الباحث ان يكون واضحا قدر الإمكان في صياغة الأسئلة حتى لا تتم توجيه إجابات المشاركين، لان الهدف هو الحصول على بيانات أقرب إلى الواقع من أجل الحصول على نتائج ذات صلة بموضوع البحث.
- تحديد الأشخاص او المؤسسات التي ستتم عليها الدراسة: وهو امر مهم جدا حتى نتمكن من الحصول على نتائج جيدة وذات صلة. وهناك عدة طرق تم التطرق اليها في مقدمة هذه المطبوعة.
- اختبار الاستبيان لمعرفة ما إذا كان المحتوى يوفر النتائج المطلوبة، ويبدا هذا باختبار الاستبيان مع بعض الاشخاص ثم تتم مراجعة المحتوى وفقا لملاحظاتهم. كما يتم النسخة المعدلة من الاستبيان مع ثلاثة إلى عشرة مستجوبين في ظل ظروف مماثلة لتلك الخاصة بالتقييم الرئيسي.
- البدا في عملية الاستطلاع: يمكن للباحث البدا في عملية ادارة الاستبيان اما وجها لوجه او عن طريق البريد او الهاتف. كما يمكنه اطلاق الاستطلاع عبر الإنترنت والذي يتيح نشر الاستبيان على نطاق واسع عن طريق البريد الإلكتروني، أو في المنتديات ، أو الشبكات الاجتماعية ، أو حتى عن طريق دمجه مباشرة في موقع الويب.

- إدخال البيانات: وذلك بإدراج اجوبة المشاركين في قواعد بيانات. وتجدر الاشارة الى ان تطوير استبيان عبر حل عبر الإنترنت يجعل من الممكن حوسبة الاجوبة. هذا يزيل مخاطر الخطأ عند إدخال البيانات.
- تحليل النتائج: بمجرد استكمال عملية الاستطلاع تبدا مرحلة يجب تجميع البيانات وتحليلها وهنا يقوم الباحث بمعالجة البيانات وتحليل النتائج اعتمادا على اجوبة المشاركين التي تم الحصول عليها

محتوى الاستبيان

الهدف من استخدام الاستبيان عامة هو تحقيق أهداف الدراسة والتي توجه كتابة الاستبيان وتساعد الباحث في الحصول على المعلومات اللازمة. لذا لا بد ان يشمل الاستبيان اسئلة قصيرة يسهل على الجميع فهمها. ويجب أن تتبع الأسئلة صيغة تسلسلية وتكون مرتبطة بعضها البعض ، مما سيسمح بالحصول على اجوبة أكثر دقة. ويكون تصميم الاسئلة بطريقة تجعل الاستبيان يلبي مستوى القياس الناتج (الاسمي، الترتيبي، الفاصل الزمني، النسبة، ...الخ) و الاحتياجات من المعلومات.

لذا سيؤثر اختيار نوع الأسئلة أيضا على نتائج الحصر (المسح)، وفيما يلي الأنواع الأكثر شيوعا:

النوع	الوصف	مثال
أسئلة مغلقة بإجابات واحدة	يتم عرض قائمة على المشارك ويجب عليه اختيار اجابة واحدة.	مثال :الجنس: ذكر/انثى المستوى الدراسي: متسوط/ثانوي/ جامعي
أسئلة مغلقة بإجابات متعددة	يتم عرض قائمة على المشارك ويمكن ان يختار عدة اجابات.	القراءة بالنسبة لك هي (عدو اجابات محتملة) ترفيه/ نشاط انفرادي/ طريقة مفيدة لتنمية قدراتك الفكرية/ محاربة الملل
الاسئلة المفتوحة	يكون المشارك حرا في ادراج اجابته، ولا يتم تقيده باختيارات محددة	ما رايك في استخدام الاطفال لمواقع التواصل الاجتماعي؟ .....
الاسئلة نصف المفتوحة (او نصف المغلقة)	يتم إنشاؤها في البداية كسؤال مغلق (أي مع الإجابات المقترحة)، ولكن إحدى الإجابات المقترحة هي الإجابة اجابة يطلب فيها من المشارك تحديدها.	البرنامج المستخدم في تحليل البيانات هو: /Statistica / Excel /SPSS برامج اخرى اذكرها: .....
أسئلة بتصنيف تسلسلي	الأسئلة تدعو الشخص إلى تحديد ترتيب للمقترحات المقدمة	أي من الصفات التالية تتوقعها من الهاتف؟ (ترتيب تنازلي حسب الأهمية من 1 إلى 7 الصفات التالية) الشكل/ جودة الصوت/ حجم الشاشة/

