الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة جيجل
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم العلوم المالية والمحاسبة
السنة الثالثة محاسبة
الأستاذة: عميور عزيزة

تحليل المركبات الرئيسية (PCA)
Analysis Component Principal

١ تعريف تحليل المركبات الرئيسية

.1 **تعريف الأسلوب:** هو أسلوب رياضي يقوم على أساس تحويل مجموعة من المتغيرات التفسيرية المترابطة فيما بينها إلى مجموعة جديدة من المتغيرات المستقلة (المتعامدة) تدعى **المركبات الرئيسية.**

حيث تعبر كل مركبة رئيسية عن توليفة خطية للمتغيرات الأصلية وفق الصيغة:

$$PC_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p \qquad (١)$$

.2 **الهدف من التقنية:** هي تقنية لتقليل الأبعاد لمجموعة بيانات كبيرة، حيث تحولها إلى مجموعة أصغر مع الاحتفاظ بأهم المعلومات في المجموعة الأصلية. الهدف الرئيسي هو تقليل تعقيد البيانات دون فقدان جوهر ما تتضمنه.

٢ أنواع تحليل المركبات الرئيسية

أ- الطريقة غير المعيارية (ACP Non-normé)

تستعمل في حالة **تجانس المتغيرات** (أي أن كل المتغيرات لها نفس وحدة القياس). في هذه الحالة يتم طرح متوسط المتغير من كل عناصر العمود للحصول على **مصفوفة ممركزة.**

ب- الطريقة المعيارية (ACP Normé)

تستعمل في حالة **عدم تجانس المتغيرات** (اختلاف وحدات القياس). يتم طرح متوسط المتغير والقسمة على الانحراف المعياري، للحصول على **مصفوفة ممركزة ومعيارية** (مختزلة)، مما يجعل كل المتغيرات بدون وحدة قياس.

ملاحظة هامة:

- في حالة البيانات **الممركزة فقط**، يتم الاعتماد على مصفوفة التباين والتباين المشترك لاستخراج المركبات الرئيسية والقيم الذاتية.
- في حالة البيانات **الممركزة والمعيارية**، يتم الاعتماد على مصفوفة الارتباط (Correlation Matrix).

٣ خطوات إجراء طريقة المركبات الرئيسية

لإجراء تحليل المركبات الرئيسية، نتبع الخطوات التالية:

1. تشكيل جدول البيانات ومعرفة البيانات

تستخدم هذه الطريقة مع الجداول التي تحتوي على بيانات كمية، وتأخذ الشكل التالي:

المتغيرات /الأفراد	X_1	X_2	$\dots$	X_p
ind_1	n_{11}	n_{12}	$\dots$	n_{1p}
ind_2	n_{21}	n_{22}	$\dots$	n_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
ind_n	n_{n1}	n_{n2}	$\dots$	n_{np}

حيث أن كل فرد لديه قيمة n_{ij} حسب المتغير X_j.

- X_j: يمثل المتغير j حيث $(j = 1, 2, \dots, p)$.
- n_{ij}: تمثل قيمة المتغير X_j عند الفرد i.
- الأفراد من 1 إلى n تمثل الأفراد الإحصائيين.

2. مصفوفة البيانات الأولية X

نكتب مصفوفة البيانات الأولية X ذات البعد $(n \times p)$ كما يلي:

$$X_{(n\times p)} = \begin{pmatrix} n_{11} & n_{12} & \dots & n_{1p} \\ n_{21} & n_{22} & \dots & n_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ n_{n1} & n_{n2} & \dots & n_{np} \end{pmatrix}$$

ملاحظات هندسية:

- كل فرد i يمثل في فضاء ذو بعد p بالشكل: $X_i = (n_{i1}, n_{i2}, \dots, n_{ip})$.
- كل متغير j يمثل في فضاء ذو بعد n بالشكل العمودي: $X_j = (n_{1j}, n_{2j}, \dots, n_{nj})'$.

مثال توضيحي:

يمثل الجدول التالي علامات 5 طلبة موزعة على ثلاثة مقاييس:

الأفراد / المتغيرات	X_1 (اقتصاد)	X_2 (إحصاء)	X_3 (تحليل البيانات)
1	11	10	13
2	11	8	11
3	14	13	15
4	10	8	11
5	9	11	10

ومنه، مصفوفة البيانات تُكتب بالشكل:

$$X_{(5,3)} = \begin{pmatrix} 11 & 10 & 13 \\ 11 & 8 & 11 \\ 14 & 13 & 15 \\ 10 & 8 & 11 \\ 9 & 11 & 10 \end{pmatrix}$$

2. مصفوفة الأوزان وإحداثيات مركز ثقل سحابة النقاط

1-2 مصفوفة الأوزان:

فبإعطاء كل فرد نفس الثقل أي $p_i = \frac{1}{n}$ حيث $\sum_{i=1}^{n} p_i = 1$ (نفس الوزن). فإن مصفوفة الأوزان تُعطى بالعلاقة:

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

حيث I_n هي مصفوفة الوحدة.

