وزارة التعليم العالي والبحث العلمي
جامعة لونيسي علي – البليدة 02 –
كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير
– الشهيد طالب عبد الرحمان –

قسم العلوم: الاقتصادية

دروس في مقياس تحليل معطيات معمق

السنة الأولى ماستر (تخصص: تحليل اقتصادي واستشراف)
في ميدان "العلوم الاقتصادية والتجارية وعلوم التسيير"

السداسي الثاني

من إعداد الأستاذ: حوشين يوسف

فهرس المحتويات:

I. المحور الأول: تحليل المركبات الرئيسية (ACP)

1. مفهوم طريقة تحليل المركبات الرئيسية:

أ. تعريف طريقة تحليل المركبات الرئيسية:

تحليل المركبات الرئيسية هو أسلوب رياضي يقوم على أساس تحويل مجموعة من المتغيرات التوضيحية المترابطة فيما بينها إلى مجموعة جديدة من المتغيرات غير المترابطة (أو المتعامدة – Orthogonal) تدعى المركبات الرئيسية. حيث كل مركبة رئيسية هي عبارة عن توليفة خطية (Combinaison linéaire) للمتغيرات الأصلية.

يحتوي هذا الجدول على n فرد وp متغير، فكل فرد يمثل في فضاء ذو p بعد، وكل متغير يمثل في فضاء ذو n بعد.

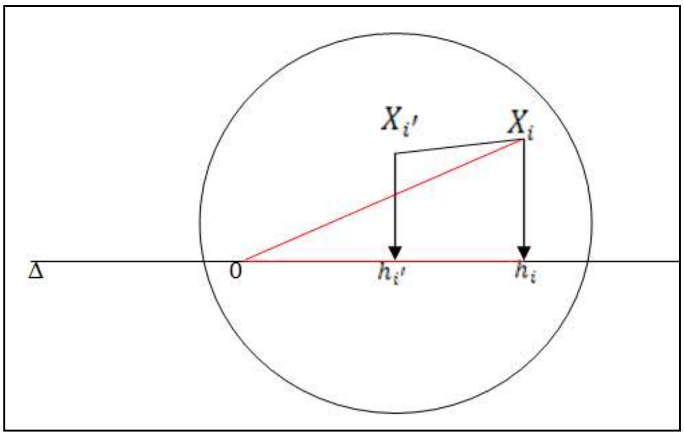
مثال:

جدول بيانات يحتوي على نقاط 20 طالب في 10 مواد دراسية، فيكون: $n = 20$، و $p = 10$.

ب. مبدأ طريقة تحليل المركبات الرئيسية:

يتمثل مبدأ ACP في إسقاط سحابة نقاط الأفراد والمتغيرات من فضاء ذو p (n) بُعد إلى محاور ومستويات من 2 إلى 3 أبعاد فقط، مع الحفاظ على المسافات بين هذه الأفراد (والمسافات بين المتغيرات) قدر الإمكان، أي العمل على عدم تشويه سحابة النقاط الأصلية لجميع المتغيرات عند اسقاطها (الحصول على تمثيل أقل تشوها لسحابة النقاط)، وبالتالي المحافظة بأفضل شكل ممكن على المسافات الأصلية، وتحقيق أفضل تمثيل للتنوع والتباين الأصلي. وذلك بالاعتماد على التباين الكلي (Inertie).

يعتمد الإسقاط على مبدأ المحافظة على المسافات بين الأفراد في المتوسط قدر المستطاع عند الإسقاط:



أي: $d(X_i\ ,\ X_{i'}) \approx d(h_i\ ,\ h_{i'})$.

حيث: $X_i = (x_{i1} \quad x_{i2} \quad \dots \quad x_{ij} \quad \dots \quad x_{ip})$ (X_i هي نقطة تمثيلية للفرد i في الفضاء ذو البعد p، أي لها p إحداثية) و h_i هي إسقاطها العمودي على المحور Δ.

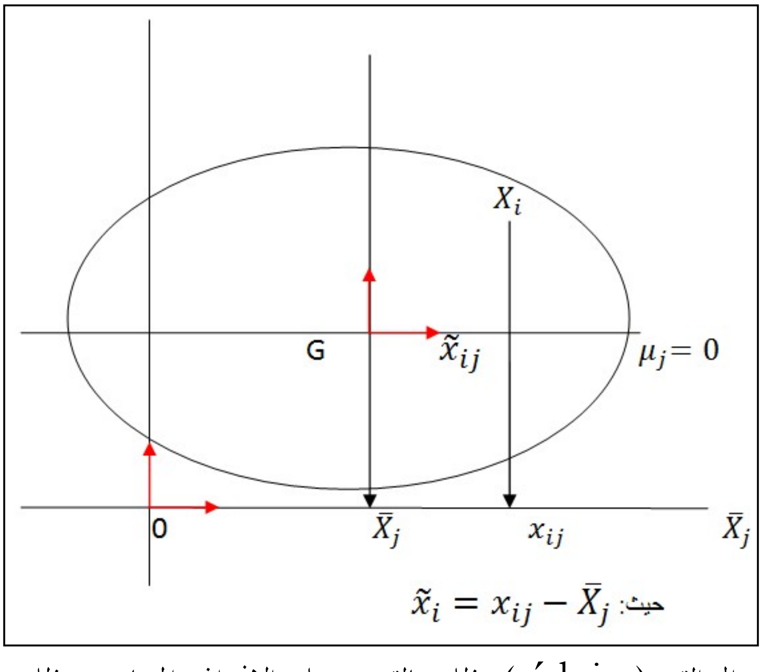
ملاحظة: إذا كان هذان الفردان متقاربان (المسافة بينهما صغيرة)، فهذا يدل على أنهما متشابهان، أي يأخذان قيما متقاربة (إحداثياتهما على المعالم قريبة).

وحتى نتحصل على تمثيل جيد عند الإسقاط، يجب أن نجعل مسافة الإسقاط $0h_i$ أقرب ما يمكن من المسافة الحقيقة $0X_i$ (مع $0h_i \leq 0X_i$)، أي جعل $0h_i$ أكبر ما يمكن، وبالتالي فقدان أقل قدر ممكن من المعلومات.

عند الإسقاط على المستقيم Δ وللحصول على تمثيل أقل تشوه لسحابة النقاط، نأخذ مركز ثقل سحابة النقاط (G) كمركز للإسقاط، والذي يمثل متوسط المتغيرات (المبدأ 0 لا يعطي تمثيل جيد لتشتت سحابة النقاط).

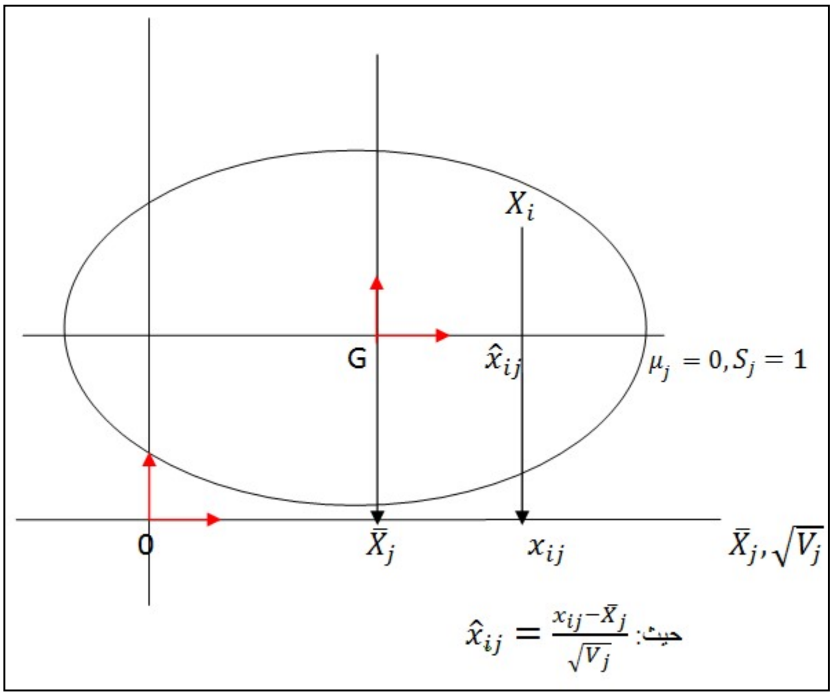
عند جعل المتغيرات ممركزة (Centré) وذلك بطرح كل القيم من متوسطاتها، نكون قد حولنا المبدأ من 0 إلى مركز ثقل سحابة النقاط (G).

ويمكن توضيح ذلك بالشكل التقريبي التالي:



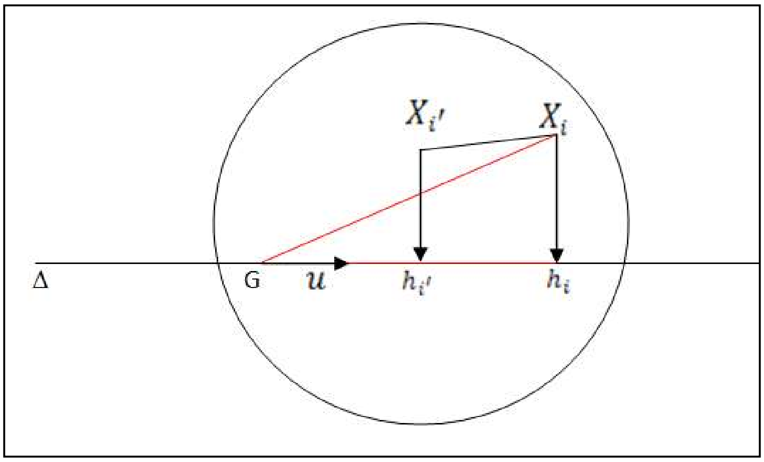
كما نقوم باختزال القيم (réduire) وذلك بالقسمة على الانحراف المعياري، وذلك لجعل القيم متجانسة (إهمال وحدات القياس)، حيث يصبح تباين القيم المختزلة هو 1 بالنسبة لجميع المتغيرات.

ويمكن توضيح ذلك بالشكل التقريبي التالي:



فيتضح من هذا التمثيل أن نقطة الإسقاط على المحور الجديد ما هي في الحقيقة إلا تمثيل للقيمة $\hat{x}_{ij}$، أي أن: $h_i = \hat{x}_{ij}$.

وبشكل عام يمكن توضيح الإسقاط بجعل مركز الإسقاط هو النقطة G كما يلي:



بما أن الإسقاط يكون عمودي، فحسب نظرية فيثاغورس:

$d^2(G, X_i) = d^2(X_i, h_i) + d^2(G, h_i)$ (أي: $\|\overrightarrow{GX_i}\|^2 = \|\overrightarrow{X_i h_i}\|^2 + \|\overrightarrow{Gh_i}\|^2$).

وبما أن عدد الأفراد التي سنقوم بإسقاطها على المستقيم Δ هي n، فإن:

$$\sum_{i=1}^{n} d^2(G, X_i) = \sum_{i=1}^{n} d^2(X_i, \Delta) + \sum_{i=1}^{n} d^2(G, h_i).$$

نقسم جميع حدود المعادلة على n، فتصبح العلاقة:

$\frac{1}{n}\sum_{i=1}^{n} d^2(G\,,X_i) = \frac{1}{n}\sum_{i=1}^{n} d^2(X_i\,,\Delta) + \frac{1}{n}\sum_{i=1}^{n} d^2(G\,,h_i)$.

أي أن: $I_G(N) = I_\Delta(N) + I_{\Delta''}(N)$.

حيث:

- $I_G(N)$: هو التباين الكلي لسحابة نقاط (Inertie du nuage de points des individus)، والتي تمثل المسافة أو البعد بين كل فرد والمركز G (يمثل تشتت سحابة النقاط حول المركز G).
- $I_\Delta(N)$: هو تباين سحابة النقاط لبعد كل فرد عن المستقيم Δ (يمثل تشتت سحابة النقاط حول المستقيم Δ).
- $I_{\Delta''}(N)$: هو تباين سحابة النقاط لطول المستقيم Δ ‏(Inertie de long de Δ) (يمثل تشتت النقاط المسقطة على المستقيم Δ حول المركز G).

وأفضل إسقاط للنقاط على المستقيم Δ هو جعل المسافة بين النقطة X_i والمستقيم Δ أقرب ما يمكن (البحث عن أفضل محور)، وهذا يكافئ أن المسافة بين الإسقاط h_i والمركز G أبعد ما يمكن (لأن الإسقاط يكون عمودي فالمثلث يكون قائم)، وهذا يعني أن تكون نقاط الإسقاط أبعد ما يمكن عن المركز G.

أي: $\begin{cases} Min(\sum_{i=1}^{n} d^2(X_i\,,\Delta)) \\ Max(\sum_{i=1}^{n} d^2(G\,,h_i)) \end{cases}$، والتي تكافئ: $\begin{cases} Min(I_\Delta) \\ Max(I_{\Delta''}) \end{cases}$.

لدينا: $I_{\Delta''} = \frac{1}{n}\sum_{i=1}^{n} d^2(G\,,h_i) = \frac{1}{n}\sum_{i=1}^{n}\left\|\overrightarrow{Gh_i}\right\|^2 = \frac{1}{n}\sum_{i=1}^{n}[(Gh_i)^t \times Gh_i]$.

وكما أشرنا سابقا إلى أن: $h_i = \hat{x}_{ij}$. فيكون: $Gh_i = \sum_{j=1}^{p} \hat{x}_{ij} \times u_j = \hat{X}_i u$ حيث: $i = 1,2,\dots,n$.

حيث: $\hat{X}_i = (\hat{x}_{i1} \quad \hat{x}_{i2} \quad \dots \quad \hat{x}_{ip})$، و $\hat{x}_{ij} = \frac{x_{ij}-\bar{X}_j}{\sqrt{V_j}}$. و: u هو شعاع توجيه المحور Δ.

إذن: $I_{\Delta''} = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\hat{X}_i u\right)^t \times \left(\hat{X}_i u\right)\right] = \frac{1}{n}\left(\hat{X}u\right)^t \times \hat{X}u = \frac{1}{n}u^t\hat{X}^t\hat{X}u = u^tRu$.

حيث: $R = \frac{1}{n}\hat{X}^t\hat{X}$ هي مصفوفة الارتباط.

إذن النظام الذي نبحث عن تعظيمه هو: $\begin{cases} Max(u^tRu) \\ S/C\colon u^tu = 1 \end{cases}$.

حيث أن: $u^tu = 1$ تعني أن الشعاع $\vec{u}$ معياري (Normé)، أي: $\|u\| = 1$.

ومنه دالة لاقرانج (LANGRANGE) هي: $L(u) = u^tRu - \lambda(u^tu - 1)$.

لإيجاد الحل: $\frac{\partial L(u)}{\partial u} = 0$.

أي: $\frac{\partial\left(u^tRu-\lambda\left(u^tu-1\right)\right)}{\partial u} = 0 \Longrightarrow 2Ru - 2\lambda u = 0$.

ومنه: $Ru = \lambda u$.

إذن: λ هي قيمة ذاتية لـ R و u هي الشعاع الذاتي الموافق للقيمة الذاتية λ.

لدينا: $Ru = \lambda u \Longrightarrow u^t Ru = \lambda u^t u = \lambda$.

ومنه: $I_{\Delta''}\big(N(I)\big) = u^t Ru = \lambda$. أي أن القيم الذاتية تعبر عن التباين.

ومنه نستنتج أن تعظيم التباين الكلي لطول المستقيم (Δ) يعني البحث عن أكبر القيم الذاتية للمصفوفة R.

أي: $\boldsymbol{Max(I_{\Delta''}(N)) \Leftrightarrow Max(\lambda_i)}$.

2. خطوات إجراء طريقة تحليل المركبات الرئيسية:

أ. تشكيل جدول البيانات الكمية X:

تستخدم طريقة تحليل المركبات الرئيسية مع جدول البيانات الكمية، والذي يأخذ الشكل التالي:

المتغيرات / الأفراد	X_1	…	X_j	…	X_p
1	x_{11}				
2					
…					
i			x_{ij}		
…					
n					x_{np}

حيث: X_j يمثل المتغير j $(j = 1,2,\dots,p)$.

و: الأعداد من 1 إلى n تمثل الأفراد.

و: x_{ij} تمثل قيمة المتغير X_j عند الفرد i.

نرمز لجدول البيانات الكمية بالرمز X، حيث: $X_{n\times p} = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$.

كل فرد i يُمثّل في فضاء ذو بعد p $\mathbb{R}^p$: $X_i = (x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip})$.

وكل متغير j يُمثّل في فضاء ذو بعد n $\mathbb{R}^n$: $X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$.

ب. حساب المتوسطات وتشكيل جدول البيانات الكمية المركز $\tilde{X}$ (Tableau Centré):

جدول البيانات الكمية المركز $\tilde{X}$ (X Tilde):

المتغيرات / الأفراد	$\tilde{X}_1$	…	$\tilde{X}_j$	…	$\tilde{X}_p$
1	$\tilde{x}_{11}$				
2					
…					
i			$\tilde{x}_{ij}$		
…					
n					$\tilde{x}_{np}$
G	$\bar{X}_1$		$\bar{X}_j$		$\bar{X}_p$

حيث: $\tilde{X}_j = X_j - \bar{X}_j$ $(j = 1,2, \dots, p)$. أي: $\tilde{x}_{ij} = x_{ij} - \bar{X}_j$.

و G هي نقطة مركز ثقل سحابة النقاط (إحداثياتها هي متوسطات المتغيرات) .

نرمز لجدول البيانات الكمية المركز بالرمز $\tilde{X}$ ، حيث:

$$\tilde{X}_{n\times p} = \begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1j} & \dots & \tilde{x}_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \tilde{x}_{i1} & \dots & \tilde{x}_{ij} & \dots & \tilde{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \tilde{x}_{n1} & \cdots & \tilde{x}_{nj} & \dots & \tilde{x}_{np} \end{pmatrix}.$$

ملاحظات:

- النقاط القريبة من G تمثل المعلومات المفقودة عند تقليص الأبعاد، لأنها ستكون ممثلة أفضل في محاور أخرى (المحور الرابع أو المحور الخامس مثلا).
- جعل البيانات ممركزة لا يؤثر على شكل سحابة النقاط.

ج. حساب التباينات وتشكيل جدول البيانات الكمية المعياري $\hat{X}$ (Tableau normé):

جدول البيانات الكمية المعياري $\hat{X}$ (X Chapeau):

المتغيرات / الأفراد	$\hat{X}_1$	…	$\hat{X}_j$	…	$\hat{X}_p$

1	$\hat{x}_{11}$				
2					
…					
i			$\hat{x}_{ij}$		
…					
n					$\hat{x}_{np}$
G	$\bar{X}_1$		$\bar{X}_j$		$\bar{X}_p$
التباين	V_1		V_j		V_p

حيث: $\hat{X}_j = \frac{X_j - \bar{X}_j}{\sqrt{V_j}}$ $(j = 1,2, \ldots, p)$. أي: $\hat{x}_{ij} = \frac{x_{ij} - \bar{X}_j}{\sqrt{V_j}}$.

نرمز لجدول البيانات الكمية المعياري بالرمز $\hat{X}$، حيث:

$$\hat{X}_{n\times p} = \begin{pmatrix} \hat{x}_{11} & \cdots & \hat{x}_{1j} & \cdots & \hat{x}_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \hat{x}_{i1} & \cdots & \hat{x}_{ij} & \cdots & \hat{x}_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ \hat{x}_{n1} & \cdots & \hat{x}_{nj} & \cdots & \hat{x}_{np} \end{pmatrix}.$$

ملاحظات:

- ✓ المتغيرات التي لها تباينات كبيرة هي المسؤولة عن تشتت (تبعثر) الأفراد في سحابة النقاط (المستوي العاملي).
- ✓ يسمح تحويل البيانات إلى بيانات مختزلة بتوحيد وحدات القياس.

د. حساب مصفوفة التباين المشترك V أو مصفوفة الارتباط R:

يمكن القيام بالتحليل بالمركبات الرئيسية إما بالاعتماد على مصفوفة التباين المشترك (Matrice de covariance)، أو مصفوفة الارتباط للمتغيرات التوضيحية، وإن نوع المصفوفة المفضل استخدامها يعتمد في الغالب على طبيعة المتغيرات قيد التحليل. فإذا كانت للمتغيرات نفس وحدات القياس (تجانس الوحدات)، ونسمي في هذه الحالة التحليل بالمركبات الرئيسية غير معياري (ACP Non normé)، أو بالاعتماد على مصفوفة الارتباط R (Matrice de corrélation) وهذا إن لم تكن للمتغيرات نفس وحدات القياس (عدم تجانس الوحدات)، ونسمي في هذه الحالة التحليل بالمركبات الرئيسية معياري (ACP Normé).

1. مصفوفة التباين المشترك V:

مصفوفة التباين المشترك V تحسب بالاعتماد على الصيغة المصفوفية التالية:

$$.V_{p\times p} = D_{p\times p}.\tilde{X}^t_{p\times n}.\tilde{X}_{n\times p}$$

حيث D هي مصفوفة الثقل (الأوزان): $D_{p\times p} = \frac{1}{n}.I_p$. و I هي المصفوفة الأحادية.

أي: $$.V = \begin{pmatrix} \frac{1}{n} & \cdots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \cdots & \frac{1}{n} & \cdots & 0 \\ \vdots & \dots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \dots & \frac{1}{n} \end{pmatrix}.\begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{i1} & \dots & \tilde{x}_{n1} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \tilde{x}_{1j} & \dots & \tilde{x}_{ij} & \dots & \tilde{x}_{nj} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \tilde{x}_{1p} & \cdots & \tilde{x}_{ip} & \dots & \tilde{x}_{np} \end{pmatrix}.\begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1j} & \dots & \tilde{x}_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \tilde{x}_{i1} & \dots & \tilde{x}_{ij} & \dots & \tilde{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \tilde{x}_{n1} & \cdots & \tilde{x}_{nj} & \dots & \tilde{x}_{np} \end{pmatrix}$$

أي: $$.V = \begin{pmatrix} \frac{1}{n} & \cdots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \cdots & \frac{1}{n} & \cdots & 0 \\ \vdots & \dots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \dots & \frac{1}{n} \end{pmatrix}.\begin{pmatrix} \sum_{i=1}^n \tilde{x}_{i1}^2 & \cdots & \sum_{i=1}^n \tilde{x}_{i1}.\tilde{x}_{ij} & \dots & \sum_{i=1}^n \tilde{x}_{i1}.\tilde{x}_{ip} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \sum_{i=1}^n \tilde{x}_{ij}.\tilde{x}_{i1} & \dots & \sum_{i=1}^n \tilde{x}_{ij}^2 & \dots & \sum_{i=1}^n \tilde{x}_{ij}.\tilde{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \sum_{i=1}^n \tilde{x}_{ip}.\tilde{x}_{i1} & \cdots & \sum_{i=1}^n \tilde{x}_{ip}.\hat{x}_{ij} & \dots & \sum_{i=1}^n \tilde{x}_{ip}^2 \end{pmatrix}$$

أي: $$.V = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \tilde{x}_{i1}^2 & \cdots & \frac{1}{n}\sum_{i=1}^n \tilde{x}_{i1}.\tilde{x}_{ij} & \dots & \frac{1}{n}\sum_{i=1}^n \tilde{x}_{i1}.\tilde{x}_{ip} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ij}.\tilde{x}_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ij}^2 & \dots & \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ij}.\tilde{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ip}.\tilde{x}_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ip}.\tilde{x}_{ij} & \dots & \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ip}^2 \end{pmatrix}$$

تكون عناصر القطر الرئيسي من الشكل: $V_j = \frac{1}{n}\sum_{i=1}^n \tilde{x}_{ij}^2 = \frac{1}{n}\sum_{i=1}^n\left(x_{ij} - \bar{X}_j\right)^2$. أي تباين المتغير X_j.

والعناصر الأخرى من الشكل (مثلا السطر 1 والعمود j):

$$.\frac{1}{n}\sum_{i=1}^n \tilde{x}_{i1}.\tilde{x}_{ij} = \frac{1}{n}\sum_{i=1}^n(x_{i1} - \bar{X}_1)\left(x_{ij} - \bar{X}_j\right) = Cov(X_1, X_j)$$

أي التباين المشترك (التغاير) بين X_1 و X_j.

ومنه يمكن كتابة V على الشكل المصفوفي التالي:

$$.V = \begin{pmatrix} V_1 & \cdots & Cov(X_1, X_j) & \dots & Cov(X_1, X_p) \\ \vdots & \dots & \vdots & \dots & \vdots \\ Cov(X_j, X_1) & \dots & V_j & \dots & Cov(X_j, X_p) \\ \vdots & \dots & \cdots & \cdots & \cdots \\ Cov(X_p, X_1) & \cdots & Cov(X_p, X_j) & \dots & V_p \end{pmatrix}$$

وبما أن $Cov(X_i, X_j) = Cov(X_j, X_i)$، فإنه يمكن كتابة المصفوفة V على الشكل التالي:

$$.V = \begin{pmatrix} V_1 & \cdots & Cov(X_1, X_j) & \dots & Cov(X_1, X_p) \\ \vdots & \dots & \vdots & \dots & \vdots \\ Cov(X_1, X_j) & \dots & V_j & \dots & Cov(X_j, X_p) \\ \vdots & \dots & \cdots & \cdots & \cdots \\ Cov(X_1, X_p) & \cdots & Cov(X_j, X_p) & \dots & V_p \end{pmatrix}$$

2. مصفوفة الارتباط R:

مصفوفة الارتباط R تحسب بالاعتماد على الصيغة المصفوفية التالية:

$$R_{p\times p} = D_{p\times p}.\hat{X}^t_{p\times n}.\hat{X}_{n\times p}.$$

أي:

$$R = \begin{pmatrix} \frac{1}{n} & \cdots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \cdots & \frac{1}{n} & \cdots & 0 \\ \vdots & \dots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \dots & \frac{1}{n} \end{pmatrix}.\begin{pmatrix} \hat{x}_{11} & \cdots & \hat{x}_{i1} & \dots & \hat{x}_{n1} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \hat{x}_{1j} & \dots & \hat{x}_{ij} & \cdots & \hat{x}_{nj} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \hat{x}_{1p} & \cdots & \hat{x}_{ip} & \dots & \hat{x}_{np} \end{pmatrix}.\begin{pmatrix} \hat{x}_{11} & \cdots & \hat{x}_{1j} & \dots & \hat{x}_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \hat{x}_{i1} & \dots & \hat{x}_{ij} & \dots & \hat{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \hat{x}_{n1} & \cdots & \hat{x}_{nj} & \dots & \hat{x}_{np} \end{pmatrix}.$$

أي:

$$R = \begin{pmatrix} \frac{1}{n} & \cdots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \cdots & \frac{1}{n} & \cdots & 0 \\ \vdots & \dots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \dots & \frac{1}{n} \end{pmatrix}.\begin{pmatrix} \sum_{i=1}^{n} \hat{x}_{i1}^2 & \cdots & \sum_{i=1}^{n} \hat{x}_{i1}.\hat{x}_{ij} & \dots & \sum_{i=1}^{n} \hat{x}_{i1}.\hat{x}_{ip} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \sum_{i=1}^{n} \hat{x}_{ij}.\hat{x}_{i1} & \dots & \sum_{i=1}^{n} \hat{x}_{ij}^2 & \dots & \sum_{i=1}^{n} \hat{x}_{ij}.\hat{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \hat{x}_{ip}.\hat{x}_{i1} & \cdots & \sum_{i=1}^{n} \hat{x}_{ip}.\hat{x}_{ij} & \dots & \sum_{i=1}^{n} \hat{x}_{ip}^2 \end{pmatrix}.$$

أي:

$$R = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{i1}^2 & \cdots & \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{i1}.\hat{x}_{ij} & \dots & \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{i1}.\hat{x}_{ip} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ij}.\hat{x}_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ij}^2 & \cdots & \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ij}.\hat{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ip}.\hat{x}_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ip}.\hat{x}_{ij} & \dots & \frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ip}^2 \end{pmatrix}.$$

تكون عناصر القطر الرئيسي من الشكل:

$$\frac{1}{n}\sum_{i=1}^{n} \hat{x}_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{ij}-\bar{X}_j}{\sqrt{V_j}}\right)^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-\bar{X}_j\right)^2}{V_j} = \frac{V_j}{V_j} = 1.$$

أي معامل الارتباط بين X_j و X_j (بين المتغير ونفسه) .

والعناصر الأخرى من الشكل (مثلا السطر 1 والعمود j):

$$\frac{1}{n}\sum_{i=1}^{n} \hat{x}_{i1}.\hat{x}_{ij} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{i1}-\bar{X}_1}{\sqrt{V_1}}\right)\left(\frac{x_{ij}-\bar{X}_j}{\sqrt{V_j}}\right) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{X}_1).(x_{ij}-\bar{X}_j)}{\sqrt{V_1}.\sqrt{V_j}} = \frac{Cov(X_1,X_j)}{\sigma_1.\sigma_j} = r_{1j}.$$

أي معامل الارتباط بين X_1 و X_j.

ومنه يمكن كتابة R على الشكل المصفوفي التالي:

$$R = \begin{pmatrix} 1 & \cdots & r_{1j} & \cdots & r_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ r_{j1} & \dots & 1 & \cdots & r_{jp} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ r_{p1} & \cdots & r_{pj} & \dots & 1 \end{pmatrix}.$$

وبما أن $r_{ij} = r_{ji}$، فإنه يمكن كتابة مصفوفة الارتباط R على الشكل التالي:

$$R = \begin{pmatrix} 1 & \cdots & r_{1j} & \cdots & r_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ r_{1j} & \cdots & 1 & \cdots & r_{jp} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ r_{1p} & \cdots & r_{jp} & \cdots & 1 \end{pmatrix}.$$

ملاحظة: مصفوفة الارتباط هي مصفوفة متناظرة ومن الدرجة $p \times p$.

3. التفسير:

تسمح مصفوفة الارتباط بدراسة الارتباطات بين مختلف المتغيرات، ويمكن تفسير الارتباطات على نوعين:

✓ **الارتباط بشكل عام**:

- قيم الارتباطات يمكن أن تعطيها بشكل عام قوة أو ضعف الارتباط بين المتغيرات المدروسة.
- إذا كانت الارتباطات بشكل عام قوية، فإن الانتقال من p بعد إلى k بعد لن يفقدنا الكثير من المعلومات (فقدان المعلومات يكون في حده الأدنى).

✓ **الارتباط بشكل مفصل**:

- نقوم بدراسة الارتباط بين كل متغيرين بشكل مفصل.

ملاحظة: اختزال الأبعاد بطريقة تحليل المركبات الرئيسية هو ممكن في حالة وجود تكرار بين المتغيرات (Redondance) أي أن المتغيرات مترابطة فيما بينها، لكن إن كانت هذه المتغيرات مستقلة، فإن طريقة تحليل المركبات الرئيسية غير فعالة في اختزال الأبعاد.

هـ. حساب القيم الذاتية والأشعة الذاتية لمصفوفة الارتباط R:

1. حساب القيم الذاتية:

لحساب القيم الذاتية لمصفوفة الارتباط R، نعتمد على العلاقة التالية:

$Ru = \lambda u$، حيث $\vec{u}$ هو الشعاع الذاتي للمصفوفة R.

النظام يمكن كتابته كما يلي:

$$Ru = \lambda u \Leftrightarrow (R - \lambda I)u = 0$$

ولكي يقبل النظام حل غير صفري لابد أن يكون: $|\mathrm{R} - \lambda \mathrm{I}| = 0$

وهي العلاقة التي تسمح بحساب القيم الذاتية للمصفوفة R.

2. إيجاد الأشعة الذاتية:

كل قيمة ذاتية يقابلها شعاع ذاتي، ولتحديد الأشعة الذاتية نعتمد على العلاقة التالية:

- ✓ إذا كانت λ_1 هي القيمة الذاتية الأولى، فإن الشعاع الذاتي الأول u_1 يحدد من العلاقة:
$$(R - \lambda_1 I)u_1 = 0$$
ويجب أن يتحقق في هذا الشعاع: $u_1^t . u_1 = 1$.
- ✓ إذا كانت λ_2 هي القيمة الذاتية الثانية، فإن الشعاع الذاتي الثاني u_2 يحدد من العلاقة: $(R - \lambda_2 I)u_2 = 0$
ويجب أن يتحقق في هذا الشعاع: $u_2^t . u_2 = 1$، و $u_1^t . u_2 = 0$.
- ✓ إذا كانت λ_3 هي القيمة الذاتية الثالثة، فإن الشعاع الذاتي الثالث u_3 يحدد من العلاقة: $(R - \lambda_3 I)u_3 = 0$
ويجب أن يتحقق في هذا الشعاع: $u_3^t . u_3 = 1$، و $u_1^t . u_3 = 0$، و $u_2^t . u_3 = 0$.
- ✓ ... وهكذا لبقية القيم الذاتية والأشعة الذاتية.

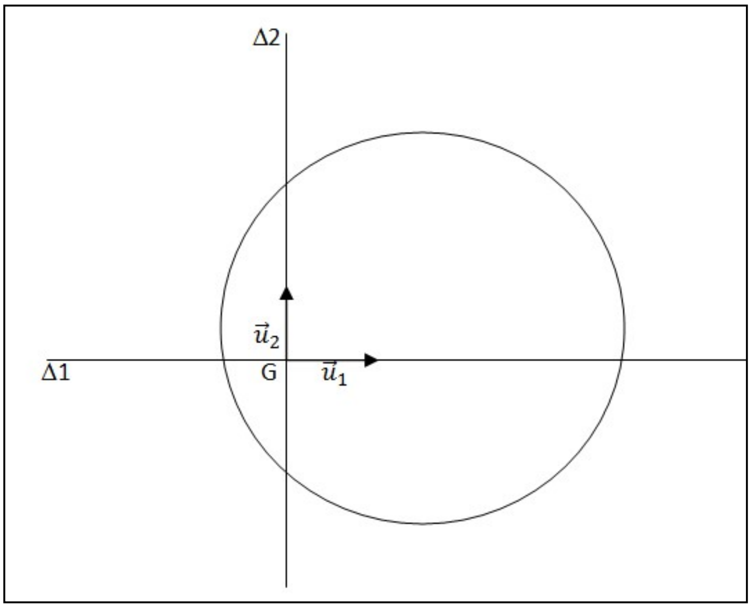
ملاحظة:

بعد ترتيب القيم الذاتية من الأكبر إلى الأصغر، القيمة الذاتية الأولى λ_1 تعطينا الشعاع الذاتي الأول u_1، والذي يعطينا بدوره المحور العاملي الأول (المستقيم: Δ_1)، وهذا الأخير يعطينا المركبة الرئيسية الأولى (C_1).

وبنفس الطريقة، القيمة الذاتية الثانية λ_2 تعطينا الشعاع الذاتي الثاني u_2، والذي يعطينا بدوره المحور العاملي الثاني (المستقيم: Δ_2)، وهذا الأخير يعطينا المركبة الرئيسية الثانية (C_2).

ثم المحورين الأوليين 1 و2 يشكلان لنا المستوي العاملي الأول (حيث يكون الشعاع الذاتي الأول متعامد مع الشعاع الذاتي الثاني، أي: $\vec{u}_1 \perp \vec{u}_2$، وهذا يعني: $\vec{u}_1 . \vec{u}_2 = 0$)

ونشير إلى أن المركبات الرئيسية هي توليفة خطية للمتغيرات الأصلية (مع مركبات الأشعة الذاتية)، وعدد المركبات الرئيسية هو نفسه عدد المتغيرات الأصلية، إلا أننا لا نستخدم إلا المركبة الأولى والثانية فقط، وفي بعض الأحيان نضيف الثالثة.



3. تشكيل جدول القيم الذاتية:

بعد حساب القيم الذاتية نقوم بترتيبها تنازليا (من الأكبر إلى الأصغر)، ثم نشكل جدول القيم الذاتية التالي:

الرقم	القيم الذاتية (λ_i)	النسبة	النسبة المئوية (%)	النسبة المئوية الصاعدة (%↑)
1	λ_1	$\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}*100$	$\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}*100$
2	λ_2	$\frac{\lambda_2}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_2}{\sum_{i=1}^{p}\lambda_i}*100$	$\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i}*100$
3	λ_3	$\frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i}*100$	$\frac{\lambda_1+\lambda_2+\lambda_3}{\sum_{i=1}^{p}\lambda_i}*100$
…	…	…	…	…
I	λ_i	$\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i}*100$	$\frac{\lambda_1+\lambda_2+\cdots+\lambda_i}{\sum_{i=1}^{p}\lambda_i}*100$
…	…	…	…	…
P	λ_p	$\frac{\lambda_p}{\sum_{i=1}^{p}\lambda_i}$	$\frac{\lambda_p}{\sum_{i=1}^{p}\lambda_i}*100$	$\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{p}\lambda_i}*100=100$
المجموع	$\sum_{i=1}^{p}\lambda_i=p$	1	100	

4. التفسير:

- ✓ كل قيمة ذاتية λ_i تمثل المساهمة المطلقة للمحور i في التباين الكلي (Inertie)
- ✓ كل نسبة $\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i}$ تمثل المساهمة النسبية للمحور i في التباين الكلي (Inertie)
- ✓ تمثل القيمة $\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}$ نسبة التباين الكلي (% d'inertie) المفسر بالمحور العاملي الأول (1er Axe Factoriel). وهي تمثل الأهمية النسبية للمركبة الرئيسية الأولى.
- ✓ تمثل القيمة $\frac{\lambda_2}{\sum_{i=1}^{p}\lambda_i}$ نسبة التباين الكلي المفسر بالمحور العاملي الثاني (2ème Axe Factoriel).
- ✓ تمثل القيمة $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i}$ نسبة التباين الكلي المفسر بالمستوي العاملي الأول (1er Plan Factoriel) (المحور الأول والمحور الثاني).
- ✓ تمثل القيمة $\frac{\lambda_1+\lambda_3}{\sum_{i=1}^{p}\lambda_i}$ نسبة التباين الكلي المفسر بالمستوي العاملي الثاني (2ème Plan Factoriel) (المحور الأول والمحور الثالث).
- ✓ إذا كان تحليل المركبات الرئيسية معياري (ACP Normé)، فإن $\sum_{i=1}^{p}\lambda_i=tr(R)=p$.