		جودة الصورة/ سعة التخزين/ مقاومة الماء والصدمات/ الاتصال بالأنترنت
أسئلة بتصنيف تسلسلي مع سلم (Echelle)	تدعو الشخص إلى تخصيص قياس للمقترحات. هناك مقاييس مختلفة والاكثر شيعا هو مقياس ليكرت Likert	الخدمة المقدمة ذات جودة عالية: 5 موافق جدا/ 4 موافق/ 3 محايد/ 2 غير موافق/ 1 غير موافق جدا

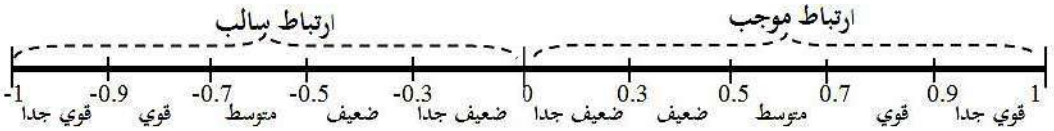
ملحق رقم 02: الارتبـاط وطرق قياسه

في معظم التطبيقات العملية نجد أن هناك علاقة بين متغيرين (أو أكثر)، وتقاس درجة الارتباط بين المتغيرين من خلال تحليل الارتباط. إن الغرض من تحليل الارتباط هو تحديد نوع وقوة العلاقة بين متغيرين. يقاس الارتباط بين متغيرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r.

نوع العلاقة: وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

- إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس
- إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .
- إذا كان معامل الارتباط قيمته صفرا ($r = 0$) نقول انه لا توجد علاقة بين المتغيرين .

قوة العلاقة: ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (±1)، حيث أن قيمة معامل الارتباط محصورة بين: $-1 \leq r \leq +1$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها في الشكل التالي:



ويمكننا قياس الارتباط كما يلي:

معامل الارتباط الخطي البسيط لبيرسون Pearson: من أكثر معاملات الارتباط استخداما خاصة في العلوم الإنسانية والاجتماعية. ومستوى القياس المطلوب عند تطبيق معامل بيرسون للارتباط هو أن تكون بيانات كلا المتغيرين (الظاهرتين) كمية. ويساوي :

$$r = \frac{cov(x,y)}{\sigma_x \, \sigma_y} = \frac{\sum(x - \bar{x}) \times (y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2}\sqrt{\sum(y - \bar{y})^2}}$$

حيث أن :

$cov(x,y)$: هو التباين المشترك بين $(x \ , \ y)$، $cov\,(,y) = \frac{\sum(x-\bar{x})\times(y-\bar{y})}{n-1}$

σ_x: هو الانحراف المعياري لقيم (x)،

σ_Y : هو الانحراف المعياري لقيم (y).

معامل ارتباط الرتب لسبيرمان Spearman: في بعض الحالات نقوم بحساب معامل الارتباط بين رتب متغيرين وليس قيمهما، أو ما يعرف بمعامل ارتباط الرتب. ولحساب معامل سبيرمان نتبع الخطوات التالية:

- ترتيب المتغيرين تصاعديا أو تنازليا
- حساب الرتب
- حساب الفرق بين رتب كل من المتغيرين (ونرمز لها بالرمز d)
- حساب $\sum d^2$
- حساب معامل سبيرمان بالعلاقة: $r_s = 1 - \frac{6(\sum d^2)}{n(n^2 - 1)}$

