تحليل المركبات الأساسية ACP
Analyse des correspondances principales

تنتمي طريقة تحليل المركبات الرئيسية إلى طرق الاحصاء الوصفي المتعدد الأبعاد (multidimensionnelle) والتي تسمى أيضا بالطرق العاملية (Factorielle). لا تعتمد هذه الطرق على النماذج الاحتمالية بل على نماذج هندسية تنطلق من جدول مستطيل الشكل يحتوي قياسات كمية لعينة أو مجتمع (معطيات قابلة للتكميم وليست نوعية) ذو بعد (nxp) ، بحيث يمثل p عدد الاعمدة (المتغيرات) و n عدد الأسطر (الأفراد).

نحاول من خلال هذا التمثيل الهندسي الكشف عن الارتباطات الموجودة بين المتغيرات والتعرف على الأفراد المتشابهة وكذا التعرف على الاحتياطات الواجب مراعاتها عند تحليل هاته الاشكال الهندسية من أجل الوصول إلى النتائج ،كما نشير إلى أن طريقة التحليل مرتبطة بمدى معرفتنا بمجال الدراسة.

وتعتبر أيضا طريقة استكشافية حيث نستطيع من الدراسة للخروج بفرضيات أولية يمكن أن تساعد في تحسين النتائج من خلال استخدام طرق اخرى.

1-المعطيات:

تعتبر المعطيات قياسات على الافراد $\{u_1, u_2, \dots . u_n\}$ والتي تمثل بدورها p متغير كمي $\{v_1, v_2, \dots . v_n\}$ يمكن وضع هذه القياسات في جدول من الشكل التالي:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \dots & x_{1j} & x_{1p} \\ x_{21} & x_{22} & x_{23} \dots & x_{2j} & x_{2p} \\ . & . & \dots\dots\dots\dots & . & . \\ x_{i1} & x_{i2} & x_{i3} \dots\dots & x_{ij} & x_{ip} \\ x_{n1} & x_{n2} & x_{n3} \dots & x_{nj} & x_{np} \end{bmatrix}$$

يمكن تمثيل كل فرد بشعاع يمثل قياسات p متغير: $u_i \in R^p$

$$^t u_i = [x_{i1} \quad x_{i2} \quad x_{i3} \quad \dots x_{ij} \quad \dots\dots x_{ip}];\ u_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ . \\ x_{ij} \\ . \\ x_{ip} \end{bmatrix}$$

بصفة مماثلة نمثل كل متغير بشعاع من R^n ويمثل مركبات قيم المتغير لـn فرد: $v_j \in R^n$

$$v_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ . \\ x_{ij} \\ . \\ x_{nj} \end{bmatrix}$$

مثال: يمثل الجدول التالي عينة من نقاط الطلبة:

	رياضيات	احصاء	منهجية	تاريخ الوقائع
علي	12	15	11	7
عمر	9	10	14	13
منال	10	10	8	9
خديجة	13	15	9	8
رحاب	8	9	14	16
ألاء	15	17	9	10

على سبيل المثال u_4 تمثل نقاط الطالبة خديجة أما v_2 فيمثل نقاط كل الطلبة في مادة الاحصاء.

$$u_4 = \begin{bmatrix} 13 \\ 15 \\ 9 \\ 8 \end{bmatrix}; \quad v_2 = \begin{bmatrix} 15 \\ 10 \\ 10 \\ 15 \\ 9 \\ 17 \end{bmatrix}$$

بالنسبة للتمثيل الهندسي فيمكن تمثيل أي فرد في معلم بـ04 محاور والتي تمثل المتغيرات (04 مواد) أما المتغيرات فتمثل في 06 محاور والتي تمثل الأفراد (الطلبة علي، عمر.....)

2- مفهوم المسافة:

من أجل تمثيل الأفراد في الفضاء يجب التطرق إلى المسافة المستخدمة في طريقة تحليل المركبات الرئيسية والتي تكون مسافة أقليديه (Distance Euclidienne):

تعطى المسافة الإقليدية بين فردين u_i و $u_{i'}$ بالعلاقة التالية:

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

استخدام المسافة الإقليدية يمنح نفس الدور لكل متغير ونعمل في أساس متعامد. يمكن أن نرفق هذه المسافة بجداء سلمي يعطى:

$$\langle \overrightarrow{ou}_i, \overrightarrow{ou}_{i'} \rangle = \sum_{j=1}^{p} x_{ij} x_{i'j} = u_i^t u_{i'}$$

وطاولة الشعاع: $\|\overrightarrow{ou}_i\|^2 = \sum_{j=1}^{p} \left(x_{ij}\right)^2 = u_i^t u_i$

ومنه نستطيع حساب الزاوية بين شعاعين:

$$\cos\alpha = \frac{\langle \overrightarrow{ou}_i, \overrightarrow{ou}_{i'} \rangle}{\|\overrightarrow{ou}_i\|.\|\overrightarrow{ou}_{i'}\|} = \frac{u_i^t u_{i'}}{(u_i^t u_i).(u_{i'}^t u_{i'})}$$

3- السحابة النقطية:

عند تمثيل الافراد في معلم متعامد ومتجانس فإن المحاور تمثل المتغيرات ونحصل على شكل يسمى بالسحابة النقطية ،فعندما يفوق عدد المتغيرات 2 يصبح التمثيل في الفضاء ويستحيل في هذه الحالة قراءة أو إعطاء تفسير لهذه السحابة.

