A.C.P

المحور الثالث:

التحليل بالمركبات الأساسية Principal component analysis

تمهيد:

يمكن تجميع الظاهرة المدروسة الممثلة في قيم كمية في جدول، وتفسير هذا الجدول كمصفوفة، مثلا إذا اعتبرنا المصفوفة $X(p,q)$ حيث:

$$X(p,q)=\begin{array}{c} \quad j\rightarrow \\ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & \boldsymbol{x_{ij}} & \cdots & x_{iq} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pj} & \cdots & x_{pq} \end{pmatrix} \end{array} \boldsymbol{i}\downarrow$$

التفسير: يمكن التعبير عن المصفوف $X(p,q)$ بـ: p سطر $L_i(\mathrm{i}=\overline{1,\mathrm{p}})$ و q عمود $C_j(\mathrm{J}=\overline{1,\mathrm{q}})$.

بمعنى أن:

$$X(p,q)=\begin{pmatrix} L_1 \\ L_2 \\ L_3 \\ \vdots \\ L_p \end{pmatrix}=\begin{pmatrix} C_1 & C_2 & C_3 & \cdots & C_q \end{pmatrix}$$

وبالتالي: المصفوفة المنقولة لها هي:

$$X^t(q,p)=\begin{pmatrix} L_1 & L_2 & L_3 & \cdots & L_q \end{pmatrix}=\begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_p \end{pmatrix}$$

عموما:

- **السطر** $L_i(\mathrm{i}=\overline{1,\mathrm{q}})$ **يمثل قيم المتغيرات الـ (q) بالنسبة لـ: ($\boldsymbol{i}$) مشاهدة** $L_{i=}(x_{i1},x_{i2},\cdots,x_{iq})$
- **العمود** $C_j(\mathrm{j}=\overline{1,\mathrm{p}})$ **يمثل قيم المتغيرات الـ (p) بالنسبة لـ: ($\boldsymbol{j}$) مشاهدة** $C_{j=}(x_{1j},x_{2j},\cdots,x_{pj})$

تسمى الأسطر : المفردات (الأفراد)؛

وتسمى الأعمدة: المتغيرات.

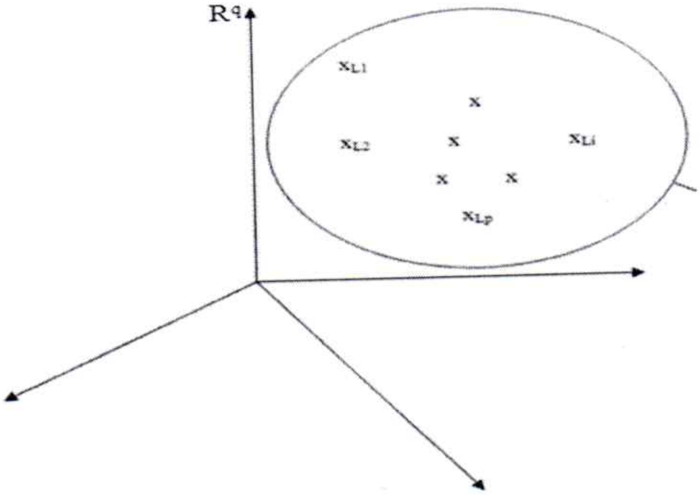
- يمكن تمثيل السطر L_i بواسطة نقطة في فضاء ذي بعد q $(\mathbb{R}^q)$؛
- يمكن تمثيل العمود C_j بواسطة نقطة في فضاء ذي بعد p $(\mathbb{R}^p)$.

اعداد: د/ إبراهيم رحيم

وعليه، يمكن تفسير المصفوفة X كبيانات من خلال:

1) q نقطة C_j في الفضاء $\mathbb{R}^q$، بمعنى أن كل نقطة تقبل p مركبة.
2) P نقطة L_I في الفضاء $\mathbb{R}^P$، بمعنى أن كل نقطة تقبل q مركبة.

الـ: ACP هندسيا، هي النقط على محور أو مستوى أو مستوى متعدد، كما هو موضح بالشكل التالي:



وعليه، فإن الـ: ACP تبحث عن دراسة اسقاطات سحابة النقط على محور (Axe) أو مستوى $(Plan)$ أو مستوى متعدد $(Hyper\ Plan)$.

رياضيا: الـ: ACP تعتبر أفضل تعديل (اسقاط) لسحابة النقط على فضاء جزئي لـ: $\mathbb{R}^q$.

