تحليل البيانات بطريقة المركبات الأساسية

Principal Component Analysis

ACP

تمرين: يحتوي الجدول التالي على قياسات ضغط الدم الانقباضي وضغط الدم الانبساطي لستة أفراد:

	SBP	DBP
عادل	126	78
أمين	128	80
جود	128	82
أفنان	130	82
رشيد	130	84
وسيم	132	86

قم بتحليل الجدول بطريقة المركبات الأساسية مع احترام جميع الخطوات.

حل التمرين:

لتحليل البيانات في الجدول بطريقة المركبات الرئيسية ACP يجب إتباع الخطوات الأساسية التالية:

① - كتابة أو استخراج المصفوفة الأصلية للبيانات في الجدول =

$$A=\begin{pmatrix}126 & 78\\ 128 & 80\\ 128 & 82\\ 130 & 82\\ 130 & 84\\ 132 & 86\end{pmatrix}$$

② - إيجاد المصفوفة المركزة، العيارية B: وذلك بحساب المتوسط $\bar{X}$ وطرحه من المتغيرات في المصفوفة A:

$$B=\begin{pmatrix}x_1-\bar{x} & y_1-\bar{y}\\ \vdots & \vdots\\ x_n-\bar{x} & y_n-\bar{y}\end{pmatrix}=\begin{pmatrix}\cdots\\ \vdots\\ \cdots\end{pmatrix}$$

$$B=\begin{pmatrix}126-\bar{x} & 78-\bar{y}\\ \vdots & \vdots\\ 132-\bar{x} & 86-\bar{y}\end{pmatrix}$$

$$\bar{x}=\frac{1}{n}\sum x_i \qquad \bar{y}=\frac{1}{n}\sum y_i$$

$$\bar{x}=129 \qquad \bar{y}=82$$

$$B=\begin{pmatrix}126-129 & 78-82\\ 128-129 & 80-82\\ 128-129 & 82-82\\ 130-129 & 82-82\\ 130-129 & 84-82\\ 132-129 & 86-82\end{pmatrix}$$

$$B=\begin{pmatrix}-3 & -4\\ -1 & -2\\ -1 & 0\\ 1 & 0\\ 1 & 2\\ 3 & 4\end{pmatrix}\Leftarrow$$ المصفوفة المركزة.

③- ايجاد مصفوفة التباين والتباين المشترك

نسميها المصفوفة S : حيث:

$$S = \frac{1}{n} A^t \cdot A$$

أو

حساب Var ، cov لكل من xy:

$$S = \begin{bmatrix} Var(x) & cov(x,y) \\ cov(y,x) & Var(y) \end{bmatrix}$$

$$Var(x) = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

دقيقة =

$$S = \frac{1}{6} \begin{pmatrix} -3 & -1 & -1 & 1 & 1 & 3 \\ -4 & -2 & 0 & 0 & 2 & 4 \end{pmatrix} \cdot \begin{pmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$S = \begin{bmatrix} 3,667 & 4,667 \\ 4,667 & 6,667 \end{bmatrix}$$

أو يمكننا حسابها بالتباين، التباين المشترك سنجد نفس النتيجة.

④- حساب القيم الذاتية والمتجهات الذاتية

انطلاقا من مصفوفة التباين والتباين المشترك

- حساب القيم الذاتية: تحقيق العلاقة التالية: $\det(S - \lambda I) = 0$

نجد: $\lambda_1 = 10,068$ ، $\lambda_2 = 0,265$

نقوم بإنشاء جدول لإختيار المركبات الأساسية وإيجاد عدد المحاور.

لاختيار عدد المحاور نطبق إحدى الطريقتين:

ط1: إذا كانت نسبة التباين الإجمالي التي تفسرها هذه المركبات تفوق 75٪.

ط2: استخدام معيار كايزر Kaiser حيث تحد المركبات الرئيسية المستخرجة نوعية تباينية لتقدير تمثيل دقيق لسحابة النقاط في المخطط العاملي

إذا كانت القيمة الذاتية المواقفة أكبر من معدل القيم الذاتية:

إذا كان: $\lambda_1 > \bar{\lambda}$ ⟸ نحتفظ بالمركب CP_1

إذا كان: $\lambda_1 < \bar{\lambda}$ ⟸ نتخلى عن المركب CP_1

حيث: $\bar{\lambda} = \frac{1}{n} \sum \lambda_i$

القيمة الذاتية	النسبة المئوية ٪	النسبة المتراكمة ٪
$CP_1 = \lambda_1 =$ 10,068	$\frac{\lambda_1}{\sum \lambda_i} \times 100 =$ 97,437	$\frac{\lambda_i}{\sum \lambda_i} \times 100 =$ 97,437
$CP_2 = \lambda_2 =$ 0,265	$\frac{\lambda_2}{\sum \lambda_i} \times 100 =$ 2,563	$\frac{\lambda_1 + \lambda_2}{\sum \lambda_i} \times 100 =$ 100