3-1-مركز السحابة النقطية:

نقطة مبدأ المعلم المتعامد والمتجانس O تمثل نقطة المبدأ ذات الإحداثيات المعدومة ،في حالة الافراد ذات القياسات الكبيرة (قيم كبيرة) تكون بعيدة عن المبدأ وهذا يصعب قراءة السحابة ،نقوم في هذه الحالة بتحول المبدأ إلى مركز السحابة النقطية والذي يمثل مركز الثقل.

في الحالات العامة تكون الافراد في تحليل الاحصائي ليس لها نفس الأثر بل كل فرد يؤثر حسب وزنه أو أهميته . نعطي في تحليل المركبات الرئيسية نفس الوزن للأفراد:

$$P_1 = P_2 = P_3 = \dots P_n = 1/n$$

$$\sum_{j=1}^{p} p_i \overrightarrow{Gu_i} = \vec{0}$$

$$G = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{pmatrix} = \begin{pmatrix} x_{.1} \\ x_{.2} \\ \vdots \\ x_{.j} \\ \vdots \\ x_{.p} \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_j \\ \bar{X}_p \end{pmatrix} = \bar{X}_j$$

أخذ كمركز الثقل يعني التعامل مع جدول ممركز والتي تسمى المصفوفة الممركزة:

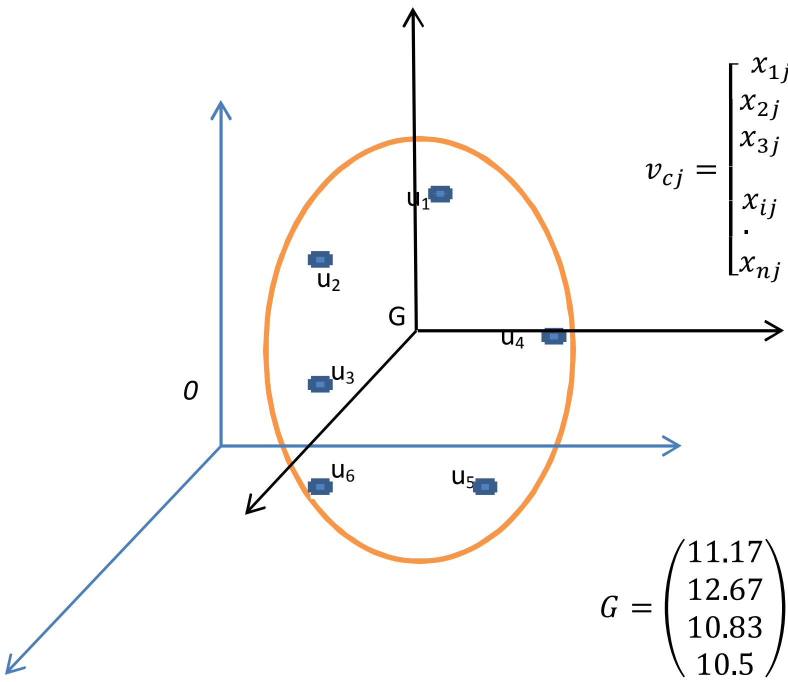
$$X_c = \begin{bmatrix} x_{11} - x_{.1} & x_{12} - x_{.2} & \cdots & x_{1j} - x_{.j} & \cdots & x_{1p} - x_{.p} \\ x_{21} - x_{.1} & x_{22} - x_{.2} & \cdots & x_{2j} - x_{.j} & \cdots & x_{2p} - x_{.p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} - x_{.1} & x_{i2} - x_{.2} & \cdots & x_{ij} - x_{.j} & \cdots & x_{ip} - x_{.p} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ x_{n1} - x_{.1} & x_{n2} - x_{.2} & \cdots & x_{nj} - x_{.j} & \cdots & x_{np} - x_{.p} \end{bmatrix}$$

حيث يمثل كل عمود الاحداثيات الممركزة للفرد i.

$$u_{ci} = \begin{bmatrix} x_{i1} - x_{.1} \\ x_{i2} - x_{.2} \\ x_{i3} - x_{.3} \\ \vdots \\ x_{ij} - x_{.j} \\ \vdots \\ x_{ip} - x_{.p} \end{bmatrix}$$