2-2 مركز ثقل سحابة النقاط:

مع العلم أن المبدأ الصفري لا يؤدي إلى الحصول على تمثيل جيد لسحابة النقاط، لذلك نقوم بأخذ مركز ثقل السحابة كمبدأ، وعليه فإن مركز ثقل السحابة هو متوسط المتغيرات:

$$G_X = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{pmatrix}$$

تُكتب عن طريق المصفوفات بالعلاقة:

$$\boxed{G_X = X^T D_p 1_n} \qquad (٢)$$

حيث:

- 1_n: متجه الوحدة.
- D_p: مصفوفة الأوزان.
- X^T: منقول مصفوفة البيانات.

مثال: بأخذ المثال السابق، فإن مصفوفة الثقل هي:

$$D_p = \begin{pmatrix} 1/5 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 0 & 0 & 0 \\ 0 & 0 & 1/5 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 1/5 \end{pmatrix}$$

حساب متوسطات المتغيرات الثلاثة

	X_1	X_2	X_3
المتوسط $\bar{X}_j$	$\frac{11+11+14+10+9}{5} = \frac{55}{5} = 11$	10	12

أي أن: $G_X = \begin{pmatrix} 11 & 10 & 12 \end{pmatrix}$

3. المصفوفة المركزة (La matrice centrée)

يتم حساب مصفوفة البيانات المركزة X_c بطرح الأوساط الحسابية لكل متغير من القيم الأصلية وفق الصيغة التالية:

$$X_c = X - 1_n \cdot \bar{G}_x^T \qquad (٣)$$

حيث أن:

- X: مصفوفة البيانات الأولية.
- 1_n: متجه الوحدة.
- $\bar{G}_x^T$: منقول مصفوفة (متجه) الأوساط الحسابية.

ويمكننا كذلك حسابها مباشرة بطرح الوسط الحسابي لكل متغير من قيم العمود المقابل له، حيث: $x_{cij} = x_{ij} - \bar{X}_j$. **ملاحظة:** جعل البيانات مركزة لا يؤثر على شكل سحابة النقاط، بل يؤدي فقط إلى نقل مركز ثقل السحابة إلى مبدأ الإحداثيات.

ملاحظة: جعل البيانات مركزة لا يؤثر على شكل سحابة النقاط.

مثال: بأخذ المثال السابق لعلامات 5 طلبة في 3 مقاييس

لدينا: $\bar{X}_1 = 11$ ، $\bar{X}_2 = 10$ ، $\bar{X}_3 = 12$ أي أن: $G_X = \begin{pmatrix} 11 & 10 & 12 \end{pmatrix}$

ومن ثم، تكون مصفوفة البيانات المركزة X_c كالتالي:

$$X_c = \begin{pmatrix} 11-11 & 10-10 & 13-12 \\ 11-11 & 8-10 & 11-12 \\ 14-11 & 13-10 & 15-12 \\ 10-11 & 8-10 & 11-12 \\ 9-11 & 11-10 & 9-12 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & -1 \\ 3 & 3 & 3 \\ -1 & -2 & -1 \\ -2 & 1 & -3 \end{pmatrix}$$

حساب المصفوفة المركزة باستعمال الطريقة الثانية X_c:

$$X_c = X - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{pmatrix} 11 & 10 & 12 \end{pmatrix}$$

$$= \begin{pmatrix} 11 & 10 & 13 \\ 11 & 8 & 11 \\ 14 & 13 & 15 \\ 10 & 8 & 11 \\ 9 & 11 & 9 \end{pmatrix} - \begin{pmatrix} 11 & 10 & 12 \\ 11 & 10 & 12 \\ 11 & 10 & 12 \\ 11 & 10 & 12 \\ 11 & 10 & 12 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & -1 \\ 3 & 3 & 3 \\ -1 & -2 & -1 \\ -2 & 1 & -3 \end{pmatrix}$$

4. المصفوفة المعيارية (المركزة والمختزلة):

لاختزال البيانات، نقسم كل عمود في المصفوفة المركزة على الانحراف المعياري المقابل له، وعليه نحسب أولاً الانحراف المعياري:

$$\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2} \quad , \quad j = 1, 2, \ldots, p$$

نقوم بحساب مصفوفة الجدول المركز والمختزل مباشرة بقسمة إحداثيات كل عمود في الجدول المركز على الانحراف المعياري المقابل له وفقاً للعلاقة:

$$(x_{cr})_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

$$X_{cr} = \begin{pmatrix} \frac{x_{11}-\bar{x}_1}{\sigma_1} & \cdots & \frac{x_{1p}-\bar{x}_p}{\sigma_p} \\ \vdots & \ddots & \vdots \\ \frac{x_{n1}-\bar{x}_1}{\sigma_1} & \cdots & \frac{x_{np}-\bar{x}_p}{\sigma_p} \end{pmatrix}$$

ملاحظة: يسمح تحويل البيانات إلى بيانات مختزلة بتوحيد وحدات القياس.