ملحق رقم 03: التذكير بالجبر الخطي والمصفوفات

1. المصفوفات

تعريف المصفوفة: المصفوفة عبارة عن تنظيم عددي (مجموعة من الأعداد الحقيقية) مؤلف من: m . n عنصرا، مرتبة في جدول مكون من m صف ، n عمودا، حيث m ، n عددان طبيعيان ومحصورة بين قوسين من الشكل [ ]. اذا كانت المصفوفة تحتوي صفوفا عددها m وأعمدة عددها n نقول عنها إنها مصفوفة من الرتبة n X m. والشكل العام للمصفوفة هو:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

انواع المصفوفات

مصفوفة مستطيلة: (مصفوفة من النوع $m \times n$ حيث $m \neq n$)

المصفوفة المربعة: هي المصفوفة من النوع $n \times n$ أي أن عدد صفوفها يساوي عدد أعمدتها، ونجد فيها:

- المصفوفة القطرية : جميع عناصرها أصفار ما عدا العناصر الواقعة على القطر الاساسي فيكون أحدها على الأقل مغايرا للصفر.
- مصفوفة الوحدة: هي مصفوفة قطرية يكون فيها كل من العناصر الواقعة على القطر مساويا للواحد.

مصفوفة صفرية: المصفوفة من الرتبة $m \times n$ وجميع عناصرها أصفار (امثلة مصفوفة صفرية مستطيلة، مصفوفة صفرية مربعة).

العمليات على المصفوفات

جمع (طرح) مصفوفتين: إذا كانت A و B مصفوفتان فانه يمكننا حساب A+B (أو A–B)، اذا تحقق ما يلي: لهما نفس الرتبة وهنا نجمع (ناو طرح) المدخلات المتناظرة.

جداء مصفوفتين: قبل اجراء عملية ضرب مصفوفتين علينا التحقق من الشرط التالي: عدد الاعمدة في المصفوفة الاولى تساوي عدد الصفوف في المصفوفة الثانية. ونحصل على النتيجة اذا ضربنا عناصر صفوف المصفوفة A في عناصر الاعمدة المناظرة لها من المصفوفة B وجمعنا نواتج الضرب. رتبة المصفوفة A.B يتحدد تماماً من : عدد صفوف A وعدد أعمدة B.

جداء مصفوفة في ثابت: حاصل ضرب مصفوفة بعدد حقيقي هو مصفوفة من النوع نفسه، ولكن العناصر تغيرت حيث نضرب كل منها بالعدد الثابت. ونشير هنا الى ان قسمة المصفوفة على عدد حقيقي نفس مفهوم ضربها بعدد حقيقي.

حساب المحدد

تعريف: يمكن تعريف المحدد على أنه تطبيق خطي معرف من الفضاء الشعاعي $\mathcal{M}_n(K)$ نحو K ، نرمز لمحدد المصفوفة بالرمز $|A|$ أو $\det(A)$.

امثلة :

$|3|$ $\quad\quad$ $\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$ $\quad\quad$ $\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}$

محدد من الدرجة: الاولى الثانية الثالثة

مثال: لتكن المصفوفة المربعة من الدرجة الثانية المعرفة بالشكل التالي: $A = \begin{pmatrix} 1 & -1 \\ 2 & 5 \end{pmatrix}$ نحسب محدد المصفوفة كما يلي:

$$|A| = \begin{vmatrix} 1 & -1 \\ 2 & 5 \end{vmatrix} = (1*5) - ((-1)*2) = 5-(-2) = 5+2 = 7$$

مثال :لتكن المصفوفة المربعة من الدرجة الثالثة المعرفة بالشكل التالي: $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & -5 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ نحسب محدد المصفوفة كما يلي:

$$\det(A) = 2\begin{vmatrix} -5 & 1 \\ 1 & 1 \end{vmatrix} - (-1)\begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} + 0\begin{vmatrix} -1 & -5 \\ 0 & 1 \end{vmatrix} = -13$$

2. القيم الذاتية والاشعة الذاتية

القيم الذاتية للمصفوفة

نقول عن λ انها قيمة ذاتية للمصفوفة A اذا كان:

$$|A - \lambda I| = 0 \quad \text{أو} \quad |\lambda I - A| = 0$$

وتسمى هذه المعادلة بالمعادلة المميزة للمصفوفة A. ولكي تكون λ قيمة ذاتية يجب ان يكون لهذه المعادلة حل غير صفري.