تعريف الـACP: (اقترحها $(Spearman)$ وقد توسع استخدامها نتيجة التطور الحاصل في الإعلام الآلي)

هذه المنهجية (ACP) تسمح بتشخيص المعلومات مع السماح بفقد بعض المعلومات، وهي قائمة على مجموعة من الحسابات تعتمد على طريقتين:

1) طريقة مصفوفاتية$(Matrix\ Method)$: مبنية على الجبر الخطي.
2) طريقة تتابعية $(Iterative\ Method)$: مبنية على استخدام برامج تطبيقية للإعلام الآلي.

خطوات إجراء الـ: ACP:

يمكن تلخيصها في الخطوات التالية:

1) بناء جدول للقيم الممركزة أو القيم الممركزة المختصرة، بناء على جدول القيم الأصلي، مع الإشارة إلى أن القيم المتضمنة هي قيم كمية $(Quantitative\ Data)$.

2) الحصول على المصفوفة M (مصفوفة التباين–التباين المشترك، حيث: $M = X^t * X$.

3) نقوم بتقطير المصفوفة M للحصول على المصفوفة القطرية D حيث:
$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

علما أن القيم $\lambda_{i(i=\overline{1.n})}$ هي القيم الذاتية للمصفوفة M، وهذه القيم الذاتية تمثل نسبة التباين المفسر بواسطة العامل (المحور) (i).

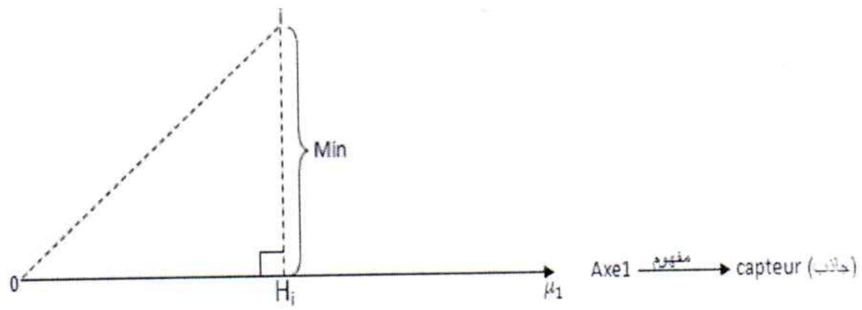
4) يتم إيجاد الأشعة الذاتية لكل القيم الذاتية $\lambda_{i(i=\overline{1.n})}$ المحددة للمركبات (العوامل).

5) حساب إحداثيات نقط الأسطر (المفردات) وإحداثيات الأعمدة (المتغيرات) لإجراء التمثيل البياني.

6) تفسير المحاور (المركبات) المتحصل عليها: حيث مفهوم المحور هو مفهوم المتغير **الكامن**.

فكرة أساسية: طريقة ACP تبحث عن أفضل إسقاط، أي بأفضل صورة ممثلة للسحابة. ويتم ذلك من خلال:

1) إيجاد العامل (المحور) الذي يمثل نوعا ما السحابة كما هو موضح في الشكل:



نريد البحث عن:

$$(iH_i)^2 Petit\ avec\ \in (Axe\ u_1) \Leftrightarrow (OH_i)^2 Grand \Rightarrow \left(\text{نظرية فيثاغورس}\right)$$

$$(Maximisation\ la\ dispertion\ des\ points) \Rightarrow \left(\sum (OH_i)^2\ Grand\right)$$

2) إيجاد أفضل مستوى، بمعنى البحث عن: $Max[\sum(OH_i)^2] avec\ H_i \in Plan$

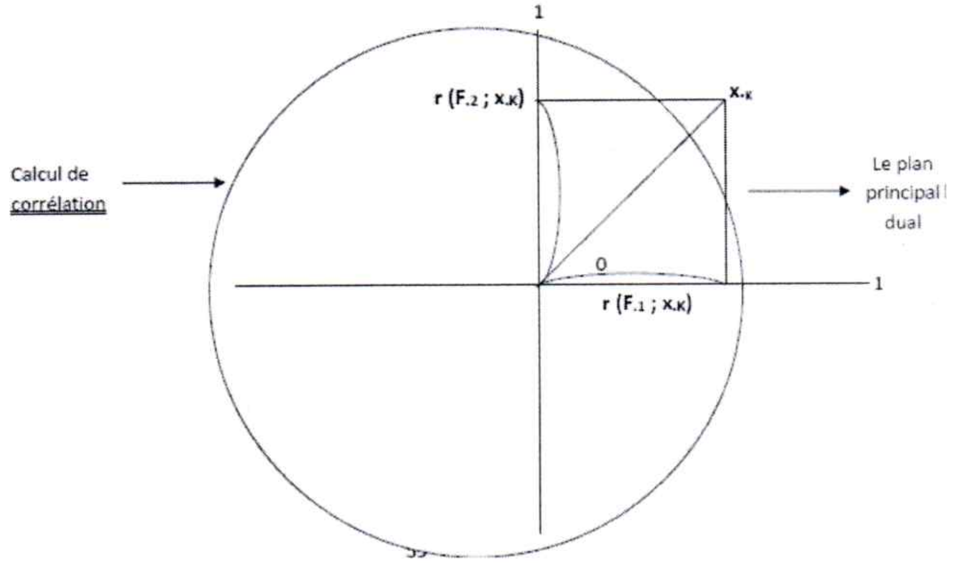
أفضل مستوى يتطلب أفضل محور، وعليه نبحث عن:

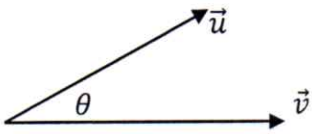
$$[u_2(Axe\ 2) \perp u_1] \wedge Max\left[\sum (OH_i)^2\right]$$

3) تكرار إيجاد المحاور u_3 ، u_4 ، ... مع $Max[\sum(OH_i)^2]$. لاحظ الشكل:

اعداد: د/ إبراهيم رحيم

3





مع ملاحظة أن: $\boxed{cos\theta \approx correlation\ (x_{.K}; F_{.1})}$

مع العلم أن $COS(\theta) = \frac{\vec{u}*\vec{v}}{\|\vec{u}\|\cdot\|\vec{v}\|} = \frac{COV(U.V)}{\sigma_u \sigma_v}$

بدءا من ثلاثة متغيرات أو ثلاثة أفراد، الـ: ACP تقوم بإجراء اسقاطات لنقط الأفراد ونقط المتغيرات على مستوى، بمعنى تقوم بإجراء تصوير فوتوغرافي، تسمى هذه الصورة الفوتوغرافية أحيانا الخريطة العاملية. هذه الصورة يجب أن تعكس بصفة دقيقة **المسافات بين الأفراد** وبصفة دقيقة **الارتباطات بين المتغيرات.**

تطبيق توضيحي:

الجدول التالي يمثل ثلاث علامات تجارية للسيارات، وتقدير ستة أفراد لهذه العلامات (تعطى نقطة من 0 إلى 10، أين النقطة 10 تفسر بتقدير جيد للعلامة على عكس النقطة 0)

Marques / Individus	B MW	Citroën	Renault
A	9	4	7
B	4	8	6
C	8	6	5
D	5	7	8
E	10	3	4
F	3	8	9
Moyennes	6.5	6	6.5

المطلوب: إجراء A.C.P على الجدول أعلاه

الحل:

<u>المرحلة 01</u>: إيجاد القيم الممركزة (أو القيم الممركزة المختصرة)

في هذا التطبيق، يكفي إيجاد القيم الممركزة لأن الجدول الأصلي يتضمن متغيرات من نفس الطبيعة، وعليه فإن إيجاد القيم الممركزة المختصرة في هذه الحالة ليس ضروريا.

$$X=\begin{pmatrix}2.5 & -2 & 0.5\\ -2.5 & 2 & -0.5\\ 1.5 & 0 & -1.5\\ -1.5 & 1 & 1.5\\ 3.5 & -3 & -2.5\\ -3.5 & 2 & 2.5\end{pmatrix}\Rightarrow X^t=\begin{pmatrix}2.5 & -2.5 & 1.5 & -1.5 & 3.5 & -3.5\\ -2 & 2 & 0 & 1 & -3 & 2\\ 0.5 & -0.5 & -1.5 & 1.5 & -2.5 & 2.5\end{pmatrix}$$

<u>المرحلة 02</u>: إيجاد مصفوفة التباين–التباين المشترك M حيث:

$$M=X^t*X=\begin{pmatrix}2.5 & -2.5 & 1.5 & -1.5 & 3.5 & -3.5\\ -2 & 2 & 0 & 1 & -3 & 2\\ 0.5 & -0.5 & -1.5 & 1.5 & -2.5 & 2.5\end{pmatrix}*\begin{pmatrix}2.5 & -2 & 0.5\\ -2.5 & 2 & -0.5\\ 1.5 & 0 & -1.5\\ -1.5 & 1 & 1.5\\ 3.5 & -3 & -2.5\\ -3.5 & 2 & 2.5\end{pmatrix}$$

ومنه:

$$M=X^t*X=\begin{pmatrix}\boxed{41.5} & -29 & -19.5\\ -29 & \boxed{22} & 12\\ -19.5 & 12 & \boxed{17.5}\end{pmatrix}$$