مثال: نعتمد في حساب المصفوفة المعيارية على أعمدة المصفوفة المركزة X_c:

$$X_c = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & -1 \\ 3 & 3 & 3 \\ -1 & -2 & -1 \\ -2 & 1 & -3 \end{pmatrix}$$

1. المتغير الأول X_1:

$$V_1 = \frac{0^2 + 0^2 + 3^2 + (-1)^2 + (-2)^2}{5} = \frac{14}{5} = 2.8$$

$$\sigma_1 = \sqrt{2.8} \approx 1.673$$

2. المتغير الثاني X_2:

$$V_2 = \frac{0^2 + (-2)^2 + 3^2 + (-2)^2 + 1^2}{5} = \frac{18}{5} = 3.6$$

$$\sigma_2 = \sqrt{3.6} \approx 1.897$$

3. المتغير الثالث X_3:

$$V_3 = \frac{1^2 + (-1)^2 + 3^2 + (-1)^2 + (-3)^2}{5} = \frac{21}{5} = 4.2$$

$$\sigma_3 = \sqrt{4.2} \approx 2.049$$

حساب المصفوفة المركزة والمختزلة X_{cr}:

تُحسب إحداثيات الجدول المركز والمختزل وفق العلاقة:

$$(x_{cr})_{ij} = \frac{x_{ij} - \bar{X}_j}{\sigma_j}$$

$$X_{cr} = \begin{pmatrix} \frac{0}{1.673} & \frac{0}{1.897} & \frac{1}{2.049} \\ \frac{0}{1.673} & \frac{-2}{1.897} & \frac{-1}{2.049} \\ \frac{3}{1.673} & \frac{3}{1.897} & \frac{3}{2.049} \\ \frac{-1}{1.673} & \frac{-2}{1.897} & \frac{-1}{2.049} \\ \frac{-2}{1.673} & \frac{1}{1.897} & \frac{-3}{2.049} \end{pmatrix} = \begin{pmatrix} 0.00 & 0.00 & 0.49 \\ 0.00 & -1.05 & -0.49 \\ 1.79 & 1.58 & 1.46 \\ -0.60 & -1.05 & -0.49 \\ -1.20 & 0.53 & -1.46 \end{pmatrix}$$

5. حساب مصفوفة التباين والتباين المشترك أو مصفوفة الارتباط

5-1. مصفوفة التباين والتباين المشترك V

في حالة ما إذا كانت المتغيرات متجانسة (لها نفس وحدة القياس)، نستخدم هذه المصفوفة التي تُحسب بالعلاقة التالية:

$$V = X_c^T \cdot D_p \cdot X_c = \frac{1}{n} X_c^T \cdot X_c \tag{٤}$$

حيث أن:

- X_c: مصفوفة البيانات المركزة.
- X_c^T: منقول مصفوفة البيانات المركزة.
- D_p: مصفوفة الأوزان (في حالة الأوزان المتساوية $D_p = \frac{1}{n} I_n$).

تكتب المصفوفة V وهي مصفوفة مربعة متناظرة ذات بعد $(p \times p)$ على الشكل التالي:

$$V = \begin{pmatrix} V_1 & Cov(X_1, X_2) & \dots & Cov(X_1, X_p) \\ Cov(X_2, X_1) & V_2 & \dots & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & Cov(X_p, X_2) & \dots & V_p \end{pmatrix}$$

حيث أن عناصر القطر الرئيسي تمثل التباينات:

$$V_j = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{X}_j)^2 \tag{٥}$$

والعناصر خارج القطر تمثل التباينات المشتركة:

$$Cov(X_j, X_k) = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{X}_j)(x_{ik} - \bar{X}_k) \quad j, k = 1, ..., p \tag{٦}$$

ملاحظة: مصفوفة التباين والتباين المشترك هي دائماً مصفوفة متناظرة ($V^T = V$).