مثال: سوف نقوم بحساب القيم الذاتية للمصفوفة التالية: $A = \begin{bmatrix} 5 & -3 \\ -6 & 2 \end{bmatrix}$

$$|A-\lambda I|=0 \rightarrow \left|\begin{bmatrix}5 & -3\\ -6 & 2\end{bmatrix}-\lambda\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right|$$

$$=\left|\begin{bmatrix}5 & -3\\ -6 & 2\end{bmatrix}-\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}\right|=\left|\begin{bmatrix}5-\lambda & -3-0\\ -6-0 & 2-\lambda\end{bmatrix}\right| \quad =\begin{vmatrix}5-\lambda & -3\\ -6 & 2-\lambda\end{vmatrix}$$

$$=\begin{vmatrix}5-\lambda & -3\\ -6 & 2-\lambda\end{vmatrix}$$

$$=(5-\lambda)(2-\lambda)-(6)(3)$$
$$=10-5\lambda-2\lambda+\lambda^2-18$$
$$=\lambda^2-7\lambda-8$$

معادلة من الدرجة 2 ← حلها بحساب المميز

بحل المعادلة نجد: $\lambda=-1 \quad ; \quad \lambda=8$

الاشعة الذاتية

اذا كانت A مصفوفة وكانت $\lambda \in R$ ، λ هي قيمة ذاتية للمصفوفة A .

اذا وجد $X \in R^n$ و $X \neq 0$ حيث $AX=\lambda X$ وفي هذه الحالة يسمى X بالشعاع الذاتي للقيمة الذاتية λ، ونكتب : $|A-\lambda I|x=0$

ان الاشعة الذاتية للمصفوفة A المناظرة للقيمة الذاتية λ هي الاشعة الذاتية التي تحقق العلاقة $AX=\lambda X$ بطريقة متكافئة. الاشعة الذاتية المناظرة للقيمة الذاتية λ هي الاشعة الصفرية في فضاء الحلول للمعادلة :

$$|A-\lambda I|x=0$$

والذي يسمى بفضاء الحلول للفضاء الذاتي للمصفوفة A المناظرة للقيمة الذاتية λ.

من المثال السابق سنحاول ايجاد الاشعة الذاتي:

لما: $\lambda=-1$

$$\begin{pmatrix}5-\lambda & -3\\ -6 & 2-\lambda\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}5-(-1) & -3\\ -6 & 2-(-1)\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}-6 & -3\\ -6 & -3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