إذن من خلال الجدول وباستعمال كلتا الطريقتين نجد أن CP_1 تمثل 97,4٪ بالتالي نحتفظ بها ونتخلى عن المركبة CP_2 لأنها تمثل 2,6٪ فقط

- ايجاد المتجهات الذاتية :

بتطبيق العلاقة $(S-\lambda I)(U)=(0)$

لـ $\lambda_1 = 10,068$:

$$\begin{bmatrix} -6,402 & 4,667 \\ 4,667 & -3,402 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

لدينا نظام متجانس من المعادلات الخطية نقوم بحلها بواسطة حذف تاوس من خلال عمليتين هما :

- أولاً نقسم العنصر المتغير الأول في المصفوفة ($a_{11} = -6,402$) على نفسه ليصير ساوي (1)

- ثانيا نعدم السطر الثاني :

الصف الثاني ← الصف الثاني $\mp a_{21}$ × الصف الأول

، فنجد :

- نقسم السطر الأول على $-6,402$ فنحصل :

$$\begin{bmatrix} 1 & -0,729 \\ 4,667 & -3,402 \end{bmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- نعدم السطر الثاني بضرب السطر الأول في $(-4,667)$ ثم جمعه مع السطر الثاني :

$$\begin{array}{cc} -4,667 & 3,402 \\ + & + \\ 4,667 & -3,402 \\ \hline 0 & 0 \end{array}$$

ملاحظة : نعكس اشارة a_{21} من أجل تحقيق النتيجة الصفرية بحيث إذا كان a_{21} موجب يحول سالب والعكس ، ثم يضرب في السطر الأول ويجمع مع السطر الثاني ، هكذا.

نتحصل على النتائج التالية :

$$\begin{bmatrix} 1 & -0,729 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

، فنجد :

$$U_1 + (-0,729)U_2 = 0 \Rightarrow U_1 = 0,729\, U_2$$

، فمن الشعاع الذاتي لـ $\lambda_1 = 10,068$ يعطينا الشعاع الذاتي $U_1 = \begin{pmatrix} 0,729 \\ 1 \end{pmatrix}$

- حساب شعاع الوحدة المرافق للقيمة الذاتية الأولى : $\lambda_1 = 10,068$

$$\|U_1\| = \sqrt{(U_1)^2 + (U_2)^2}$$

$$\|U_1\| = \sqrt{(0,729)^2 + (1)^2} = 1,237$$

ثم تقوم بقسمة :

العنصرة على الحاصل $(1,237)$ فنجد :

$$U = \begin{bmatrix} \frac{0,729}{1,237} \\ \frac{1}{1,237} \end{bmatrix} = \begin{bmatrix} 0,589 \\ 0,808 \end{bmatrix}$$

لـ $\lambda = 0,265$:

بنفس الطريقة السابقة ، نقس الخطوات

$$\begin{bmatrix} 3,402 & 4,667 \\ 4,667 & 6,402 \end{bmatrix} * \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

تقوم لحل النظام بواسطة حذف تاوس

$$U = \begin{pmatrix} \frac{-1,372}{1,697} \\ \frac{1}{1,697} \end{pmatrix} = \begin{pmatrix} -0,808 \\ 0,589 \end{pmatrix}$$

تكتب مصفوفة الأشعة الذاتية U:

$$U = \begin{pmatrix} 0,589 & -0,808 \\ 0,808 & 0,589 \end{pmatrix}$$

① حساب المركبات الرئيسية للأفراد على المحاور:

من خلال تطبيق العلاقة: $F = B.U$

بحيث مصفوفة المركبات الرئيسية تساوي المصفوفة المركزة B مضروبة في مصفوفة الأشعة U

$$F = \begin{pmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0,589 & -0,808 \\ 0,808 & 0,589 \end{pmatrix}$$

$$F = \begin{pmatrix} -5 & -0,068 \\ -2,205 & 0,370 \\ -0,589 & -0,808 \\ 0,589 & 0,808 \\ 2,205 & -0,370 \\ 5 & 0,068 \end{pmatrix}$$