- ✓ إذا كان تحليل المركبات الرئيسية غير معياري (ACP Non normé)، فإن $\sum_{i=1}^{p}\lambda_i = tr(V) = \sum_{i=1}^{p}V_i$.

ملاحظة: إذا كان تحليل المركبات الرئيسية غير معياري بالاعتماد على المصفوفة التباين المشترك V، فإن القيم الذاتية والأشعة الذاتية تستخرج من المصفوفة V.

و. تحديد عدد المحاور التي تؤخذ في التحليل (Nombre d'axes à retenir):

هناك عدة معايير تسمح بتحديد عدد المحاور التي تؤخذ في التحليل، من بينها:

- ✓ **معيار كيزر Kaiser**: نأخذ كل القيم الذاتية الأكبر من 1 .
- ✓ **معيار نسبة المستوي العاملي الأول**: إذا كانت قيمة النسبة $100 * \frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i}$ أكبر من 80%، ففقدان المعلومات صغير، فلا داعي لإنشاء المستوي العاملي الثاني.
- ✓ **معيار نسبة المحور العاملي الثالث**: إذا كانت قيمة النسبة $100 * \frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i}$ أكبر من 15%، فلا بد من إنشاء المستوي العاملي الثاني.
- ✓ **معيار التمثيل البياني للقيم الذاتية**: بعد إنشاء التمثيل البياني بالأعمدة للقيم الذاتية نقوم برسم مستقيم يربط بين أكبر عدد من القيم الذاتية، والقيم الكبيرة التي لا تنتمي للمستقيم تمثل المحاور التي تؤخذ في التحليل.

أو نربط بين رؤوس الأعمدة بخطوط مستقيمة، وأين تشكل لنا شكل مرفق (Coude) ، فنتوقف عند تلك القيمة الذاتية (المرفق يؤخذ).

ز. الإسقاط على المحاور والمستويات:

1. إسقاط الأفراد:

- ➢ **حالة تحليل المركبات الرئيسية معياري:**

إذا كان تحليل المركبات الرئيسية معياري (بالاعتماد على المصفوفة R)، فإن العلاقة التي تسمح بإسقاط الأفراد على المحاور العاملية الجديدة هي العلاقة التالية:

$$\boldsymbol{F_{n\times k} = \widehat{X}_{n\times p}.u_{p\times k}}.$$

حيث: n هو عدد الأفراد. p عدد المتغيرات. و k عدد المحاور التي تؤخذ في التحليل. و $u_{p\times k}$ هي الأشعة الذاتية للمصفوفة $R_{p\times p}$. إذا أخذنا محورين في التحليل: $u_{p\times 2} = (u_1 \quad u_2)$.

- ➢ **حالة تحليل المركبات الرئيسية غير معياري:**

إذا كان تحليل المركبات الرئيسية غير معياري (بالاعتماد على المصفوفة V)، فإن العلاقة التي تسمح بإسقاط الأفراد على المحاور العاملية الجديدة هي العلاقة التالية:

$$\boldsymbol{F_{n\times k} = \widetilde{X}_{n\times p}.\, u_{p\times k}}.$$

حيث: n هو عدد الأفراد. p عدد المتغيرات. و k عدد المحاور التي تؤخذ في التحليل.

و $u_{p\times k}$ هي الأشعة الذاتية للمصفوفة $V_{p\times p}$.

2. إسقاط المتغيرات:

➢ **حالة تحليل المركبات الرئيسية معياري:**

بالنسبة للمتغيرات، نقوم بحساب المصفوفة A (بدلا من حساب المصفوفة R)، وفق العلاقة التالية:

$$A_{n\times n} = D_{n\times n}.\, \hat{X}_{n\times p}.\, \hat{X}^t_{p\times n}.$$

نعتمد على نفس مبدأ إسقاط الأفراد. والنظام الذي نبحث عن تعظيمه هو: $\begin{cases} Max(v^tAv) \\ S/C: v^tv = 1 \end{cases}$.

حيث v تمثل الأشعة الذاتية للمصفوفة A.

وباستخدام دالة لاقرانج (LANGRANGE) نصل إلى نفس النتيجة التي توصلنا إليها في حالة الأفراد، أي البحث عن أكبر القيم الذاتية (نرمز لها μ) للمصفوفة A. ثم نستنتج الأشعة الذاتية وهي: v.

➢ ثم نعتمد للإسقاط على العلاقة التالية: $\boldsymbol{P_{p\times k} = \hat{X}^t_{p\times n}.\, v_{n\times k}}$.

حيث: n هو عدد الأفراد. p عدد المتغيرات. و k عدد المحاور التي تؤخذ في التحليل.

إذا أخذنا محورين في التحليل فإن: $v_{n\times 2} = (v_1 \quad v_2)$. حيث: $v_1 \perp v_2$.

➢ كما يمكن الاعتماد على العبارة التالية للإسقاط: $\boldsymbol{P_{p\times k} = (P_1 \quad P_2 \quad \cdots \quad P_k)}$.

حيث $\boldsymbol{P_i = \sqrt{\lambda_i}.\, u_i}$ $\quad (i = 1,2,\dots,k)$.

حيث $\boldsymbol{\lambda_i}$ هي القيم الذاتية للمصفوفة $\boldsymbol{R}$، و $\boldsymbol{u_i}$ هي الأشعة الذاتية الموافقة لها.

➢ **حالة تحليل المركبات الرئيسية غير معياري:**

إذا كان تحليل المركبات الرئيسية غير معياري بالاعتماد على المصفوفة التباين المشترك V، فإن اسقاط المتغيرات يكون وفق نفس العبارة السابقة:

$$\boldsymbol{P_i = \sqrt{\lambda_i}.\, u_i} \quad (i = 1,2,\dots,k).$$

مع $\boldsymbol{\lambda_i}$ هي القيم الذاتية للمصفوفة $\boldsymbol{V}$، و $\boldsymbol{u_i}$ هي الأشعة الذاتية الموافقة لها.

لكن بعد ذلك، وللحصول على معاملات الارتباط بين المتغيرات والمحاور العاملية نقوم بقسمة الإسقاطات على الإنحرافات المعيارية للمتغيرات، أي: $\boldsymbol{P'_i = {P_i}/{\sqrt{V}}}$ حيث: $i = 1,2,\dots,k$.

3. عبارات الانتقال:

توجد بعض العبارات الرياضية تسمح لنا بالانتقال من الصيغ الرياضية لحساب إسقاط الأفراد إلى الصيغ الرياضية لحساب إسقاط المتغيرات أو العكس.

- عبارة أول انتقال: هي $v_i = \frac{1}{\sqrt{\lambda_i}}.\hat{X}.u_i$ حيث: $i = 1,2,\dots,k$.

وهذا يستلزم أن: $v_i = \frac{1}{\sqrt{\lambda_i}}.F_i$.

ومنه: $\boldsymbol{F_i = \sqrt{\lambda_i}.v_i}$ حيث: $i = 1,2,\dots,k$.

- عبارة ثاني انتقال: هي: $u_i = \frac{1}{\sqrt{\lambda_i}}.\hat{X}^t.v_i$

وهذا يستلزم أن: $u_i = \frac{1}{\sqrt{\lambda_i}}.P_i$

ومنه: $\boldsymbol{P_i = \sqrt{\lambda_i}.u_i}$ $(i = 1,2,\dots,k)$. فتكون: $\boldsymbol{P_{p\times k}} = (\boldsymbol{P_1} \quad \boldsymbol{P_2} \quad \cdots \quad \boldsymbol{P_k})$.

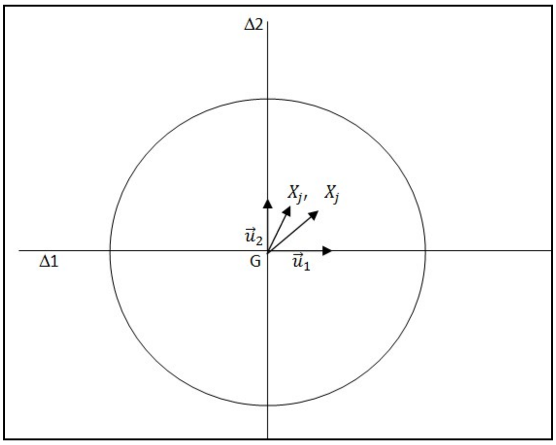
ح. التمثيل البياني للمتغيرات وللأفراد:

بعد تحديد المحاور والقيام بالإسقاط، نقوم بالتمثيل البياني للمتغيرات والأفراد.

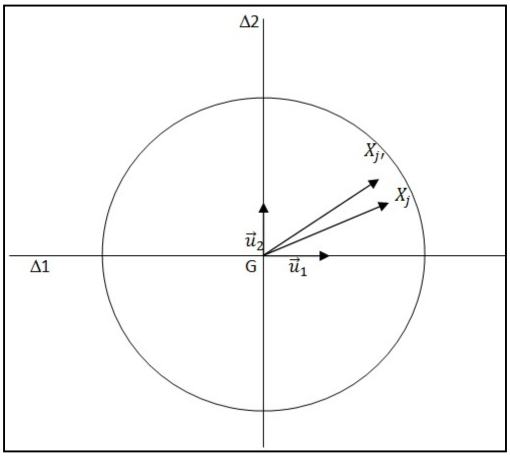
1. التمثيل البياني للمتغيرات:

تمثل المتغيرات في دائرة مثلثية، ومن هذه الدائرة المثلثية يمكن استنتاج ما يلي:

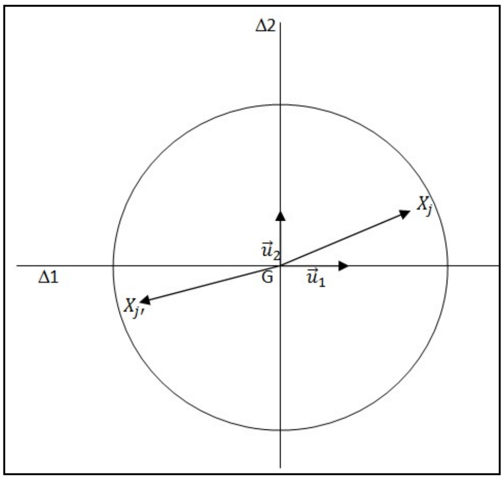
- المسافات بين المتغيرات تدل على الارتباط (corrélation).
- الارتباط بين المتغيرات يكون أكثر دلالة إحصائية كلما كانت النقاط الممثلة لهذه المتغيرات بعيدة عن المبدأ (G).
- إذا كان تمثيل المتغيرات في الدائرة المثلثية قريب من المبدأ (أي لم تكن قريبة من محيط الدائرة) ، **فلا يمكن تفسير علاقة الارتباط** بينهم، لأنهم يمثلون أفضل في محاور عاملية أخرى.



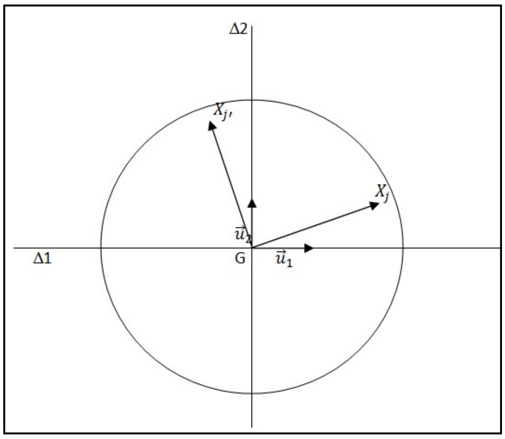
- إذا كان المتغير j قريب من المتغير 'j فهذا يدل على أن هذان المتغيران **مترابطان ايجابيا** (الزاوية المحصورة بينهما بالنسبة للمبدأ صغيرة وبالتالي تجب (cos) هذه الزاوية يكون قريب من 1).



- إذا كان المتغير j في الجهة المقابلة من المتغير 'j فهذا يدل على أن هذان المتغيران **مترابطان سلبيا** (الزاوية المحصورة بينهما بالنسبة للمبدأ كبيرة (قريبة من 180°) وبالتالي تجب (cos) هذه الزاوية يكون قريب من 1-).



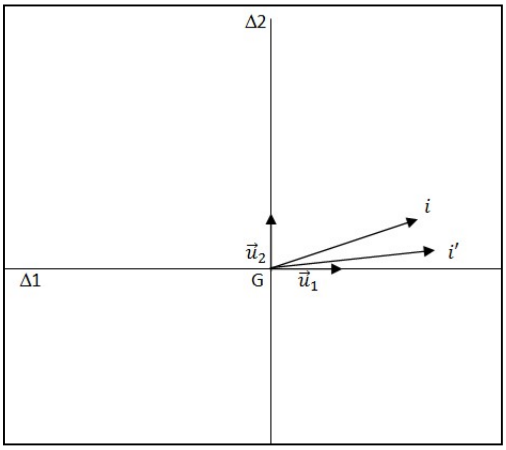
- إذا كان المتغير j بعيد عن المتغير j' فهذا يدل على أن هذان المتغيران **غير مترابطان** (الزاوية المحصورة بينهما بالنسبة للمبدأ قائمة (قريبة من 90°) وبالتالي تجب (cos) هذه الزاوية يكون قريب من 0).



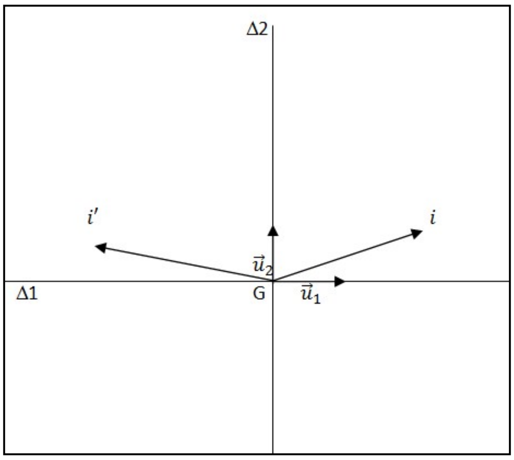
2. التمثيل البياني للأفراد:

من التمثيل البياني للأفراد يمكن استنتاج ما يلي:

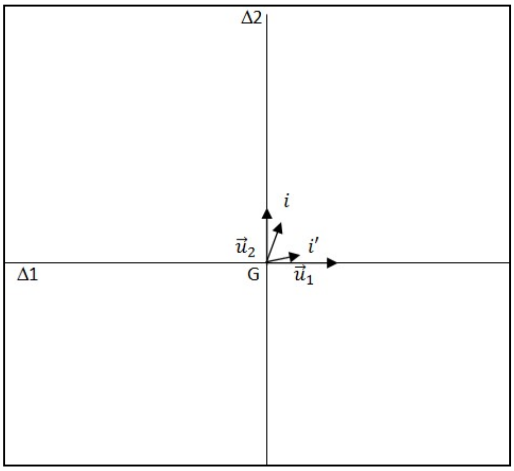
- المسافات بين الأفراد تدل على التشابه (Ressemblance).
- إذا كان الفرد i قريب من الفرد i' فهذا يدل على أن هذان الفردان متشابهان، أي يأخذان قيما متقاربة على كل المتغيرات بشكل عام.



- إذا كان الفرد i بعيد عن الفرد i' فهذا يدل على أن هذان الفردان مختلفان، أي يأخذان قيما متباعدة على كل المتغيرات بشكل عام.



- الأفراد القريبة من المبدأ (G) هي الأفراد التي قيمها عند جميع المتغيرات قريبة من المتوسط بشكل عام.

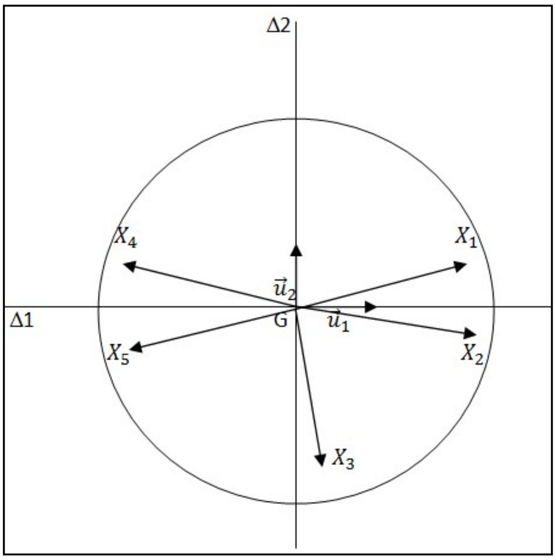


3. التمثيل البياني المشترك للأفراد والمتغيرات:

من التمثيل البياني المشترك للأفراد والمتغيرات يمكن استنتاج ما يلي:

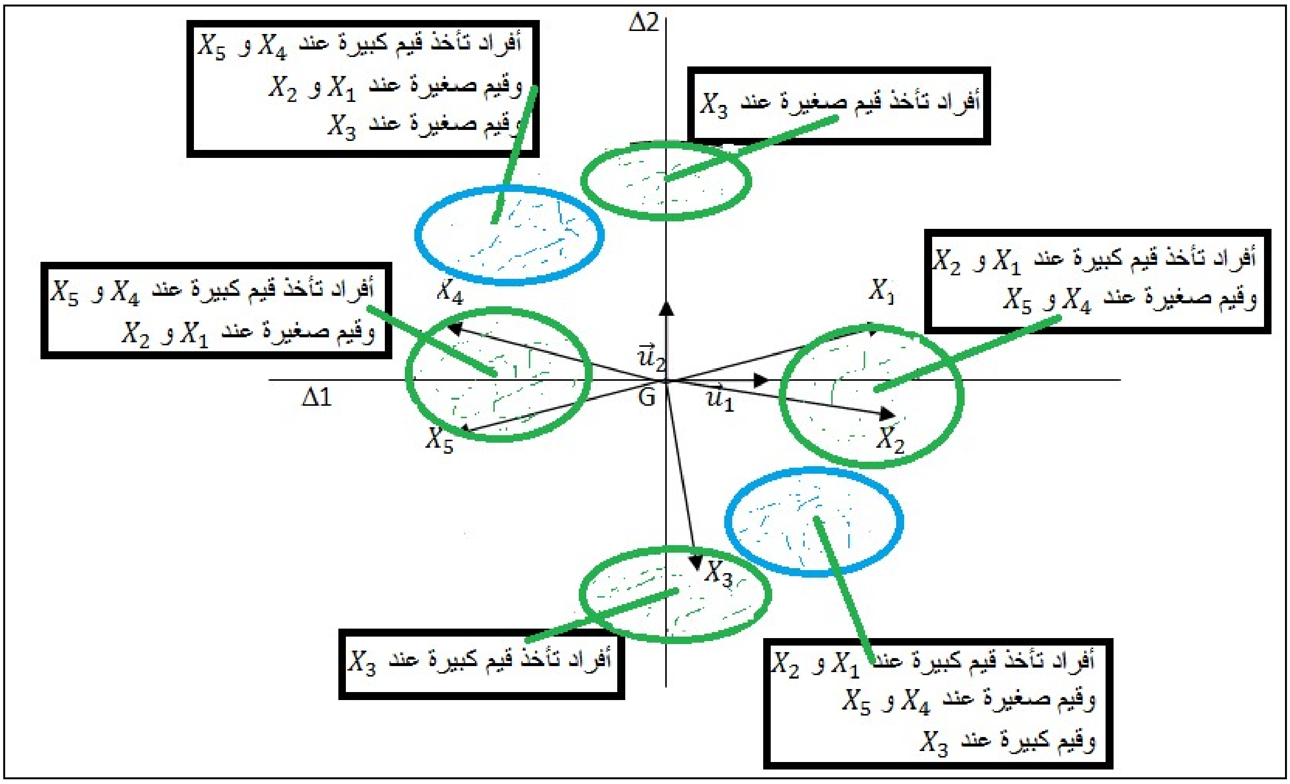
- الفرد سيكون قريب (نفس الجانب) من المتغيرات التي له قيم كبيرة عندها، وبالعكس سيكون بعيد (من الجانب المقابل) من المتغيرات التي له قيم صغيرة عندها.
- تقارب نقطة فرد من نقطة متغير في نفس المحور العاملي، تدل على أن هذا المتغير له دور كبير في تفسير سلوك ذلك الفرد بشكل إيجابي، أما إذا كانا متقابلين على نفس المحور العاملي، فهذا يدل على أن هذا المتغير له دور كبير في تفسير سلوك ذلك الفرد كذلك لكن بشكل سلبي (عكسي).
- ويمكن تلخيص العلاقة بين المتغيرات والأفراد وفق التمثيل الموالي:

ليكن لدينا خمس متغيرات ممثلة في الدائرة المثلثية التالية:



نستنتج من التمثيل السابق أن X_1 و X_2 مترابطان إيجابيا، كما أن X_4 و X_5 مترابطان إيجابيا كذلك.
لكن مثلا X_1 و X_5 مترابطان سلبيا. كما يتبين أنه لا يوجد ترابط بين كل هذه المتغيرات السابقة (X_1 و X_2 و X_4 و X_5) والمتغير X_3 (كل المتغيرات غير مرتبطة بالمتغير X_3).

تفسير العلاقات بين الأفراد والمتغيرات تبعا لتمركزهم على المحاور العاملية وفق الشكل الموالي:



مما يلاحظ كذلك أنه مثلا الأفراد الموجودة أسفل الشكل (عندX_3)، لا نفسر قيمها عند المتغيرات (X_1 و X_2 و X_4 و X_5) لأن هذه الأفراد تأخذ قيما متوسطة (valeurs moyennes) عند هذه المتغيرات (قيم قريبة من G).

3. تطبيق طريقة التحليل بالمركبات الرئيسية على برنامج SPSS

المثال التطبيقي يخص نفقات التجهيز (بالدينار الجزائري) لمختلف القطاعات (7 قطاعات)، الخاصة بالجزائر، خلال الفترة الممتدة من سنة 1985 إلى سنة 2018 (34 سنة).

أ. الخطوات على برنامج SPSS:

للقيام بـ ACP على برنامج SPSS نتبع الخطوات التالية:

1) بعد فتح برنامج SPSS على الصفحة التي تحتوي بيانات الدراسة، نذهب إلى شريط القوائم

2) في شريط القوائم نختار بالترتيب التالي:

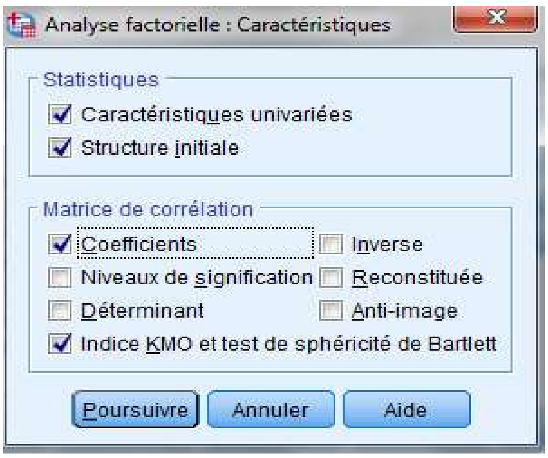
Analyse ⟹ Réduction des dimensions ⟹ Analyse factorielle

3) فيظهر لنا النافذة التالي:

4) من خلال هذه النافذة نختار:

❖ **المتغيرات (Variables)**: نقوم بنقل المتغيرات التي سنقوم بتطبيق طريقة ACP عليها من اليسار إلى مستطيل المتغيرات (Variables)

❖ **Caractéristiques**: عند الضغط عليها يظهر لنا النافذة التالي:



نظلل في هذه النافذة الخانات التالية:

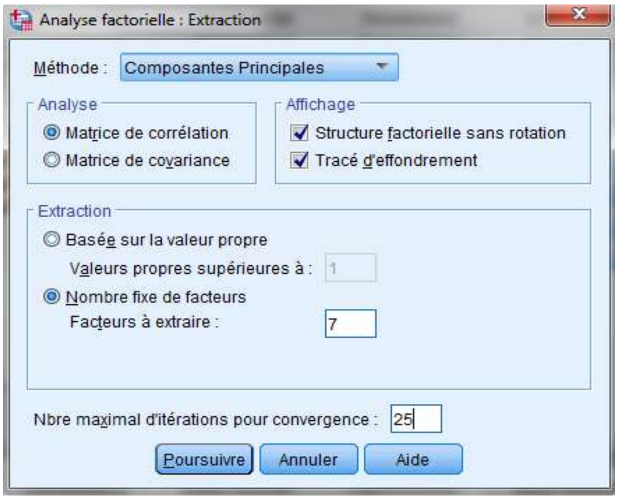
- الإحصائيات (Statistiques): نظلل على (Caractéristiques univariées) و/أو (Structure initiale)
- مصفوفة الارتباط (Matrice de corrélation): نظلل خاصة على المعاملات (Coefficients) ومؤشر Kaiser-Meyer-Olkin (Indice KMO et test de sphéricité de Bartlett)، وهو مؤشر لقياس جودة التحليل.

بالنسبة لمؤشر KMO، إذا كانت قيمة المؤشر أكبر من 0,7 فالتحليل صالح (valide)، وإذا كان بين 0,6 و 0,69 فالتحليل ضعيف الصلاحية، أما إذا كان أصغر من 0,5 فالتحليل غير صالح.

أما بالنسبة لاختبار Bartlett، فهو يعتمد هذا الاختبار على الفرضيتين التاليتين:

- الفرضية الصفرية: لا توجد ارتباطات معنوية بين المتغيرات ($R = I$).
- الفرضية البديلة: يوجد على الأقل ارتباط معنوي بين المتغيرات، أي مصفوفة الارتباط مختلفة عن مصفوفة الوحدة ($R \neq I$).

❖ **Extraction**: عند الضغط عليها يظهر لنا النافذة التالي:



نختار أو نظلل في هذه النافذة الخانات التالية:

- الطريقة (Méthode): ونختار منها طريقة المركبات الرئيسية (Composantes principales)
- تحليل (Analyse): ونختار هل يكون التحليل بمصفوفة الارتباط (Matrice de corrélation)، أو بمصفوفة التباين المشترك (Matrice de covariance)
- العرض Affichage:

- ✓ Structure factorielle sans rotation: نختارها لعرض المحاور بدون تدوير.
- ✓ Tracés d'effondrement: نختارها للعرض البياني للقيم الذاتية.

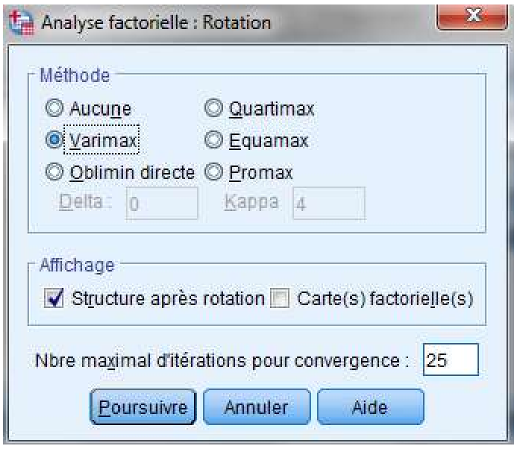
➢ استخراج (Extraction): نختار:

- ✓ إما Basée sur la valeur propre: نعتمد على القيم الذاتية لاختيار المحاور العاملية (مثلا نختار المحاور الموافقة للقيم الذاتية الأكبر من 1 فقط).
- ✓ أو Nombre fixe de facteurs: بتحديد عدد العوامل بغض النظر عن القيم الذاتية.

➢ Nombre maximal d'itérations pour convergence: تسمح لنا بتحديد القيمة القصوى لعدد التكرارات التي يقوم بها البرنامج (خوارزمية الحساب) لإيجاد الحل.

➢ ثم نضغط على متابعة (Poursuivre).

❖ **التدوير (Rotation)**: عند الضغط عليها يظهر لنا النافذة التالي:



نظلل في هذه النافذة الخانات التالية:

➢ الطريقة (Méthode): ونختار منها إحدى طرق تدوير العوامل، وهي:

- ✓ None: وهي الطريقة الافتراضية للتحليل
- ✓ Varimax: وهي طريقة تدوير عمودية (Orthogonale)، والتي تصغر عدد المتغيرات التي لها تغيرات كبيرة على كل محور عاملي.
- ✓ Oblimin directe: وهي طريقة تدوير غير عمودية بل مائلة (oblique)، فتكون الحلول المقترحة هي الأكثر مَيَالاً.
- ✓ Quartimax: وهي طريقة تدوير تختزل عدد المحاور العاملية المطلوبة لتفسير كل متغير.
- ✓ Equamax: وهي طريقة تدوير تجمع بين طريقة Varimax وطريقة Quartimax.

✓ Promax: وهي طريقة تدوير مائلة (oblique)، والتي تسمح للمحاور العاملية أن تكون مترابطة.

➢ العرض (Affichage): نختار منها:

✓ Structure après rotation: لعرض البنية الجديدة بعد التدوير.

✓ Carte(s) factorielle(s): لعرض الخريطة العاملية.

➢ ثم نضغط على متابعة (Poursuivre).

❖ **Scores**: عند الضغط عليها يظهر لنا النافذة التالي:

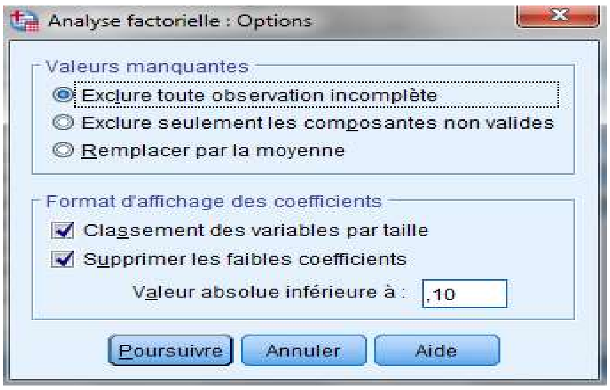
نظلل في هذه النافذة الخانات التالية:

➢ حفظ العوامل كمتغيرات (Enregistrer dans des variables): تسمح بإنشاء متغير جديد لكل محور عاملي. وتوجد ثلاث طرق مختلفة لحساب قيم هذه المتغيرات الجديدة: Régression، Bartlett، Anderson-Rubin.

➢ Afficher la matrice des scores factoriels: تسمح بعرض المعاملات المضروبة في المتغيرات لحساب قيم المتغيرات الجديدة (الموافقة للمحاور العاملية).

➢ ثم نضغط على متابعة (Poursuivre).

❖ **Options**: عند الضغط عليها يظهر لنا النافذة التالي:



نظلل من هذه النافذة الخانات التالية:

- القيم المفقودة (Valeurs manquantes): تسمح لنا تحديد طريقة التعامل مع القيم المفقودة في جدول البيانات، ونختار منها إحدى الخيارات التالية:
  - ✓ exclure toute observation incomplète: حذف (إقصاء) كل المشاهدات غير الكاملة.
  - ✓ Exclure seulement les composantes non valides: نحذف فقط المركبات غير الصالحة.
  - ✓ Remplacer par la moyenne: استبدال القيمة المفقودة بالمتوسط.
- شكل عرض المعاملات (Format d'affichage des coefficients): نختار:
  - ✓ Classement des variables par taille: لترتيب المتغيرات بحسب الحجم في المصفوفة
  - ✓ و/أو Supprimer les faibles coefficients: حذف المعاملات ذات القيم الصغيرة (إذا كانت قيمة المعامل أقل من 0,10 مثلا فإنها تحذف).
- ثم نضغط على متابعة (Poursuivre).

ب. مخرجات المثال التطبيقي والتعليق عليها:

1) الإحصائيات الوصفية (Statistiques descriptives):

Statistiques descriptives

	Moyenne	Ecart type	Analyse N
نفقات تجهيز قطاع الصناعة	2,04E+9	3,015E+9	34
نفقات تجهيز دعم الخدمات المنتجة	1,71E+10	1,94E+10	34
نفقات تجهيز المنشآت القاعدية الاجتماعية والثقافية	6,80E+10	9,05E+10	34
نفقات تجهيز قطاع التربية والتكوين	1,01E+11	1,17E+11	34
نفقات تجهيز قطاع الفلاحة والري	1,17E+11	1,25E+11	34
نفقات تجهيز دعم الحصول على سكن	1,22E+11	1,34E+11	34
نفقات تجهيز المنشآت القاعدية الاقتصادية والإدارية	3,11E+11	3,76E+11	34

يُظهر الجدول السابق المتوسطات والانحرافات المعيارية للمتغيرات السبعة المدروسة بالإضافة إلى عدد المشاهدات (34 مشاهدة).

يتبين من المتوسطات أن نفقات التجهيز لقطاع الصناعة هي الأضعف، وأكبر النفقات هي لقطاع المنشآت القاعدية الاقتصادية والإدارية.

كما يتضح من الانحرافات المعيارية أن نفقات التجهيز لقطاع المنشآت القاعدية الاقتصادية والإدارية هي المسؤولة عن تشتت (تبعثر) الأفراد في سحابة النقاط (المستوي العاملي)، لأن لها أكبر تباين.

2) مصفوفة الارتباط (**Matrice de corrélation**):

Matrice de corrélation

		نفقات تجهيز قطاع الصناعة	نفقات تجهيز قطاع الفلاحة والري	نفقات تجهيز دعم الخدمات المنتجة	نفقات تجهيز المنشآت القاعدية الاقتصادية والإدارية	نفقات تجهيز قطاع التربية والتكوين	نفقات تجهيز المنشآت القاعدية الاجتماعية والثقافية	نفقات تجهيز دعم الحصول على سكن
Corrélation	نفقات تجهيز قطاع الصناعة	1,000	,379	,348	,519	,165	,211	,544
	نفقات تجهيز قطاع الفلاحة والري	,379	1,000	,761	,935	,864	,843	,836
	نفقات تجهيز دعم الخدمات المنتجة	,348	,761	1,000	,795	,686	,680	,831
	نفقات تجهيز المنشآت القاعدية الاقتصادية والإدارية	,519	,935	,795	1,000	,856	,875	,846
	نفقات تجهيز قطاع التربية والتكوين	,165	,864	,686	,856	1,000	,982	,656
	نفقات تجهيز المنشآت القاعدية الاجتماعية والثقافية	,211	,843	,680	,875	,982	1,000	,636
	نفقات تجهيز دعم الحصول على سكن	,544	,836	,831	,846	,656	,636	1,000

تسمح مصفوفة الارتباط بدراسة الارتباطات بين مختلف المتغيرات، ويمكن تفسير الارتباطات على نوعين:

✓ **الارتباط بشكل عام:**

- يتضح أن الارتباطات بشكل عام قوية (غالبها أكبر من 0,6)، فيكون الانتقال من 7 أبعاد إلى 2 أو 3 بُعد جيد، حيث أن هذا الانتقال لن يؤدي إلى فقدان معلومات كثيرة.
- بما أن الارتباطات بشكل عام قوية بين المتغيرات، فإن المستوي العاملي الأول (المحور 1 والمحور 2) سيكون كافيا للاحتفاظ بأكبر قدر من المعلومات.

✓ **الارتباط بشكل مفصل:**

- نلاحظ وجود ارتباط قوي موجب بين غالب المتغيرات، فكل متغيرين مترابطين بقوة، وهذا ما يجعل هذه المتغيرات تتجمع في نفس المحور العاملي، ماعدا متغير نفقات التجهيز لقطاع الصناعة، فهو مرتبط بدرجة ضعيفة مع كل المتغيرات الأخرى، فيكون ممثل في محور عاملي أخر.

3) مؤشر **KMO** واختبار **Bartlett** ونوعية التمثيل:

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,765
Test de sphéricité de Bartlett	Khi-deux approx.	333,730
	ddl	21
	Signification	,000

✓ مؤشر KMO: يقيس هذا المؤشر دقة المعاينة (précision de l'échantillonnage).
نلاحظ أن قيمة مؤشر KMO والتي تساوي 0,765 أكبر من 0,6، وهذا يعني أن القياس ذو جودة جيدة. أي أن هذا التحليل العاملي قام باختزال العوامل بجودة عالية.

✓ اختبار Bartlett: يعتمد هذا الاختبار على الفرضية الصفرية: لا توجد ارتباطات معنوية بين المتغيرات ($R = I$).

