UNIVERSITÉ LARBI BEN M'HIDI-OUM EL BOUAGHI

Département de Mathématique et Informatique

M1 VA   Corrigé type Contrôle N 01 TSD

Exercice 01 (12 points) : Analyse en Composantes Principales ACP

Soit le tableau de données de type $(3, 2)$ suivant :

$$X = \begin{pmatrix} 2 & 2 \\ 6 & 2 \\ 6 & 4 \\ 10 & 4 \end{pmatrix}$$

1. Les moyennes de X_1 et X_2 respectivement :
$\overline{x_1} = \frac{\sum_j x_{i1}}{4} = \frac{24}{4} = 6$ (0.25 pt), $\overline{x_2} = \frac{\sum_i x_{i2}}{n} = \frac{12}{4} = 3$ (0.25 pt).
La variance de X_1 :
$$var(X_1) = \frac{\sum_i xi1^2}{n} - \overline{x_1}^2 = \frac{176}{4} - 6^2 = 8 \ \ (0.5pt)$$
donc : $\sigma_{X_1} = \sqrt{var(X_1)} = 2.83.$ (0.25 pt)
La variance de X_2 :
$$var(X_2) = \frac{\sum_i xi2^2}{n} - \overline{x_2}^2 = \frac{40}{4} - 3^2 = 1 \ \ (0.5pt)$$
donc : $\sigma_{X_2} = \sqrt{var(X_2)} = 1.$ (0.25 pt)
On déduit la matrice centrée réduite X_{cr} comme suite :
$$X_{cr} = \begin{pmatrix} -1.41 & -1 \\ 0 & -1 \\ 0 & 1 \\ 1.41 & 1 \end{pmatrix} (1.25 \ pt)$$

2. Trouver R :
$$X_{cr}^t \times X_{cr} = \begin{pmatrix} -1.41 & 0 & 0 & 1.41 \\ -1 & -1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} -1.41 & -1 \\ 0 & -1 \\ 0 & 1 \\ 1.41 & 1 \end{pmatrix} = \begin{pmatrix} 3.98 & 2.82 \\ 2.82 & 4 \end{pmatrix}$$
Alors la matrice R de corrélations est :
$$R = \frac{1}{4} \times X_{cr}^t \times X_{cr} = \begin{pmatrix} 1 & 0.71 \\ 0.71 & 1 \end{pmatrix} (01pt)$$
D'après la matrice R, on a le coefficient de corrélation linéaire entre X_1 et X_2 :
$r_{X_1\,X_2} = 0.71 = \frac{Cov(X_1, X_2)}{\sigma_{X_1}\,\sigma_{X_2}}$, donc la liaison linéaire entre les deux variables est moyenne et négativement. (0.5 pt)
La droite d'ajustement linéaire (de régression) de X_2 en X_1 : $x_2 = a\ x_1 + b$ (0.25 pt)tels que $a = \frac{Cov(X_1, X_2)}{var(X_1)} = 0.25\ \ (0.25pt)$ et $b = \overline{x_2} - a\,\overline{x_1} = 1.5$ (0.25 pt).

3. Les valeurs propres λ_i sont les racines de l'équation $det(R - \lambda_i\ I) = 0$

$$R - \lambda_i\ I = \begin{pmatrix} 1 & 0.71 \\ 0.71 & 1 \end{pmatrix} - \lambda_i \times \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1-\lambda_i & 0.71 \\ 0.71 & 1-\lambda_i \end{pmatrix}$$

Alors :

$$det(R - \lambda_i\ I) = (1-\lambda_i)^2 - 0.71^2 = (1 - \lambda_i + 0.71) \times (1 - \lambda_i - 0.71) = 0\ \ (0.5pt)$$

donc : $1 - \lambda_i + 0.71 = 0 \Rightarrow \lambda_i = 1.71 = \lambda_1$

et $1 - \lambda_i - 0.71 = 0 \Rightarrow \lambda_i = 0.29 = \lambda_2$. (0.5 pt)

4. Soient $u_1 = (0.71, 0.71)^t$ et $u_2 = (0.71, -0.71)^t$ les vecteurs propres associés aux valeurs λ_1 et λ_2 respectivement.

Les coordonnées des individus : $C_i = X_{cr}\ \times\ u_i$. (1 pt)

$$F_1 = X_{cr}\ \times\ u_1 = \begin{pmatrix} -1.41 & -1 \\ 0 & -1 \\ 0 & 1 \\ 1.41 & 1 \end{pmatrix} \times \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix} = \begin{pmatrix} -1.71 \\ -0.71 \\ 0.71 \\ 1.71 \end{pmatrix}$$

et

$$F_2 = X_{cr}\ \times\ u_2 = \begin{pmatrix} -1.41 & -1 \\ 0 & -1 \\ 0 & 1 \\ 1.41 & 1 \end{pmatrix} \times \begin{pmatrix} 0.71 \\ -0.71 \end{pmatrix} = \begin{pmatrix} -0.29 \\ 0.71 \\ -0.71 \\ 0.29 \end{pmatrix}$$

donc :