نقل $3x_2$ الى الطرف الثاني مع تغيير الاشارة

$$-6x_1+3x_2=0$$

نعوض في أحد المتغيرات x_1 بأي رقم ماعدا 0

$$-6x_1=-3x_2$$
$$x_1=1$$
$$(-6)(1)=-3x_2$$

بقسمة الطرفين على -3

$$\frac{-6}{-3}=\frac{-3x_2}{-3}$$
$$x_2=2$$

ومنه الشعاع الذاتي هو: $\begin{pmatrix}1\\ 2\end{pmatrix}$

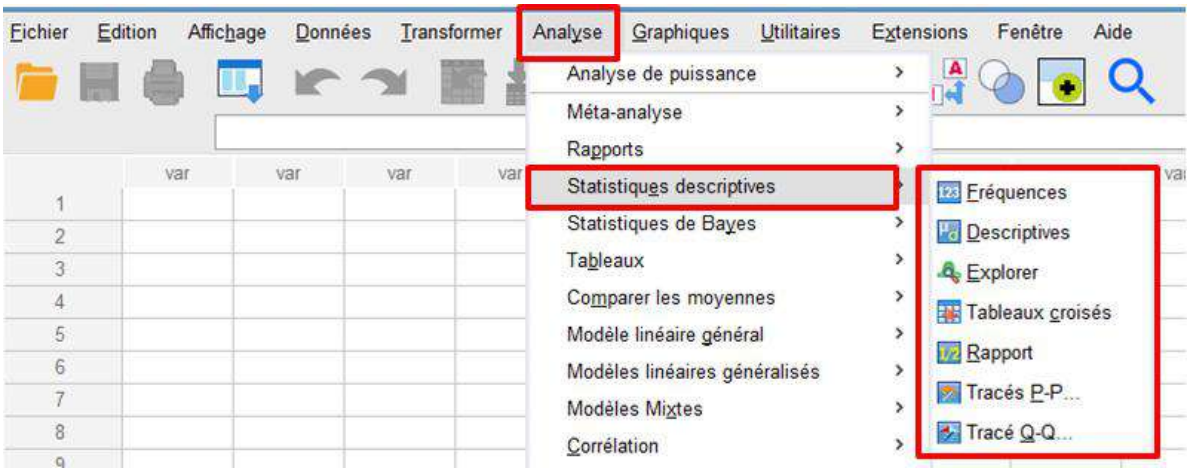
بنفس الطريقة نحسب الشعاع الذاتي للقيمة الذاتية الاخرى $\lambda=8$ ونجد:

$$\begin{pmatrix}1\\ -1\end{pmatrix}$$

ملحق 04: مراحل التحليل الوصفي باستخدام SPSS

كما راينا من خلال الفصول السابقة فان الخطوة الاولى في تحليل البيانات هي إجراء تحليل الاحصاء الوصفي، حيث يوفر هذا التحليل ا فكرة عن شكل توزيع البيانات ويسمح بالحصول على مقاييس النزعة المركزية (المتوسط، والوسيط ، ... الخ) ، ومقاييس التشتت (التباين، والانحراف المعياري ، وغيرها) ، والنسب المئوية، بالإضافة إلى التمثيلات البيانية (الدائرة، المدرج التكراري، ...).

ويوفر SPSS العديد من الأدوات لإجراء هذا الوصف التحليلي، وكلها ضمن قائمة Analyse والتي يندرج منها خيار الإحصاء الوصفي (Statistiques descriptives)، والتي تشمل عدة اجراءات منها: التكرارات (Fréquences)، الوصفية (Descriptives)، الاستكشاف (Explorer)، وغيرها. كما توضحه الصورة الموالية.



حيث يمكن تلخيص الاجراءات الثلاث الاولى وما توفره من تطبيقات للتحليل الوصفي من خلال الجدول الموالي

Fréquences	يوفر مقاييس إحصائية وتمثيلات بيانية مفيدة لوصف أنواع مختلفة من المتغيرات. يسمح بالحصول على وصف للمتغير من جداول التكرارات، التمثيلات البيانية ، النسب المئوية ، مقاييس النزعة المركزية ومقاييس التشتت.
Descriptives	يمكن هذا الاجراء حساب مقاييس النزعة المركزية والتشتت والتوزيع للعديد من المتغيرات، ويعرضها في جدول واحد.
Explorer	ينشئ هذا الإجراء مقاييس موجزة وتمثيلات بيانية مثل مخططات الصندوق، والمدرج التكراري ، والتوزيع الطبيعي، وغيرها.

بالإضافة إلى الرسوم البيانية الناتجة عن الإجراءات الوصفية السابقة، يحتوي برنامج SPSS على قائمة مخصصة للحصول على نتائج في شكل تمثيلات بيانية، والتي تعتبر بمثابة عرض موجز للخصائص العامة في التعامل مع التمثيلات البيانية. ومكن استعمال هذه التمثيلات من خلال قائمة Graphiques ومن ثم اختيار الشكل البياني المناسب (من بين التمثيلات الموضحة في الفصل الاول) لمعطيات الدراسة.