الاحتمال (Signification=0,00) أصغر من 0,05، وبالتالي نرفض الفرضية الصفرية. وهذا ما يدل على أنه يوجد على الأقل ارتباط معنوي بين المتغيرات.

✓ نوعية التمثيل (Qualité de représentation):

Qualités de représentation

	Initiales	Extraction
نفقات تجهيز قطاع الصناعة	1,000	,910
نفقات تجهيز قطاع الفلاحة والري	1,000	,911
نفقات تجهيز دعم الخدمات المنتجة	1,000	,741
نفقات تجهيز المنشآت القاعدية الاقتصادية والإدارية	1,000	,956
نفقات تجهيز قطاع التربية والتكوين	1,000	,950
نفقات تجهيز المنشآت القاعدية الاجتماعية والثقافية	1,000	,924
نفقات تجهيز دعم الحصول على سكن	1,000	,860

Méthode d'extraction : Analyse en composantes principales.

يتبين من الجدول أن قيم جميع المتغيرات بعد تطبيق ACP (العمود extraction) أكبر من 0,5 وقريبة من 1، وهذا ما يدل على أن الإسقاط بطريقة المركبات الرئيسية ذو نوعية جيدة.

ملاحظة1: إذا كان أحد المتغيرات غير ممثل بجودة جيدة (القيمة أصغر من 0,5)، فمن الأفضل حذف المتغير، ثم إعادة تطبيق طريقة ACP.

ملاحظة2: ليعطي ACP نتائج جيدة (أو بعبارة أخرى يمكن تطبيق ACP على البيانات بشكل جيد) يجب التأكد من ثلاث شروط:

- يجب أن تكون قيم غالب معاملات الارتباط في مصفوفة الارتباط أكبر من 0,5.
- يجب أن تكون قيمة مؤشر Indice KMO أكبر من 0,5.
- يجب أن يكون اختبار Bartlett معنوي (Sign < 0,5).

4) التباين الكلي المفسر (Variance totale expliquée)

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	5,222	74,597	74,597	5,222	74,597	74,597	4,491	64,162	64,162
2	1,030	14,717	89,314	1,030	14,717	89,314	1,761	25,151	89,314
3	,431	6,156	95,470						
4	,181	2,587	98,056						
5	,085	1,216	99,273						
6	,042	,594	99,867						
7	,009	,133	100,000						

Méthode d'extraction : Analyse en composantes principales.

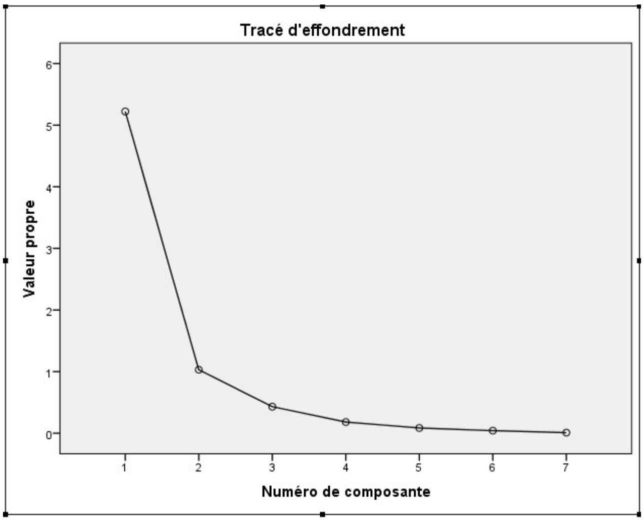
نستنتج من خلال الجدول السابق ما يلي:

- ✓ القيم الذاتية هي: $\lambda_1 = 5{,}222$، $\lambda_2 = 1{,}030$، $\lambda_3 = 0{,}431$، ...
- ✓ النسبة $\frac{\lambda_1}{\sum_{i=1}^{7}\lambda_i} * 100 = 74{,}597$ تمثل المساهمة النسبية للمحور 1 في التباين الكلي. أو نسبة التباين الكلي المفسر بالمحور العاملي الأول. أو الأهمية النسبية للمركبة الرئيسية الأولى.
- ✓ تمثل القيمة $\frac{\lambda_2}{\sum_{i=1}^{7}\lambda_i} * 100 = 14{,}717$ نسبة التباين الكلي المفسر بالمحور العاملي الثاني.
- ✓ تمثل القيمة $\frac{\lambda_3}{\sum_{i=1}^{7}\lambda_i} * 100 = 6{,}156$ نسبة التباين الكلي المفسر بالمحور العاملي الثالث.
- ✓ تمثل القيمة $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{7}\lambda_i} * 100 = 89{,}314$ نسبة التباين الكلي المفسر بالمستوي العاملي الأول.
- ✓ تمثل القيمة $\frac{\lambda_1+\lambda_3}{\sum_{i=1}^{7}\lambda_i} * 100 = 80{,}757$ نسبة التباين الكلي المفسر بالمستوي العاملي الثاني.
- ✓ تمثل القيمة $\frac{\lambda_1+\lambda_2+\lambda_3}{\sum_{i=1}^{7}\lambda_i} * 100 = 95{,}470$ نسبة التباين الكلي المفسر بالمحاور العاملية الثلاثة الأولى.
- ✓ بما أن تحليل المركبات الرئيسية معياري (ACP Normé)، فإن: $\sum_{i=1}^{7}\lambda_i = tr(R) = 7$.

5) تحديد عدد المحاور التي تؤخذ في التحليل (Nombre d’axes à retenir):

هناك عدة معايير تسمح بتحديد عدد المحاور التي تؤخذ في التحليل، من بينها:

- ✓ **معيار كيزر Kaiser**: لو اعتمدنا على هذا المعيار فسنأخذ محورين فقط 1 و2 (القيم الذاتية الأكبر من 1).
- ✓ **معيار نسبة المستوي العاملي الأول:** بما أن قيمة النسبة $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{7}\lambda_i} * 100 = 89{,}314$ أكبر من 80%، ففقدان المعلومات صغير، فلا داعي لإنشاء المستوي العاملي الثاني (المحورين العامليين الأول والثاني يحافظان على أكثر من 89% من المعلومات الموجودة في البيانات الأولية).
- ✓ **معيار نسبة المحور العاملي الثالث**: بما أن قيمة النسبة $\frac{\lambda_3}{\sum_{i=1}^{7}\lambda_i} * 100 = 6{,}156$ أقل من 15%، فلا داعي لإنشاء المستوي العاملي الثاني.
- ✓ **معيار التمثيل البياني للقيم الذاتية:**



لو قمنا برسم مستقيم يربط بين أكبر عدد من القيم الذاتية، تكون القيمتين الأولى والثانية هي التي لا تنتمي للمستقيم. كما نلاحظ تشكل شكل مرفق (Coude) عند القيمة الذاتية الثانية، فنتوقف عند هذه القيمة الذاتية.

6) إسقاط المتغيرات:

نجد قيم الإسقاط للمتغيرات في الجدول التالي (مصفوفة المركبات):

Rotation de la matrice des composantes[a]

	Composante 1	Composante 2
نفقات تجهيز قطاع التربية والتكوين	,974	
نفقات تجهيز قطاع المنشآت القاعدية الاجتماعية والثقافية	,959	
نفقات تجهيز قطاع الفلاحة والري	,893	,337
نفقات تجهيز قطاع المنشآت القاعدية الاقتصادية والإدارية	,867	,453
نفقات تجهيز قطاع دعم الخدمات المنتجة	,762	,401
نفقات تجهيز قطاع دعم الحصول على سكن	,697	,611
نفقات تجهيز قطاع الصناعة		,950

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.

a. Convergence de la rotation dans 3 itérations.

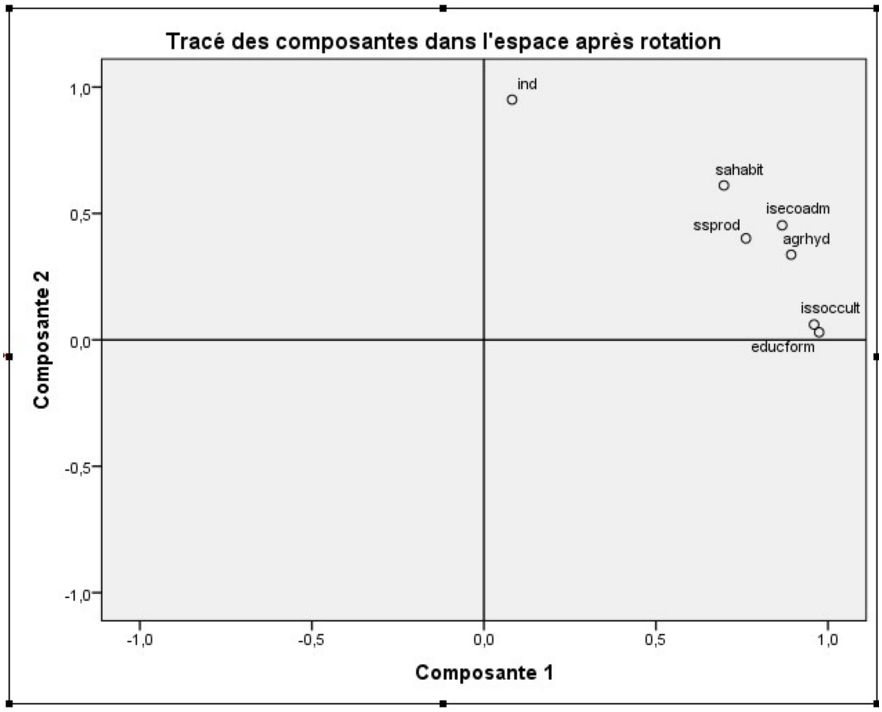
يتبين من الجدول السابق أن كل المتغيرات (نفقات تجهيز قطاع التربية والتكوين، نفقات تجهيز قطاع المنشآت القاعدية الاجتماعية والثقافية، نفقات تجهيز قطاع الفلاحة والري، نفقات تجهيز قطاع المنشآت القاعدية الاقتصادية والإدارية، نفقات تجهيز قطاع دعم الخدمات المنتجة، نفقات تجهيز قطاع دعم الحصول على سكن) تمثل بشكل أفضل في المحور العاملي 1، أي أكثر ارتباطا بالمحور العاملي 1 (أكبر قيمة لها تأخذها في المحور العاملي 1)، ماعدا متغير "نفقات تجهيز قطاع الصناعة" فإنه يمثل بشكل أفضل في المحور العاملي 2.

ومما سبق نستنتج أن:

- ❖ المركبة الرئيسية الأولى لها علاقة قوية مع 6 متغيرات من أصل 7 متغيرات.
- ❖ المركبة الرئيسية الثانية لها علاقة قوية مع متغير واحد. وعلاقة متوسطة مع متغيرين.

ملاحظة: أيّ مركبة رئيسية (محور عاملي) له قيمة ارتباط أكبر من 0,30 مع أكثر من 3 متغيرات، فهي مركبة جيدة يجب أخذها في التحليل العاملي.

7) التمثيل البياني للمتغيرات بعد الإسقاط:



نلاحظ أن غالب المتغيرات مرتبطة ارتباطا قويا بالمحور العاملي 1، ماعدا متغير "نفقات تجهيز قطاع الصناعة" (ind)، فهو غير مرتبط بالمحور العاملي 1، لكنه يرتبط ارتباطا قويا بالمحور العاملي 2. وبما أن المحورين العامليين متعامدان، فإن هذا المتغير (ind) مستقل (غير مرتبط) عن المتغيرات الأخرى.

كما يتأكد من التمثيل البياني العلاقة الموجبة لهذه المتغيرات مع المحاور المرتبطة بها، فكل المتغيرات في الجانب الموجب للمحورين.

المسافات بين المتغيرات (الارتباط): بما أن المسافة بين المتغيرين نفقات التربية والتكوين (educform) ونفقات البنى التحتية الاجتماعية والثقافية (issoccult) صغيرة، فهذا يدل على أنه هناك ارتباط كبير في جميع السنوات (بشكل عام) لهذين المتغيرين (معامل الارتباط يساوي 0,98). وعلى العكس من ذلك، فبما أن المسافة بين نفقات التربية والتكوين (educform) ونفقات القطاع الصناعي (ind) كبيرة، فهذا يدل على ضعف الارتباط بينهما في جميع السنوات (بشكل عام) لهذين المتغيرين (معامل الارتباط يساوي 0,16).

8) إسقاط الأفراد (السنوات):

عند حفظ العوامل كمتغيرات يقوم برنامج SPSS بإنشاء متغير جديد لكل محور عاملي. فنجد أن إسقاطات الأفراد على المحورين الأول والثاني محفوظة في نافذة البيانات.

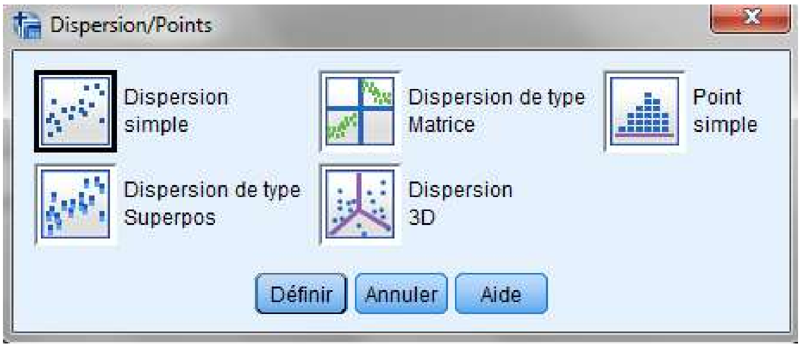
9) التمثيل البياني للأفراد (السنوات) بعد الإسقاط:

برنامج SPSS لا يعرض لنا التمثيل البياني لإسقاطات الأفراد بشكل تلقائي مع المخرجات، بل يجب اتباع الخطوات التالية على برنامج:

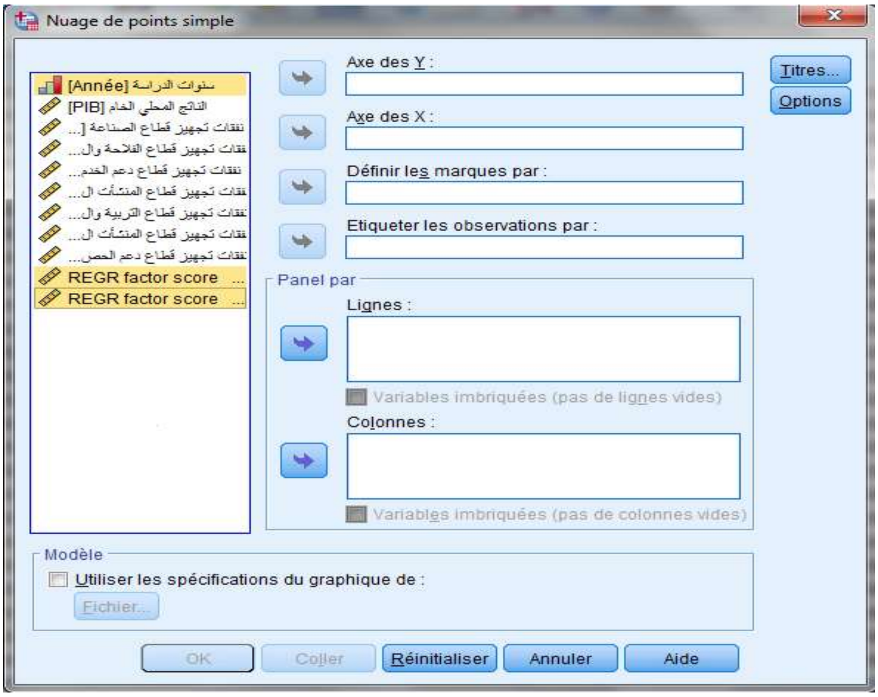
- في نافذة مخرجات الـ ACP على برنامج SPSS، نذهب إلى شريط القوائم
- في شريط القوائم نختار بالترتيب التالي:

Graphiques ⟹ Boites de dialogue ancienne version ⟹ Dispersion/points

- فيظهر لنا النافذة التالي:

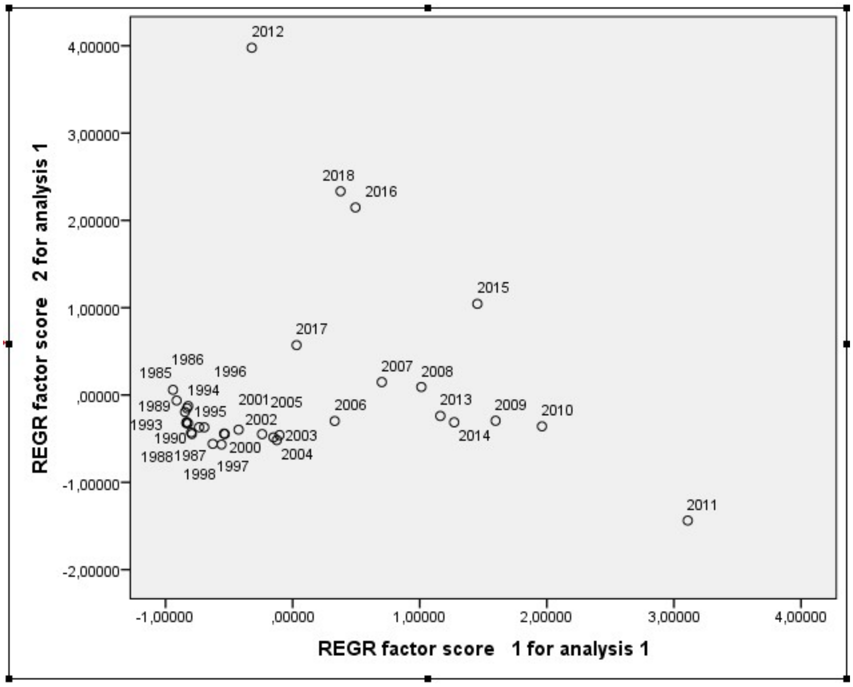


- من خلال هذه النافذة نختار "Dispersion simple"، ثم نضغط على Définir
- فيظهر لنا النافذة التالي:

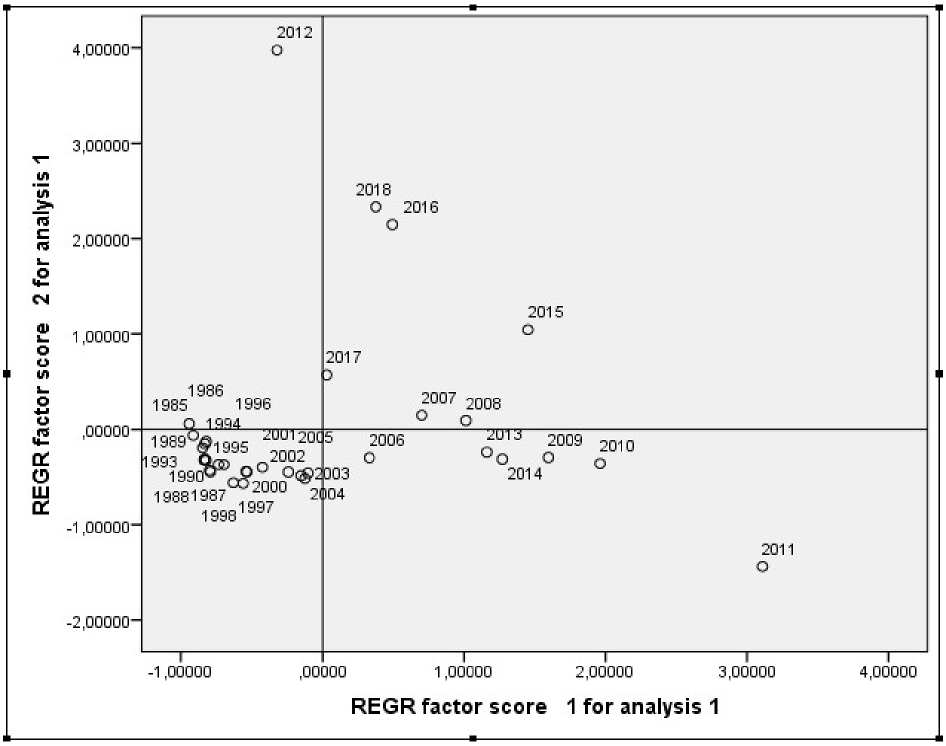


❖ من خلال هذه النافذة نختار:

- على محور الفواصل (Axe des X): ندخل المحور العاملي الأول (FAC1)
- على محور التراتيب (Axe des Y): ندخل المحور العاملي الثاني (FAC2)
- المشاهدات (Etiqueter les observations par): ندخل سنوات الدراسة
- العناوين (Titres): عنوان التمثيل البياني والعناوين الفرعية
- الخيارات (Options): نظلل على عرض التمثيل البياني بأسماء المشاهدات (Afficher le graphique avec les libellés d'observations).
- عند الضغط على "OK" يظهر لنا التمثيل البياني التالي:



نضغط مرتين على التمثيل البياني لإدخال إضافات عليه: أين نقوم بإضافة المحور الأفقي والمحور العمودي عند الصفر (0). فيصبح التمثيل البياني بالشكل التالي:



يتبين من التمثيل البياني للسنوات (الأفراد) أن:

- معظم السنوات الأخيرة للدراسة (2006 إلى 2015 –ماعدا سنة 2012–) مرتبطة ارتباطا قويا وموجبا بالمحور العاملي الأول.

- غالب السنوات الأولى للدراسة (من 1985 إلى 2000) مرتبطة ارتباطا قويا وسالبا بالمحور العاملي الأول.
- السنوات من 2016 إلى 2018 -بالإضافة إلى سنة 2012- مرتبطة ارتباطا قويا وموجبا بالمحور العاملي الثاني.
- السنوات من 2000 إلى 2005 مرتبطة ارتباطا قويا وسالبا بالمحور العاملي الثاني.
- المسافات بين الأفراد (التشابه): يمكن تفسير العلاقات بين الأفراد على هذا النحو:
  - ✓ بما أن المسافة بين السنتين 2016 و 2018 صغيرة، فهذا يدل على وجود تشابه كبير للنفقات (بشكل عام) في هاتين السنتين.
  - ✓ وعلى العكس من ذلك، فبما أن المسافة بين 2001 و1985 كبيرة، فهذا يدل على الاختلاف الكبير في النفقات (بشكل عام) لهاتين السنتين.

10) التمثيل البياني المشترك للأفراد والمتغيرات بعد الإسقاط:

برنامج SPSS لا يعرض التمثيل البياني المشترك للأفراد والمتغيرات بعد الإسقاط (في تمثيل بياني واحد)، لكن يمكن التفسير بالاعتماد على التمثيلين البيانيين المستقلين للمتغيرات والأفراد على النحو التالي:

- نفقات التجهيز لمعظم القطاعات كانت مرتفعة في السنوات الممتدة من 2006 إلى 2015، ومنخفضة في السنوات من 1985 إلى 2000، لارتباطها القوي إما بشكل موجب أو سالب بالمحور العاملي الأول.
- نفقات التجهيز لقطاع الصناعة كان مرتفعا في السنوات الممتدة من 2016 إلى 2018-بالإضافة إلى سنة 2012-، ومنخفضة في السنوات من 2000 إلى 2005، لارتباطها القوي إما بشكل موجب أو سالب بالمحور العاملي الثاني.

4. تطبيق طريقة التحليل بالمركبات الرئيسية على برنامج XL-STAT

سنحاول في هذا العنصر تقديم أهم الخطوات المتبعة على برنامج XL-STAT للقيام بتحليل مركبات رئيسية، بالإضافة إلى مثال تطبيقي.

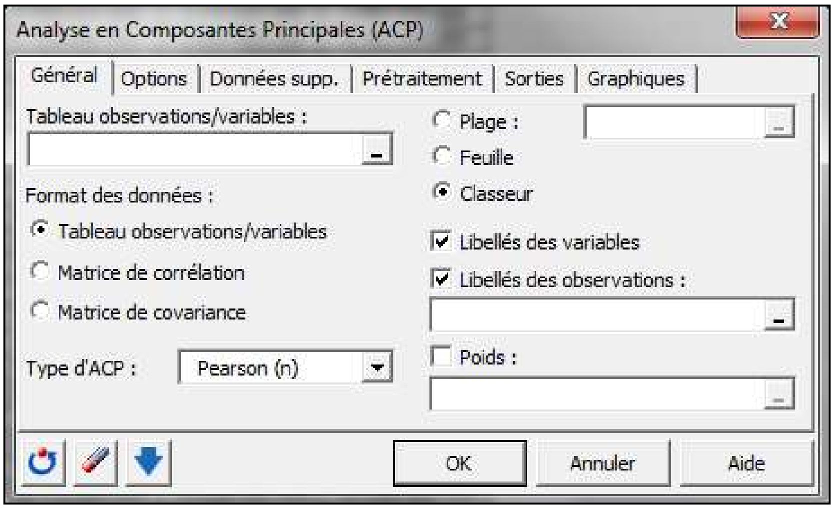
أ. الخطوات على برنامج XL-STAT:

للقيام بـ ACP على برنامج XL-STAT نتبع الخطوات التالية:

1) بعد فتح برنامج XL-STAT نقوم بإدخال بيانات الدراسة، ثم نذهب إلى شريط القوائم
2) في شريط القوائم نختار بالترتيب التالي:

XL-STAT ⟹ Analyse des données ⟹ Analyse en composantes principales

3) فيظهر لنا النافذة التالي:

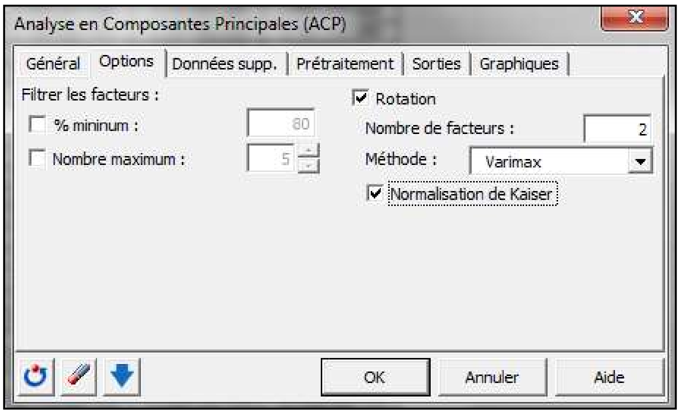


4) من خلال هذه النافذة نختار:

❖ **من عام Général**: نقوم باختيار:

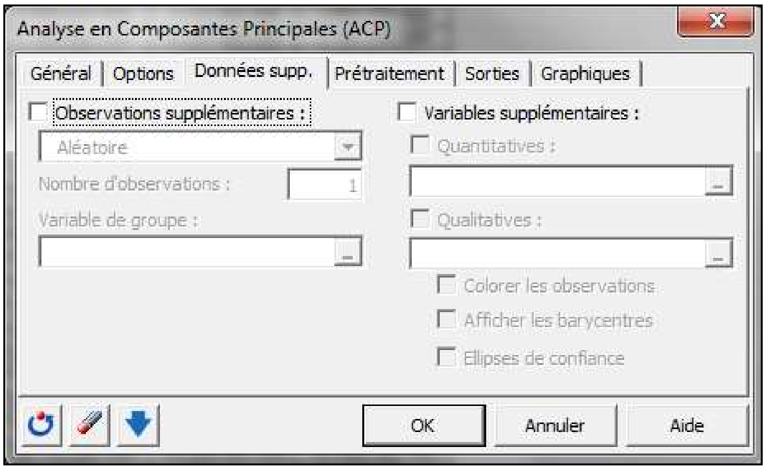
- مكان حفظ المخرجات (plage، feuille، classeur)
- شكل البيانات (Format des données): هل هو جدول البيانات والمتغيرات X، أو مصفوفة الارتباط R، أو مصفوفة التباين المشترك V.
- نوع ACP (Type ACP): pearson n، pearson n-1، covariance، spearman، ....
- تعيين أسماء المتغيرات (Libellés des variables)
- تعيين أسماء المشاهدات (Libellés des observations)
- الأوزان (poids) -إن وجدت-

❖ **من خيارات Option**: نقوم باختيار:



- ➢ Filtrer les facteurs: نختار إما % الحد الأدنى (minimum) (مثلا المحاور التي تضمن أكثر من 80% من التباين الكلي) ، أو الحد الأعلى لعدد العوامل (maximum) التي تؤخذ في التحليل.
- ➢ **التدوير (Rotation)**: نظلل على Rotation، ثم نحدد عدد العوامل، ثم الطريقة: ونختار منها إحدى طرق التدوير (Varimax، Oblimin، Quartimax، Equamax، Promax) (نلاحظ عدم وجود اختيار None مقارنة ببرنامج SPSS).
- ➢ كما نظلل Normalisation de Kaiser

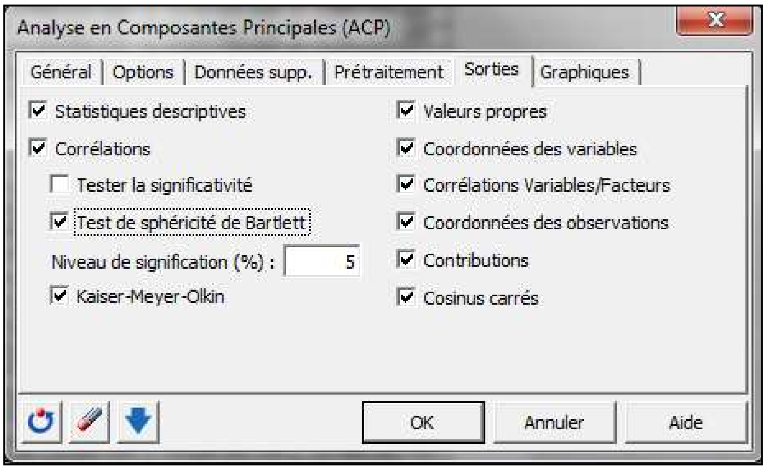
❖ **بيانات إضافية (Données supplémentaires)**: تسمح هذه النافذة بإدخال البيانات الإضافة في حالة وجودها، إما أفراد أو متغيرات (كمية أو كيفية).



❖ **معالجة أولية (prétraitement)**: عند الضغط عليها يظهر لنا النافذة التالي:

وهي نافذة خاصة بمعالجة البيانات المفقودة (Données manquantes) بالإضافة إلى المجموعات (Groupes)

❖ **المخرجات (Sortie)**: عند الضغط عليها يظهر لنا النافذة التالي:



نظلل على المخرجات التي نرغب في عرضها (الإحصائيات الوصفية، اختبار KMO، الارتباطات، ...).

❖ **التمثيلات البيانية (Graphiques)**: عند الضغط عليها يظهر لنا النافذة التالي:

في هذه النافذة ندخل التعليمات الخاصة بالتمثيلات البيانية للمتغيرات والأفراد والتمثيل البياني المشترك بينهما (وهو غير متوفر في برنامج SPSS).

بعد الانتهاء من إدخال كل هذه الاختيارات والتعليمات، نضغط على OK، فتظهر لنا المخرجات.

ب. مخرجات ACP على برنامج XL-STAT:

نطبق طريقة ACP على برنامج XL-STAT بنفس المثال المطبق على برنامج SPSS. وبالنسبة للتعليق على الجداول والأشكال، فهو نفس التعليق المذكور في العنصر السابق الخاص ببرنامج SPSS، فلا داعي لتكرار ذلك.

1) الإحصائيات الوصفية (**Statistiques descriptives**):

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Industrie	34	100000000	15567000000	2044970558,8235	3015121452,7452
Ag hyd	34	6825000000	393748000000	117481867647,0590	124800294294,4220
SS Prodct	34	135000000	80309269000	17087943117,6471	19438738232,5242
infr éco adm	34	5085000000	1095942000000	311234023823,5290	376389930257,6410
Educ form	34	7100000000	540754000000	100621829294,1180	116781031536,3930
Infr socio cult	34	2703000000	363062000000	67962501882,3529	90536380849,1481
SA habit	34	343000000	469781674000	121911367470,5880	134374519559,4100

2) مصفوفة الارتباط (**Matrice de corrélation**):

Variables	Industrie	Ag hyd	SS Prodct	infr éco adm	Educ form	Infr socio cult	SA habit
Industrie	**1**	0,3788	0,3482	0,5188	0,1653	0,2109	0,5443
Ag hyd	0,3788	**1**	0,7609	0,9354	0,8635	0,8426	0,8356
SS Prodct	0,3482	0,7609	**1**	0,7945	0,6865	0,6800	0,8306
infr éco adm	0,5188	0,9354	0,7945	**1**	0,8557	0,8752	0,8463
Educ form	0,1653	0,8635	0,6865	0,8557	**1**	0,9816	0,6562
Infr socio cult	0,2109	0,8426	0,6800	0,8752	0,9816	**1**	0,6361
SA habit	0,5443	0,8356	0,8306	0,8463	0,6562	0,6361	**1**

3) مؤشر **KMO** واختبار **Bartlett** ونوعية التمثيل:

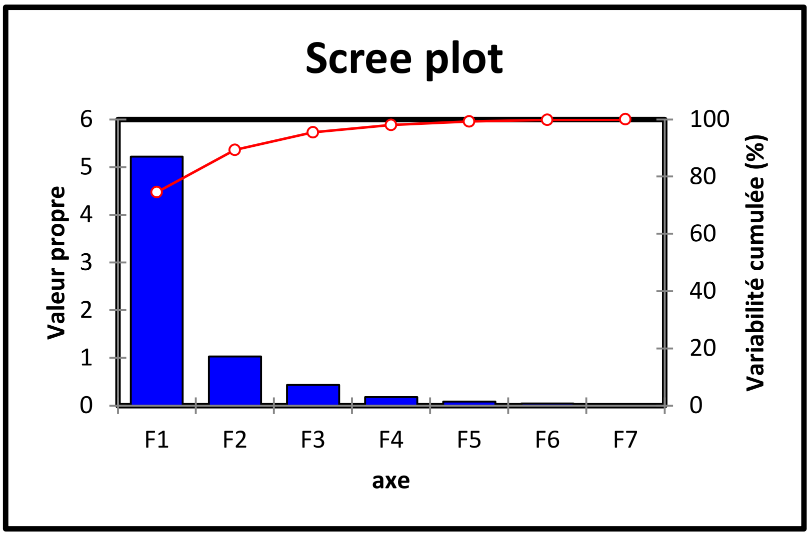
Khi² (Valeur observée)	333,7299
Khi² (Valeur critique)	32,6706

DDL	21
p-value	< 0,0001
alpha	0,05
KMO	0,7654

4) القيم الذاتية (Valeurs propres):

	F1	F2	F3	F4	F5	F6	F7
Valeur propre	5,2218	1,0302	0,4309	0,1811	0,0851	0,0416	0,0093
Variabilité (%)	74,5968	14,7169	6,1560	2,5868	1,2162	0,5941	0,1332
% cumulé	74,5968	89,3137	95,4697	98,0564	99,2726	99,8668	100,0000

التمثيل البياني للقيم الذاتية:



5) الأشعة الذاتية (Vecteurs propres): (وهي غير متوفرة على برنامج SPSS)

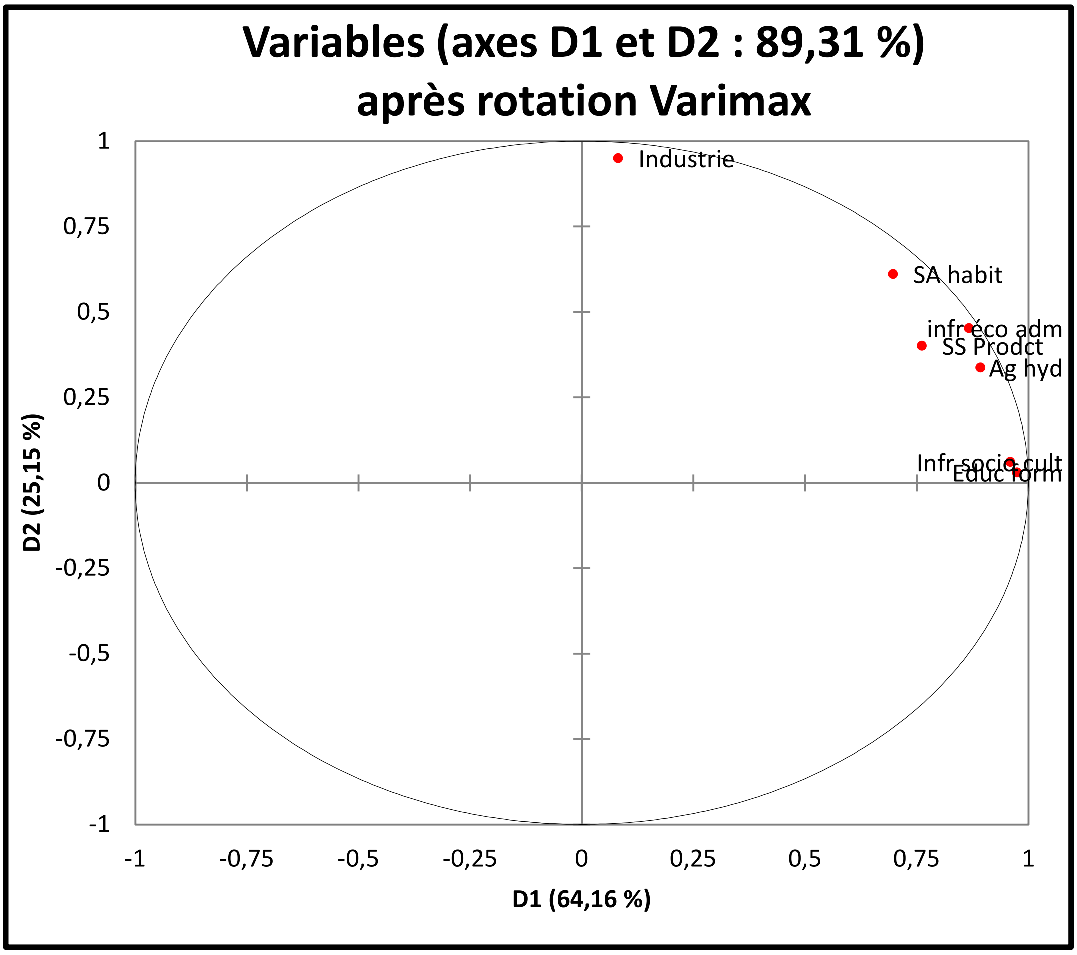
	F1	F2	F3	F4	F5	F6	F7
Industrie	0,2061	0,8171	-0,4240	0,2454	0,0119	-0,2228	0,0120
Ag hyd	0,4167	-0,0655	-0,0368	-0,5490	-0,5581	-0,4082	0,2033
SS Prodct	0,3762	0,0462	0,6687	0,5784	-0,2397	-0,1306	0,0068
infr éco adm	0,4274	0,0490	-0,1508	-0,0561	-0,3028	0,7661	-0,3323
Educ form	0,3928	-0,3741	-0,2642	0,1213	0,2952	-0,3899	-0,6182
Infr socio cult	0,3926	-0,3405	-0,3461	0,2866	0,2208	0,1372	0,6785
SA habit	0,3889	0,2602	0,3992	-0,4515	0,6351	0,0950	0,0739