مثال (حسب المثال السابق):

بناءًا على مصفوفة البيانات المركزة X_c التي حصلنا عليها سابقاً:

$$V = \frac{1}{n} X_c^T \cdot X_c = \frac{1}{5} \begin{pmatrix} 0 & 0 & 3 & -1 & -2 \\ 0 & -2 & 3 & -2 & 1 \\ 1 & -1 & 3 & -1 & -3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & -1 \\ 3 & 3 & 3 \\ -1 & -2 & -1 \\ -2 & 1 & -3 \end{pmatrix}$$

$$V = \begin{pmatrix} 2.8 & -1.8 & 3.2 \\ -1.8 & 3.8 & 2 \\ 3.2 & 2 & 4.2 \end{pmatrix}$$

1-1-5. التباين الكلي لسحابة الأفراد (I_t)

يمثل **التباين الكلي Inertie)**) للسحابة تشتت النقاط بالنسبة لمركز ثقلها. كلما كان هذا التباين كبيراً، كلما كانت الأفراد مشتتة وبعيدة عن مركز ثقلها، والعكس صحيح. وهو عبارة عن مجموع العناصر القطرية لمصفوفة التباين والتباين المشترك، ويُعبر عنه بـ "أثر المصفوفة" :(Trace):

$$I_t = Trace(V) = \sum_{j=1}^{p} V_j \qquad (٧)$$

2-5. مصفوفة الارتباط R

في حالة ما إذا كانت المتغيرات **غير متجانسة**، نستخدم مصفوفة الارتباط بدلاً من مصفوفة التباين، حيث يتم حسابها باستخدام مصفوفة البيانات المركزة والمعيارية X_{cr}:

$$R = X_{cr}^T \cdot D_p \cdot X_{cr} = \frac{1}{n} X_{cr}^T \cdot X_{cr} \qquad (٨)$$

تأخذ المصفوفة R الشكل التالي:

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

حيث أن:

- العناصر على القطر تساوي دائماً **1** (ارتباط المتغير مع نفسه).
- العناصر خارج القطر r_{ij} هي معامل ارتباط "بيرسون" بين المتغيرين X_i و X_j:

$$r_{ij} = \frac{Cov(X_i, X_j)}{\sigma_i \cdot \sigma_j} \qquad (٩)$$

حيث σ يمثل الانحراف المعياري للمتغير.

ملاحظة: مصفوفة الارتباط هي مصفوفة مربعة ومتناظرة ذات بعد $(p \times p)$.

تطبيق على مصفوفة الارتباط R (من المثال السابق): تُحسب مصفوفة الارتباط انطلاقاً من المصفوفة الممركزة والمختزلة X_{cr} وفق العلاقة:

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

$$R = \frac{1}{5} \begin{pmatrix} 0 & 0 & 01.79 & -0.6 & -1.2 \\ 0 & -1.05 & 1.58 & -1.05 & 0.53 \\ 0.49 & -0.49 & 1.46 & -0.49 & -1.46 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0.49 \\ 0 & -1.05 & -0.49 \\ 1.79 & 1.58 & 1.46 \\ -0.6 & -1.05 & -0.49 \\ -1.2 & 0.53 & -1.46 \end{pmatrix}$$

=

$$R = \begin{pmatrix} 1.00 & 0.56 & 0.93 \\ 0.56 & 1.00 & 0.51 \\ 0.93 & 0.51 & 1.00 \end{pmatrix}$$

2-5 التباين الكلي لسحابة الأفراد: عند استخدام مصفوفة الارتباط، يُحسب التباين الكلي (I_t) بالعلاقة:

$$I_t = \text{trace}(R) = p$$

حيث أن p هو عدد المتغيرات.

ملاحظة هامة: اختيار نوع التحليل (PCA)

طريقة ممركزة ومختزلة (PCA Normé)	طريقة بسيطة (PCA Non-normé)
وحدات غير متجانسة	وحدات متجانسة
$S > 5$	$S < 5$
نعتمد على مصفوفة الارتباط R	نعتمد على مصفوفة التباين V

حيث أن $S = \frac{\max(\sigma_j)}{\min(\sigma_j)}$، فإذا كانت $S > 5$ نلجأ حتماً للطريقة المعيارية.

6. حساب القيم الذاتية والمتجهات الذاتية

1-6 حساب القيم الذاتية (λ):

يتم حساب القيم الذاتية انطلاقاً من مصفوفة التباين V أو مصفوفة الارتباط R من خلال إيجاد حلول المعادلات التالية (المعادلة المميزة):

$$P_V(\lambda) = \det(V - \lambda I) = 0$$

أو

$$P_R(\lambda) = \det(R - \lambda I) = 0$$

2-6 إيجاد المتجهات الذاتية ($\vec{u}$):