مؤثر $(Trace)$ المصفوفة M: هو مجموع قيم القطر الرئيسي للمصفوفة M حيث :

$$Trace(M)=41.5+22+17.5=81$$

<u>المرحلة 03</u>: تقطير المصفوفة M

بمعنى نقوم بحل المعادلة: $(*)$ ... $det(M-\lambda I_3)=0$

$$(*)\Leftrightarrow\begin{vmatrix}\mathbf{41.5-\lambda} & \mathbf{-29} & \mathbf{-19.5}\\ \mathbf{-29} & \mathbf{22-\lambda} & \mathbf{12}\\ \mathbf{-19.5} & \mathbf{12} & \mathbf{17.5-\lambda}\end{vmatrix}=\mathbf{0}\Leftrightarrow\mathbf{-\lambda^3+81\lambda^2-659\lambda+490.5=0}$$

وبعد الحل نحصل على ما يلي:

$\boldsymbol{\lambda_1=71.93}\ \rightarrow \boldsymbol{plus\ forte\ valeur}$

$\boldsymbol{\lambda_2=8.24}\ \rightarrow\ \boldsymbol{valeur}$ $Intermidiaire$

$\boldsymbol{\lambda_3=0.83}\ \rightarrow \boldsymbol{plus\ faible\ valeur}$

إذن المصفوفة القطرية هي: $D=\begin{pmatrix}71.5 & 0 & 0\\ 0 & 8.24 & 0\\ 0 & 0 & 0.83\end{pmatrix}$

ملاحظة: نلاحظ أن $TraceM \leftarrow 81 = 17.5 + 22 + 41.5 = \sum \boldsymbol{\lambda}_i$

ملاحظة هامة: القيم الذاتية تمثل النسب المئوية التالية:

$\boldsymbol{\lambda_1} = \frac{71.93}{81} = \mathbf{88.8\%}$ بمعنى أن المحور الأول يفسر $\mathbf{88.8\%}$ من التباين.

$\boldsymbol{\lambda_2} = \frac{8.24}{81} = \mathbf{10.2\%}$ بمعنى أن المحور الأول يفسر $\mathbf{10.2\%}$ من التباين.

$\boldsymbol{\lambda_3} = \frac{0.83}{81} = \mathbf{1\%}$ بمعنى أن المحور الأول يفسر $\mathbf{1\%}$ من التباين.

يمكن الاحتفاظ فقط بالمحورين 1 و2 لأنهما يفسران 99% من التباين.

المرحلة 04: حساب الأشعة الذاتية المرافقة للقيم الذاتية:

لدينا: $(M - \lambda I_3)\mathrm{u} = 0$ حيث: λ، u هما القيمة الذاتية والشعاع الذاتي الملحق بها للمصفوفة M.

- من أجل: $\boldsymbol{\lambda_1} = \mathbf{71.93}$ و $u_1 = \begin{pmatrix} x \\ y \\ w \end{pmatrix}$ يكون لدينا:

$$(M - \lambda_1 I_3)\mathrm{u} = 0 \Leftrightarrow \begin{pmatrix} -30.43 & -29 & -19.5 \\ -29 & -49.93 & 12 \\ -19.5 & 12 & 16.67 \end{pmatrix}\begin{pmatrix} x \\ y \\ w \end{pmatrix} = 0 \Leftrightarrow u_1 = (1 \quad -0.7 \quad -0.51)$$

- من أجل: $\boldsymbol{\lambda_2} = \mathbf{8.24}$ و $u_2 = \begin{pmatrix} x \\ y \\ w \end{pmatrix}$ يكون لدينا:

$$(M - \lambda_2 I_3)\mathrm{u} = 0 \Leftrightarrow \begin{pmatrix} 33.69 & -29 & -19.5 \\ -29 & 13.76 & 12 \\ -19.5 & 12 & 9.26 \end{pmatrix}\begin{pmatrix} x \\ y \\ w \end{pmatrix} = 0 \Leftrightarrow u_2 = (-1 \quad 2.08 \quad -0.51)$$

- من أجل: $\boldsymbol{\lambda_3} = \mathbf{0.93}$ و $u_3 = \begin{pmatrix} x \\ y \\ w \end{pmatrix}$ يكون لدينا:

$$(M - \lambda_3 I_3)\mathrm{u} = 0 \Leftrightarrow \begin{pmatrix} 40.67 & -29 & -19.5 \\ -29 & 21.17 & 12 \\ -19.5 & 12 & 16.67 \end{pmatrix}\begin{pmatrix} x \\ y \\ w \end{pmatrix} = 0 \Leftrightarrow u_3 = (2 \quad 2.39 \quad 0.62)$$