6) إسقاط المتغيرات:

إحداثيات المتغيرات بعد التدوير (Coordonnées des variables après rotation Varimax):

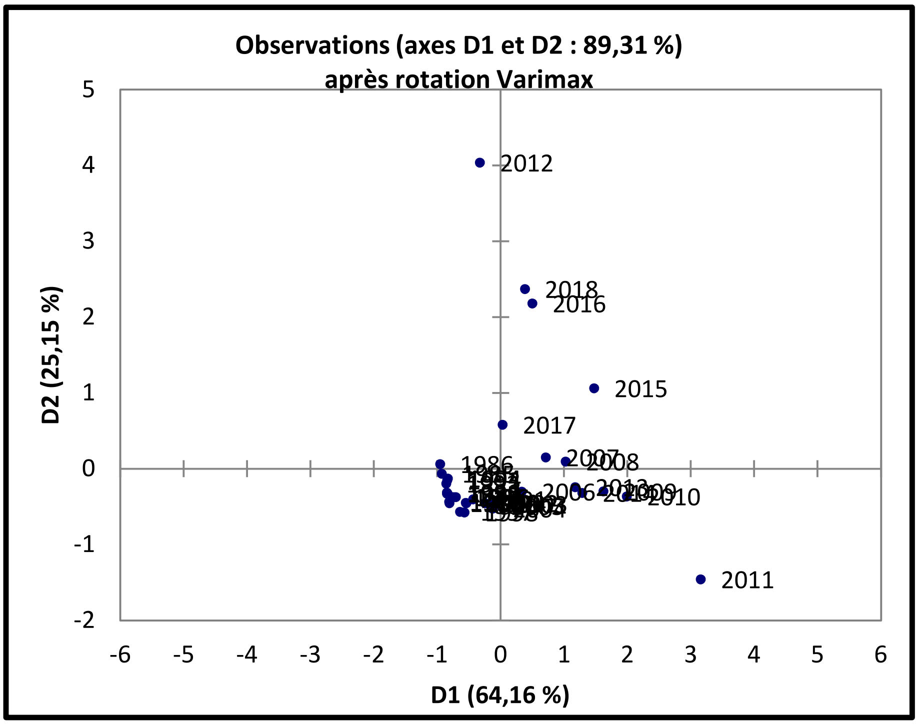
	D1	D2
Industrie	0,0817	0,9502
Ag hyd	0,8930	0,3371
SS Prodct	0,7615	0,4014
infr éco adm	0,8668	0,4529
Educ form	0,9742	0,0297
Infr socio cult	0,9595	0,0605
SA habit	0,6974	0,6110

7) التمثيل البياني للمتغيرات بعد الإسقاط:



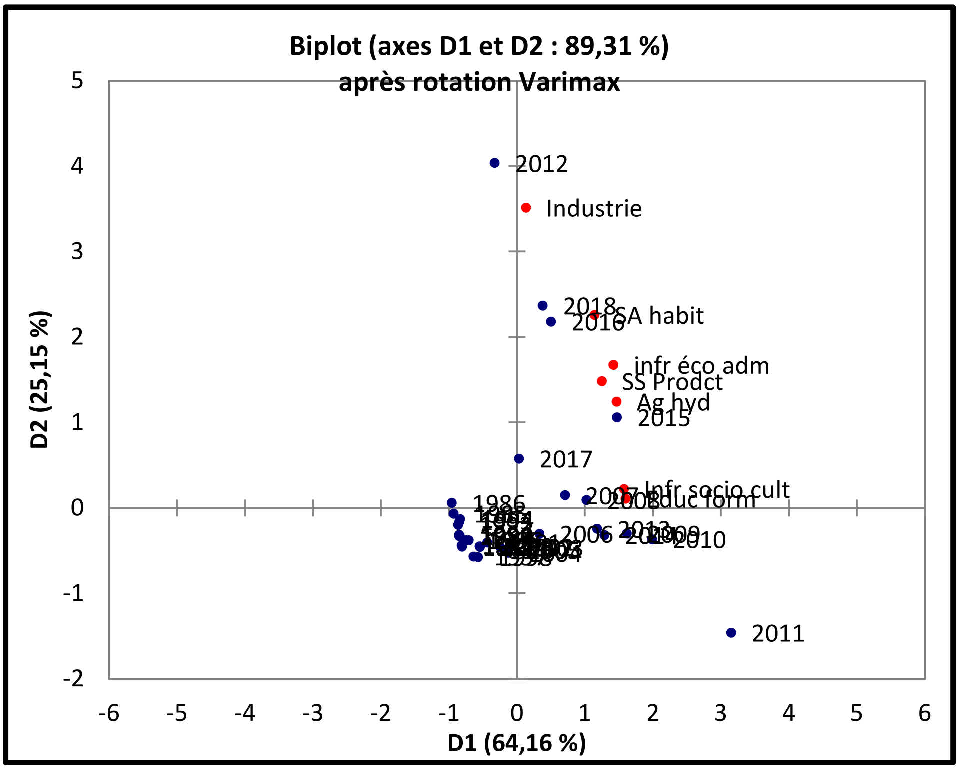
8) التمثيل البياني للأفراد (السنوات) بعد الإسقاط:

على عكس برنامج SPSS، التمثيل البياني للأفراد يكون ضمن المخرجات مباشرة.



9) التمثيل البياني للأفراد (السنوات) والمتغيرات بعد الإسقاط في نفس الشكل:

على عكس برنامج SPSS، فإن برنامج XLSTAT يعطينا ضمن المخرجات التمثيل البياني المشترك للمتغيرات والأفراد بالمحاور العاملية في نفس الشكل البياني.



II. المحور الثاني: التحليل العاملي التقابلي (AFC)

1. مفهوم طريقة التحليل العاملي التقابلي (أو التوافقي):

تم تطوير هذه الطريقة لدراسة التوافق – Correspondance (الارتباط Liaison) بين متغيرين كيفيين، لكل منهما فئات معينة، ولكل فئة تكرارات. ويتم جمع بيانات المتغيرين في جدول مزدوج (جدول التوافق Tableau de contingence).

ليكن X و Y متغيرين كيفيين. فئات المتغير X هي $\{M_1, M_2, \dots, M_n\}$، وفئات المتغير Y هي $\{N_1, N_2, \dots, N_p\}$.

أ. تعريف طريقة التحليل العاملي التقابلي:

طريقة التحليل العاملي التقابلي هي طرقة تسمح بتحليل البيانات الخاصة بمتغيرين كيفيين من خلال البحث عن الارتباط (أو التوافق) بينهما. لكل من هذين المتغيرين مجموعة من الفئات (خاصيات)، وتكون هذه البيانات مدرجة في جدول مزدوج (جدول التوافق)، وقيمه تعبر عن تكرارات فئات المتغيرين.

وغالبا ما نستخدم مصطلح الارتباط (corrélation) عندما يكون المتغيران كميين، ومصطلح التوافق (correspondance) عندما يكون المتغيران كيفيين.

مثال: ليكن المتغير X هو المستوى التعليمي، وفئاته هي: إبتدائي، متوسط، ثانوي، جامعي.

والمتغير Y هو الجنس، وفئاته هي: ذكر، أنثى.

ب. الهدف من طريقة التحليل العاملي التقابلي:

الهدف من طريقة التحليل العاملي التقابلي هو تقريبا نفس هدف جميع الطرق العاملية، وهو اختزال البيانات وتبسيط جدول بأبعاد كبيرة (جدول له p عمود وn سطر) للحصول على تمثيل لبياني مبسط يوضح مختلف العلاقات بين فئات المتغيرين الكيفيين.

فهذه الطريقة تهدف إلى التحليل والتمثيل البياني للجداول المزدوجة، وذلك من خلال تحليل التوافق بين متغيرين كيفيين من حيث: التمركز/المحيط (centre/périphérie)، القرب/التباعد (éloignement/proximité)، التشابه/الاختلاف (ressemblance/dissemblance)، التجاذب/التنافر (attraction/répulsion). وهذا بالاعتماد على حساب المسافات (distances) بين الفئات. ومن خلال البحث عن الأسطر (فئات المتغير الأول) المتشابهة والأسطر المختلفة، وكذلك البحث عن الأعمدة (فئات المتغير الثاني) المتشابهة والأعمدة المختلفة، وبالتالي البحث عن المجموعات المتجانسة في الأسطر (وفي الأعمدة)، بالإضافة إلى البحث في العلاقات بين الأسطر والأعمدة.

ج. مبدأ طريقة التحليل العاملي التقابلي:

تقوم طريقة التحليل العاملي التقابلي على مبدأ اسقاط سحابة النقاط من فضاء متعدد الأبعاد (فضاء ذو n أو p بعد) على فضاء ذو بعد صغير (بعد أو بعدين أو ثلاث أبعاد)، مع المحافظة على المسافات الأصلية قدر الإمكان، وبالتالي المحافظة على أكبر قدر ممكن من المعلومات المحتوات في جدول البيانات الأصلية (الجدول المزدوج)، مع تحليل جانب الأسطر وتحليل جانب الأعمدة (الإسقاطات والتمثيلات البيانية).

2. خطوات إجراء طريقة التحليل العاملي التقابلي:

أ. جمع البيانات وتشكيل الجداول الكاملة (أو جدول البيانات الخام):

بعد جمع البيانات حول موضوع الدراسة (إما بأسلوب مباشر أو غير مباشر)، نقوم بإدخالها إلى البرامج الإحصائية الجاهزة (كبرنامج SPSS أو برنامج XL-STAT). وذلك على شكل الجداول المنفصلة الكاملة (Tableau disjonctif complet)، أو جداول البيانات الخام (Tableau des données brutes).

ب. تشكيل الجدول المزدوج:

بالاعتماد على الجداول الكاملة أو الجداول الخام، نقوم بتشكيل الجدول المزدوج (أو جدول التوافق أو جدول التبعية) (Tableau de contingence ou tableau de dépendance ou tableau Croisé) والذي يضم التكرارات المطلقة لكل فئة من الفئات كما يلي:

Y / X	N_1	…	N_j	…	N_p	المجموع
M_1	n_{11}				n_{1p}	$n_{1.}$
…		…				…
M_i			n_{ij}			$n_{i.}$
…				…		…
M_n	n_{n1}				n_{np}	$n_{n.}$
المجموع	$n_{.1}$	…	$n_{.j}$	…	$n_{.p}$	$n_{..}$

حيث: n_{ij} يمثل عدد الأفراد (عدد التكرارات) التي تنتمي للفئة M_i من X والفئة N_j من Y في نفس الوقت.

و: $n_{i.} = \sum_{j=1}^{p} n_{ij}$ يمثل عدد الأفراد التي تنتمي للفئة M_i من X $(i = 1,2, \dots , n)$ (المجموع الهامشي للأسطر).

و: $n_{.j} = \sum_{i=1}^{n} n_{ij}$ يمثل عدد الأفراد التي تنتمي للفئة N_j من Y ($j = 1,2, \dots , p$). (المجموع الهامشي للأعمدة).

و: $n_{..} = \sum_{i=1}^{n} \sum_{j=1}^{p} n_{ij}$ العدد الكلي للأفراد (المجموع الكلي).

نرمز لهذا الجدول بالرمز N، حيث: $N_{n\times p} = \begin{pmatrix} n_{11} & \cdots & n_{1j} & \cdots & n_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ n_{i1} & \dots & n_{ij} & \dots & n_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ n_{n1} & \cdots & n_{nj} & \dots & n_{np} \end{pmatrix}$.

كل فئة i من المتغير X تُمثّل في فضاء ذو بعد p $\mathbb{R}^p$: $M_i = (n_{i1} \quad n_{i2} \quad \cdots \quad n_{ip})$.

كل فئة j من المتغير Y تُمثّل في فضاء ذو بعد n $\mathbb{R}^n$: $N_j = \begin{pmatrix} n_{1j} \\ n_{2j} \\ \vdots \\ n_{nj} \end{pmatrix}$.

ملاحظة: على برنامج SPSS يمكن القيام بتحليل عاملي توافقي بالاعتماد على جدول البيانات الخام فقط، ولا يمكن ذلك من خلال الجدول المزدوج. أما على برنامج XL-STAT فيمكن القيام بتحليل عاملي توافقي إما بالاعتماد على جدول البيانات الخام أو من خلال الجدول المزدوج.

ج. تشكيل جدول التكرارات النسبية المشاهدة:

لتشكيل جدول التكرارات النسبية المشاهدة (Tableau des fréquences relatives observées) نقوم بقسمة كل تكرار في الجدول المزدوج (n_{ij}) على المجموع الكلي للتكرارات ($n_{..}$).

فتتحصل على التكرار النسبي f_{ij}، حيث: $f_{ij} = \frac{n_{ij}}{n_{..}}$.

ويسمى كذلك بجدول الاحتمالات (Tableau des probabilités).

الجدول الموالي يمثل جدول التكرارات النسبية:

Y / X	N_1	…	N_j	…	N_p	المجموع
M_1	f_{11}				f_{1p}	$f_{1.}$
…		…				…
M_i			f_{ij}			$f_{i.}$
…				…		…
M_n	f_{n1}				f_{np}	$f_{n.}$

المجموع	$f_{.1}$	…	$f_{.j}$	…	$f_{.p}$	$f_{..} = 1$

حيث: f_{ij} يمثل التكرار النسبي للفئة M_i من X والفئة N_j من Y في نفس الوقت.

وتمثل الاحتمال: $P\left[(X = M_i) \cap P\left(Y = N_j\right)\right]$.

و: $f_{i.}$ يمثل التكرار النسبي للفئة M_i من X ($i = 1,2, \dots, n$).

وتمثل الاحتمال الهامشي: $P(X = M_i) = f_{i.}$.

و: $f_{.j}$ يمثل التكرار النسبي للفئة N_j من Y ($j = 1,2, \dots, p$).

وتمثل الاحتمال الهامشي: $P\left(Y = N_j\right) = f_{.j}$.

و: $f_{..}$ يمثل التكرار النسبي الكلي (مجموع جميع الاحتمالات).

نرمز لهذا الجدول بالرمز F، حيث: $F_{n\times p} = \begin{pmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ f_{n1} & \cdots & f_{nj} & \cdots & f_{np} \end{pmatrix}$.

د. تشكيل جدول التكرارات النسبية النظرية:

لتشكيل جدول التكرارات النسبية النظرية (Tableau des fréquences relatives théoriques) نقوم بضرب مجموع التكرارات النسبية للأسطر ($f_{i.}$) في مجموع التكرارات النسبية للأعمدة ($f_{.j}$)، فنتحصل على التكرار النظري (المتوقع) (t_{ij})، أي: $t_{ij} = f_{i.} \times f_{.j}$.

هذه القيم لجدول التكرارات النسبية النظرية تفترض الاستقلالية بين الأسطر والأعمدة (المتغيرين X وY مستقلين).

تذكير: إذا كان A وB حدثان مستقلان ($A \perp B$)، فإن: $P(A \cap B) = P(A) \times P(B)$.

وعليه، فتحت فرضية استقلالية الأسطر عن الأعمدة فإن:

$$P\left[(X = M_i) \cap P\left(Y = N_j\right)\right] = P(X = M_i) \times P\left(Y = N_j\right).$$

ومنه: $t_{ij} = P(X = M_i) \times P\left(Y = N_j\right) = f_{i.} \times f_{.j}$.

الجدول الموالي يمثل جدول التكرارات النسبية النظرية:

Y / X	N_1	…	N_j	…	N_p	المجموع
M_1	t_{11}				t_{1p}	$f_{1.}$

$\dots$		$\dots$				$\dots$
M_i			t_{ij}			$f_{i.}$
$\dots$				$\dots$		$\dots$
M_n	t_{n1}				t_{np}	$f_{n.}$
المجموع	$f_{.1}$	$\dots$	$f_{.j}$	$\dots$	$f_{.p}$	$f_{..} = 1$

نرمز لهذا الجدول بالرمز T، حيث: $T_{n\times p} = \begin{pmatrix} t_{11} & \cdots & t_{1j} & \cdots & t_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ t_{i1} & \cdots & t_{ij} & \cdots & t_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ t_{n1} & \cdots & t_{nj} & \cdots & t_{np} \end{pmatrix}$.

ملاحظة: هذا التحويل للقيم لا يؤثر على المجاميع الهامشية للأسطر $(f_{i.})$ ولا الأعمدة $(f_{.j})$ (نفس المجاميع).

لأن: - الأسطر: $\sum_{j=1}^{p} t_{ij} = \sum_{j=1}^{p}(f_{i.} \times f_{.j}) = f_{i.}\sum_{j=1}^{p}(f_{.j}) = f_{i.} \times 1 = f_{i.}$.

- الأعمدة: $\sum_{i=1}^{n} t_{ij} = \sum_{i=1}^{n}(f_{i.} \times f_{.j}) = f_{.j}\sum_{i=1}^{n}(f_{.j}) = f_{.j} \times 1 = f_{.j}$.

هـ. اختبار كاي تربيع:

يسمح اختبار كاي تربيع ((χ^2(**Chi–square**) بالكشف عن وجود علاقة بين فئات متغيرين كيفيين. أي يجيب عن التساؤل: هل الفئة i من المتغير X تدل على الفئة j من المتغير Y؟ فهو يستخدم لدراسة الارتباط بين الأعمدة والأسطر في الجداول المزدوجة. ويقوم هذا الاختبار على مقارنة القيمة المشاهدة (f_{ij}) في كل خلية (في الجدول المزدوج) من خلايا تقاطع فئات المتغيرين بالقيمة المتوقعة (النظرية) t_{ij}.

- ✓ **فرضيات الاختبار**:

الفرضية الصفرية (H_0): لا توجد علاقة ارتباط ذات دلالة بين المتغيرين (المتغيرين مستقلين: $f_{ij} = t_{ij}$).

الفرضية البديلـــة (H_1): توجد علاقة ارتباط ذات دلالة بين المتغيرين (المتغيرين مترابطين: $f_{ij} \neq t_{ij}$).

- ✓ **احصائية الاختبار**: إحصائية اختبار كاي تربيع هي: $\chi^2 = n_{..}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(f_{ij}-t_{ij})^2}{t_{ij}}$.

درجة الحرية هي: (عدد الأسطر – 1)*(عدد الأعمدة – 1).

- ✓ **قرار الاختبار**: إذا كانت إحصائية كاي تربيع المحسوبة χ_c^2 أصغر من المجدولة $\chi_\alpha^2((n-1) * (p-1))$ (أو القيمة الاحتمالية أكبر من مستوى المعنوية α) فإنا نقبل الفرضية الصفرية H_0، والتي تقضي بعدم وجود علاقة بين المتغيرين (المتغيرين مستقلين).

- ✓ **ملاحظات**:

- يقيس اختبار كاي تربيع البعد عن الاستقلالية l'écart à l'indépendances $(f_{ij} - t_{ij})$.
- اختبار كاي تربيع يسمح فقط باختبار الارتباط بين متغيرين كيفيين، لكنه لا يبين نوع وقوة العلاقة بينهما. لكن التحليل العاملي التقابلي يسمح بالتفصيل في تحليل العلاقة بينهما. ولهذا فإن اختبار كاي تربيع لا يكفي لتحليل العلاقة بين متغيرين كيفيين، فحتى وإن كان غير معنوي، فلا بد من إكمال وتعزيز التحليل بطريقة التحليل العاملي التقابلي. فمثلا في حالة تباين كلي (Inertie) صغير فإن اختبار كاي تربيع لا يستطيع اكتشاف العلاقة بين المتغيرين، لكن يمكن ذلك من خلال التحليل العاملي التقابلي.
- التباين الكلي لسحابة النقاط هو:

$$I = \sum_{i=1}^{n} f_{i.} d^2(i,G) = \sum_{j=1}^{p} f_{.j} d^2(j,G) = \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}}.$$

- صيغة أخرى لإحصائية كاي تربيع هي: $\boldsymbol{\chi^2 = n_{..} \times I}$ (حيث $I = \sum_{\alpha=1}^{p-1} \lambda_\alpha$ هي التباين الكلي).

و. **اختبار التجاذب/التنافر:**

نسمي القيمة $d_{ij} = \frac{f_{ij}}{t_{ij}} = \frac{f_{ij}}{f_{i.} \times f_{.j}}$ بمعامل التجاذب/التنافر (attraction/répulsion).

واستعملت هذه العلاقة من طرف كل من Escofier et Pagés (1988) وتلعب هذه الأخيرة دور أساسي في AFC.

يمكن استنتاج وجود تجاذب أو تنافر بين فئات المتغيرين من خلال حساب معامل التجاذب كما يلي:

- ✓ إذا كان $d_{ij} = 1$ (أي: $f_{ij} = t_{ij}$)، نقول أنه لا يوجد ارتباط بين الفئة M_i من المتغير X والفئة N_j من المتغير Y (أي الفئتين مستقلتين).
- ✓ إذا كانت $d_{ij} < 1$ (أي: $f_{ij} < t_{ij}$)، نقول أن الفئتين M_i و N_j متنافرتان.
- ✓ إذا كانت $d_{ij} > 1$ (أي: $f_{ij} > t_{ij}$)، نقول أن الفئتين M_i و N_j متجاذبتان.

ز. **تحليل جانب الأسطر (جدول التكرارات النسبية للأسطر):**

لتحليل جانب الأسطر (**Profils lignes**) نقوم بقسمة جميع أسطر جدول التكرارات النسبية (f_{ij}) على مجموع التكرار النسبي للسطر الموافق لها $(f_{i.})$، أي نحسب: $l_{ij} = \frac{n_{ij}}{n_{i.}} = \frac{f_{ij}}{f_{i.}}$.

وهي تمثل الاحتمالات الشرطية التالية: $P\left(Y = N_j / X = M_i\right) = l_{ij}$.

يسمح هذا الجدول بالمقارنة بين الأسطر فيما بينها (فئات متغير الأسطر).

يمثل الجدول التالي التكرارات النسبية للأسطر:

Y	N_1	…	N_j	…	N_p	المجموع

X						
M_1	$\frac{f_{11}}{f_{1.}}$					1
$\cdots$		$\cdots$				$\cdots$
M_i			$\frac{f_{ij}}{f_{i.}}$			1
$\cdots$				$\cdots$		$\cdots$
M_n					$\frac{f_{np}}{f_{n.}}$	1
	$f_{.1}$		$f_{.j}$	$\cdots$	$f_{.p}$	

نرمز لهذا الجدول بالرمز L (مصفوفة التكرارات النسبية للأسطر)، حيث:

$$L_{n\times p} = \begin{pmatrix} l_{11} & \cdots & l_{1j} & \cdots & l_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ l_{i1} & \cdots & l_{ij} & \cdots & l_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ l_{n1} & \cdots & l_{nj} & \cdots & l_{np} \end{pmatrix}.$$

يمكن القيام بهذه الحسابات من خلال الجداء المصفوفي التالي: $L_{n\times p} = D_n^{-1}.F_{n\times p}$.

حيث D_n هي مصفوفة قطرية: $D_n = \begin{pmatrix} f_{1.} & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & f_{i.} & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & f_{n.} \end{pmatrix}$.

أي: $L = D_n^{-1}.F = \begin{pmatrix} \frac{1}{f_{1.}} & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \frac{1}{f_{i.}} & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & \frac{1}{f_{n.}} \end{pmatrix} . \begin{pmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ f_{n1} & \cdots & f_{nj} & \cdots & f_{np} \end{pmatrix}$

ملاحظة: مجموع قيم الأسطر تساوي1. ($\sum_{j=1}^{p} l_{ij} = \sum_{j=1}^{p}\left(\frac{f_{ij}}{f_{i.}}\right) = \frac{1}{f_{i.}}\sum_{j=1}^{p}(f_{ij}) = \frac{f_{i.}}{f_{i.}} = 1$).

ح. تحليل جانب الأعمدة (جدول التكرارات النسبية للأعمدة) (Profils colonnes):

لتحليل جانب الأعمدة نقوم بقسمة جميع أعمدة جدول التكرارات النسبية (f_{ij}) على مجموع التكرار النسبي للعمود الموافق لها ($f_{.j}$)، أي نحسب: $c_{ij} = \frac{n_{ij}}{n_{.j}} = \frac{f_{ij}}{f_{.j}}$.

وهي تمثل الاحتمالات الشرطية التالية: $P(X = M_i / Y = N_j) = c_{ij}$.

يسمح هذا الجدول بالمقارنة بين الأعمدة فيما بينها (فئات متغير الأعمدة).

يمثل الجدول التالي التكرارات النسبية للأعمدة:

Y / X	N_1	$\cdots$	N_j	$\cdots$	N_p	
M_1	$\frac{f_{11}}{f_{.1}}$					$f_{1.}$
$\cdots$		$\cdots$				$\cdots$
M_i			$\frac{f_{ij}}{f_{.j}}$			$f_{i.}$
$\cdots$				$\cdots$		$\cdots$
M_n					$\frac{f_{np}}{f_{.p}}$	$f_{n.}$
المجموع	1		1	$\cdots$	1	

نرمز لهذا الجدول بالرمز C (مصفوفة التكرارات النسبية للأعمدة)، حيث:

$$C_{n\times p} = \begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ c_{i1} & \cdots & c_{ij} & \cdots & c_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ c_{n1} & \cdots & c_{nj} & \cdots & c_{np} \end{pmatrix}.$$

يمكن القيام بهذه الحسابات من خلال الجداء المصفوفي التالي: $C_{n\times p} = F_{n\times p}.D_p^{-1}$.

حيث D_n هي مصفوفة قطرية: $D_p = \begin{pmatrix} f_{.1} & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & f_{.j} & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & f_{.p} \end{pmatrix}$.

أي:

$$C = F.D_p^{-1} = \begin{pmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ f_{n1} & \cdots & f_{nj} & \cdots & f_{np} \end{pmatrix} \begin{pmatrix} \frac{1}{f_{.1}} & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \frac{1}{f_{.j}} & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & \frac{1}{f_{.p}} \end{pmatrix}$$

ملاحظة: مجموع قيم الأعمدة تساوي 1. ($\sum_{i=1}^{n} c_{ij} = \sum_{i=1}^{n}\left(\frac{f_{ij}}{f_{.j}}\right) = \frac{1}{f_{.j}}\sum_{i=1}^{n}(f_{ij}) = \frac{f_{.j}}{f_{.j}} = 1$).

ط. سحابة نقاط الأسطر في الفضاء $\mathbb{R}^p$:

لدينا كل فئة i من X لها p إحداثية: $\left\{l_{ij} = \frac{f_{ij}}{f_{i.}}, j = 1,2,\ldots,p\right\}$، وبالتالي تمثل في الفضاء $\mathbb{R}^p$.

مركز ثقل سحابة نقاط جانب الأسطر G هي المتوسط المرجح للأسطر، مركبتها j (مع j=1,2,…,p) هي $f_{.j}$ (المجموع الهامشي للأعمدة)، لأن

$$g_j = \sum_{i=1}^{n} f_{i.} \times l_{ij} = \sum_{i=1}^{n} f_{i.} \times \frac{f_{ij}}{f_{i.}} = \sum_{i=1}^{n} f_{ij} = f_{.j}.$$

لدينا:
$$G = L_{p\times n}^t . D_n . 1_n = \begin{pmatrix} \frac{f_{11}}{f_{1.}} & \ldots & \frac{f_{i1}}{f_{i.}} & \ldots & \frac{f_{n1}}{f_{n.}} \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ \frac{f_{1j}}{f_{1.}} & \ldots & \frac{f_{ij}}{f_{i.}} & \ldots & \frac{f_{nj}}{f_{n.}} \\ \vdots & \ldots & \cdots & \cdots & \cdots \\ \frac{f_{1p}}{f_{1.}} & \ldots & \frac{f_{ip}}{f_{i.}} & \ldots & \frac{f_{np}}{f_{n.}} \end{pmatrix} . \begin{pmatrix} f_{1.} & \cdots & 0 & \ldots & 0 \\ \vdots & \ldots & \vdots & \cdots & \vdots \\ 0 & \cdots & f_{i.} & \cdots & 0 \\ \vdots & \ldots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \ldots & f_{n.} \end{pmatrix} . \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

ومنه:
$$G = \begin{pmatrix} f_{11} & \cdots & f_{i1} & \cdots & f_{n1} \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ f_{1j} & \cdots & f_{ij} & \cdots & f_{nj} \\ \vdots & \ldots & \cdots & \cdots & \cdots \\ f_{1p} & \cdots & f_{ip} & \cdots & f_{np} \end{pmatrix} . \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} f_{.1} \\ \vdots \\ f_{.j} \\ \vdots \\ f_{.p} \end{pmatrix}.$$

نقوم بالإسقاط بالاعتماد على نقاط الأسطر (l_{ij})، مع الأخذ مركز الاسقاط مركز ثقل سحابة النقاط (G).

لدينا تحت فرضية الاستقلالية فإن: $f_{ij} = t_{ij} = f_{i.} \times f_{.j} \Longrightarrow \frac{f_{ij}}{f_{i.}} = f_{.j} \Longrightarrow l_{ij} = g_j$.

وهذا يعني أنه كلما اقتربت التكرارات النسبية للأسطر من متوسط العمود (المركز) (أي: $l_{ij} = g_j$)، كلما اقتربنا من الاستقلالية بين الفئات، والعكس فكلما ابتعدنا عن المركز فهذا يدل على الارتباط (وكأننا نقوم بقياس البعد عن الاستقلالية l'écart à l'indépendances).

ي. سحابة نقاط الأعمدة في الفضاء $\mathbb{R}^n$:

لدينا كل فئة j من Y لها n إحداثية: $\left\{c_{ij} = \frac{f_{ij}}{f_{.j}}, i = 1,2,\ldots,n\right\}$، وبالتالي تمثل في الفضاء $\mathbb{R}^n$.

مركز ثقل سحابة نقاط جانب الأعمدة G هي المتوسط المرجح للأعمدة، مركبتها i (مع i=1,2,…,n) هي $f_{i.}$ (المجموع الهامشي للأسطر)، لأن:

$$g_i = \sum_{j=1}^{p} f_{.j} \times c_{ij} = \sum_{j=1}^{p} f_{.j} \times \frac{f_{ij}}{f_{.j}} = \sum_{j=1}^{p} f_{ij} = f_{i.}.$$

لدينا: $\begin{pmatrix}1\\ \vdots\\ 1\\ \vdots\\ 1\end{pmatrix} \cdot \begin{pmatrix} f_{.1} & \cdots & 0 & \dots & 0\\ \vdots & \dots & \vdots & \dots & \vdots\\ 0 & \dots & f_{.j} & \cdots & 0\\ \vdots & \dots & \cdots & \cdots & \cdots\\ 0 & \cdots & 0 & \dots & f_{.p}\end{pmatrix} \cdot \begin{pmatrix} \frac{f_{11}}{f_{.1}} & \cdots & \frac{f_{1j}}{f_{.j}} & \cdots & \frac{f_{1p}}{f_{.p}}\\ \vdots & \dots & \vdots & \dots & \vdots\\ \frac{f_{i1}}{f_{.1}} & \dots & \frac{f_{ij}}{f_{.j}} & \dots & \frac{f_{ip}}{f_{.p}}\\ \vdots & \dots & \cdots & \cdots & \cdots\\ \frac{f_{n1}}{f_{.1}} & \dots & \frac{f_{nj}}{f_{.j}} & \dots & \frac{f_{np}}{f_{.p}}\end{pmatrix} = G = C_{n\times p}.D_p.1_p$.

ومنه: $\begin{pmatrix}f_{1.}\\ \vdots\\ f_{i.}\\ \vdots\\ f_{n.}\end{pmatrix} = \begin{pmatrix}1\\ \vdots\\ 1\\ \vdots\\ 1\end{pmatrix} \cdot \begin{pmatrix} f_{11} & \cdots & f_{1j} & \dots & f_{1p}\\ \vdots & \dots & \vdots & \dots & \vdots\\ f_{i1} & \dots & f_{ij} & \dots & f_{ip}\\ \vdots & \dots & \cdots & \cdots & \cdots\\ f_{n1} & \cdots & f_{nj} & \dots & f_{np}\end{pmatrix} = G$.

نقوم بالإسقاط بالاعتماد على نقاط جانب الأعمدة (c_{ij})، مع الأخذ مركز الاسقاط مركز ثقل سحابة النقاط (G).

لدينا تحت فرضية الاستقلالية فإن: $f_{ij} = t_{ij} = f_{i.} \times f_{.j} \Longrightarrow \frac{f_{ij}}{f_{.j}} = f_{i.} \Longrightarrow c_{ij} = g_i$.

وهذا يعني أنه كلما اقتربت التكرارات النسبية للأعمدة من متوسط السطر (المركز) (أي: $c_{ij} = g_i$)، كلما اقتربنا من الاستقلالية بين الفئات، والعكس فكلما ابتعدنا عن المركز فهذا يدل على الارتباط (وكأننا نقوم بقياس البعد عن الاستقلالية l'écart à l'indépendances).

ك. حساب المسافات بين الأسطر وبين الأعمدة:

التوافق بين سطرين (أو عمودين) تعرف بالمسافة بينهما، مقياس المسافة المستعمل هو إما مقياس كاي تربيع (χ^2) أو المقياس الإقليدي (euclidienne).

➢ **المسافة بمقاس كاي تربيع:**

✓ المسافة بين فئتين في الأسطر i و i' هي: $d_{\chi^2}(i,i') = \sqrt{\sum_{j=1}^{p} \frac{1}{f_{.j}} \left(l_{ij} - l_{i'j}\right)^2}$.

✓ المسافة بين فئة في الأسطر i والمركز G هي: $d_{\chi^2}(i,G) = \sqrt{\sum_{j=1}^{p} \frac{1}{f_{.j}} \left(l_{ij} - f_{.j}\right)^2}$.

✓ المسافة بين فئتين في الأعمدة j و j' هي: $d_{\chi^2}(j,j') = \sqrt{\sum_{i=1}^{n} \frac{1}{f_{i.}} \left(c_{ij} - c_{ij'}\right)^2}$.

✓ المسافة بين فئة في الأعمدة j والمركز G هي: $d_{\chi^2}(j,G) = \sqrt{\sum_{j=1}^{p} \frac{1}{f_{i.}} \left(c_{ij} - f_{i.}\right)^2}$.

➢ **المسافة بمقياس إقليدي:**

- ✓ المسافة بين فئتين في الأسطر i و 'i هي: $d_e(i,i') = \sqrt{\sum_{j=1}^{p}\left(l_{ij} - l_{i'j}\right)^2}$.
- ✓ المسافة بين فئة في الأسطر i والمركز G هي: $d_e(i,G) = \sqrt{\sum_{j=1}^{p}\left(l_{ij} - f_{.j}\right)^2}$.
- ✓ المسافة بين فئتين في الأعمدة j و 'j هي: $d_e(j,j') = \sqrt{\sum_{i=1}^{n}\left(c_{ij} - c_{ij'}\right)^2}$.
- ✓ المسافة بين فئة في الأعمدة j والمركز G هي: $d_e(j,G) = \sqrt{\sum_{j=1}^{p}\left(c_{ij} - f_{i.}\right)^2}$.