يتم إيجاد المتجه الذاتي $\vec{u}$ المرتبط بكل قيمة ذاتية λ من خلال حل الجملة:

$$(V - \lambda I)\vec{u} = 0$$

أو

$$(R - \lambda I)\vec{u} = 0$$

حيث أن $\vec{u}$ هو متجه ذاتي غير صفري، عدد عناصره يساوي البعد p.
بعد الحصول على الشعاع الذاتي $\vec{u_\alpha}$ فإنه يجب ترجيحه:

$$u_\alpha = \frac{u_\alpha}{\|u_\alpha\|}$$

حيث:

$$\|u_\alpha\| = \sqrt{a^2 + b^2 + c^2}$$

يجب أن تكون طويلة كل شعاع ذاتي تساوي 1 و إذا لم تكن كذلك نقوم بقسمة كل عنصر من الشعاع الذاتي على طويلته حتى نحصل على 1.
ملاحظة:

- **في حالة الإعتماد على مصفوفة التباينات:** $I_t = \sum_{j=1}^{p} \lambda_j = \sum_{j=1}^{p} V(x_j)$
- **في حالة الإعتماد مصفوفة الارتباطات:** $I_t = \sum_{i=1}^{p} \lambda_i = p$

7. معايير تحديد عدد المحاور (المركبات الرئيسية)

أ. قاعدة نسبة التباين المفسر:

تقاس المساهمة النسبية لكل محور (أو قيمة ذاتية) في تفسير التباين الكلي للسحابة بالعلاقة التالية:

$$Cr(\lambda_\alpha) = \frac{\lambda_\alpha}{\sum_{i=1}^{p} \lambda_i} \times 100$$

حيث تمثل هذه النسبة مقدار المعلومات التي يحتفظ بها المحور α. نحسب نسبة التباين المفسر بالعلاقة التالية:

$$\%\ \text{Cumulative} = \frac{\sum_{i=1}^{k} \lambda_i}{\text{Trace}(V)}$$

يتم ترتيب القيم الذاتية تنازلياً وتشكيل الجدول التالي:

	F_1	F_2	F_3	$\dots$	F_p
القيم الذاتية	λ_1	λ_2	λ_3	$\dots$	λ_p
نسبة التباين	$\frac{\lambda_1}{\sum \lambda_i}$	$\frac{\lambda_2}{\sum \lambda_i}$	$\frac{\lambda_3}{\sum \lambda_i}$	$\dots$	$\frac{\lambda_p}{\sum \lambda_i}$
النسبة التراكمية	$\frac{\lambda_1}{\sum \lambda_i}$	$\frac{\lambda_1+\lambda_2}{\sum \lambda_i}$	$\frac{\lambda_1+\lambda_2+\lambda_3}{\sum \lambda_i}$	$\dots$	1

تكون المركبات الرئيسية المتحصل عليها جديدة و كافية لإعطاء صورة واضحة عن سحابة النقاط في المخطط العاملي إذا كانت نسبة التباين المفسر بالمركبات الرئيسية اكبر من 75%.

مثال:

تشكيل جدول القيم الذاتية وبيان موثوقية المركبات الرئيسية: نقوم بترتيب القيم الذاتية تنازلياً من الأكبر إلى الأصغر، ونحسب المساهمة النسبية (التباين المفسر) لكل محور:

الرقم	**القيمة الذاتية**	**النسبة المئوية**	**النسبة المئوية المتصاعدة**
1	λ_1	$\frac{\lambda_1}{\sum_{i=1}^p \lambda_i}$	$\frac{\lambda_1}{\sum_{i=1}^p \lambda_i}$
2	λ_2	$\frac{\lambda_2}{\sum_{i=1}^p \lambda_i}$	$\frac{\lambda_1+\lambda_2}{\sum_{i=1}^p \lambda_i}$
3	λ_3	$\frac{\lambda_3}{\sum_{i=1}^p \lambda_i}$	$\frac{\lambda_1+\lambda_2+\lambda_3}{\sum_{i=1}^p \lambda_i}$

بناءً على ترتيب القيم الذاتية تنازلياً، نحصل على النتائج التالية:

F	**القيمة الذاتية**	**النسبة المئوية**	**النسبة المئوية المتصاعدة**
F1	0.666	57.12	57
F2	0.5	42.88	100
F3	0	0	100

تحليل موثوقية المركبات الرئيسية: تكون المركبات الرئيسية المتحصل عليها جيدة وكافية لإعطاء صورة واضحة عن سحابة النقاط في المخطط العاملي إذا كانت نسبة التباين الكلي المفسر بالمركبات الرئيسية أكبر من 75%. كل قيمة ذاتية تمثل المساهمة المطلقة للمحور في التباين الكلي, و عليه نجد:

- المساهمة المطلقة للمحور العاملي الأول في التباين الكلي تساوي 0.666.
- المساهمة المطلقة للمحور العاملي الثاني في التباين الكلي تساوي 0.5.