ملاحة هامة: ت1كر أنه يوجد عدد لا نهائي من الأشعة الذاتية المقابلة لـ $\boldsymbol{\lambda_i}$ $(i = 1.2.3)$، لذلك يفضل حساب شعاع الوحدة الذاتي على النحو التالي:

- $\|u_1\| = \sqrt{1^2 + (-0.7)^2 + (-0.51)^2} = 1.32$ ومنه: $u_{1\,norme} = (0.76. -9.53. -0.38\ )$؛
- $\|u_2\| = \sqrt{(-1)^2 + (2.08)^2 + (-4.82)^2} = 5.34$ ومنه: $u_{2\,norme} = (-0.19.0.39. -0.9\ )$؛
- $\|u_3\| = \sqrt{2^2 + (2.39)^2 + (0.62)^2} = 3.18$ ومنه: $u_{3\,norme} = (0.63\ .\ 0.75\ .\ \ 0.19\ )$؛

المرحلة 05: حساب نقط الإحداثيات

اعداد: د/ إبراهيم رحيم

أولا–نحسب إحداثيات نقط الأفراد، حيث نقوم بحساب الجداء المصفوفاتي التالي:

$$X * P = \begin{pmatrix} 2.5 & -2 & 0.5 \\ -2.5 & 2 & -0.5 \\ 1.5 & 0 & -1.5 \\ -1.5 & 1 & 1.5 \\ 3.5 & -3 & -2.5 \\ -3.5 & 2 & 2.5 \end{pmatrix} \begin{pmatrix} 0.76 & -0.19 & 0.63 \\ -0.53 & 0.39 & 0.75 \\ -0.38 & -0.9 & 0.19 \end{pmatrix}$$

$$X * P = \begin{array}{cc} \begin{pmatrix} 2.77 & -1.705 & 0.17 \\ -2.77 & 1.705 & -0.17 \\ 1.71 & 1.065 & 0.66 \\ -2.24 & -0.675 & 0.09 \\ 5.2 & 0.415 & -0.52 \\ -4.67 & -0.805 & -0.23 \end{pmatrix} & \begin{array}{l} \to M^rA \\ \to M^rB \\ \to M^rC \\ \to M^rD \\ \to M^rE \\ \to M^rF \end{array} \\ \begin{array}{ccc} Axe1 & Axe2 & Axe3 \end{array} & \end{array}$$

ثانيا–نحسب إحداثيات نقط المتغيرات، بمعنى علامات السيارات، نقوم بإجراء الجداء المصفوفاتي التالي:

$$P * \sqrt{D} = \begin{pmatrix} 0.76 & -0.19 & 0.63 \\ -0.53 & 0.39 & 0.75 \\ -0.38 & -0.9 & 0.19 \end{pmatrix} \begin{pmatrix} \sqrt{71.9} & 0 & 0 \\ 0 & \sqrt{8.23} & 0 \\ 0 & 0 & \sqrt{0.83} \end{pmatrix}$$

ومنه:

$$P * \sqrt{D} = \begin{array}{cc} \begin{pmatrix} 6.44 & -054 & 0/57 \\ -4.49 & 1.12 & 0/68 \\ -3.2 & -2.6 & 0.17 \end{pmatrix} & \begin{array}{l} \to BMW \\ \to Citroen \\ \to Renault \end{array} \\ \begin{array}{ccc} Axe1 & Axe2 & Axe3 \end{array} & \end{array}$$

المرحلة 06: التمثيل البياني

نحتفظ بالمحورين 1 و2 من أجل التمثيل في مستوى، إحداثيات النقط التي تم حسابها أعلاه في الجدول التالي:

Rubrique		$Axe\ (F_1)$	$Axe\ (F_1)$
Coordonnees Variables	BMW Citroen Renault	6.44 −4.49 −3.2	−0/54 1.12 −2.6

Coordonnees Variables	M^rA	2.77	-1.705
	M^rB	-2.77	1.705
	M^rC	1.71	1.065
	M^rD	-2.24	-0.675
	M^rE	5.2	0.415
	M^rF	-4.67	-0.805

المرحلة 07: التعليق على الخريطة العاملية (تسمية المحاور)

من خلال الخريطة العاملية نلاحظ أن العلامتين الفرنسيتين قريبتين من بعضهما، وعلى العكس فهما مناظرتين للماركة الألمانية. هذه المناظرة يمكن لها تفسير العامل الأول (1)، فالعلامتين: Citroën وRenault تجيب على احتياجات عائلية (سيارات عائلية)، بينما BMW تجيب على احتياجات رياضية.