ل. تحديد عدد محاور التحليل:

- عدد المحاور التي تأخذ في التحليل العاملي التقابلي تساوي أصغر قيمة بين عدد الأسطر وعدد الأعمدة نطرح منها واحد ($Min(n,p) - 1$).
- يمكن أن يتم تحديد عدد المحاور العاملية في التحليل بنفس الطرق المبينة في تحليل المركبات الرئيسية (ACP) (نسبة المستوي العاملي الأول، نسبة المحور العاملي الثالث، ...).
- في الغالب المستوي العاملي الأول (1er plan) يكفي للقيام بتحليل عاملي توافقي.
- كل القيم الذاتية في AFC تكون أصغر من أو تساوي 1 ($0 \leq \lambda_\alpha \leq 1$).

م. اسقاط جانب الأسطر:

في فضاء $\mathbb{R}^p$:

نبحث عن تعظيم الدالة التالية تحت القيود: $\begin{cases} Max(\mu^t.D_p^{-1}.S.\mu) \\ s/c: \mu^t.D_p^{-1}.\mu = 1 \end{cases}$.

مع: $\boldsymbol{S_{p\times p} = L^t.C}$ (أي: $S_{p\times p} = F^t.D_n^{-1}.F.D_p^{-1}$) .

و μ هو شعاع ذاتي للمصفوفة S الموافق للقيمة الذاتية λ لهذه المصفوفة.

عناصر المصفوفة S تكتب: $s_{jj'} = \sum_{i=1}^{n}\frac{f_{ij}\times f_{ij'}}{f_{i.}\times f_{.j}}$.

والمحاور العاملية تكتب: $S_{\mu_\alpha} = \lambda_\alpha.\mu_\alpha$.

والاسقاطات على المحاور العاملية هي: $\boldsymbol{\psi_\alpha = L.D_p^{-1}.\mu_\alpha}$ (أي: $\psi_\alpha = D_n^{-1}.F.D_p^{-1}.\mu_\alpha$).

(مع: $\psi_{\alpha_i} = \sum_{j=1}^{p}\frac{f_{ij}}{f_{i.}\times f_{.j}}.\mu_{\alpha_j}$).

ن. اسقاط جانب الأعمدة:

في فضاء $\mathbb{R}^n$: نفس طريقة التحليل، مع استبدال الشعاع الذاتي μ بالشعاع الذاتي v.

نبحث عن تعظيم الدالة التالية تحت القيود: $\begin{cases} Max(\nu^t.D_n^{-1}.A.\nu) \\ s/c: \nu^t.D_n^{-1}.\nu = 1 \end{cases}$.

لدينا: $\boldsymbol{A_{n\times n} = C.L^t}$ (أي: $A_{n\times n} = F.D_p^{-1}.F^t.D_n^{-1}$).

و ν هو شعاع ذاتي للمصفوفة A الموافق للقيمة الذاتية لهذه المصفوفة.

والاسقاطات على المحاور العاملية هي: $\boldsymbol{\varphi_\alpha = D_p^{-1}.L^t.\nu_\alpha}$ (أي: $\varphi_\alpha = D_p^{-1}.F^t.D_n^{-1}.\nu_\alpha$).

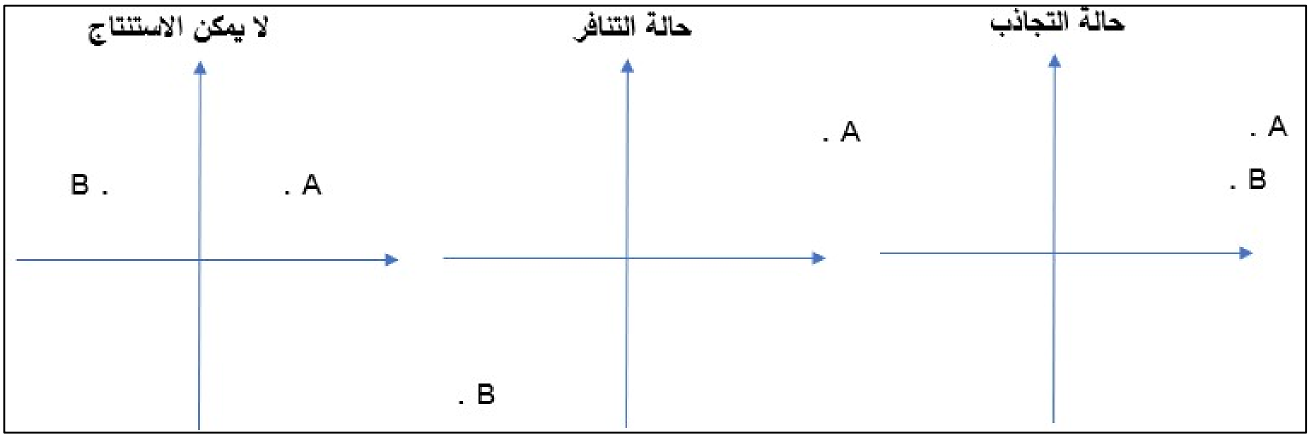
(مع: $\varphi_{\alpha_j} = \sum_{i=1}^{n} \frac{f_{ij}}{f_{i.}\times f_{.j}}.\nu_{\alpha_i}$).

ملاحظات:

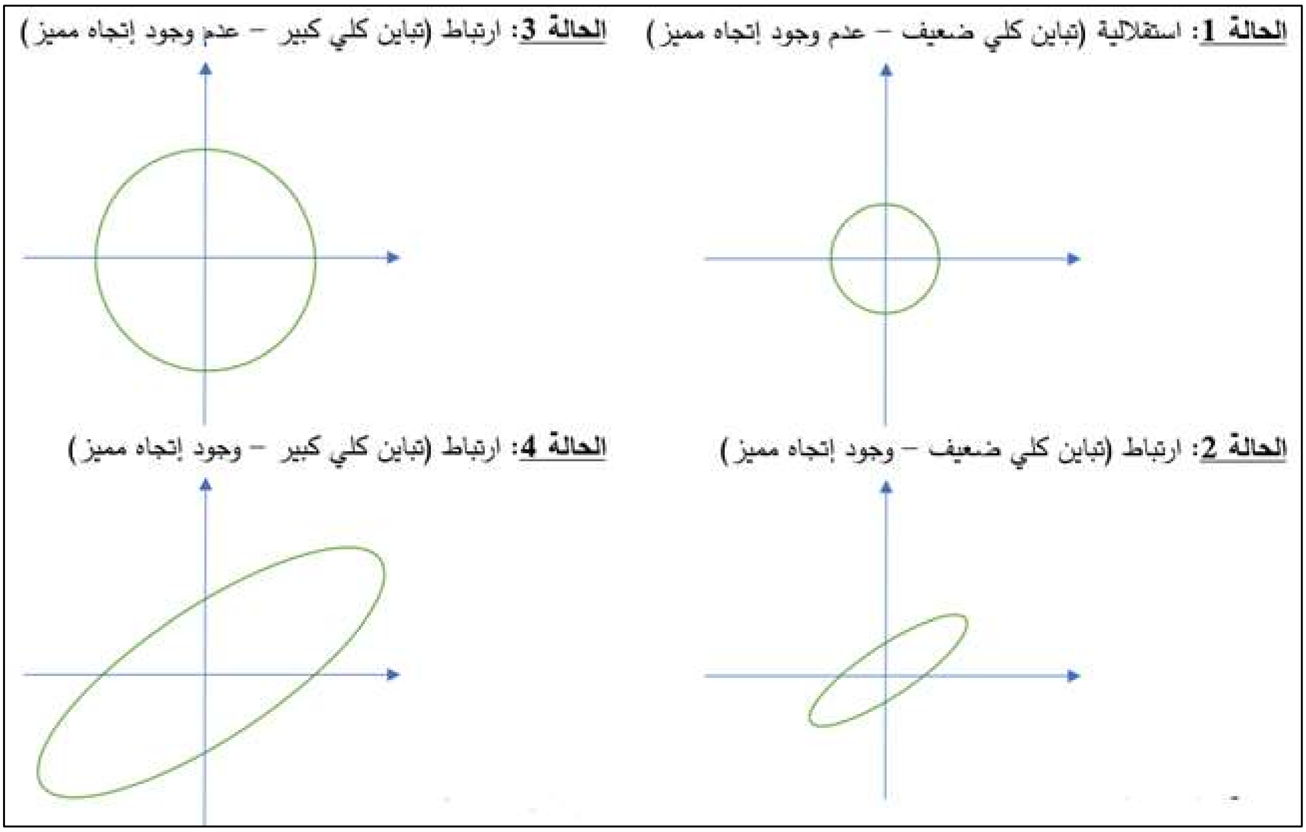
- صيغة علاقة الانتقال بين الفضاءين ($\mathbb{R}^p$ و $\mathbb{R}^n$): $\begin{cases} \mu_\alpha = \frac{1}{\sqrt{\lambda_\alpha}}.F^t.D_n^{-1}.\nu_\alpha \\ \nu_\alpha = \frac{1}{\sqrt{\lambda_\alpha}}.F.D_p^{-1}.\mu_\alpha \end{cases}$.
- صيغة علاقة الاسقاط بين الفضاءين ($\mathbb{R}^p$ و $\mathbb{R}^n$): $\begin{cases} \psi_{\alpha_i} = \frac{1}{\sqrt{\lambda_\alpha}}.\sum_{j=1}^{p} \frac{f_{ij}}{f_{i.}}.\varphi_{\alpha_j} \\ \varphi_{\alpha_j} = \frac{1}{\sqrt{\lambda_\alpha}}.\sum_{i=1}^{n} \frac{f_{ij}}{f_{.j}}.\psi_{\alpha_i} \end{cases}$.

س. تحليل التمثيلات البيانية للأسطر والأعمدة:

- يتم تفسير النتائج من التمثيلات البيانية للأسطر أو للأعمدة بالاعتماد على المسافات بين نقاط الإسقاط، فالمسافات القريبة بين الفئات (سواء بالنسبة للأسطر أو الأعمدة) تدل على التوافق (التشابه)، والبعيدة تدل على عدم التوافق (الاختلاف).
- يتم تفسير النتائج من التمثيلات البيانية المشتركة للأسطر وللأعمدة بالاعتماد كذلك على المسافات بين نقاط الإسقاط، فالمسافة القريبة بين فئة من المتغير الأول وفئة من المتغير الثاني تدل على أن اجتماع الفئتين له تكرار كبير، والعكس.
- التمثيل البياني يسمح كذلك بتحليل الارتباطات بين المتغيرين من حيث التجاذب/التنافر (attraction/répulsion). كما هو مبين في الأشكال أدناه:



- مركز المستوي يمثل النقاط الحيادية، أي الاستقلالية التامة للفئات (وكأنها التكرارات النسبية النظرية).
- الفئات القريبة من المركز تكون قريبة من المتوسط بشكل عام. وهي التي تجعل من التباين الكلي لسحابة النقاط ضعيف. على العكس من الفئات البعيدة عن المركز.
- المتغيرات المستقلة تعطي شكل دائري لسحابة النقاط، وبالتالي عدم وجود إتجاه مميز (direction privilégiée) للمحاور العاملية.
- المتغيرات المترابطة تعطي شكلا أكثر توسعا لسحابة النقاط، مع وجود إتجاه مميز (direction privilégiée) للمحاور العاملية.
- يكون المتغيران مستقلان إذا كان التباين الكلي ضعيف (كلما اقتربت التكرارات النسبية للأسطر (أو للأعمدة) من متوسط السطر (المركز) (أي: $l_{ij} = g_i$)، كلما اقتربنا من الاستقلالية بين الفئات)، ولم يكن للتمثيل البياني إتجاه معين (يكون التمثيل البياني في شكل دائري forme sphérique)، وتكون جميع النقاط ممركزة حول مركز ثقل سحابة النقاط (**الحالة 1**).
- يكون المتغيران مترابطان إذا كان التباين الكلي كبير –بغض النظر عن وجود إتجاه مميز أو لا– (**الحالتين 3 و4**)، أو كان التباين الكلي ضعيف مع وجود إتجاه معين (يكون التمثيل البياني في شكل غير دائري forme non sphérique) (**الحالة 2**).
- اختبار كاي تربيع يسمح بالكشف عن الحالتين 3 و 4 فقط، ولكن لا يسمح بتبيان الحالة 2 وهذا ما يبرز أهمية التحليل العاملي التقابلي.



3. تطبيق طريقة التحليل العاملي التقابلي على برنامج SPSS:

المثال التطبيقي الذي سنعتمد عليه لتطبيق طريقة AFC على برنامج SPSS، يخص بيانات دراسة ميدانية قمت بها على عينة حجمها 48 طالب من طلبة جامعة البليدة 2 سنة 2018، وكانت تهدف إلى دراسة العلاقة بين الظروف الاجتماعية والاقتصادية للأسرة والنجاح الدراسي للأبناء، وقد اخترت من هذه الدراسة متغيرين كيفيين هما: المستوى التعليمي للأب (أمي، ابتدائي، متوسط، ثانوي، جامعي) والحالة المهنية للأب (عاطل عن العمل، عامل بسيط، عامل متوسط التأهيل، عامل عالي التأهيل).

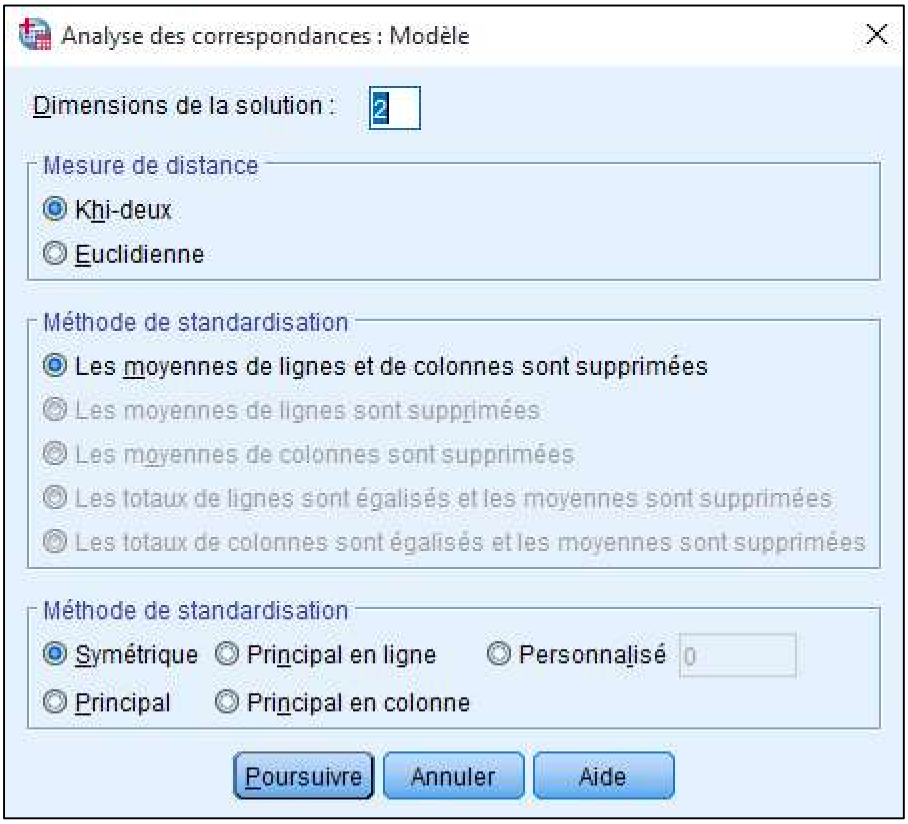
أ. الخطوات على برنامج SPSS:

للقيام بـ AFC على برنامج SPSS نتبع الخطوات التالية:

1) بعد فتح برنامج SPSS على الصفحة التي تحتوي بيانات الدراسة، نذهب إلى شريط القوائم.
2) في شريط القوائم نختار بالترتيب: Analyse ثم Réduction des dimensions ثم Analyse des correspondances.
3) فتظهر لنا النافذة التالية:

4) من خلال هذه النافذة نختار:

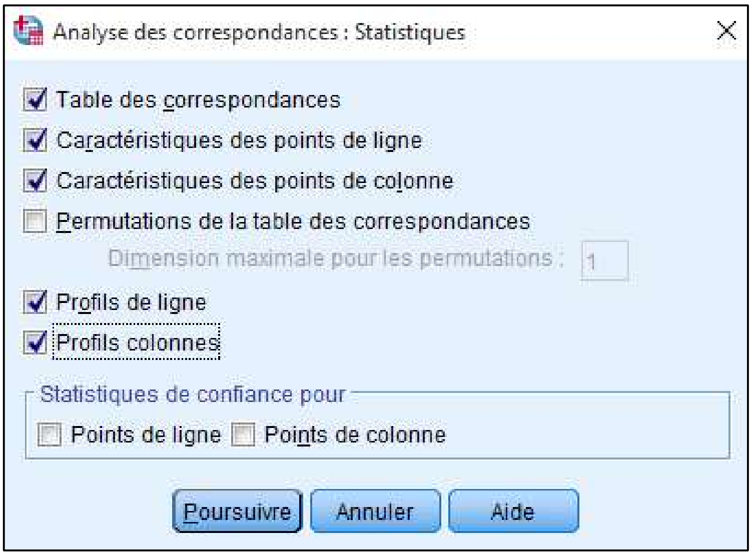
- ➢ **متغير الأسطر (ligne)**: نقوم بنقل متغير المستوى التعليمي إلى الأسطر مع تعريف فئاته (définir plage) من 1 إلى 5.
- ➢ **متغير الأعمدة (colonne)**: نقوم بنقل متغير الحالة المهنية للأب إلى الأعمدة مع تعريف فئاته (définir plage) من 1 إلى 4.
- ➢ **ثم نضغط على Modèle**: فتظهر لنا النافذة التالية:



من خلال هذه النافذة نختار:

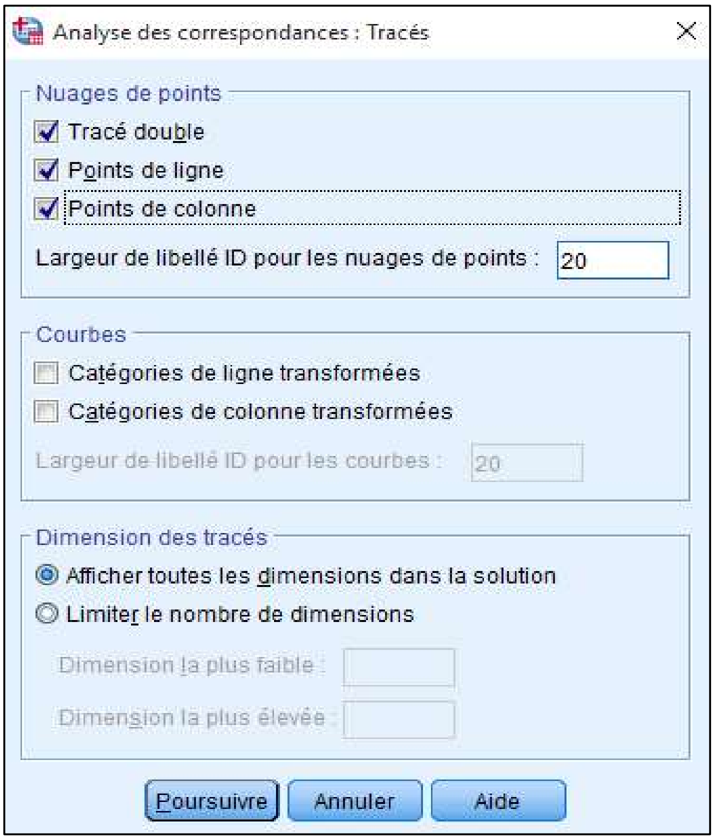
- ✓ أبعاد الحلول: نختار 2.
- ✓ مقياس المسافة: إما مقياس كاي تربيع أو مقياس إقليدي.
- ✓ نختار طريقة التناظر symétrique.

➢ ثم نضغط على **Statistiques**: فتظهر لنا النافذة التالية:



من خلال هذه النافذة نظلل على: جدول التوافقات، خواص نقاط الأسطر، خواص نقاط الأعمدة، احصائيات جانب الأسطر، إحصائيات جانب الأعمدة.

➢ ثم نضغط على **Tracés**: فتظهر لنا النافذة التالية:



من خلال هذه النافذة نظلل: سحابة نقاط الأسطر، سحابة نقاط الأعمدة، سحابة النقاط المشتركة للأسطر والأعمدة.

ب. مخرجات المثال التطبيقي والتعليق عليها:

1) الجدول المزدوج:

Table des correspondances

	الحالة المهنية للأب				
المستوى التعليمي للأب	عاطل عن العمل	عامل بسيط	عامل متوسط التأهيل	عامل عالي التأهيل	Marge active
أمي	2	1	0	0	3
ابتدائي	3	6	2	0	11
متوسط	1	5	8	0	14
ثانوي	2	0	4	5	11
جامعي	0	0	6	3	9
Marge active	8	12	20	8	48

من بين النقاط التي يمكن ملاحظتها من الجدول السابق أن:

➢ الفئة الأكثر تكرارا هي الفئة التي تجمع بين الصفتين "عامل متوسط التأهيل" و"المستوى التعليمي متوسط" وعددها 8 من أصل 48 فرد.

- عدد الأفراد الذين مستوى تعليم أبائهم "متوسط" هو 14 فرد (بغض النظر عن الحالة المهنية).
- عدد الأفراد الذين الحالة المهنية لأبائهم "متوسط التأهيل" هو 20 فرد (بغض النظر عن المستوى التعليمي).

2) جدول التكرارات النسبية للأسطر (Profils-ligne):

Profils lignes

	الحالة المهنية للأب				
المستوى التعليمي للأب	عاطل عن العمل	عامل بسيط	عامل متوسط التأهيل	عامل عالي التأهيل	Marge active
أمي	,667	,333	,000	,000	1,000
ابتدائي	,273	,545	,182	,000	1,000
متوسط	,071	,357	,571	,000	1,000
ثانوي	,182	,000	,364	,455	1,000
جامعي	,000	,000	,667	,333	1,000
Masse	,167	,250	,417	,167	

من بين النقاط التي يمكن ملاحظتها من الجدول السابق أن:

- 7,1 % من الأفراد الذين مستوى تعليم أبائهم "متوسط"، حالة أبائهم المهنية هي "عاطل عن العمل".
- 35,7 % من الأفراد الذين مستوى تعليم أبائهم "متوسط"، حالة أبائهم المهنية هي "عامل بسيط".
- 57,1 % من الأفراد الذين مستوى تعليم أبائهم "متوسط"، حالة أبائهم المهنية هي "عامل متوسط التأهيل".
- 0 % من الأفراد الذين مستوى تعليم أبائهم "متوسط"، حالة أبائهم المهنية هي "عامل عالي التأهيل".

3) جدول التكرارات النسبية للأعمدة (Profils-colonne):

Profils colonnes

	الحالة المهنية للأب				
المستوى التعليمي للأب	عاطل عن العمل	عامل بسيط	عامل متوسط التأهيل	عامل عالي التأهيل	Masse
أمي	,250	,083	,000	,000	,063
ابتدائي	,375	,500	,100	,000	,229
متوسط	,125	,417	,400	,000	,292
ثانوي	,250	,000	,200	,625	,229
جامعي	,000	,000	,300	,375	,188
Marge active	1,000	1,000	1,000	1,000	

من بين النقاط التي يمكن ملاحظتها من الجدول السابق أن:

- 0 % من الأفراد الذين حالة أبائهم المهنية "عامل متوسط التأهيل"، مستوى تعليم أبائهم هو "أمي".
- 10 % من الأفراد الذين حالة أبائهم المهنية "عامل متوسط التأهيل"، مستوى تعليم أبائهم هو "ابتدائي".
- 40 % من الأفراد الذين حالة أبائهم المهنية "عامل متوسط التأهيل"، مستوى تعليم أبائهم هو "متوسط".

- 20 % من الأفراد الذين حالة أبائهم المهنية "عامل متوسط التأهيل"، مستوى تعليم أبائهم هو "ثانوي".
- 30 % من الأفراد الذين حالة أبائهم المهنية "عامل متوسط التأهيل"، مستوى تعليم أبائهم هو "جامعي".

4) جدول القيم الذاتية والتباين المفسر:

Récapitulatif

Dimension	Valeur singulière	Inertie	Khi-deux	Sig.	Proportion d'inertie: Représentation	Proportion d'inertie: Cumulé	Valeur singulière de confiance: Ecart type	Valeur singulière de confiance: Corrélation 2
1	,694	,481			,686	,686	,069	-,197
2	,446	,199			,284	,970	,125	
3	,145	,021			,030	1,000		
Total		,701	33,658	,001[a]	1,000	1,000		

a. 12 degrés de liberté

من بين النقاط التي يمكن ملاحظتها من الجدول السابق أن:

- ترتيب القيم الذاتية الثلاثة هي: $\lambda_1 = 0{,}481$، $\lambda_2 = 0{,}199$، $\lambda_3 = 0{,}021$.
- التباين الكلي هو: $I = \sum_{i=1}^{3} \lambda_i = 0{,}701$.
- إحصائية كاي تربيع المحسوبة هي: $\chi^2 = n_{..} \times I = 48 \times 0{,}701 = 33{,}65$.

- بما أن قيمة إحصائية كاي تربيع المحسوبة $\chi_c^2 = 33{,}65$ أكبر من المجدولة $\chi_{0,05}^2(4*3) = 21{,}03$ (كما أن $Sig = 0{,}01 < 0{,}05$)، فإنا نرفض الفرضية الصفرية H_0، ونقبل الفرضية البديلة التي تقضي بوجود علاقة بين المتغيرين (المتغيران "المستوى التعليمي للأب" و"الحالة المهنية للأب" مترابطان).
- المحور العاملي الأول يفسر 68,6 % من التباين الكلي.
- المحور العاملي الثاني يفسر 28,4 % من التباين الكلي.
- المستوي العاملي الأول يفسر 97 % من التباين الكلي.

5) اسقاط نقاط الأسطر:

Présentation des points de ligne[a]

المستوى التعليمي للأب	Masse	Score de la dimension 1	Score de la dimension 2	Inertie	Contribution: Du point vers l'inertie de la dimension 1	Du point vers l'inertie de la dimension 2	De la dimension vers l'inertie du point 1	De la dimension vers l'inertie du point 2	Total
أمي	,063	-1,157	1,510	,132	,121	,320	,440	,482	,922
ابتدائي	,229	-,987	,144	,164	,322	,011	,944	,013	,957
متوسط	,292	-,302	-,757	,095	,038	,375	,196	,788	,984
ثانوي	,229	,880	,690	,173	,256	,245	,710	,281	,992
جامعي	,188	,987	-,346	,138	,263	,050	,922	,073	,995
Total actif	1,000			,701	1,000	1,000			

a. Normalisation symétrique

من بين النقاط التي يمكن ملاحظتها من الجدول السابق أن:

- المجموع الهامشي للفئة "أمي" هي: $f_{1.} = 0{,}063$ (احتمال أن يكون المستوى التعليمي للأب "أمي").
- احداثيات الفئة "أمي" على المستوي العاملي الأول هي: $M_1(-1{,}157 \quad , \quad 1{,}510)$.

6) اسقاط نقاط الأعمدة:

Présentation des points de colonne[a]

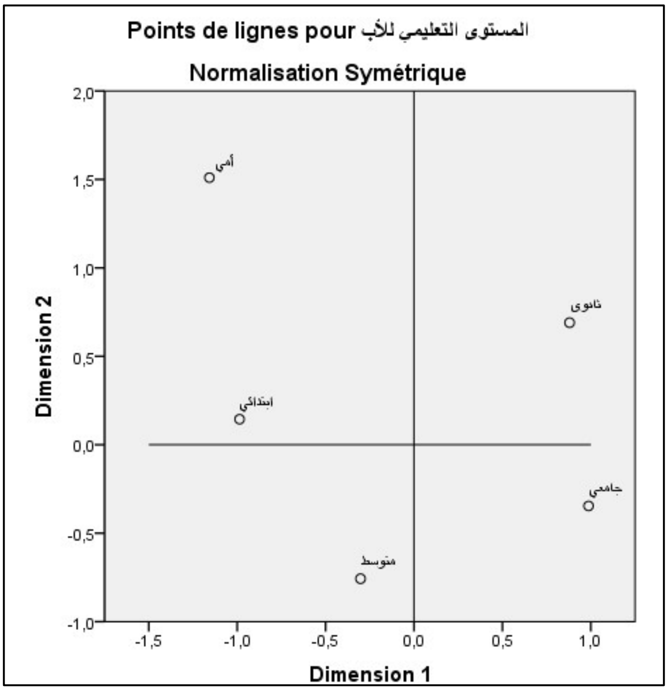
الحالة المهنية للأب	Masse	Score de la dimension		Inertie	Contribution				
					Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point		
		1	2		1	2	1	2	Total
عاطل عن العمل	,167	-,688	1,142	,157	,114	,487	,349	,619	,969
عامل بسيط	,250	-1,032	-,263	,199	,384	,039	,927	,039	,965
عامل متوسط التأهيل	,417	,364	-,570	,103	,080	,303	,372	,587	,959
عامل عالي التأهيل	,167	1,326	,676	,242	,423	,171	,839	,140	,979
Total actif	1,000			,701	1,000	1,000			

a. Normalisation symétrique

من بين النقاط التي يمكن ملاحظتها من الجدول السابق أن:

- المجموع الهامشي للفئة "عاطل عن العمل" هي: $f_{.1} = 0{,}167$ (احتمال أن تكون الحالة المهنية للأب "عاطل عن العمل").
- احداثيات الفئة "عاطل عن العمل" على المستوي العاملي الأول هي: $N_1(-0{,}688 \quad , \quad 1{,}142)$.

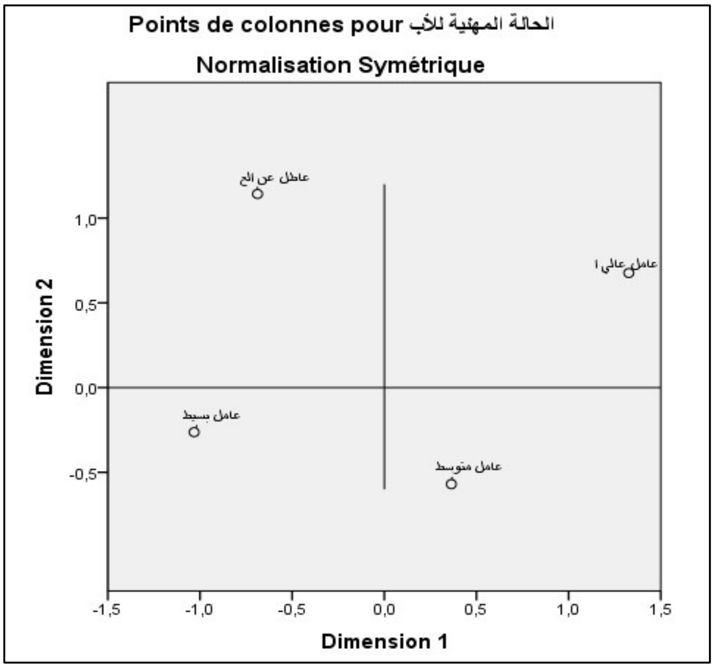
7) التمثيل البياني لنقاط الأسطر:



يُظهر التمثيل البياني السابق أن:

- المسافة بين الفئتين "جامعي" و"ثانوي" قريبة وفي نفس الجهة للمحور العاملي، وهذا يدل على تشابه هاتين الفئتين، وهما مختلفتين عن الفئتين "أمي" و"ابتدائي" اللتان تمثلان بالجهة المقابلة للمحور العاملي.
- قرب الفئة "متوسط" من المركز، يدل على أن هذه الفئة قريبة من المتوسط بشكل عام.

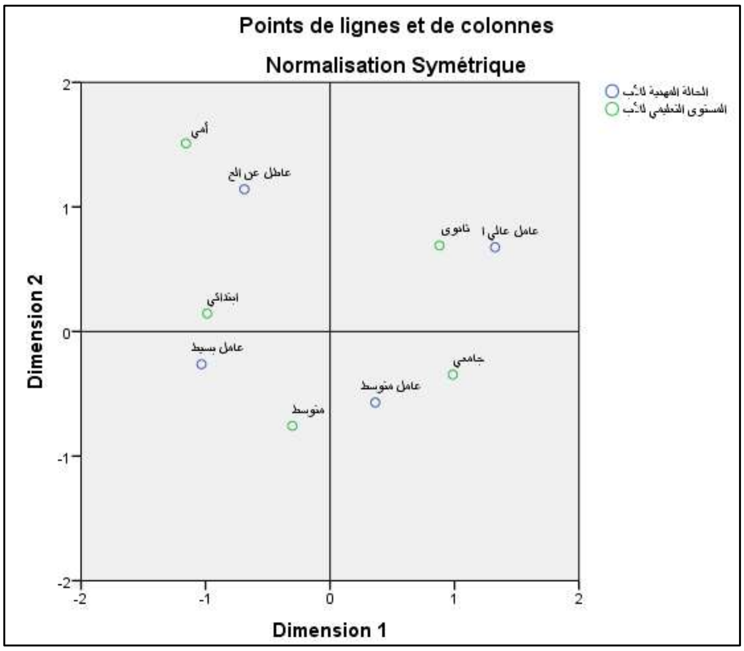
8) التمثيل البياني لنقاط الأعمدة:



يُظهر التمثيل البياني السابق أن:

- المسافة بين الفئة "عامل عالي التأهيل" والفئتين "عاطل عن العمل" و"عامل بسيط" بعيدة وفي جهتين مختلفتين من المحور العاملي، وهذا يدل على الاختلاف بين فئة "عالي التأهيل" وهاتين الفئتين.
- قرب الفئة "عامل متوسط التأهيل" من المركز، يدل على أن هذه الفئة قريبة من المتوسط بشكل عام.

9) التمثيل البياني المشترك لنقاط الأسطر ونقاط الأعمدة:



يُظهر التمثيل البياني السابق أن:

- المسافة بين فئة "عاطل عن العمل" والفئة "أمي" قريبة وفي نفس الجهة من المحور العاملي، وهذا يدل على أن غالب العاطلين عن العمل أميين.
- المسافة بين فئة "عامل بسيط" والفئة "ابتدائي" قريبة وفي نفس الجهة من المحور العاملي، وهذا يدل على أن غالب العمال البسطاء مستواهم التعليمي "ابتدائي".
- المسافة بين فئة "عامل عالي التأهيل" والفئة "ثانوي" قريبة وفي نفس الجهة من المحور العاملي، وهذا يدل على أن غالب العمال عاليي التأهيل مستواهم التعليمي "ثانوي".
- المسافة بين فئة "عامل متوسط التأهيل" والفئة "جامعي" قريبة وفي نفس الجهة من المحور العاملي، وهذا يدل على أن غالب العمال متوسطي التأهيل مستواهم التعليمي "جامعي".
- هاتين الملاحظتين الأخيرتين يبرزان نوع من عدم التناسب في سوق العمل، أين يتقلد العمال الذين مستواهم ثانوي مهام عالية (ربما بحكم الأقدمية والخبرة)، بينما يتقلد العمال الذين مستواهم جامعي مهام متوسطة التأهيل (ربما بسبب عدم توفر فرص عمل تناسب مستواهم، فيضطرون إلى قبول مناصب أقل من مستواهم).
- بشكل عام نلاحظ توافق (ارتباط) بين المستوى التعليمي والحالة المهنية، فكلما ارتفع المستوى التعليمي تحسنت نوعية العمل.
- شكل سحابة النقاط أقرب للدائري مع تباين كلي كبير نسبيا (الحالة 3)، وهذا ما يؤكد على أن المتغيرين مترابطان.

4. تطبيق طريقة التحليل العاملي التقابلي على برنامج XL-STAT:

سأقدم في هذا العنصر أهم الخطوات المتبعة على برنامج XL-STAT للقيام بتحليل عاملي توافقي، على نفس بيانات المثال التطبيقي على برنامج SPSS. وسأكتفي بالتعليق فقط على بعض الإضافات غير المتوفرة على برنامج SPSS.

أ. الخطوات على برنامج XL-STAT:

للقيام بـ AFC على برنامج XL-STAT نتبع الخطوات التالية:

1) بعد فتح برنامج XL-STAT نقوم بإدخال بيانات الدراسة، ثم نذهب إلى شريط القوائم.

2) في شريط القوائم نختار بالترتيب: XL-STAT ثم Analyse des données ثم Analyse factorielle des correspondances.