Individus	Axe principal 1 Composante principale 1 Facteur principal 1	Axe principal 2 Composante principale 2 Facteur principal 2
I_1	-1.71	-0.29
I_2	-0.71	0.71
I_3	0.71	-0.71
I_4	1.71	0.29

Les coordonnées des variables : $C_j = u_j\ \times\ \sqrt{\lambda_j}$. (1 pt)

$$C_1 = u_1\ \times\ \sqrt{\lambda_1} = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix} \times \sqrt{1.71} = \begin{pmatrix} 0.93 \\ 0.93 \end{pmatrix}$$

et

$$C_2 = u_2\ \times\ \sqrt{\lambda_2} = \begin{pmatrix} 0.71 \\ -0.71 \end{pmatrix} \times \sqrt{0.29} = \begin{pmatrix} 0.39 \\ -0.39 \end{pmatrix}$$

donc :

Variables	Composante principale 1	Composante principale 2
X_1	0.93	0.39
X_2	0.93	-0.39

5. L'inertie total $I_G = \lambda_1 + \lambda_2 = 2$ (0.25 pt).
   L'inertie expliquée $I_e = \lambda_1 + \lambda_2 = 2$ (0.25 pt).
   L'inertie expliquée de l'axe principal 1 est : $I_1 = \lambda_1 = 1.71$. L'inertie de l'axe 2 est : $I_2 = \lambda_2 = 0.29$. (0.25 pt)
6. La qualité globale des représentations $Q_g = \dfrac{I_e}{I_Q} = 1$. (0.25 pt)
   La qualité de l'axe 1 est : $Q_1 = \dfrac{\lambda_1}{I_Q} = 0.85$ (0.25 pt), et la qualité de l'axe 2 est : $Q_g = \dfrac{\lambda_2}{I_Q} = 0.15$ (0.25 pt).
7. La qualité de la représentation des individus 2 et 4 par l'axe 1 :
   $cos^2\alpha_{i,1} = \dfrac{F_1^2}{X_{cr}^2}$ tel que $X_{cr}^2 = \sum_{j=1}^{2} x_{ij}^2$ (0.5 pt) donc : $cos^2\alpha_{i=2,1} = \dfrac{(-0.71)^2}{0^2+(-1)^2} = 0.5$ (0.25 pt) et $cos^2\alpha_{i=4,1} = 0.98$. (0.25 pt)
8. $cos^2(X_1, C_1) = r(X_1, C_1) = \dfrac{C_1}{\sigma_{X_1}}$ (0.5 pt), on trouve : $(0.33, 0.33)^t$.

Exercice 02 (03 points) : L'AFCM

1. Le nombre des individus $n = 4$, des variables $q = 2$ (0.25 pt),
   le nombre de modalités de chaque variable $m_1 = 2$ et $m_2 = 2$, (0.25 pt)
   et le nombre total des modalités $m = 4$ (0.25 pt). L'inertie totale $R = \dfrac{m}{q} - 1 = 1$ (0.25 pt).
2. Le tableau disjonctif complet Z :

	Licence	Master	$n_{i.}$
Lettre	55	06	61
Sc.Ex	06	10	16
Gestion	03	20	23
$n_{.j}$	64	36	$n = 100$

   (01 pt)
3. Le tableau de Burt B :

	Licence	Master	$n_{i.}$
Lettre	55	06	61
Sc.Ex	06	10	16
Gestion	03	20	23
$n_{.j}$	64	36	$n = 100$

   (01 pts)

Exercice 03 (05 points) : deux variables qualitatives

1. La population étudiée : L'ensemble de personnes. (0.25 pt)
   L'effectif total $n = 100$. (0.25 pt)
   Les caractères : X couleur des yeux (0.25pt) et Y couleur des cheveux (0.25 pt), et Les deux variables sont qualitatives (0.5 pt).

	Licence	Master	$n_{i.}$
Lettre	55	06	61
Sc.Ex	06	10	16
Gestion	03	20	23
$n_{.j}$	64	36	$n = 100$

(0.5 pts)

2. Les coefficients de la matrice X_{th} s'écrivent comme suit : $n^*_{ij} = \dfrac{n_{i.} \times n_{.j}}{n}$. (01 pt)

$$X_{th} = \begin{pmatrix} 39.04 & 21.96 \\ 10.24 & 5.76 \\ 14.72 & 8.28 \end{pmatrix}$$

On a : $n^*_{ij} = \dfrac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}}$. (01 pt)

$$X^*_{th} = \begin{pmatrix} 6.52 & 11.6 \\ 1.76 & 3.12 \\ 9.33 & 16.59 \end{pmatrix}$$

Alors le coefficient **Khi-deux** est :

$$\mathcal{X}^2 = \sum_i \sum_j \frac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}} = 48.92 \ \ (0.5\ pt)$$

Et le coefficient de Cramer C est :

$$C = \sqrt{\frac{\mathcal{X}^2}{n\,(d-1)}} = \sqrt{\frac{48.92}{100\,(2-1)}} = 0.69 \ \ (0.25\ pt)$$

On remarque que C proche de 0, donc les deux variables X_1 et X_2 sont liées c'est-à-dire dépendantes (0.25 pt).