المساهمة النسبية للمحاور في التباين الكلي:

النسبة: $\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$ تمثل المساهمة النسبية للمحور i في التباين الكلي. وعليه، نجد:

- المساهمة النسبية للمحور العاملي الأول في التباين الكلي تساوي 57.12 في المائة.

- المساهمة النسبية للمحور العاملي الثاني في التباين الكلي تساوي 42.88 في المائة.

كما أن النسبة $\frac{\lambda_1}{\sum_{i=1}^{p}\lambda_i}$ تمثل التباين الكلي المفسر بالمحور العاملي الأول، وهي الأهمية النسبية للمركبة الرئيسية الأولى، وهي نسبة التباين الكلي الذي تمت المحافظة عليه، في المركبة الرئيسية الأولى.

وبناء عليه، يمكن القول: إن نسبة التمثيل الخاصة بالمركبة الرئيسية الأولى F_1 تساوي 57.12 في المائة من التباين الكلي، ونسبة التمثيل الخاصة بالمركبة الرئيسية الثانية تساوي 42.88 في المائة. فتكون نسبة التمثيل على المخطط العاملي المتكون من المحورين الأول والثاني، والممثلة بنسبة تباين كلي 100 في المائة، مما يعني إعطاء صورة واضحة جدا لسحابة النقاط على المخطط العاملي.

ب. قاعدة كايزر: (Kaiser)

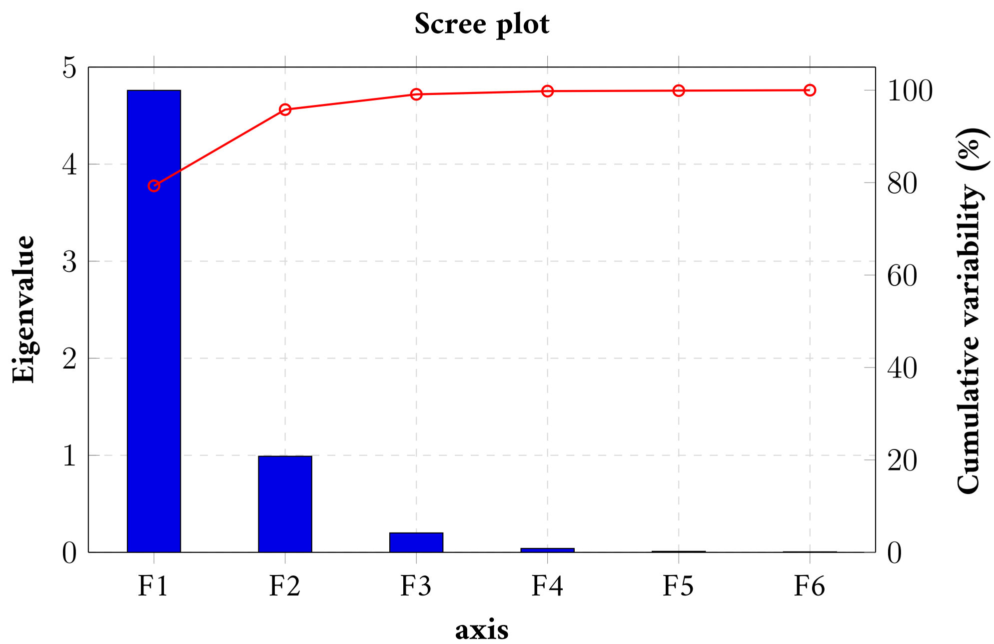
تنص على أن كل قيمة ذاتية تكون أكبر من متوسط القيم الذاتية تؤجد كمركبة رئيسية:

$$\lambda_i > \bar{\lambda} \quad \text{حيث} \; \bar{\lambda} = \frac{1}{p}\sum_{i=1}^{p}\lambda_i \qquad (١٠)$$

(في حالة المصفوفة المعيارية R، نختار $\lambda_i > 1$).

ج. معيار التمثيل البياني (Scree Plot)

بعد إنشاء التمثيل البياني بالأعمدة للقيم الذاتية، نقوم برسم منحنى يربط بين أكبر عدد من القيم الذاتية والعتبة الكبيرة التي تنتمي للمستقيم. تُمثل المحاور التي تُؤخذ في التحليل المحاور التي تسبق رؤوس الأعمدة الملتصقة بالمنحنى، وأين يتشكل لنا شكل "مرفق" (Coude) تتوقف عند تلك القيمة الذاتية.



مثال: "الانكسار" أو المرفق يقع عند المحور االثاني ، فإن عدد المركبات الرئيسية المختارة هو $\alpha = 2$.