3) فتظهر لنا النافذة التالية:

4) من خلال هذه النافذة نختار:

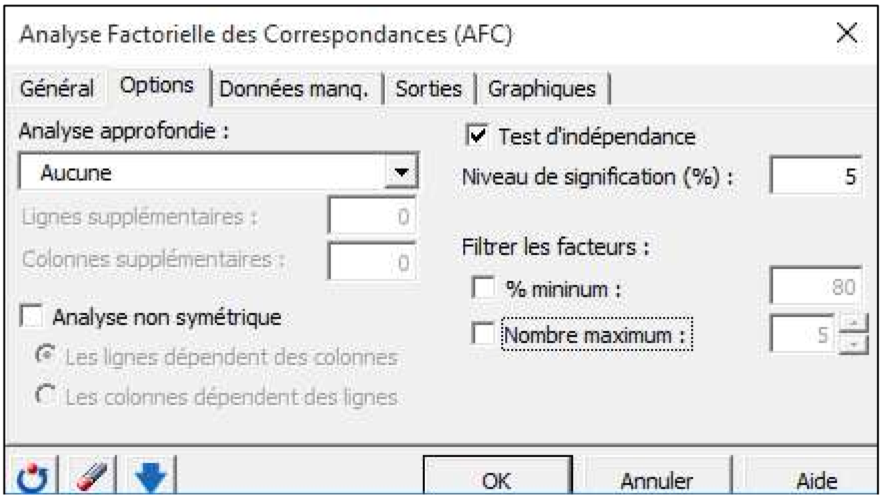
❖ **من عام Général**: نقوم باختيار:

- ➢ نوع جدول البيانات (جدول مزدوج أو جدول بيانات خام)
- ➢ مكان إدراج المخرجات (plage، feuille، classeur)
- ➢ نظلل على وصف البيانات (Libellés).

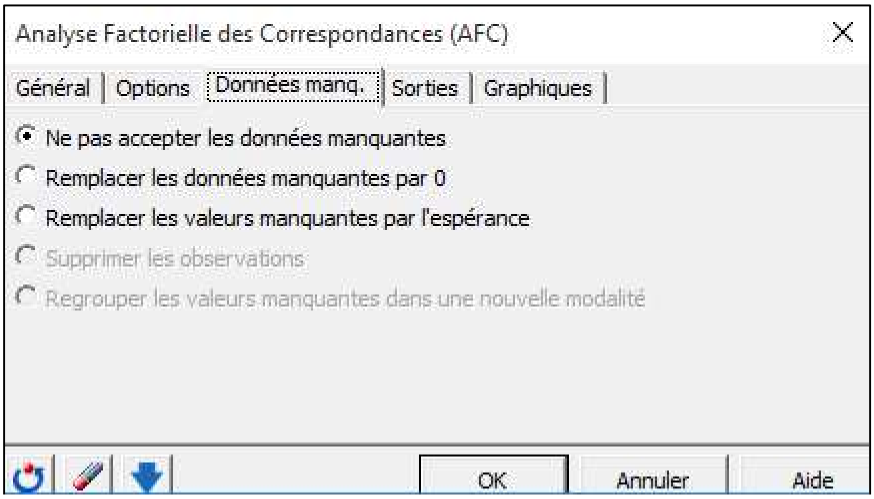
❖ **من خيارات Options**: نقوم باختيار:

- ➢ تحليل معمق (بدون تحليل معمق، بيانات إضافية، تحليل مجموعات جزئية).
- ➢ نظلل على اختبار الإستقلالية، عند مستوى المعنوية 5%.
- ➢ إمكانية تحديد عدد المحاور العاملية (إما بالنسبة المفسرة للمحور أو بعدد المحاور).

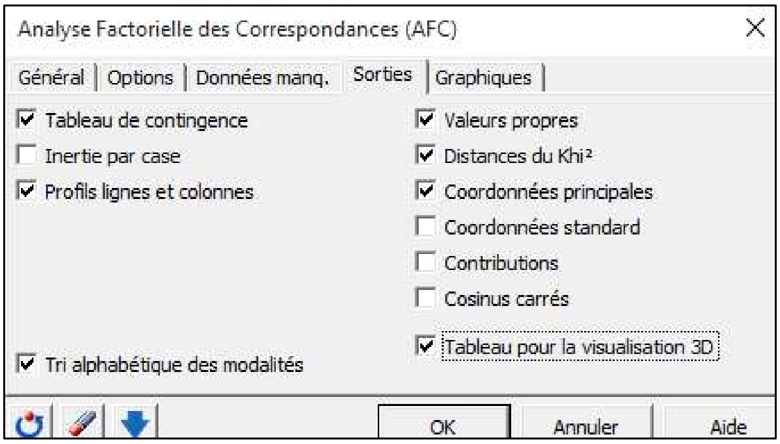
➢ إمكانية القيام بتحليل غير متناظر.



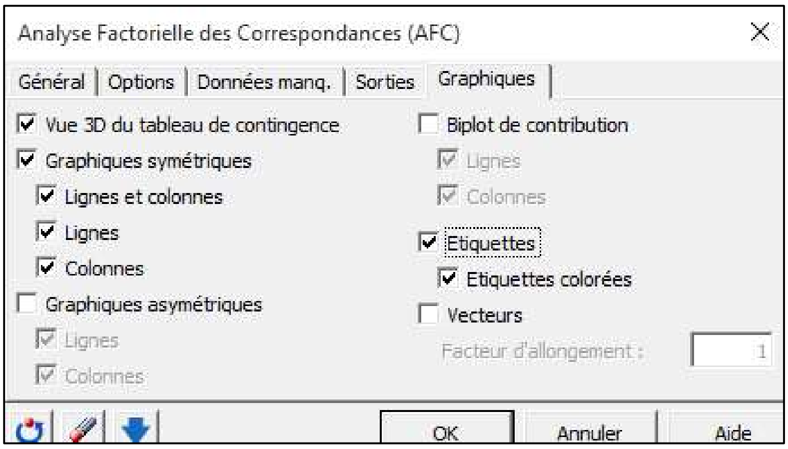
❖ من بيانات مفقودة **Données manquantes**: نقوم بتحديد إمكانية وجود قيم مفقودة وكيفية التعامل معها في حالة وجودها (استبدالها بالصفر أو بالتوقع).



❖ من مخرجات **Sorties**: نقوم بتحديد المخرجات التي نرغب من البرنامج إظهارها:



❖ **من التمثيل البياني Graphiques**: نقوم بتحديد التمثيلات البيانية التي نرغب من البرنامج إظهارها



ب. مخرجات المثال التطبيقي والتعليق عليها:

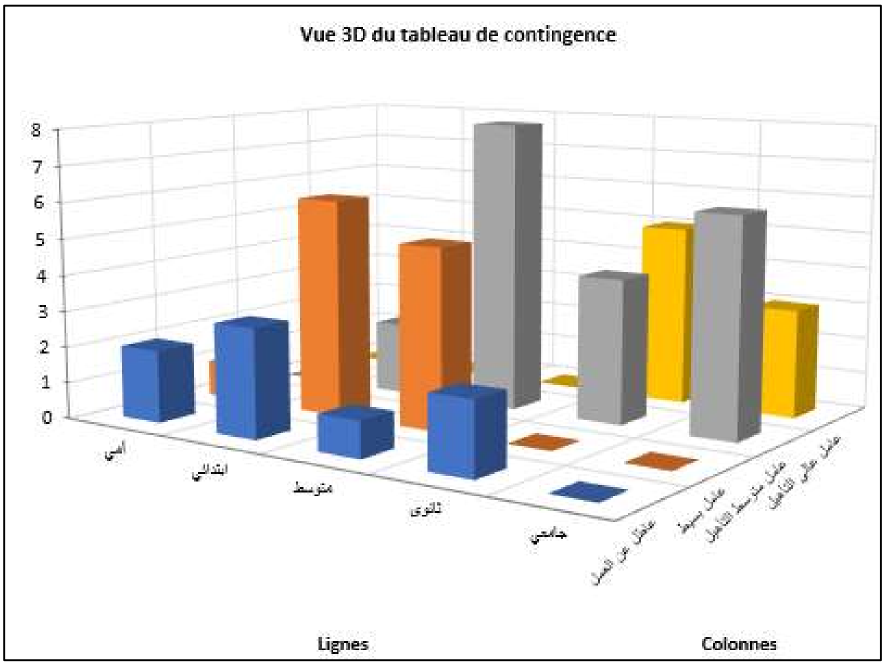
نطبق طريقة AFC على برنامج XL-STAT بنفس المثال المطبق على برنامج SPSS. وبالنسبة للتعليق على الجداول والأشكال، فهو نفس التعليق المذكور في العنصر السابق الخاص ببرنامج SPSS، فلا داعي لتكرار ذلك. ونكتفي فقط بالإشارة إلى المخرجات غير المتوفرة على برنامج SPSS.

1) الجدول المزدوج:

Tableau de contingence :				
	عاطل عن العمل	عامل بسيط	مل متوسط التأهي	عامل عالي التأهيل
أمي	2,0000	1,0000	0,0000	0,0000
ابتدائي	3,0000	6,0000	2,0000	0,0000
متوسط	1,0000	5,0000	8,0000	0,0000
ثانوي	2,0000	0,0000	4,0000	5,0000
جامعي	0,0000	0,0000	6,0000	3,0000

2) الجدول والتمثيل البياني ذو ثلاث أبعاد للجدول المزدوج (غير متوفر على برنامج SPSS):

Tableau pour la visualisation 3D :					
Modalité	Type	F1	F2	F3	Somme(Contributions)
أمي	Ligne	0,9639	1,0088	0,4054	0,9283
ابتدائي	Ligne	0,8218	0,0965	-0,1760	0,6699
متوسط	Ligne	0,2519	-0,5058	0,0724	0,4858
ثانوي	Ligne	-0,7326	0,4611	-0,0782	0,5669
جامعي	Ligne	-0,8222	-0,2310	0,0630	0,3491
عاطل عن العمل	Colonne	0,5731	0,7630	0,1712	0,8333
عامل بسيط	Colonne	0,8595	-0,1757	-0,1660	0,7500
عامل متوسط التأهيل	Colonne	-0,3031	-0,3804	0,1007	0,5833
عامل عالي التأهيل	Colonne	-1,1046	0,4517	-0,1740	0,8333



يتبين من هذا التمثيل البياني بوضوح أن الفئة الأكثر تكرارا هي الفئة التي تجمع بين الصفتين "عامل متوسط التأهيل" و"المستوى التعليمي متوسط".

3) **جدول التكرارات النسبية** (غير متوفر على برنامج SPSS):

Inertie par case :				
	عاطل عن العمل	عامل بسيط	عامل متوسط التأهي	عامل عالي التأهيل
أمي	0,09375	0,00174	0,02604	0,01042
ابتدائي	0,01547	0,08002	0,03033	0,03819
متوسط	0,01587	0,01339	0,01677	0,04861
ثانوي	0,00032	0,05729	0,00155	0,11395
جامعي	0,03125	0,04688	0,02813	0,03125

4) **اختبار الاستقلالية** (يعطي القيمة الجدولية لكاي تربيع وهذا غير متوفر على برنامج SPSS):

Test d'indépendance entre les lignes et les colonnes :

Khi² (Valeur	33,6580
Khi² (Valeur	21,0261
DDL	12
p-value	0,0008
alpha	0,05

Interprétation du test :

H0 : Les lignes et les colonnes du tableau sont indépendantes.

Ha : Il existe un lien entre les lignes et les colonnes du tableau.

Etant donné que la p-value calculée est inférieure au niveau de signification alpha=0,05, on doit rejeter l'hypothèse nulle H0, et retenir l'hypothèse alternative Ha.

Le risque de rejeter l'hypothèse nulle H0 alors qu'elle est vraie est inférieur à 0,08%.

بمقارنة قيمة إحصائية كاي تربيع المحسوبة $\chi_c^2 = 33{,}65$ بالقيمة المجدولة $\chi_{0,05}^2(4*3) = 21{,}03$ (كما أن $Sig = 0{,}00 < 0{,}05$)، فإنا نرفض الفرضية الصفرية H_0، ونقبل الفرضية البديلة التي تقضي بوجود علاقة بين المتغيرين (المتغيران "المستوى التعليمي للأب" و"الحالة المهنية للأب" مترابطان).

5) جدول التكرارات النسبية للأسطر (Profils-ligne):

Profils (lignes) :					
	عاطل عن العمل	عامل بسيط	عامل متوسط التأهي	امل عالي التأهيل	Somme
أمي	0,6667	0,3333	0,0000	0,0000	1
ابتدائي	0,2727	0,5455	0,1818	0,0000	1
متوسط	0,0714	0,3571	0,5714	0,0000	1
ثانوي	0,1818	0,0000	0,3636	0,4545	1
جامعي	0,0000	0,0000	0,6667	0,3333	1
Moyenne	0,2385	0,2472	0,3567	0,1576	1

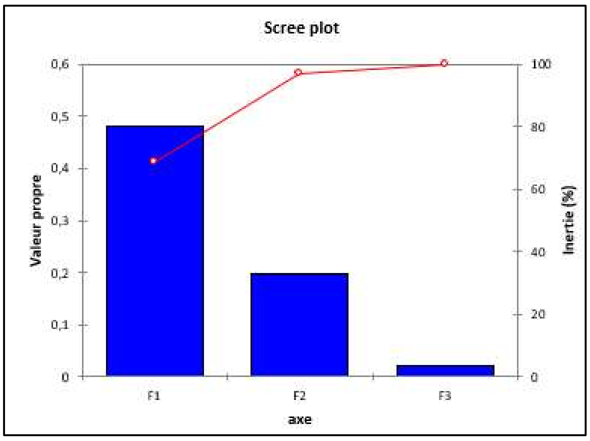
6) جدول التكرارات النسبية للأعمدة (Profils-colonne):

Profils (colonnes) :					
	عاطل عن العمل	عامل بسيط	عامل متوسط التأهي	امل عالي التأهيل	Moyenne
أمي	0,2500	0,0833	0,0000	0,0000	0,0833
ابتدائي	0,3750	0,5000	0,1000	0,0000	0,2438
متوسط	0,1250	0,4167	0,4000	0,0000	0,2354
ثانوي	0,2500	0,0000	0,2000	0,6250	0,2688
جامعي	0,0000	0,0000	0,3000	0,3750	0,1688
Somme	1	1	1	1	1

7) جدول القيم الذاتية والتباين المفسر:

Inertie totale :		0,7012	
Valeurs propres et pourcentages d'inertie :			
	F1	F2	F3
Valeur propr	0,4811	0,1991	0,0210
Inertie (%)	68,6098	28,3893	3,0009
% cumulé	68,6098	96,9991	100,0000

8) التمثيل البياني للقيم الذاتية (غير متوفر على برنامج SPSS):



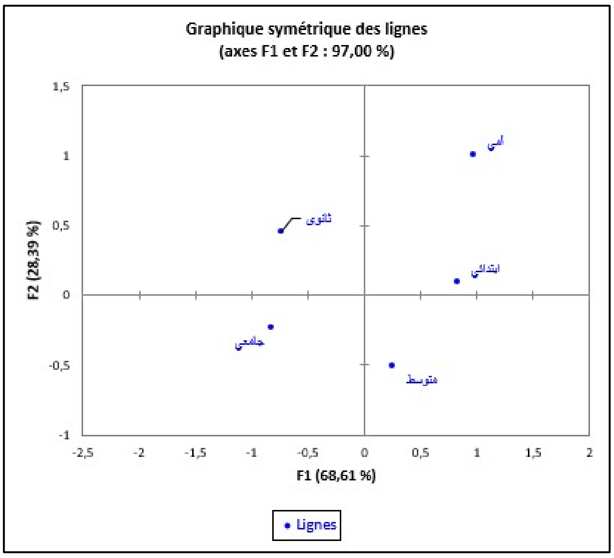
9) مسافات كاي تربيع للأسطر (غير متوفر على برنامج SPSS):

Distances du Khi² (lignes) :	أمي	ابتدائي	متوسط	ثانوي	جامعي
أمي	0	1,0911	1,7064	1,8471	2,2010
ابتدائي	1,0911	0	0,8656	1,5996	1,6933
متوسط	1,7064	0,8656	0	1,3880	1,1087
ثانوي	1,8471	1,5996	1,3880	0	0,7120
جامعي	2,2010	1,6933	1,1087	0,7120	0

10) اسقاطات الأسطر:

Coordonnées principales (lignes) :	F1	F2	F3
أمي	0,9639	1,0088	0,4054
ابتدائي	0,8218	0,0965	-0,1760
متوسط	0,2519	-0,5058	0,0724
ثانوي	-0,7326	0,4611	-0,0782
جامعي	-0,8222	-0,2310	0,0630

11) التمثيل البياني لنقاط الأسطر:



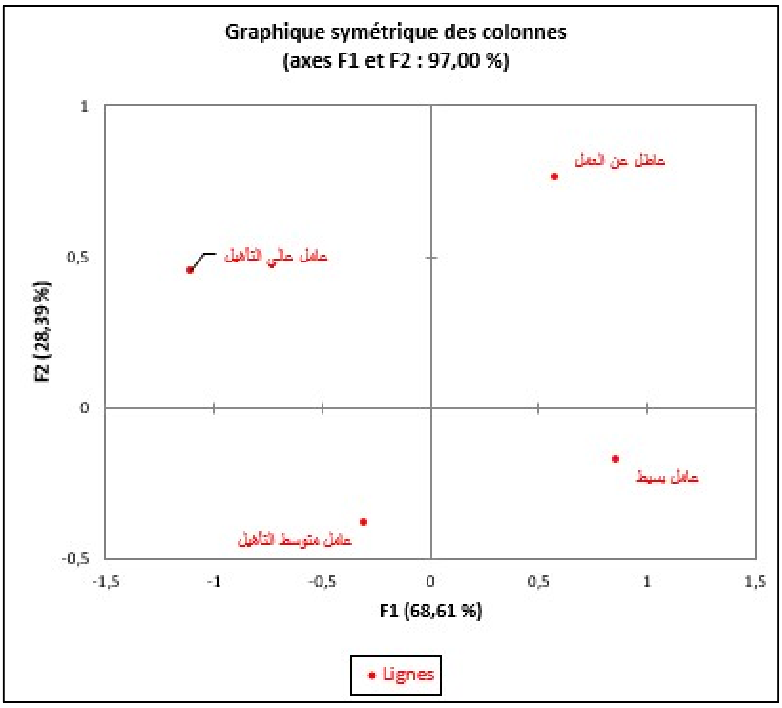
12) مسافات كاي تربيع للأعمدة (غير متوفر على برنامج SPSS):

Distances du Khi² (colonnes) :				
	عاطل عن العمل	عامل بسيط	عامل متوسط التأهيل	عامل عالي التأهيل
عاطل عن العمل	0	1,0378	1,4423	1,7409
عامل بسيط	1,0378	0	1,2103	2,0620
عامل متوسط التأهيل	1,4423	1,2103	0	1,1876
عامل عالي التأهيل	1,7409	2,0620	1,1876	0

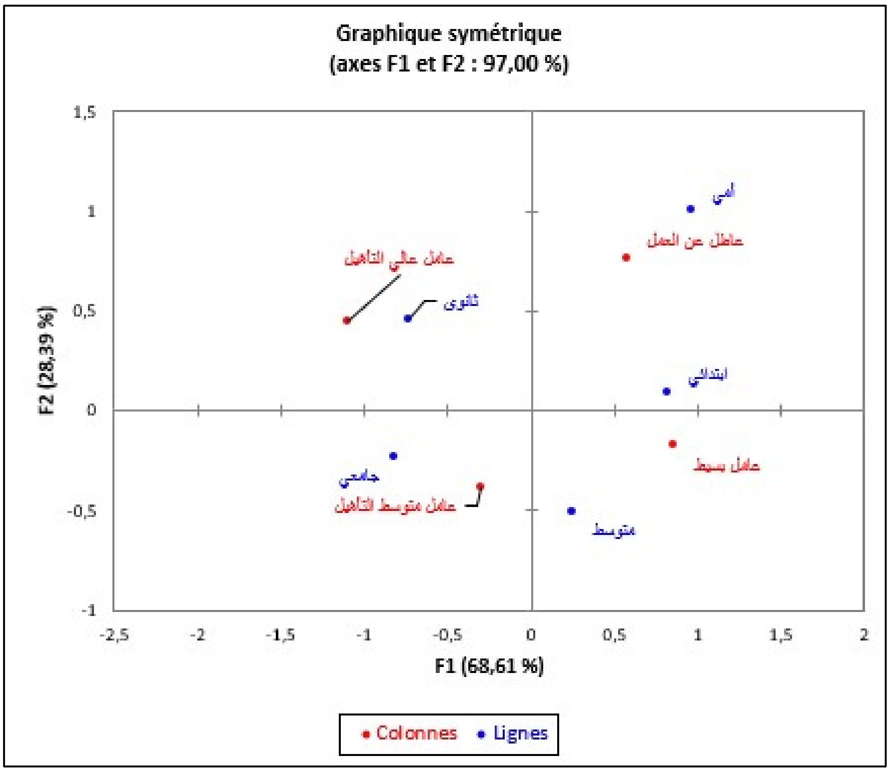
13) اسقاطات الأعمدة:

Coordonnées principales (colonnes) :			
	F1	F2	F3
عاطل عن العمل	0,5731	0,7630	0,1712
عامل بسيط	0,8595	-0,1757	-0,1660
عامل متوسط التأهيل	-0,3031	-0,3804	0,1007
عامل عالي التأهيل	-1,1046	0,4517	-0,1740

14) التمثيل البياني لنقاط الأعمدة:



15) التمثيل البياني المشترك لنقاط الأسطر ونقاط الأعمدة:



كما أشرنا إلى ذلك في برنامج SPSS، نلاحظ بشكل عام توافق (ارتباط) بين المستوى التعليمي والحالة المهنية، فكلما ارتفع المستوى التعليمي تحسنت نوعية العمل. وشكل سحابة النقاط أقرب للدائري مع تباين كلي كبير نسبيا (الحالة 3)، وهذا ما يؤكد على أن المتغيرين مترابطان.

III. المحور الثالث: التصنيف التسلسلي (CAH)

غالبا ما تركز الطرق العاملية على اختزال (تقليص) عدد المتغيرات، أما أساليب التصنيف فتركز في الغالب على تصنيف الأفراد في مجموعات (تقليص عدد الأفراد).

في الحقيقة يوجد تكامل بين الطرق العاملية وطرق التصنيف، فبعد تطبيق طريقة عاملية على البيانات الأصلية (كطريقة تحليل المركبات الرئيسية)، يتم اختزال المتغيرات في عدد محدود من المركبات (مركبتين أو ثلاث مركبات رئيسية)، وعلى أساس هذه المركبات نقوم بتصنيف الأفراد، والذي يسهل عملية التحليل والتفسير في الأخير.

يمكن تقسيم أساليب التصنيف إلى قسمين:

- **التصنيف بالتجزئة:** ويتمثل في تجزئة المجموعة الكلية للأفراد في مجموعات جزئية منفصلة. وهو تصنيف غير متسلسل. ومن بين أهل هذه الأساليب: طريقة المراكز المتحركة (Centres mobiles)، طريقة nuées dynamiques، طريقة k-means.
- **التصنيف التسلسلي:** ويتمثل في تجميع الأفراد في مجموعات متجانسة بأسلوب تسلسلي (التجميع المتتالي للأفراد المتشابهة). ومن بين أهل هذه الأساليب: طريقة **التصنيف التسلسلي التصاعدي أو التجميعي** (Classification Ascendante Hiérarchique-CAH)، طريقة التصنيف التسلسلي التنازلي أو التقسيمي (CDH).

وسنركز في هذا المحور على التصنيف التسلسلي التصاعدي (أو التجميعي)، والذي يعتبر أهم طرق التصنيف وأكثرها انتشارا واستخداما، وهو تصنيف يناسب الدراسات التي تضم عدد أفراد صغير نسبيا (العينات الصغيرة: 50-150 فرد).

1. مفهوم طريقة التصنيف التسلسلي (الهرمي – العنقودي) التصاعدي (التجميعي):

أ. تعريف طريقة التصنيف التسلسلي التصاعدي:

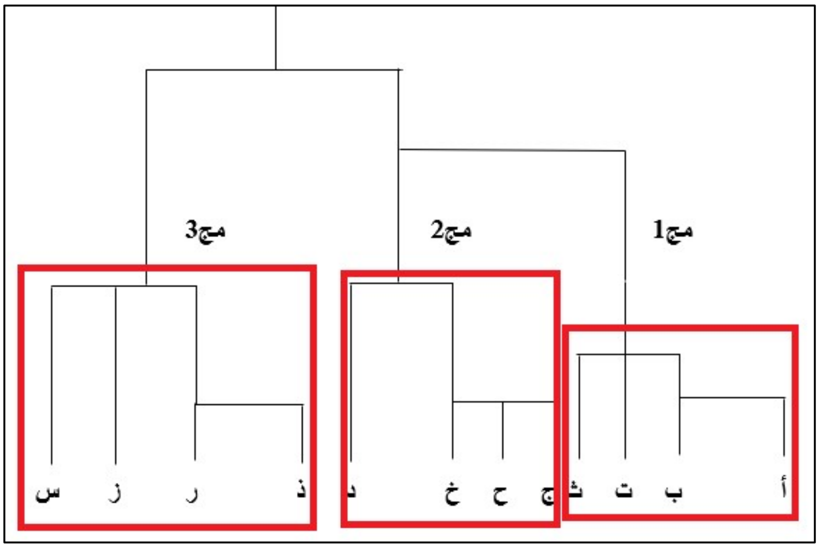
التصنيف هنا لا يقصد به الترتيب من حيث الأفضل أو الأسوء، بل تجميع الأفراد المتشابهة في مجموعات بالاعتماد على المسافات بينها. فكلما كانت المسافة بل فردين صغيرة فهذا يدل على أنهما متشابهان فيصنفان في نفس المجموعة. ويمكن أن تضم المجموعة فردا واحدا فقط، بحيث يكون غير مشابه لأي فرد من الأفراد الأخرين.

في البداية نكون أمام عدة عناصر أين يشكل كل عنصر مجموعة لوحده، فيتم ترتيبها بحيث نقوم بالجمع بين العناصر المتشابهة (القريبة) بشكل تسلسلي حتى تجتمع جميع العناصر في مجموعة واحدة فقط تمثل الجذر (la racine) (الصعود من عدة مجموعات إلى مجموعة واحدة بشكل تسلسلي).

ومع أن طرق التصنيف غالبا ما تركز على الأفراد لتصنيفها، إلا أنه يمكن كذلك تصنيف المتغيرات قصد تقليص عددها، فقد تكون نفس المتغيرات (متغيرين أو أكثر) تفسر ظاهرة معينة بنفس القدر، فبالتصنيف تجمع هذه المتغيرات في مجموعة واحدة، ثم اختيار متغير لتمثيل المجموعة. وطريقة التصنيف التسلسلي الصاعد تتناسب أفضل مع المتغيرات الكمية.

ب. مبدأ طريقة التصنيف التسلسلي التصاعدي:

يتمثل مبدأ التصنيف التسلسلي التصاعدي في التجميع التسلسلي للعناصر المتشابهة، بحيث نبدأ بتجميع الأفراد المتشابهة مثنى مثنى بالاعتماد على المسافات بينها (كل فردين قريبين هما متشابهان فيجمعان)، ثم نكرر العملية بشكل تسلسلي، وفي كل مرة نصعد إلى مستوى أعلى لضم أفراد أخرين للمجموعة، فيزيد عدد الأفراد في المجموعة الواحدة، ويقل عدد المجموعات بشكل تسلسلي تصاعدي. وتقوم هذه الطريقة على مبدأ تعظيم التجانس (التشابه) بين الأفراد المنتمين لنفس المجموعة، وفي نفس الوقت تعظيم عدم التجانس (الاختلاف) بين المجموعات. ونشير إلى أن تجميع العناصر القريبة (المتشابهة) قد يكون جمع فرد مع فرد أو جمع مجموعة مع مجموعة أو جمع فرد مع مجموعة. والتصنيف التسلسلي يمثل بواسطة الشجرة التسلسلية أو شجرة التصنيف (Dendrogramme – Arbre de classification).



لو نقوم بقطع (coupure) بخط أفقي على الشجرة التسلسلية، فإننا سنتحصل على تقسيم أو تجزئة (partition) للعناصر في مجموعات، فإننا كلما قمنا بالقطع أعلى الشجرة فسنتحصل على أقل عدد من المجموعات، وتكون هذه المجموعات أقل تجانس، وبالعكس كلما قمنا بالقطع أدنى الشجرة فسنتحصل على أكبر عدد من المجموعات، وتكون هذه المجموعات أكثر تجانسا. يمثل التمثيل البياني أعلاه تصنيف الأفراد (أ، ب، ......، س) في ثلاث مجموعات.

2. خطوات إجراء طريقة التصنيف التسلسلي التصاعدي:

أ. جمع البيانات وتشكيل جدول البيانات:

بعد جمع البيانات حول الظاهرة المراد تصنيف عناصرها، نقوم بتشكيل جدول البيانات، والذي سنعتمد عليه في عملية التصنيف.

وجدول البيانات يضم n سطر تمثل الأفراد، وp عمود تمثل المتغيرات (غالبا ما تكون كمية).

المتغيرات / الأفراد	X_1	…	X_j	…	X_p
1	x_{11}				
2					
…					
i			x_{ij}		
…					
n					x_{np}

حيث: X_j يمثل المتغير j $(j = 1,2, \dots , p)$.

و: الأعداد من 1 إلى n تمثل الأفراد (الحالات).

و: x_{ij} تمثل قيمة المتغير X_j عند الفرد i.

وتجدر الإشارة إلى أنه في حالة قيم بيانات متباينة، فيفضل تحويلها إلى بيانات مركزية (جعل متوسطها معدوم)، أما إذا كانت قيم البيانات متباينة ووحدات قياسها مختلفة فيفضل تحويلها إلى بيانات معيارية (متوسط معدوم وتباين يساوي واحد).

ينصح بتطبيق طرق التصنيف بعد التحليل العاملي، فيمكن أن تكون المجموعات متغيرات إضافية في تحليل المركبات الرئيسية (ACP)، التحليل العاملي التقابلي (AFC)، التحليل العاملي التقابلي المتعدد (ACM). وفي حالة وجود عدد كبير من المتغيرات في البيانات التي نريد تطبيق طرق التصنيف عليها، فيفضل اختيار المتغيرات الأكثر أهمية أو تطبيق طريقة تحليل المركبات الرئيسية (ACP) لاختزال عدد المتغيرات.

ب. حساب المسافات بين العناصر وتشكيل مصفوفة المسافات (مصفوفة القرب):

انطلاقا من جدول البيانات، نقوم بحساب المسافات بين العناصر ثم نشكل مصفوفة المسافات (Matrice des distances) (أو مصفوفة القرب Matrice de proximité)، ويستخدم في ذلك

عدة مقاييس، ومن أشهر هذه المقاييس والأكثر استخداما المقياس الإقليدي (Euclidienne) والمقياس الاقليدي مربع.

الصيغ الرياضية لحساب المسافات وفقا لهاذين المقياسيين هي:

- ✓ المسافة الاقليدية بين العنصرين i و i' هي: $d_e(i,i') = \sqrt{\sum_{j=1}^{p}\left(x_{ij} - x_{i'j}\right)^2}$.
- ✓ المسافة الاقليدية مربع بين العنصرين i و i' هي: $d_e^2(i,i') = \sum_{j=1}^{p}\left(x_{ij} - x_{i'j}\right)^2$.

ونشير إلى وجود العديد من المقاييس الأخرى والتي يمكن استخدامها لحساب المسافة بين عنصرين بحسب نوع المتغيرات (كمية، ترتيبية، اسمية)، من بينها: مقياس Minkowssky، مقياس City Manhattan Block، مقياس كاي تربيع، مقياس Tchebychev، ... إلى غير ذلك.

بالاعتماد على احدى المقاييس السابقة، نحسب المسافات بين جميع العناصر، ثم بعد ذلك نشكل مصفوفة المسافات، والتي تضم المسافات بين العناصر (بين كل عنصرين). وهي مصفوفة مربعة من الدرجة nxn.

ج. تحديد مؤشر التجميع:

لاحظنا أعلاه أنه يمكننا حساب المسافة بين عنصرين بالاعتماد على المقياس الإقليدي بسهولة، والسؤال الذي يجب طرحه الأن هو: ما هو المقياس (أو المؤشر) الذي يمكننا الاعتماد عليه لحساب المسافة بين عنصر ومجموعة أو بين مجموعتين؟ في الحقيقة هناك عدة مؤشرات تسمح بحساب مثل هذه المسافات، وهو ما يعرف بمؤشر التجميع (Indice d'agrégation) ويرمز له بالرمز δ.

ومن بين أهم مؤشرات التجميع:

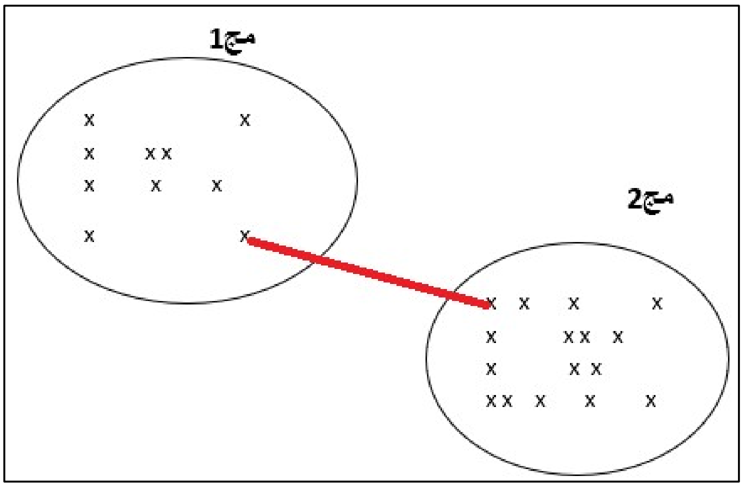
- **مؤشر التجميع لأدنى مسافة (Saut minimal)**:

بالاعتماد على المسافات بين العناصر نقوم بتجميع العناصر وفقا لهذا المؤشر عنصر بعنصر، حيث نبدأ بجمع أقرب عنصرين، ثم أقرب العناصر بأخذ أصغر مؤشر (Min)، ... وهكذا.

نحدد مؤشر التجميع كما يلي: $\delta(C_1, C_2) = \underset{\substack{x_i \in C_1 \\ x_j \in C_2}}{\mathrm{Min}}\left\{d^2(x_i, x_j)\right\}$.

حيث C_1 و C_2 هما مجموعتان من العناصر.

لضم أي عنصر من المجموعة C_2 (وليكن مثلا الفرد x_2) إلى مجموعة أخرى (ولتكن مثلا C_1)، نقوم بحساب المؤشر التالي: $\delta(\{C_1\}, x_2) = \underset{\substack{x_{i1} \in C_1 \\ x_2 \in C_2}}{\mathrm{Min}}\left\{d^2(x_{i1}, x_2)\right\}$.



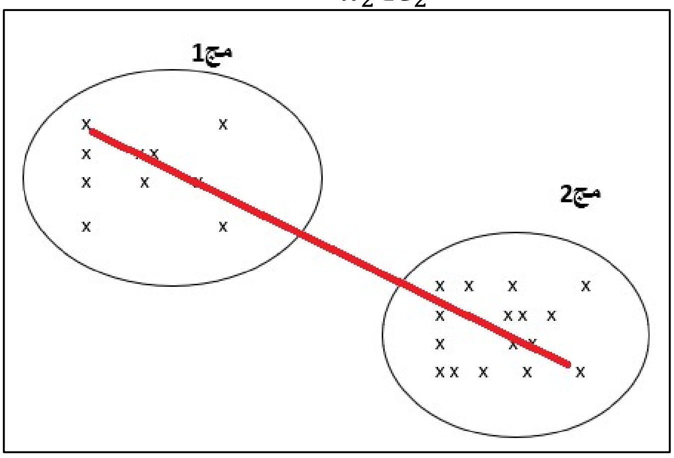
➢ **مؤشر التجميع لأقصى مسافة (Saut maximal):**

هذا المؤشر معاكس لمؤشر التجميع لأدنى مسافة، حيث نبدأ بجمع أقرب عنصرين بالاعتماد على المسافات بين العناصر، ثم أقرب العناصر بأخذ أكبر مؤشر (Max)، ... وهكذا.

نحدد مؤشر التجميع كما يلي: $\delta(C_1, C_2) = \underset{\substack{x_i \in C_1 \\ x_j \in C_2}}{\text{Max}}\{d^2(x_i, x_j)\}$.

حيث C_1 و C_2 هما مجموعتان من العناصر.

لضم أي عنصر من المجموعة C_2 (وليكن مثلا الفرد x_2) إلى مجموعة أخرى (ولتكن مثلا C_1)، نقوم بحساب المؤشر التالي: $\delta(\{C_1\}, x_2) = \underset{\substack{x_{i1} \in C_1 \\ x_2 \in C_2}}{\text{Max}}\{d^2(x_{i1}, x_2)\}$.



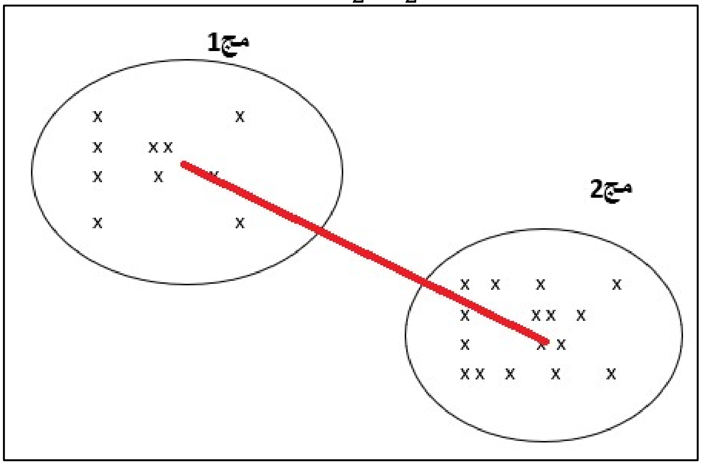
➢ **مؤشر التجميع للمسافة المتوسطة (Saut moyen):**

نقوم بتجميع العناصر وفقا لهذا المؤشر عنصر بعنصر، حيث نبدأ بجمع أقرب عنصرين، ثم أقرب العناصر بأخذ متوسط المؤشرات (Moy)، ... وهكذا. وذلك بالاعتماد على المسافات بين العناصر.