أما بالنسبة للإحداثيات الممركزة للمتغير j.

$$v_{cj} = \begin{bmatrix} x_{1j} - x_{.j} \\ x_{2j} - x_{.j} \\ x_{3j} - x_{.j} \\ \vdots \\ x_{ij} - x_{.j} \\ \vdots \\ x_{nj} - x_{.j} \end{bmatrix}$$



بالنسبة لمثال نقاط الطلبة نحصل على:

$$G = \begin{pmatrix} 11.17 \\ 12.67 \\ 10.83 \\ 10.5 \end{pmatrix} = \begin{pmatrix} x_{.1} \\ x_{.2} \\ x_{.3} \\ x_{.4} \end{pmatrix}$$

$$v_{cj}=\begin{bmatrix} x_{12}-x_{.2}\\ x_{22}-x_{.2}\\ x_{32}-x_{.2}\\ .\\ x_{i2}-x_{.2}\\ .\\ x_{n2}-x_{.2}\end{bmatrix};\quad v_{c2}=\begin{bmatrix}15-12.67\\ 10-12.67\\ 10-12.67\\ 15-12.67\\ 9-12.67\\ 17-12.67\end{bmatrix}=\begin{bmatrix}2.33\\ -2.67\\ -2.67\\ 2.33\\ -3.67\\ 4.33\end{bmatrix}$$

$$u_{ci}=\begin{bmatrix} x_{31}-x_{.1}\\ x_{32}-x_{.2}\\ x_{33}-x_{.3}\\ .\\ x_{3j}-x_{.j}\\ .\\ x_{3p}-x_{.p}\end{bmatrix};\quad u_{c4}=\begin{bmatrix}13-11.17\\ 15-12.67\\ 9-10.83\\ 8-10.5\end{bmatrix}=\begin{bmatrix}1.83\\ 2.33\\ -1.83\\ -2.5\end{bmatrix}$$

$$X_c=\begin{bmatrix} 0.83 & 2.33 & 0.167 & -3.5\\ -2.167 & -2.67 & 3.167 & 2.5\\ -1.167 & -2.67 & -2.83 & -1.5\\ 1.83 & 2.33 & -1.83 & -2.5\\ -3.167 & -3.67 & 3.167 & 5.5\\ 3.83 & 4.3 & -1.83 & -0.5\end{bmatrix}$$

4-العزوم(*Inertie*)

نرمز بـI_G لعزم سحابة الأفراد بالنسبة لمركز الثقل:

$$I_G=\frac{1}{n}\sum_{i=1}^{n}d^2(G,u_i)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(x_{ij}-x_{.j}\right)^2=\frac{1}{n}\sum_{i=1}^{n}u_{ci}^{t}\,u_{ci}$$

يقيس العزم مدى تشتت السحابة حول مركز الثقل، كلما كان قيمة العزم صغيرة كلما كانت السحابة مركزة بجوار المركز أي G والعكس صحيح، يمكن صياغة الكتابة السابقة من الشكل:

$$I_G=\sum_{j=1}^{p}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-x_{.j}\right)^2\right]=\sum_{j=1}^{p}Var\left(v_j\right)=Tr(\Sigma)$$

$$\Sigma=\frac{1}{n}X_c^tX_c$$

حيث Σ تمثل مصفوفة التباين والتباين المشترك و Tr أثر المصفوفة (مجموع قيم القطر الرئيسي للمصفوفة)، بالنسبة للمثال السابق:

$$\Sigma=\frac{1}{n}\begin{bmatrix} 0.83 & -2.167 & -1.167 & 1.83 & -3.167 & 3.83\\ 2.33 & -2.67 & -2.67 & 2.33 & -3.67 & 4.3\\ 0.167 & 3.167 & -2.83 & -1.83 & 3.167 & -1.83\\ -3.5 & 2.5 & -1.5 & -2.5 & 5.5 & -0.5\end{bmatrix}\begin{bmatrix} 0.83 & 2.33 & 0.167 & -3.5\\ -2.167 & -2.67 & 3.167 & 2.5\\ -1.167 & -2.67 & -2.83 & -1.5\\ 1.83 & 2.33 & -1.83 & -2.5\\ -3.167 & -3.67 & 3.167 & 5.5\\ 3.83 & 4.3 & -1.83 & -0.5\end{bmatrix}$$

$$\Sigma=\frac{1}{6}\begin{bmatrix} 34.83 & 43.33 & -23.83 & -30.5\\ 43.33 & 57.33 & -24.33 & -39\\ -23.83 & -24.33 & 34.83 & 34.5\\ -30.5 & -39 & 34.5 & 57.5\end{bmatrix}$$