	SBP	DBP
عادل	-5	-0,068
أمين	-2,205	0,370
جود	-0,589	-0,808
أنان	0,589	0,808
رشيد	2,205	0,370
وسيم	5	0,068

نقوم بقسمة السطر الأول على 3,402

$$\begin{bmatrix} 1 & 1,372 \\ 4,667 & 6,402 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

نقوم بجمع السطر الثاني بضرب السطر الأول في (-4,667)، بجمع السطر الثاني

$$\begin{array}{cc} -4,667 & -6,402 \\ \underline{4,667} & \underline{6,402} \\ 0 & 0 \end{array}$$

فنحصل على النتائج التالية:

$$\begin{bmatrix} 1 & 1,372 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$U_1 + 1,372\,U_2 = 0 \Rightarrow U_1 = -1,372\,U_2$$

منه الشعاع الذاتي لما $\lambda_1 = 0,265$

ليحصلنا الشعاع الذاتي: $U_2 = \begin{pmatrix} -1,372 \\ 1 \end{pmatrix}$

- حساب شعاع الوحدة المرافق للقيمة الذاتية الثانية:

$$\|U_2\| = \sqrt{(U_1)^2 + (U_2)^4} = \sqrt{(-1,372)^2 + (1)^2}$$

$$\|U_2\| = 1,697$$

نقوم بقسمة المصفوفة U_2 مع الحاصل 1,697 فنجد شعاع الوحدة المرافق

⑥- نسبة مساهمة الأفراد في تشكيل المحاور:

من خلال تطبيق العلاقة التالية:

$$C_i^a = \frac{F_a^2(i)}{n.\lambda_a}$$

حيث:

$$C_1^1 = \frac{(-5)^2}{6 \times 10,068} = 0,41376$$

$$C_2^1 = \frac{(-2,201)^2}{6 \times 10,068} = 0,805$$

وهكذا أيضا مع باقي المتغيرات.

فنتحصل على الجدول التالي:

	SBP	DBP
عادل	41,376	0,271
أمين	8,050	8,617
جود	0,574	41,092
أفنان	0,574	41,092
رشيد	8,050	8,617
وسيم	41,376	0,291
/	/	/

نلاحظ أن عادل ووسيم يساهمان بالنسبة نفسها في تشكيل المحور الأول: يساهمان معا بنسبة 82,7% في حين يساهم كل من جود وأفنان في تشكيل المحور الثاني بنفس النسبة، بمساهمة بنسبة 82,1% في تشكيل المحور الثاني.

⑦- جودة تمثيل الأفراد على المحاور:

من خلال تطبيق العلاقة التالية:

$$\cos^2\theta_{ix} = \frac{f_a^2(i)}{\sum f_a^2(i)}$$

حيث:

عادل: $\cos^2\theta_{11} = \frac{(-5)^2}{(-5)^2+(0,068)^2} \simeq 1$

أمين: $\cos^2\theta_{12} = \frac{(-2,201)^2}{(-2,201)^2+(0,370)^2} = 0,97$

أفنان: $\cos\theta_{14} = \frac{(0,589)^2}{(0,589)^2+(0,808)^2} = 0,347$

وهكذا مع جميع البيانات في الجدول فنتحصل على جدول تمثيل الأفراد على المحاور التالي:

	SBP	DBP
عادل	1	0
أمين	0,973	0,027
جود	0,347	0,653
أفنان	0,347	0,653
رشيد	0,973	0,027
وسيم	1	0

نلاحظ أن عادل، أمين، رشيد، وسيم يمثلون المحور الأول أحسن تمثيل، بينما جود، أفنان ساهما على المحور الثاني بشكل جيد.

⑧- احداثيات المتغيرات على المحاور:

(8) - ايجاد احداثيات المتغيرات مع المحاور:

وذلك بتطبيق العلاقة التالية:

$$G_\alpha = \sqrt{\lambda_i} U_i$$

إذن:

$$G_1 = \sqrt{\lambda_1} U_1 = \sqrt{10,068} \begin{pmatrix} 0,589 \\ 0,808 \end{pmatrix}$$

$$G_1 = \begin{pmatrix} 1,869 \\ 2,564 \end{pmatrix}$$

$$G_2 = \sqrt{\lambda_2} U_2 = \sqrt{0,265} \begin{pmatrix} -0,808 \\ 0,589 \end{pmatrix}$$

$$G_2 = \begin{pmatrix} -0,416 \\ 0,303 \end{pmatrix}$$

إذن نمثلها في الجدول:

	f_1	f_2
SBP	1,869	-0,416
DBP	2,564	0,303