نحدد مؤشر التجميع كما يلي: $\delta(C_1, C_2) = \underset{\substack{x_i \in C_1 \\ x_j \in C_2}}{\text{Moy}}\{d^2(x_i, x_j)\}$.

حيث C_1 و C_2 هما مجموعتان من العناصر.

لضم أي عنصر من المجموعة C_2 (وليكن مثلا الفرد x_2) إلى مجموعة أخرى (ولتكن مثلا C_1)، نقوم بحساب المؤشر التالي: $\delta(\{C_1\}, x_2) = \underset{\substack{x_{i1} \in C_1 \\ x_2 \in C_2}}{\text{Moy}}\{d^2(x_{i1}, x_2)\}$.



➢ **مؤشر التجميع لـ "وارد" Ward**:

يعتبر مؤشر التجميع لوارد من أهم وأشهر مؤشرات التجميع، وهو الأكثر استخداما ويوصى به، ويقوم على مبدأ تصغير مجموع المربعات (التباين) لكل زوجين من العناصر الممكن تشكيلها في كل مرحلة، من خلال تحليل التباين. ويحسب مؤشر التجميع من خلال مركز ثقل المجموعات.

تقوم هذه الطريقة على مبدأ تجميع الأفراد من خلال تصغير التباين داخل المجموعة وتعظيم التباين بين المجموعات. وهذه الطريقة تندرج في إطار صيغة Lance et Williams.

ليكن g_{C_1} و g_{C_2} مركزا ثقل المجموعتين C_1 و C_2 على الترتيب.

مركز ثقل لجمع المجموعتين السابقتين يعطى بالعلاقة التالية: $g_{C_1,C_2} = \frac{n_{C_1}g_{C_1} + n_{C_2}g_{C_2}}{n_{C_1} + n_{C_2}}$.

مؤشر التجميع لمجموعتين C_1 و C_2 هو: $\delta(C_1, C_2) = \frac{n_{C_1}n_{C_2}}{n_{C_1} + n_{C_2}} d^2(g_{C_1}; g_{C_2})$.

ومؤشر التجميع لعنصرين x_1 و x_2 هو: $\delta(x_1, x_2) = \frac{1}{2} d^2(x_1; x_2)$.

لضم العنصر x إلى مجموعة تضم مجموعتين جزئيتين (C_1, C_2)، نقوم بحساب:

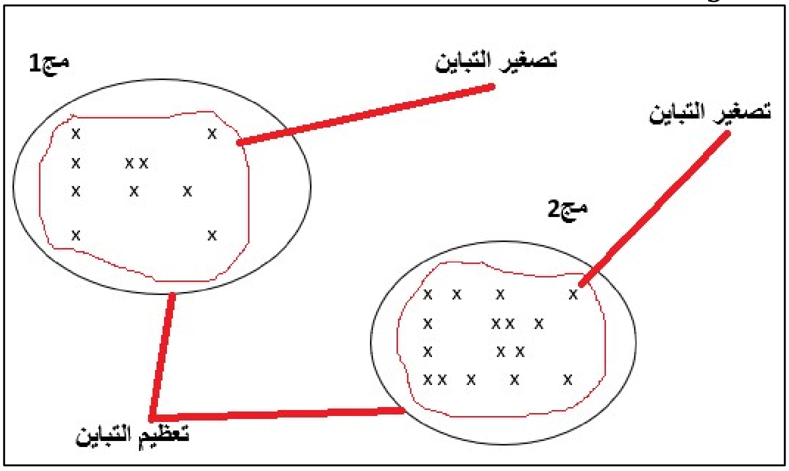
$$\delta[(C_1, C_2), x] = \frac{(n_{C_1} + n_x)\delta(C_1, x) + (n_{C_2} + n_x)\delta(C_2, x) - n_x\delta(C_1, C_2)}{n_{C_1} + n_{C_2} + n_x}.$$

لضم المجموعة C_2 التي تضم عنصرين (x_1, x_2) إلى المجموعة C_1، نقوم بحساب:

$$.\delta[C_2(x_1,x_2),C_1] = \frac{(1+n_{C_1})\delta(x_1,C_1)+(1+n_{C_1})\delta(x_2,C_1)-n_{C_1}\delta(x_1,x_2)}{n_{x_1}+n_{x_2}+n_{C_1}}$$

لضم العنصر x إلى مجموعة تضم عنصرين x_1 و x_2، نقوم بحساب:

$$.\delta[(x_1,x_2),x] = \frac{2\delta(x_1,x)+2\delta(x_2,x)-\delta(x_1,x_2)}{3}$$



د. تجميع العناصر بشكل تسلسلي تصاعدي وتكوين المجموعات:

تتم عملية التجميع بشكل تسلسلي تصاعدي كما يلي:

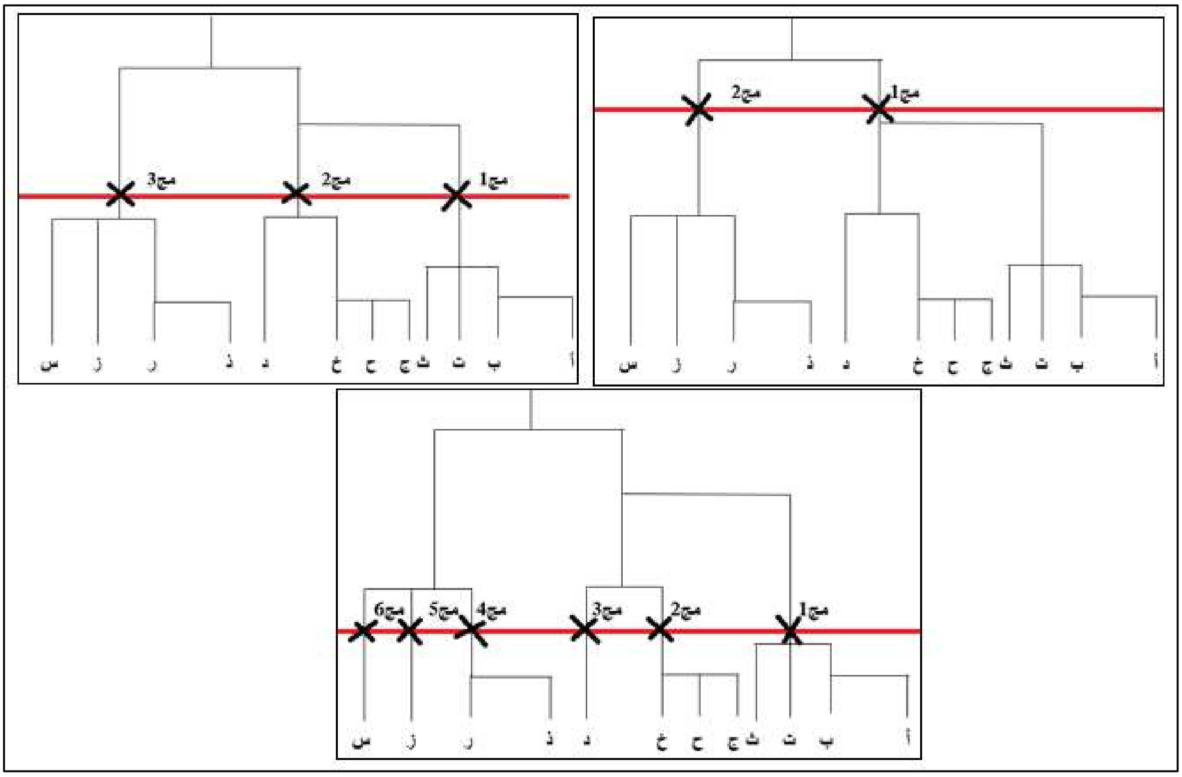
1) بالاعتماد على مصفوفة المسافات نقوم بجمع أقرب عنصرين (أصغر مسافة)، ليشكلا معا مجموعة واحدة (عنصر جديد)، فيكون هذا أول تجميع لـ n-1 مجموعة.
2) نقوم بتشكيل مصفوفة مسافات جديدة بعد تجميع العنصرين السابقين.
3) نعيد حساب المسافات بين المجموعة الجديدة وبقية العناصر بالاعتماد على مؤشر التجميع (المسافات الأولية بين العناصر خارج المجموعة المشكلة لا تتغير).
4) نقوم بجمع أقرب عنصرين (فرد أو مجموعة)، ليشكلا معا مجموعة جديدة.
5) ....
6) نكرر هذه العمليات للتجميع بشكل تسلسلي تصاعدي حتى تجمع جميع العناصر في مجموعة واحدة.
7) إذا كان هناك n عنصر فسنحتاج إلى n-1 عملية تجميع للعناصر.

هـ. تحديد مستوى القطع (العدد الأنسب من المجموعات):

نقوم في هذه الخطوة بتحديد مستوى القطع (أو التجزئة أو التقسيم) في الشجرة، أي البحث عن التقسيم الأكثر أهمية في الشجرة، وذلك بالاعتماد على التباين، فالمستوى الذي يكون عنده التباين داخل المجموعات صغير والتباين بين المجموعات كبير هو أفضل مستوى للتقسيم.

ونستعين للقيام بتقسيم العناصر في مجموعات على بعض المحددات، منها: شجرة التصنيف (Dendrogramme)، ومنحنى المؤشرات (Courbe des indices)، وحجم العينة (عدد الأفراد)، وكيفية تفسير النتائج.

فلتحديد المجموعات على شجرة التصنيف نقوم برسم خط أفقي عند مستوى القطع المرغوب فيه، فيتحدد عدد المجموعات بحسب عدد مرات التقاطع بين المستقيم وفروع الشجرة، ونستعين في ذلك بعدة معايير، من أهمها ما يعرف بالقفزة (Le saut)، والتي تشير إلى تضاعف البعد بين المجموعات، فيتم القطع عند القفز الأعلى نسبيا (Le saut le plus élevé)، والذي يشير إلى ارتفاع كبير لمؤشر التجميع مقارنة بما قبله، والقطع يتم بعد التجميع الموافق لقيم صغيرة لمؤشر التجميع، وقبل التجميع الموافق لقيم كبيرة للمؤشر (وهو المستوى الذي يكون بعده خسارة كبيرة في المعلومات).



نلاحظ من الشكل أعلاه أنه في نفس شجرة التصنيف، يختف عدد المجموعات بحسب موضع خط القطع، فقد تكون مجموعتين فقط إذا كان القطع أعلى الشجرة، وقد يرتفع إلى ستة مجموعات إذا نزلنا إلى مستويات دنيا من الشجرة.

ونشير إلى أن تحديد عدد المجموعات مهم جدا في التحليل، فإذا حددنا عدد قليل من المجموعات (أقل مما يجب)، فهذا سيجعلنا نضم في مجموعة واحدة عناصر مختلفة (مجموعات غير متجانسة)، أما إذا حددنا عدد كبير من المجموعات (أكثر مما يجب)، فهذا سيجعلنا نفصل بين عناصر متشابهة (مجموعات منفصلة متجانسة).

و. تحليل وتفسير النتائج:

نعتمد لتحليل وتفسير النتائج بشكل أساسي على قراءة الشجرة التسلسلية وجدول التصنيف، وهذا بعد تصنيف جميع العناصر في مجموعات.

ويجب أن يتم التفسير من أعلى إلى أسفل، من أجل فحص وتحليل المجموعات التي تحتوي على عدد قليل من الفئات أولاً، ثم الخوض في التحليل الأكثر تفصيلاً (عدد كبير من المجموعات) [1].

ولتفسير وتحليل النتائج نتبع الخطوات التالية:

- ✓ تحديد مستوى التقسيم (مستوى القطع في الشجرة).
- ✓ تحديد جميع المجموعات بناء على التقسيم السابق.
- ✓ تحديد جميع العناصر التي تنتمي لكل مجموعة.
- ✓ تشكيل جدول يلخص: المجموعات، الأفراد التي تنتمي لكل مجموعة، والمتغيرات الممثلة لكل مجموعة.

3. تطبيق طريقة التصنيف التسلسلي التصاعدي على برنامج SPSS:

المثال التطبيقي الذي سنعتمد عليه لتطبيق طريقة CAH على برنامج SPSS، يخص بيانات 22 دولة عربية (والتي تمثل الأفراد أو الحالات)، وثلاث متغيرات هي: نصيب الفرد من إجمالي الناتج المحلي (بالدولار الأمريكي)، إجمالي تكوين رأس المال (بالدولار الأمريكي)، نسبة البطالة (%). مع الإشارة إلى أنه تم أخذ متوسط القيم من سنة 1991 إلى سنة 2020. وقد تم الحصول على البيانات من قاعدة بيانات البنك الدولي[2].

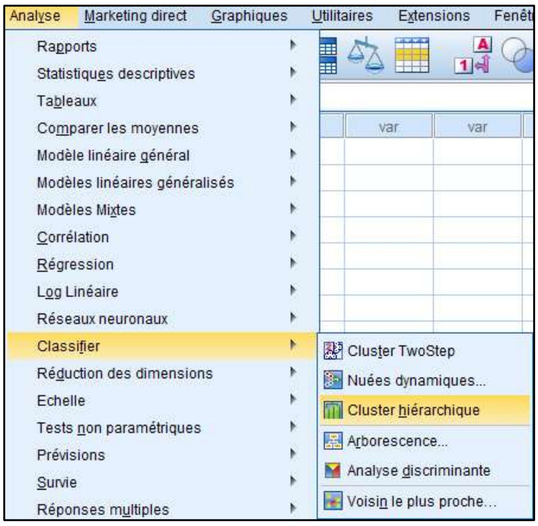
أ. الخطوات على برنامج SPSS:

للقيام بـ CAH على برنامج SPSS نتبع الخطوات التالية:

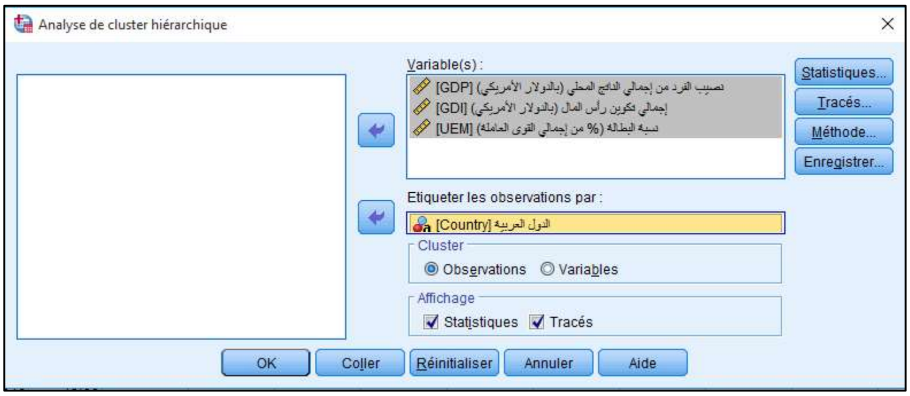
1) بعد فتح برنامج SPSS على الصفحة التي تحتوي بيانات الدراسة، نذهب إلى شريط القوائم.
2) في شريط القوائم نختار بالترتيب: Analyse ثم Classifier ثم Cluster hiérarchique.

[1] Arnaud MARTIN, Op-cit, p98.

[2] الموقع الرسمي للبنك الدولي: https://data.albankaldawli.org/country



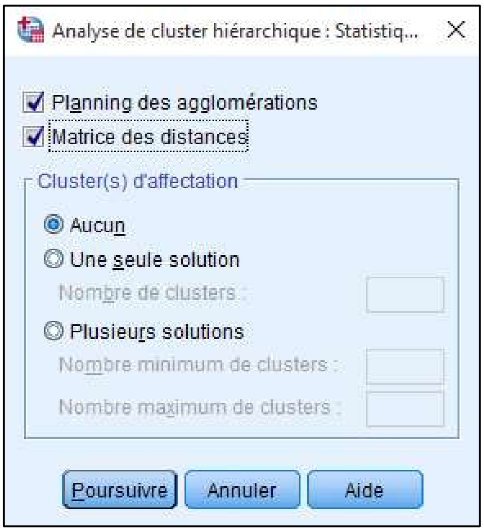
3) فتظهر لنا النافذة التالية:



✓ من خلال هذه النافذة نختار:

- متغيرات التصنيف: GDP، GDI، UEM.
- المشاهدات (الأفراد): الدول العربية Country
- التنصيف: للأفراد (للمشاهدات).
- العرض: الإحصائيات والتمثيلات البيانية.

✓ ثم نضغط على الإحصائيات (Statistiques)، فتظهر لنا النافذة التالية:



نختار من هذه النافذة:

- نظلل على مخطط التجميع (Planning des agglomérations).
- نظلل على مصفوفة المسافات (Matrice des distances).
- مجموعات التخصيص (Clusters d'affectation): تظهر ضمن المخرجات مع الإحصائيات، نختار منها إحدى الخيارات التالية:
  - بدون تخصيص مجموعات aucun (مبدئيا نختار Aucun).
  - حل وحيد (تحديد عدد ثابت للمجموعات) Une seule solution.
  - عدة حلول (تحديد عدد أعلى للمجموعات وعدد أدنى لها) Plusieurs solutions. وهي الأكثر استخداما.

✓ ثم نضغط على التمثيلات البيانية (Tracés)، فتظهر لنا النافذة التالية:

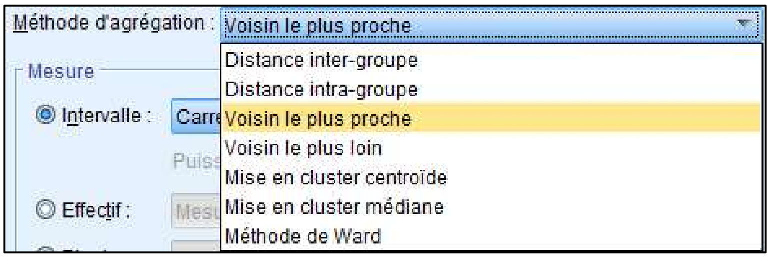
نختار من هذه النافذة:

- نظلل على شجرة التصنيف (Dendrogramme).
- التمثيل النازل (أو الألواح الجليدية) (Stalactites): نختار منها إحدى الخيارات التالية:
  - كل المجموعات (Tous les clusters).
  - تحديد المجموعات بمؤشر (Plage de clusters indiquée).
  - بدون تحديد aucun (عدم عرض التمثيل النازل).
- اتجاه التمثيل البياني (Orientation): نختار منها إحدى الخيارات التالية:
  - عمودي (Verticale).
  - أفقي (Horizontale).

✓ ثم نضغط على الطرق (Méthodes)، فتظهر لنا النافذة التالية:

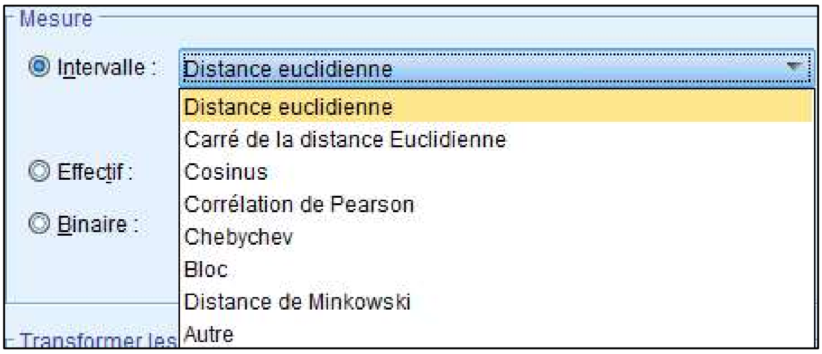
نختار من هذه النافذة:

- طريقة التجميع (Méthode d'agrégation): عند الضغط عليها تظهر لنا خيارات مؤشر التجميع (أدنى مسافة، أقصى مسافة، مسافة متوسطة، مؤشر وارد، ...)

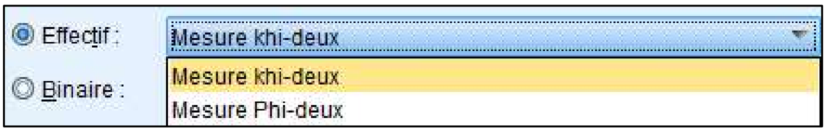


- القياس (Mesure): نختار منها احدى الخيارات التالية (بحسب نوع متغيرات الدراسة):

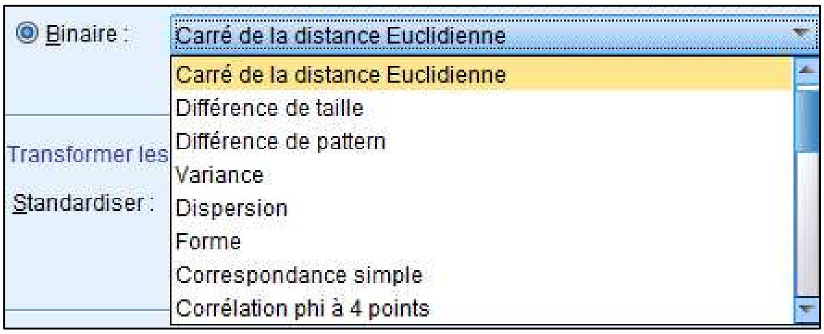
- مجال (Intervalle): نختار منها احدى طرق القياس (المسافة الاقليدية، المسافة الاقليدية مربع، التجب، ارتباط بيرسون، ...)



- التكرار (Effectif): نختار منها احدى الخيارين التالين:

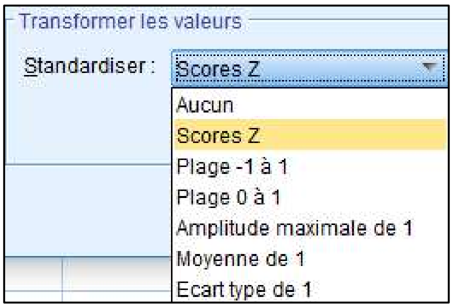


- ثنائي (Binaire): نختار منها احدى الخيارات التالية:

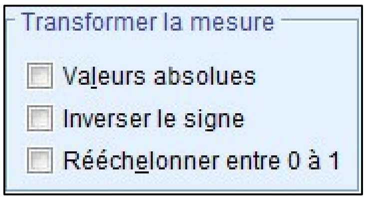


- تحويل القيم (Transformer les valeurs): يمكن تحويل القيم الأصلية (سواء للمتغيرات أو المشاهدات) إلى قيم معيارية.

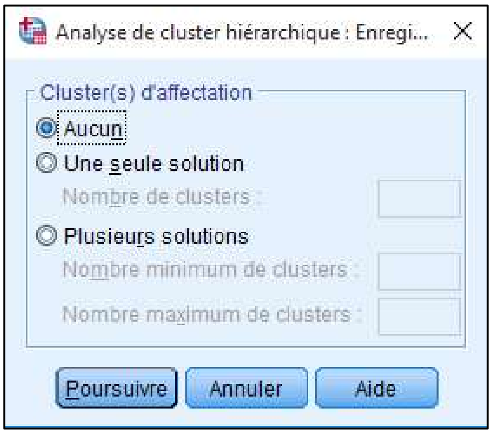
بما أن للمتغيرات المدروسة وحدات القياس مختلفة والبيانات بقيم متباينة (نسبة البطالة بالعشرات أما إجمالي رأس المال فبالملايين)، فيفضل جعل القيم معيارية.



- تحويل القياس (Transformer la mesure): يمكن تحويل القياس بأخذ القيم المطلقة أو تحويل الإشارة أو تحويل القياس بين 0 و 1.



✓ ثم نضغط على الحفظ (Enregistrer)، فتظهر لنا النافذة التالية:



نختار من هذه النافذة:

- مجموعات التخصيص (Clusters d'affectation): ويتم حفظ المجموعات في نافذة البيانات أين يتم إدراج تصنيف كل فرد (إضافة عمود جديد في النافذة)، نختار من هذه النافذة إحدى الخيارات التالية:
  - بدون تخصيص مجموعات aucun (مبدئيا نختار Aucun).
  - حل وحيد (تحديد عدد ثابت للمجموعات) Une seule solution.
  - عدة حلول (تحديد عدد أعلى للمجموعات وعدد أدنى لها) Plusieurs solutions. وهي الأكثر استخداما.

بعد الانتهاء من إدخال كل هذه الاختيارات والتعليمات، نضغط على OK، فتظهر لنا المخرجات.

ب. مخرجات المثال التطبيقي والتعليق عليها:

1) مصفوفة المسافات (القرب) (Matrice de proximité):

Observation	1:الإمارات	2:البحرين	3:جزر القمر	4:جيبوتي	5:الجزائر	6:مصر	7:العراق	8:الأردن	9:الكويت	10:لبنان	11:ليبيا
1:الإمارات	,000	3,572	4,449	5,560	4,032	3,883	3,808	4,523	3,075	4,053	4,628
2:البحرين	3,572	,000	1,424	3,810	2,799	1,801	1,726	2,169	,698	1,367	2,690
3:جزر القمر	4,449	1,424	,000	2,958	2,108	1,024	1,010	1,233	2,034	,587	1,900
4:جيبوتي	5,560	3,810	2,958	,000	1,846	2,535	2,634	1,755	4,021	2,628	1,167
5:الجزائر	4,032	2,799	2,108	1,846	,000	1,177	1,251	1,222	2,980	1,645	1,037
6:مصر	3,883	1,801	1,024	2,535	1,177	,000	,107	,940	2,169	,628	1,410
7:العراق	3,808	1,726	1,010	2,634	1,251	,107	,000	1,028	2,092	,620	1,502
8:الأردن	4,523	2,169	1,233	1,755	1,222	,940	1,028	,000	2,525	,883	,670
9:الكويت	3,075	,698	2,034	4,021	2,980	2,169	2,092	2,525	,000	1,823	2,928
10:لبنان	4,053	1,367	,587	2,628	1,645	,628	,620	,883	1,823	,000	1,492
11:ليبيا	4,628	2,690	1,900	1,167	1,037	1,410	1,502	,670	2,928	1,492	,000
12:المغرب	4,101	1,849	,926	2,351	1,185	,264	,350	,699	2,245	,537	1,230
13:موريتانيا	4,509	1,728	,534	2,425	1,699	,822	,863	,707	2,237	,469	1,377
14:عمان	3,506	,436	1,167	3,430	2,362	1,384	1,311	1,781	,880	,982	2,291
15:فلسطين	4,821	2,637	1,703	1,262	1,215	1,360	1,457	,499	2,950	1,381	,341
16:قطر	1,871	2,724	3,976	5,315	4,170	3,761	3,686	4,172	2,044	3,632	4,353
17:السعودية	1,859	3,282	3,569	4,630	2,842	2,796	2,733	3,585	3,117	3,205	3,706
18:الصومال	4,642	2,404	1,382	1,628	1,099	,994	1,094	,277	2,768	1,075	,613
19:السودان	4,975	2,749	1,771	1,189	1,326	1,481	1,580	,610	3,078	1,489	,448
20:سوريا	4,500	1,619	,353	2,606	1,844	,878	,899	,889	2,166	,463	1,558
21:تونس	4,509	2,200	1,252	1,747	1,160	,905	,997	,074	2,556	,898	,657

Observation	12:المغرب	13:موريتانيا	14:عمان	15:فلسطين	16:قطر	17:السعودية	18:الصومال	19:السودان	20:سوريا	21:تونس
1:الإمارات	4,101	4,509	3,506	4,821	1,871	1,859	4,642	4,975	4,500	4,509
2:البحرين	1,849	1,728	,436	2,637	2,724	3,282	2,404	2,749	1,619	2,200
3:جزر القمر	,926	,534	1,167	1,703	3,976	3,569	1,382	1,771	,353	1,252
4:جيبوتي	2,351	2,425	3,430	1,262	5,315	4,630	1,628	1,189	2,606	1,747
5:الجزائر	1,185	1,699	2,362	1,215	4,170	2,842	1,099	1,326	1,844	1,160
6:مصر	,264	,822	1,384	1,360	3,761	2,796	,994	1,481	,878	,905
7:العراق	,350	,863	1,311	1,457	3,686	2,733	1,094	1,580	,899	,997
8:الأردن	,699	,707	1,781	,499	4,172	3,585	,277	,610	,889	,074
9:الكويت	2,245	2,237	,880	2,950	2,044	3,117	2,768	3,078	2,166	2,556
10:لبنان	,537	,469	,982	1,381	3,632	3,205	1,075	1,489	,463	,898
11:ليبيا	1,230	1,377	2,291	,341	4,353	3,706	,613	,448	1,558	,657
12:المغرب	,000	,609	1,438	1,139	3,900	3,049	,764	1,252	,703	,667
13:موريتانيا	,609	,000	1,390	1,169	4,086	3,580	,857	1,239	,183	,727
14:عمان	1,438	1,390	,000	2,252	2,823	3,053	2,008	2,371	1,306	1,807
15:فلسطين	1,139	1,169	2,252	,000	4,503	3,862	,407	,154	1,352	,493
16:قطر	3,900	4,086	2,823	4,503	,000	3,124	4,382	4,648	4,055	4,189
17:السعودية	3,049	3,580	3,053	3,862	3,124	,000	3,610	4,004	3,587	3,545
18:الصومال	,764	,857	2,008	,407	4,382	3,610	,000	,498	1,038	,222
19:السودان	1,252	1,239	2,371	,154	4,648	4,004	,498	,000	1,419	,608
20:سوريا	,703	,183	1,306	1,352	4,055	3,587	1,038	1,419	,000	,909
21:تونس	,667	,727	1,807	,493	4,189	3,545	,222	,608	,909	,000

نلاحظ أنه تم حساب المسافات لواحد وعشرين دولة فقط (21)، ولم يتم إدخال دولة اليمن في التصنيف، لأن بها قيمة مفقودة لمتغير إجمالي تكوين رأس المال.

22	اليمن	2219	.	11,73

نلاحظ كذلك من الجدول أن أصغر مسافة بين دولتين هي 0,074 بين الأردن وتونس، وبالتالي فهاتين الدولتين ستشكلان أول مجموعة C1(8,21). ثم أقرب دولتين بعدهما هما مصر والعراق، فقد قدرت المسافة بينهما بـ 0,107، وبالتالي ثاني تجميع هو C2(6,7)، وهكذا بالنسبة لبقية التجميعات للدول.

2) مخطط التجميع (Planning des agglomérations)

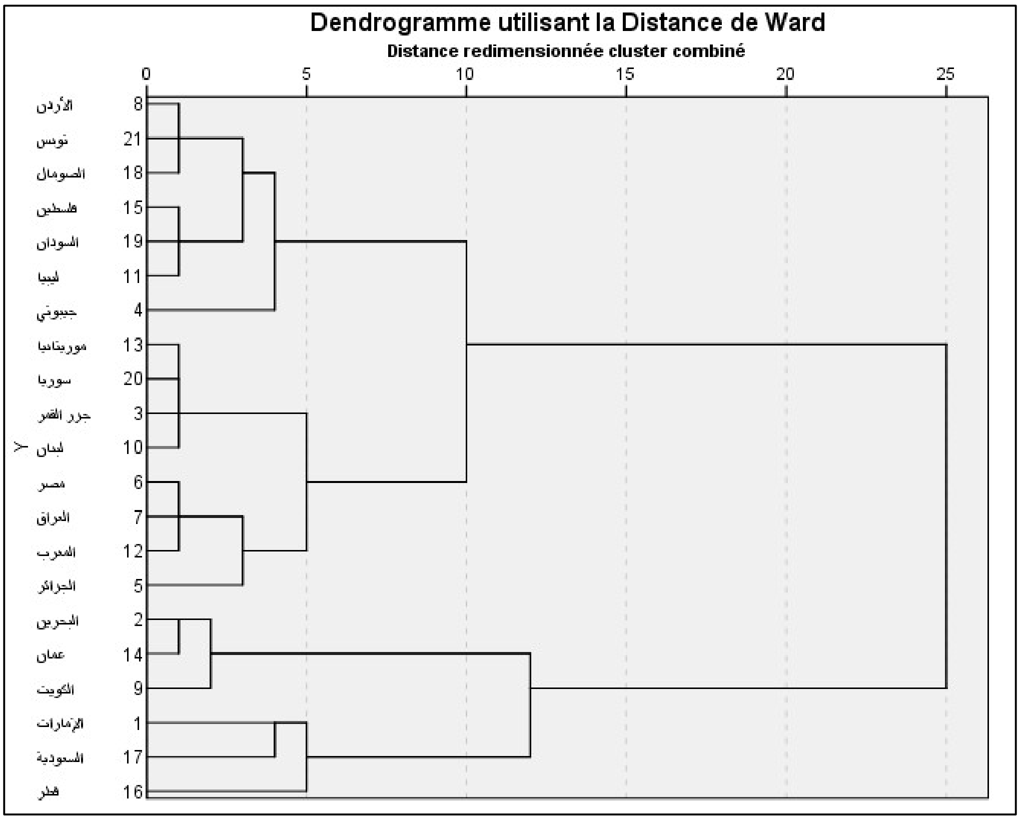
Planning des agglomérations

Etape	Cluster combiné: Cluster 1	Cluster combiné: Cluster 2	Coefficients	Etape de première apparition du cluster: Cluster 1	Etape de première apparition du cluster: Cluster 2	Etape suivante
1	8	21	,037	0	0	5
2	6	7	,090	0	0	6
3	15	19	,167	0	0	8
4	13	20	,259	0	0	9
5	8	18	,413	1	0	12
6	6	12	,600	2	0	13
7	2	14	,818	0	0	11
8	11	15	1,055	0	3	12
9	3	13	1,320	0	4	10
10	3	10	1,611	9	0	16
11	2	9	2,064	7	0	19
12	8	11	2,654	5	8	15
13	5	6	3,497	0	6	16
14	1	17	4,427	0	0	17
15	4	8	5,520	0	12	18
16	3	5	6,757	10	13	18
17	1	16	8,112	14	0	19
18	3	4	10,709	16	15	20
19	1	2	13,764	17	11	20
20	1	3	20,634	19	18	0

يبين لنا هذا الجدول جميع مراحل التجميع (الانتقال من 21 مجموعة تضم كل مجموعة عنصر واحد (دولة)، إلى مجموعة واحدة تضم جميع العناصر (الدول))، على 20 مرحلة (n-1).

نلاحظ أن أول مرحلة تجميع (أقرب دولتين) كانت بجمع رقم 8 ورقم21 في أول مجموعة (الأردن وتونس)، ثم في المرحلة الثانية تم جمع العنصرين 6 و7 في مجموعة ثانية (العراق ومصر)، ثم في المرحلة الثالثة تم ضم العنصرين 15 و19 في مجموعة ثالثة (فلسطين والسودان)، ...، وهكذا حتى أخر مرحلة.

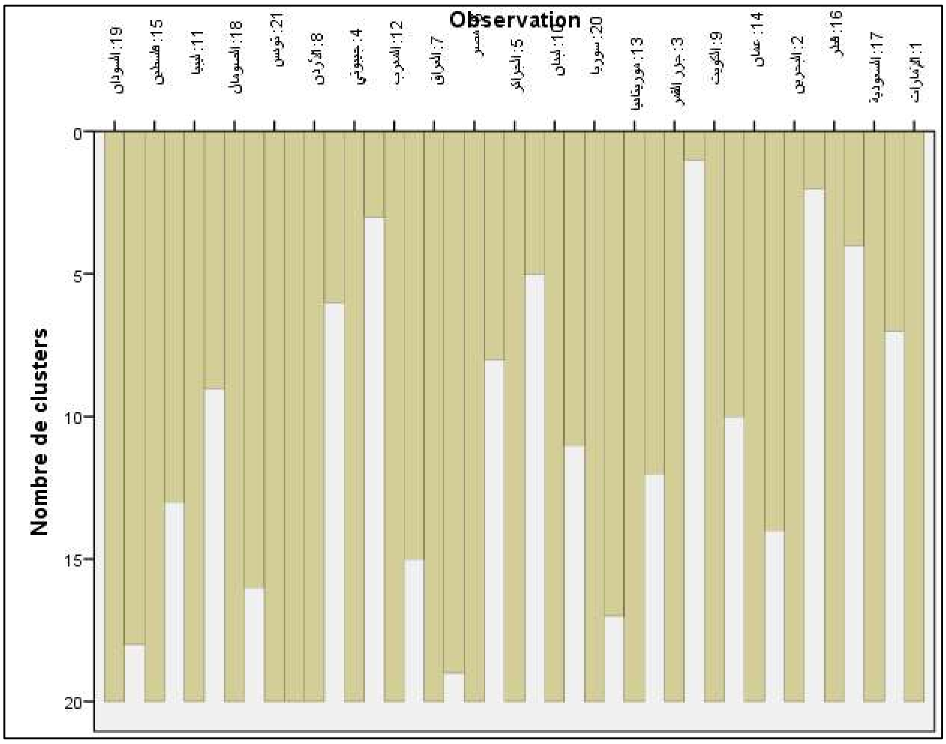
3) شجرة التصنيف (Dendrogramme):



يتبين من شجرة التصنيف وجود أربع مجموعة بارزة (التباين داخل المجموعات صغير والتباين بين المجموعات كبير):

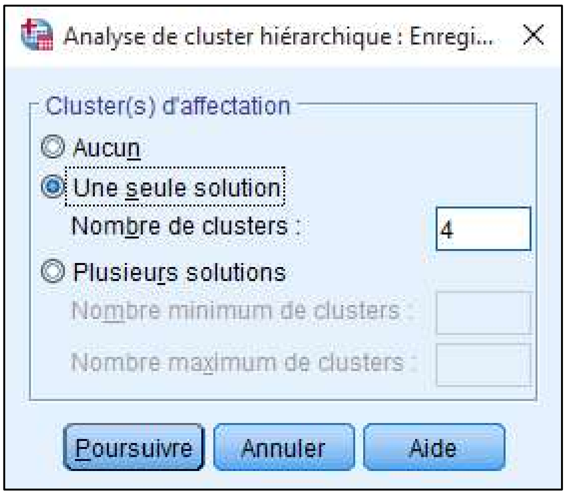
- ✓ المجموعة الأولى تضم الدول (3): الإمارات، السعودية، قطر.
- ✓ المجموعة الثانية تضم الدول (3): البحرين، عمان، الكويت.
- ✓ المجموعة الثالثة تضم الدول (8): الموريتانيا، سوريا، جزر القمر، لبنان، مصر، العراق، المغرب، الجزائر.
- ✓ المجموعة الرابعة تضم الدول (7): الأردن، تونس، الصومال، فلسطين، السودان، ليبيا، جيبوتي.