8. المركبات الرئيسية للأفراد (المحاور العاملية)

المركبات الرئيسية هي المتغيرات التي تحددها العوامل الرئيسية، وتحتوي على إحداثيات الأفراد التي يتم حسابها عن طريق الإسقاط المتعامد لصفوف المصفوفة المركزة أوالمعيارية على كل الإتجاهات المحددة بواسطة المتجهات الذاتية U_α. يتم حساب المركبات

الرئيسية من خلال تقطير مصفوفة V أو R.

8-1. حالة تحليل المركبات الرئيسية معيارياً:

إذا كان التحليل معيارياً (بناءً على المصفوفة R)، فإننا نتحصل على المصفوفة F التي تسمح بإسقاط الأفراد على المحاور العاملية الجديدة وفق العلاقة التالية:

$$F_\alpha = X_{cr} \times U_\alpha \quad (11)$$

حيث أن:

- F: المركبة الرئيسية.
- α: عدد المحاور المأخوذة في التحليل.
- U_α: المتجه الذاتي المرفق للقيمة الذاتية λ_α.

8-2. حالة تحليل المركبات الرئيسية غير معيارية:

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

$$F_\alpha = X_c \times U_\alpha \quad (12)$$

حيث أن:

- F: المركبة الرئيسية.
- α: عدد المحاور المأخوذة في التحليل.
- U_α: المتجه الذاتي المرفق للقيمة الذاتية λ_α.

9. جودة التمثيل لزيادة دقة التمثيل

يتم قياس نسبة تمثيل الأفراد على المحاور، أي جودة تمثيلها على المستوى، باستخدام جيب تمام الزاوية (Cosine Squared) وفق العلاقة:

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

حيث أن: $\sum_{\alpha=1}^{p} \cos^2 \theta_{i\alpha} = 1$.

ملاحظة: كلما اقتربت القيمة من 1، كان الفرد ممثلاً بشكل أحسن على المحور α.

- $F_\alpha(i)$: إحداثيات الفرد i على المحور α.
- $\sum F_\alpha^2(i)$: مجموع مربعات إحداثيات الفرد i على جميع المحاور.

10. نسبة مساهمة الأفراد في تشكيل المحاور

يتم حساب نسبة مساهمة الأفراد في تشكيل المحاور بغرض التعرف على الأفراد الأكثر أهمية في تشكيل المحاور، وتُحسب وفق العلاقة:

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

حيث:

- C_i^α: نسبة مساهمة الفرد i في تشكيل المحور α.
- $F_\alpha^2(i)$: مربع إحداثيات الفرد iعلى المحور α.
- λ_α: القيمة الذاتية .

11. إحداثيات المتغيرات على المحاور

يتم الحصول عليها من خلال ضرب المتجهات الذاتية u_α في جذر القيم الذاتية المرافقة لها، وهي بمثابة الإحداثيات الجديدة للمتغيرات الأصلية في الفضاء:

$$G_\alpha = \sqrt{\lambda_\alpha} \cdot u_\alpha \qquad (١٣)$$

12. الارتباط بين المتغيرات الأصلية والمركبات الأساسية

في حالة استخدام الطريقة المركزية البسيطة ACP (Non-normé)، يُحسب الارتباط كالتالي:

$$Cor(X_j, F_\alpha) = \sqrt{\lambda_\alpha} \cdot \frac{u_{\alpha j}}{\sigma_j} \qquad (١٤)$$

أما بالصيغة المصفوفية:

$$Cor = D_{\sqrt{\lambda}} \cdot U^T \cdot D_{1/\sigma} \qquad (١٥)$$

وفي حالة استخدام الطريقة المعيارية المرجحة ACP (Normé)، فإن الارتباط يساوي:

$$Cor(X_j, F_\alpha) = \sqrt{\lambda_\alpha} \cdot u_{\alpha j} \qquad (١٦)$$

وتكتب مصفوفياً على النحو التالي:

$$Cor = U \cdot D_{\sqrt{\lambda}} \qquad (١٧)$$

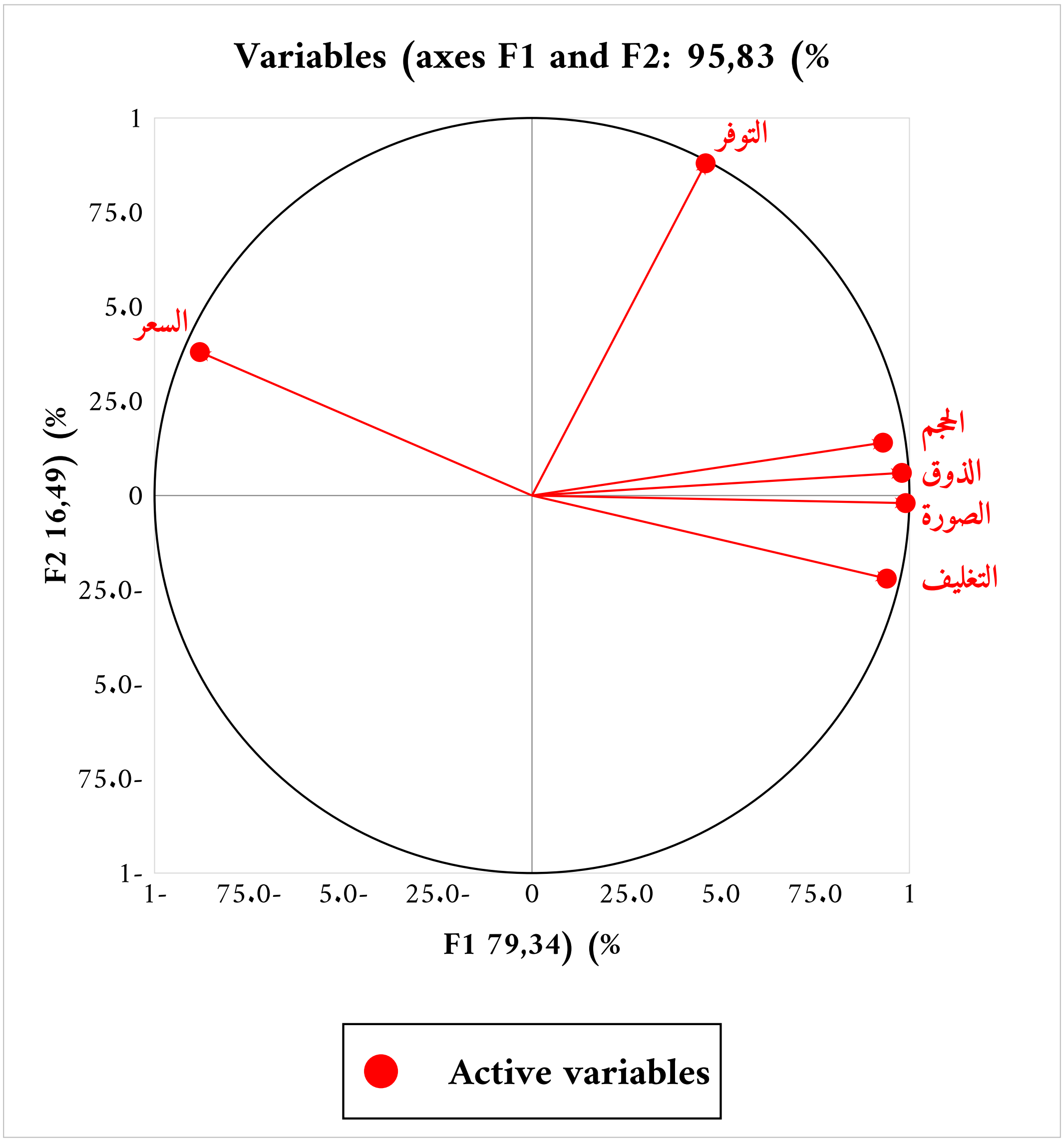
13. التمثيلات البيانية في الفضاء الجديد

13-1. دائرة الارتباطات

تُستخدم لتوضيح العلاقة بين المتغيرات الأصلية والمحاور الجديدة، حيث يتم تمثيل كل متغير بنقطة داخل دائرة الوحدة بالاعتماد على إحداثيات المتغيرات (ارتباطها مع المحاور). يتم تفسير دائرة الارتباطات كما يلي:

- **المتغيرات القريبة:** تكون مرتبطة طردياً فيما بينها (الزاوية بينها صغيرة).

- **المتغيرات المتعامدة:** تكون مستقلة تماماً (الزاوية بينها 90 درجة).
- **المتغيرات المتقابلة:** تكون مرتبطة عكسياً فيما بينها (الزاوية بينها كبيرة/منفرجة).
- **المتغيرات القريبة من محيط الدائرة:** هي المتغيرات التي لها أطول أسهم، وتكون هي الأكثر تمثيلاً وتأثيراً على تلك المحاور.



.9 التمثيل المخطط الثنائي (Biplot)

هو رسم بياني يجمع بين تمثيل الأفراد (نقاط) والمتغيرات (أسهم) في آن واحد، ويتميز بالآتي:

- **الأسهم:** تمثل اتجاه وقوة تأثير المتغيرات الأصلية.
- **النقاط:** تشير إلى تشابه أو تباعد خصائص الأفراد؛ النقاط القريبة من بعضها تعني تشابهاً في القيم.
- **الارتباط:** قرب نقطة (فرد) من سهم (متغير) يعني أن هذا الفرد يمتلك قيمة مرتفعة في ذلك المتغير.