4) التمثيل النازل (Stalactites):



هذا التمثيل كذلك يؤكد ما استنتجناه في شجرة التصنيف بخصوص المجموعات، فنلاحظ أن أول مجموعة تضم الإمارات والسعودية وقطر، وثاني مجموعة تضم البحرين وعمان والكويت، وهكذا بالنسبة لبقية الدول.

5) الأن، نعيد التصنيف ونحدد أربع مجموعات (بعدما كان بدون تحديد Aucun):



6) مجموعات التخصيص (Clusters d'affectation):

يظهر لنا ضمن المخرجات الجدول التالي: كما يظهر في نافذة البيانات العمود التالي:

Country	CLU4_1
الإمارات	1
البحرين	2
جزر القمر	3
جيبوتي	4
الجزائر	3
مصر	3
العراق	3
الأردن	4
الكويت	2
لبنان	3
ليبيا	4
المغرب	3
موريتانيا	3
عمان	2
فلسطين	4
قطر	1
السعودية	1
الصومال	4
السودان	4
سوريا	3
تونس	4
اليمن	.

Cluster(s) d'affectation

Observation	Clusters 4
1:الإمارات	1
2:البحرين	2
3:جزر القمر	3
4:جيبوتي	4
5:الجزائر	3
6:مصر	3
7:العراق	3
8:الأردن	4
9:الكويت	2
10:لبنان	3
11:ليبيا	4
12:المغرب	3
13:موريتانيا	3
14:عمان	2
15:فلسطين	4
16:قطر	1
17:السعودية	1
18:الصومال	4
19:السودان	4
20:سوريا	3
21:تونس	4

يبين الجدولين أعلاه تصنيف كل دولة في مجموعة، فالدول: الإمارات، السعودية، قطر صنفت في المجموعة الأولى (1)، والدول: البحرين، عمان، الكويت صنفت في المجموعة الثانية (2). وهكذا بالنسبة لجميع الدول. وقد تم تصنيف الدول في أربع مجموعات تبعا لما حددناه أعلاه.

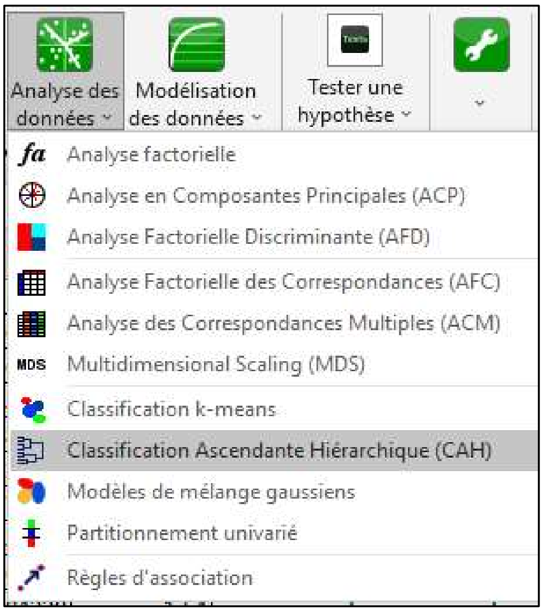
4. تطبيق طريقة التصنيف التسلسلي التصاعدي على برنامج XL-STAT:

المثال التطبيقي الذي سنعتمد عليه لتطبيق طريقة CAH على برنامج XL-STAT، هو نفس المثال المطبق على برنامج SPSS.

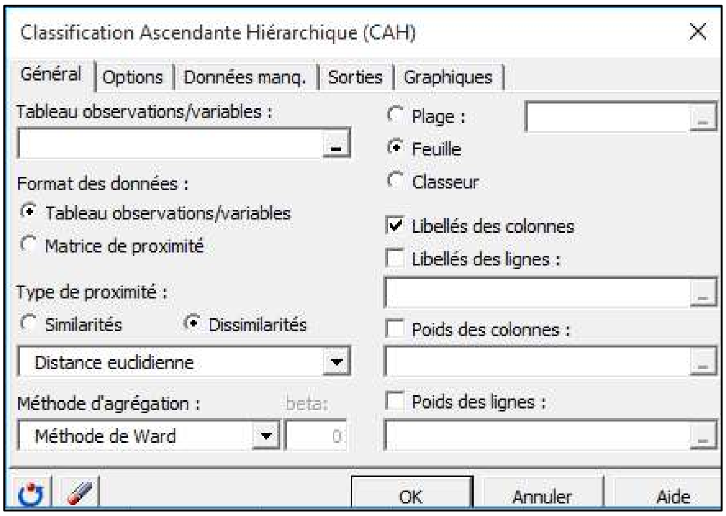
أ. الخطوات على برنامج XL-STAT:

للقيام بـ CAH على برنامج XL-STAT نتبع الخطوات التالية:

1) بعد فتح برنامج XL-STAT على الصفحة التي تحتوي بيانات الدراسة، نذهب إلى شريط القوائم.
2) في شريط القوائم نختار بالترتيب: Analyse des données ثم Classification Ascendantes Hiérarchique.



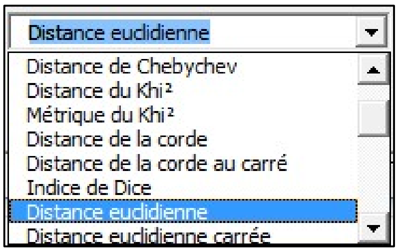
3) فتظهر لنا النافذة التالية:



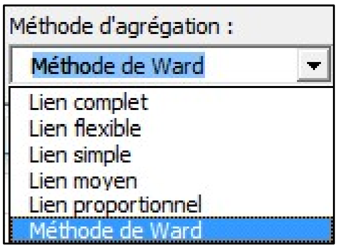
4) من خلال هذه النافذة نختار:

❖ **من عام Général**: نقوم باختيار:

- مكان حفظ المخرجات (classeur ،feuille ،plage)
- شكل البيانات (Format des données): هل هو جدول البيانات والمتغيرات، أو مصفوفة المسافات (غير متوفر على برنامج SPSS).
- نوع المسافات (Type de proximité): التشابه أو الاختلاف.
- مقياس المسافة: نختار منها احدى طرق القياس (المسافة الاقليدية، المسافة الاقليدية مربع، مسافة كاي تربيع، ...)

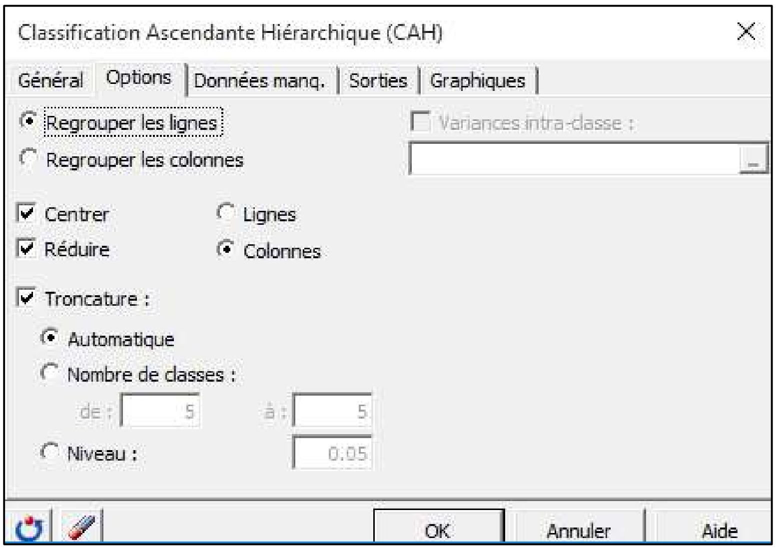


- طريقة التجميع (Méthode d'agrégation): عند الضغط عليها تظهر لنا خيارات مؤشر التجميع (أدنى مسافة، أقصى مسافة، مسافة متوسطة، مؤشر وارد، ...)



- تعيين أسماء الأعمدة (المتغيرات) (Libellés des colonnes)
- تعيين أسماء الأسطر (الأفراد) (Libellés des lignes).
- أوزان الأسطر أو أوزان الأعمدة (poids) –إن وجدت–.

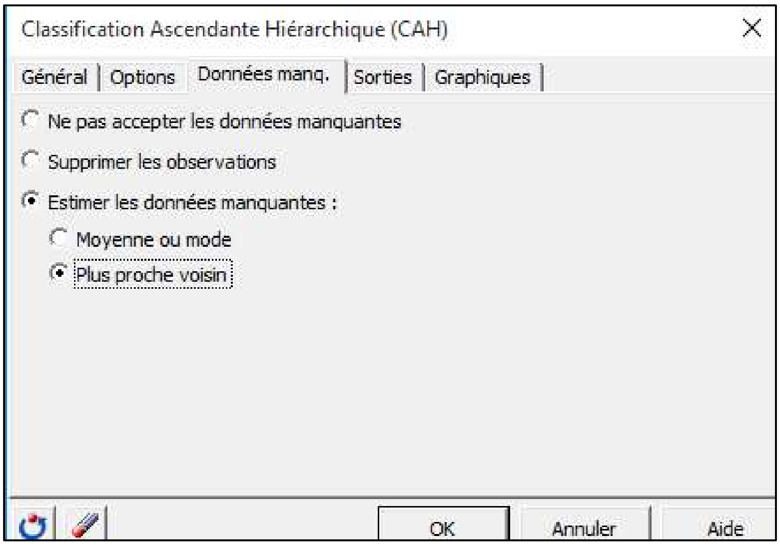
❖ من الخيارات **Option**: نقوم باختيار:



- تجميع الأسطر (Regrouper les lignes) أو تجميع الأعمدة (Regrouper les colonnes).
- تحويل البيانات إلى مركزية أو معيارية (Centrer / Réduire).
- مستوى القطع (Troncature): بشكل أوتوماتيكي (غير متوفر على SPSS)، أو تحديد عدد المجموعات (من ... و ...إلى)، أو تحديد المستوى القطع في شجرة التصنيف (غير متوفر على SPSS).

نلاحظ أن برنامج XL-STAT يسمح بالتصنيف الأوتوماتيكي، حيث يقوم بالتجزئة الأمثل للأفراد في مجموعات (من خلال اختيار automatique).

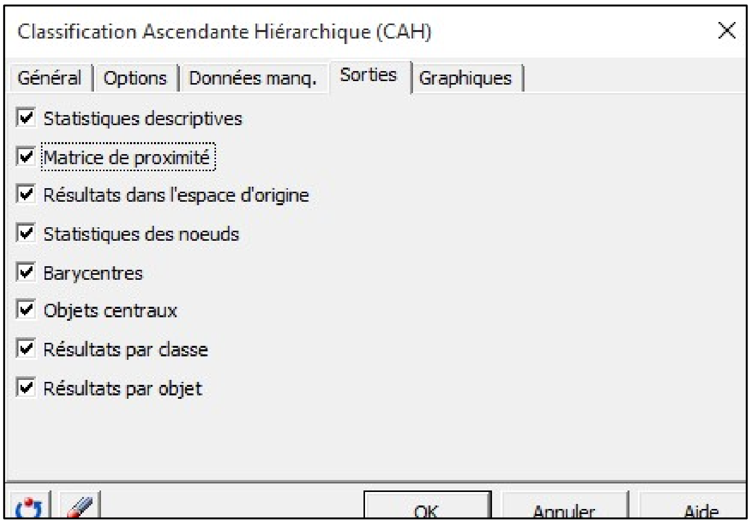
❖ **من البيانات المفقودة (Données manquantes)**: عند الضغط عليها يظهر لنا النافذة التالي (غير متوفرة على برنامج SPSS):



وهي نافذة خاصة بمعالجة البيانات المفقودة (Données manquantes)، وفق ما يلي:

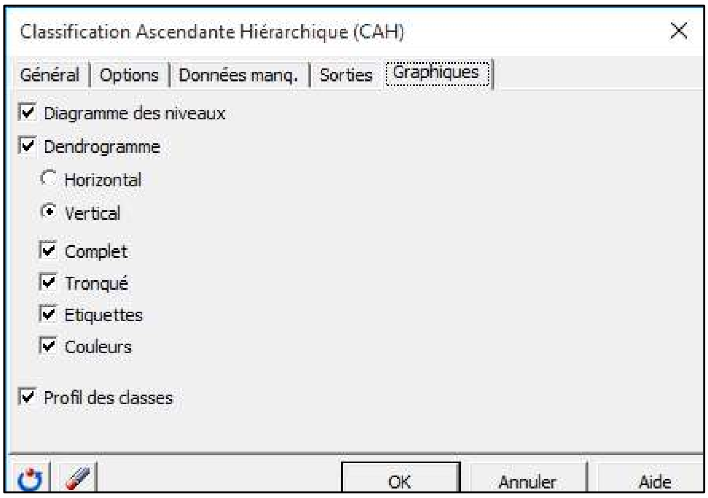
- عدم قبول البيانات المفقودة.
- حذف المشاهدات.
- تقدير البيانات المفقودة (بالمتوسط أو المنوال، أو بأقرب جار).

❖ **من المخرجات (Sortie)**: عند الضغط عليها تظهر لنا النافذة التالي:



نختار المخرجات التي نرغب في عرضها: الإحصائيات الوصفية، مصفوفة المسافات، ...). والعديد من هذه المخرجات غير متوفر على برنامج SPSS.

❖ **التمثيلات البيانية (Graphiques)**: عند الضغط عليها يظهر لنا النافذة التالي:



في هذه النافذة نختار التمثيلات البيانية التي نرغب في عرضها مع بعض خيارات العرض. والعديد من هذه الخيارات غير متوفر على برنامج SPSS.

بعد الانتهاء من إدخال كل هذه الاختيارات والتعليمات، نضغط على OK، فتظهر لنا المخرجات.

ب. مخرجات المثال التطبيقي والتعليق عليها:

بما أن هذا المثال التطبيقي هو نفس المثال مطبق على برنامج SPSS، فسأعلق خاصة على الجداول والمعلومات الجديدة والتي لم أعلق عليها عند شرح برنامج SPSS.

1) القيم المفقودة (غير متوفر على SPSS):

Observations avec des données manquantes remplacées :			
Observation	GDP	GDI	UEM
اليمن	2219,0000	12562574559,0000	11,7300

تم استبدال القيمة المفقودة لدولة اليمن لدى متغير "إجمالي تكوين رأس المال" بقيمة هذا المتغير عند أقرب جار (وهي دولة المغرب). وهذا ما سيسمح بإدخال دولة اليمن في التصنيف، على عكس برنامج SPSS أين تم إقصاد هذه الدولة من التصنيف بسبب وجود قيمة مفقودة.

2) الإحصائيات الوصفية (غير متوفر على SPSS):

Variable	Observations	données ma	données ma	Minimum	Maximum	Moyenne	Ecart-type
GDP	22	0	22	399,0000	60420,0000	11283,0455	16463,5757
GDI	22	0	22	97255837,0000	68563122016,0000	13973311858,5909	19596067247,7292
UEM	22	0	22	0,5800	27,7100	10,9045	7,1422

يقدم لنا هذا الجدول بعض الإحصائيات الوصفية لمتغيرات الدراسة (أكبر قيمة، أصغر قيمة، المتوسط، الانحراف المعياري).

3) مصفوفة المسافات (القرب) (Matrice de proximité):

	الإمارات	البحرين	جزر القمر	جيبوتي	الجزائر	مصر	العراق	الأردن	الكويت	لبنان	ليبيا
الإمارات	0	3,6540	4,5446	5,6855	4,1171	3,9640	3,8872	4,6220	3,1482	4,1408	4,7306
البحرين	3,6540	0	1,4505	3,9006	2,8637	1,8396	1,7626	2,2176	0,7110	1,3957	2,7531
جزر القمر	4,5446	1,4505	0	3,0302	2,1597	1,0492	1,0351	1,2628	2,0706	0,6003	1,9461
جيبوتي	5,6855	3,9006	3,0302	0	1,8908	2,5970	2,6985	1,7976	4,1131	2,6918	1,1950
الجزائر	4,1171	2,8637	2,1597	1,8908	0	1,2057	1,2810	1,2519	3,0451	1,6853	1,0622
مصر	3,9640	1,8396	1,0492	2,5970	1,2057	0	0,1091	0,9630	2,2105	0,6424	1,4437
العراق	3,8872	1,7626	1,0351	2,6985	1,2810	0,1091	0	1,0527	2,1326	0,6342	1,5387
الأردن	4,6220	2,2176	1,2628	1,7976	1,2519	0,9630	1,0527	0	2,5779	0,9045	0,6862
الكويت	3,1482	0,7110	2,0706	4,1131	3,0451	2,2105	2,1326	2,5779	0	1,8581	2,9932
لبنان	4,1408	1,3957	0,6003	2,6918	1,6853	0,6424	0,6342	0,9045	1,8581	0	1,5280
ليبيا	4,7306	2,7531	1,9461	1,1950	1,0622	1,4437	1,5387	0,6862	2,9932	1,5280	0
المغرب	4,1870	1,8881	0,9490	2,4080	1,2139	0,2701	0,3585	0,7159	2,2892	0,5486	1,2593
موريتانيا	4,6060	1,7638	0,5473	2,4840	1,7408	0,8425	0,8837	0,7242	2,2797	0,4791	1,4103
عمان	3,5850	0,4464	1,1885	3,5116	2,4174	1,4132	1,3387	1,8208	0,8953	1,0017	2,3451
فلسطين	4,9267	2,6974	1,7446	1,2924	1,2444	1,3936	1,4930	0,5106	3,0143	1,4144	0,3480
قطر	1,9151	2,7759	4,0497	5,4278	4,2505	3,8294	3,7523	4,2537	2,0845	3,7002	4,4413
السعودية	1,8913	3,3629	3,6542	4,7415	2,9098	2,8626	2,7983	3,6712	3,1919	3,2825	3,7954
الصومال	4,7421	2,4577	1,4158	1,6680	1,1253	1,0179	1,1207	0,2823	2,8250	1,0999	0,6265
السودان	5,0844	2,8118	1,8146	1,2182	1,3583	1,5173	1,6181	0,6245	3,1442	1,5251	0,4572
سوريا	4,5967	1,6519	0,3613	2,6691	1,8889	0,8992	0,9206	0,9100	2,2063	0,4725	1,5958
تونس	4,6072	2,2498	1,2820	1,7891	1,1885	0,9274	1,0210	0,0754	2,6088	0,9196	0,6724
اليمن	4,2145	1,9555	1,0170	2,3215	1,1427	0,3219	0,4199	0,6514	2,3431	0,6079	1,1749

	المغرب	موريتانيا	عمان	فلسطين	قطر	السعودية	الصومال	السودان	سوريا	تونس	اليمن
الإمارات	4,1870	4,6060	3,5850	4,9267	1,9151	1,8913	4,7421	5,0844	4,5967	4,6072	4,2145
البحرين	1,8881	1,7638	0,4464	2,6974	2,7759	3,3629	2,4577	2,8118	1,6519	2,2498	1,9555
جزر القمر	0,9490	0,5473	1,1885	1,7446	4,0497	3,6542	1,4158	1,8146	0,3613	1,2820	1,0170
جيبوتي	2,4080	2,4840	3,5116	1,2924	5,4278	4,7415	1,6680	1,2182	2,6691	1,7891	2,3215
الجزائر	1,2139	1,7408	2,4174	1,2444	4,2505	2,9098	1,1253	1,3583	1,8889	1,1885	1,1427
مصر	0,2701	0,8425	1,4132	1,3936	3,8294	2,8626	1,0179	1,5173	0,8992	0,9274	0,3219
العراق	0,3585	0,8837	1,3387	1,4930	3,7523	2,7983	1,1207	1,6181	0,9206	1,0210	0,4199
الأردن	0,7159	0,7242	1,8208	0,5106	4,2537	3,6712	0,2823	0,6245	0,9100	0,0754	0,6514
الكويت	2,2892	2,2797	0,8953	3,0143	2,0845	3,1919	2,8250	3,1442	2,2063	2,6088	2,3431
لبنان	0,5486	0,4791	1,0017	1,4144	3,7002	3,2825	1,0999	1,5251	0,4725	0,9196	0,6079
ليبيا	1,2593	1,4103	2,3451	0,3480	4,4413	3,7954	0,6265	0,4572	1,5958	0,6724	1,1749
المغرب	0	0,6238	1,4674	1,1665	3,9721	3,1212	0,7822	1,2828	0,7206	0,6836	0,0896
موريتانيا	0,6238	0	1,4183	1,1976	4,1626	3,6658	0,8781	1,2690	0,1876	0,7449	0,6519
عمان	1,4674	1,4183	0	2,3034	2,8757	3,1282	2,0528	2,4255	1,3310	1,8472	1,5318
فلسطين	1,1665	1,1976	2,3034	0	4,5932	3,9544	0,4169	0,1578	1,3844	0,5053	1,0876
قطر	3,9721	4,1626	2,8757	4,5932	0	3,1860	4,4669	4,7416	4,1307	4,2708	4,0066
السعودية	3,1212	3,6658	3,1282	3,9544	3,1860	0	3,6956	4,0997	3,6726	3,6301	3,1441
الصومال	0,7822	0,8781	2,0528	0,4169	4,4669	3,6956	0	0,5105	1,0632	0,2262	0,7049
السودان	1,2828	1,2690	2,4255	0,1578	4,7416	4,0997	0,5105	0	1,4533	0,6227	1,2058
سوريا	0,7206	0,1876	1,3310	1,3844	4,1307	3,6726	1,0632	1,4533	0	0,9308	0,7669
تونس	0,6836	0,7449	1,8472	0,5053	4,2708	3,6301	0,2262	0,6227	0,9308	0	0,6136
اليمن	0,0896	0,6519	1,5318	1,0876	4,0066	3,1441	0,7049	1,2058	0,7669	0,6136	0

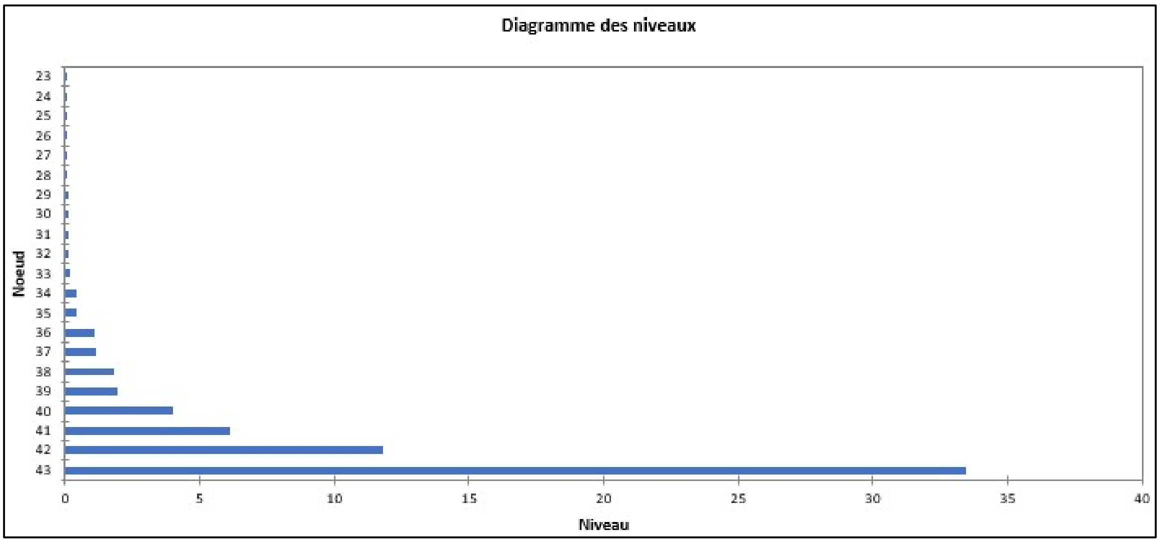
نلاحظ أنه تم حساب المسافات لاثنين وعشرين دولة (22)، فقد تم إدخال دولة اليمن في التصنيف، بعد تعويض قيمتها المفقودة.

4) إحصائيات العقد (Statistiques des noeuds) (كمخطط التجميع في SPSS):

Noeud	Niveau	Poids	Objets	Fils gauche	Fils droit
43	33,4764	22	22	40	42
42	11,8180	19	19	39	41
41	6,1407	11	11	34	36
40	4,0099	3	3	16	38
39	1,9518	8	8	4	37
38	1,7885	2	2	1	17
37	1,1683	7	7	5	35
36	1,1101	8	8	31	33
35	0,4253	6	6	28	30
34	0,4025	3	3	9	29
33	0,1645	4	4	10	32
32	0,1375	3	3	3	27
31	0,1154	4	4	24	25
30	0,1059	3	3	11	26
29	0,0996	2	2	2	14
28	0,0427	3	3	18	23
27	0,0176	2	2	13	20
26	0,0125	2	2	15	19
25	0,0059	2	2	6	7
24	0,0040	2	2	12	22
23	0,0028	2	2	8	21

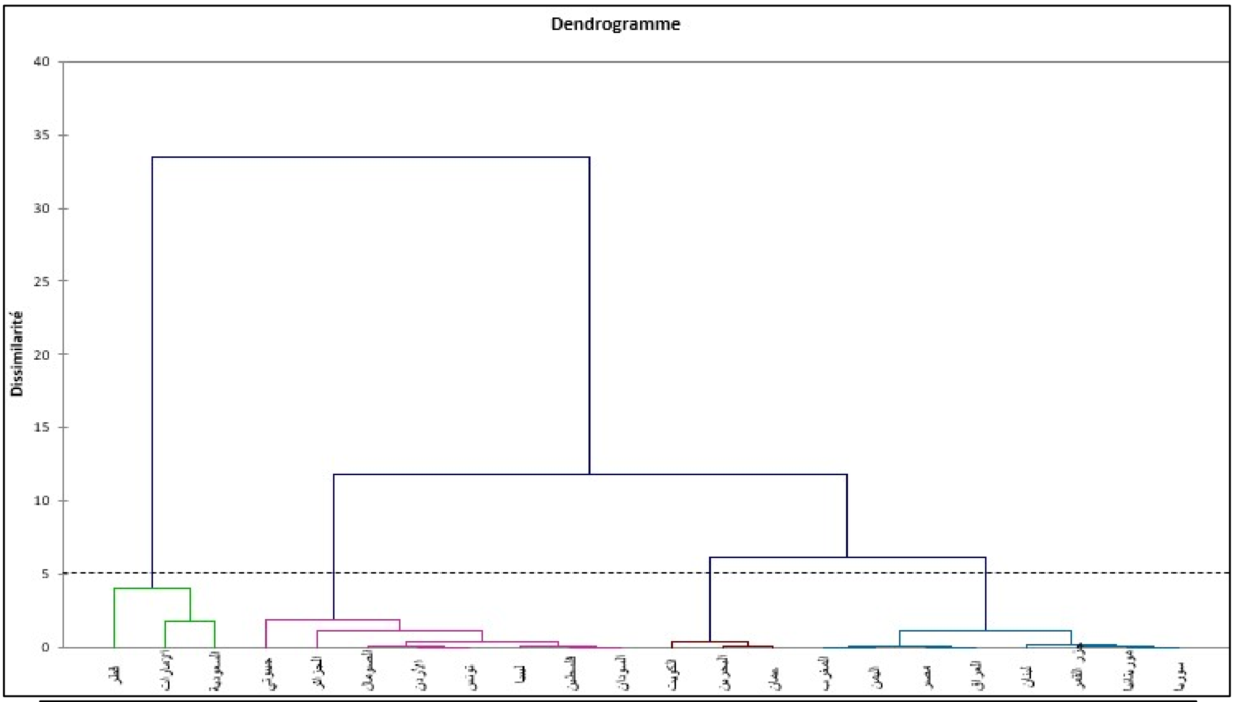
يبين لنا هذا الجدول جميع مراحل التجميع، فنلاحظ أن أول مرحلة تجميع (أقرب دولتين) كانت بجمع رقم8 ورقم21 في أول مجموعة (الأردن وتونس)، ثم في المرحلة الثانية تم جمع العنصرين 12 و22 في المجموعة الثانية (المغرب واليمن)، ثم في المرحلة الثالثة تم ضم العنصرين 6 و7 في المجموعة الثالثة (العراق ومصر)، ...، وهكذا حتى أخر مرحلة.

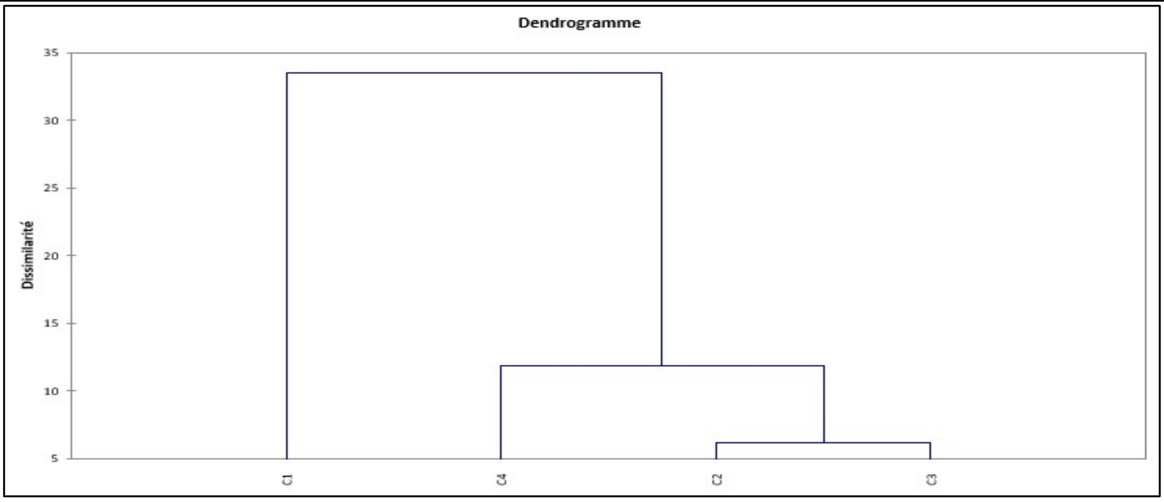
5) التمثيل البياني للمستويات (غير متوفر على SPSS):



يبين هذا المنحنى مستوى التباين داخل المجموعات، فيبدأ بأصغر مستوى وهي المجموعة 23 (ضم أول عنصرين قريبين 8 و21 (الأردن وتونس))، إلى أكبر مستوى وهي المجموعة 43 (ضم جميع الدول في مجموعة واحدة).

6) شجرة التصنيف (Dendrogramme):





يتبين من شجرة التصنيف وجود أربع مجموعة (تحديد اوتوماتيكي بحسب التباين داخل المجموعات والتباين بين المجموعات):

- ✓ المجموعة الأولى تضم الدول (3): الإمارات، السعودية، قطر.
- ✓ المجموعة الثانية تضم الدول (3): البحرين، عمان، الكويت.
- ✓ المجموعة الثالثة تضم الدول (8): الموريتانيا، سوريا، جزر القمر، لبنان، مصر، العراق، المغرب، اليمن.
- ✓ المجموعة الرابعة تضم الدول (8): الأردن، تونس، الصومال، فلسطين، السودان، ليبيا، جيبوتي، الجزائر.

7) تحليل التباين للتصنيف الأمثل (غير متوفر على برنامج SPSS):

Décomposition de la variance pour la classification optimale :		
	Absolu	Pourcentage
Intra-classe	9643731429028360000,0000	25,11%
Inter-classes	28756853728751200000,0000	74,89%
Totale	38400585157779500000,0000	100,00%

يتبين من الجدول أنه عند التصنيف الأمثل للدول (أربع مجموعات السابقة)، فإن نسبة التباين داخل المجموعات تقدر بـ 25,11%، ونسبة التباين بين المجموعات فتقدر بـ 74,89%.

8) العناصر المركزية للمجموعات والمسافات بينها (غير متوفر على برنامج SPSS):

Objets centraux :				
Classe	GDP	GDI	UEM	
1 (الإمارات)	48920,0000	67905762896,0000	2,3100	
2 (عمان)	18965,0000	7504101535,0000	3,7600	
3 (لبنان)	7057,0000	6302051983,0000	8,6900	
4 (ليبيا)	6452,0000	6359683962,0000	19,6000	
Distances entre les objets centraux :				
	1 (الإمارات)	2 (عمان)	3 (لبنان)	4 (ليبيا)
1 (الإمارات)	0	60401661361,0074	61603710913,0142	61546078934,0146
2 (عمان)	60401661361,0074	0	1202049552,0590	1144417573,0684
3 (لبنان)	61603710913,0142	1202049552,0590	0	57631979,0032
4 (ليبيا)	61546078934,0146	1144417573,0684	57631979,0032	0

يبين الجدولين العنصر المركزي لكل مجموعة من المجموعات الأربعة التي تم تحديدها، فالعنصر المركزي للمجموعة الأولى هي دولة الإمارات، والعنصر المركزي للمجموعة الثانية هي دولة عمان، والعنصر المركزي للمجموعة الثالثة هي دولة لبنان، والعنصر المركزي للمجموعة الرابعة هي دولة ليبيا. وأقرب عنصرين مركزيين في المجموعات الأربعة هما لبنان وليبيا (المجموعتين 3 و4).

9) النتائج بحسب العنصر (كمجموعات التخصيص على SPSS):

Résultats par objet :	
Observation	Classe
الإمارات	1
البحرين	2
جزر القمر	3
جيبوتي	4
الجزائر	4
مصر	3
العراق	3
الأردن	4
الكويت	2
لبنان	3
ليبيا	4
المغرب	3
موريتانيا	3
عمان	2
فلسطين	4
قطر	1
السعودية	1
الصومال	4
السودان	4
سوريا	3
تونس	4
اليمن	3

يبين الجدول أعلاه تصنيف كل دولة في مجموعة، فالدول: الإمارات، السعودية، قطر صنفت في المجموعة الأولى (1)، والدول: البحرين، عمان، الكويت صنفت في المجموعة الثانية (2). وهكذا بالنسبة لجميع الدول.

5. قائمة المراجع:

1) صواليلي صدر الدين، " تحليل المعطيات"، دار هومة للطباعة والنشر والتوزيع، 2011، الجزائر.
2) زياد رشاد الراوي، "طرق التحليل الإحصائي متعدد المتغيرات"، المعهد العربي للتدريب والبحوث الإحصائية، الطبعة الأولى، 2017.
3) المعجم الوسيط، مجمع اللغة العربية، مكتبة الشروق الدولية، 2004.
4) Alian Baccini, Philippe Besse, « Exploration Statistique », Institut National des Sciences Appliquées, Toulouse, Juin 2020.
5) Arnaud MARTIN, « L'analyse de données », Polycopié de cours ENSIETA, - Réf. : 1463, Septembre 2004.
6) Gilbert Saporta, « Probabilité, analyse des données, et statistique », Edition TECHNIP, 2ème Edition, Paris, 2006.
7) Jean Stafford, Paul Bodson, « L'analyse multivariée avec SPSS », Presse de l'Université du Québec, Canada, 2006.
8) Jean-Marie Bouruche, Gilbert Saporta, « L'analyse des données », Presses universitaires de France, 5ème Edition, 1992.
9) Ludovic Lebart, Alain Morineau, Marie Piron, « Statistique exploratoire multidimensionnelle », DUNOD, Paris, 1995.
10) الموقع الرسمي للبنك الدولي: https://data.albankaldawli.org/country