UNIVERSITÉ LARBI BEN M'HIDI-OUM EL BOUAGHI

Département de Mathématique et Informatique

M1 VA Corrigé type Contrôle N 01 STD n

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

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

$$X = \begin{pmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 1 \end{pmatrix}$$

1. Les moyennes de X_1 et X_2 respectivement : $\overline{x_1} = \frac{\sum_j x_{i1}}{3} = \frac{12}{3} = 4$ (0.5 pt), $\overline{x_2} = \frac{\sum_i x_{i2}}{n} = \frac{9}{3} = 3$ (0.5 pt).

   On déduit donc la matrice X_c, où ses coefficients sont centrées.

$$X_c = \begin{pmatrix} -2 & 0 \\ 0 & 2 \\ 2 & -2 \end{pmatrix} (0.5\ pt)$$

2. La variance de X_1 :

$$var(X_1) = \frac{\sum_i xi1^2}{n} - \overline{x_1}^2 = \frac{56}{3} - 4^2 = 2.67\ \ (0.5pt)$$

   donc : $\sigma_{X_1} = \sqrt{var(X_1)} = 1.63$. (0.25 pt)

   La variance de X_2 :

$$var(X_2) = \frac{\sum_i xi2^2}{n} - \overline{x_2}^2 = \frac{35}{3} - 3^2 = 2.67\ \ (0.5pt)$$

   donc : $\sigma_{X_2} = \sqrt{var(X_2)} = 1.63$. (0.25 pt)

   On déduit la matrice centrée réduite X_{cr} comme suite :

$$X_{cr} = \begin{pmatrix} -1.23 & 0 \\ 0 & 1.23 \\ 1.23 & -1.23 \end{pmatrix} (0.5\ pt)$$

3.

$$X_{cr}^t \times X_{cr} = \begin{pmatrix} -1.23 & 0 & 1.23 \\ 0 & 1.23 & -1.23 \end{pmatrix} \times \begin{pmatrix} -1.23 & 0 \\ 0 & 1.23 \\ 1.23 & -1.23 \end{pmatrix} = \begin{pmatrix} 3.03 & -1.51 \\ -1.51 & 3.03 \end{pmatrix} (1\ pt)$$

   Alors la matrice R de corrélations est :

$$R = \frac{1}{n} \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} (0.25pt)$$

   D'après la matrice R, on a le coefficient de corrélation linéaire entre X_1 et X_2 $r_{X_1\,X_2} = -0.5 = \frac{Cov(X_1, X_2)}{\sigma_{X_1}\,\sigma_{X_2}}$, 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 : $x_2 = a\ x_1 + b$ tels que $a = \frac{Cov(X_1, X_2)}{var(X_1)} = \frac{-0.5 \times 1.63 \times 1.63}{2.67} = -0.5\ \ (0.25pt)$ et $b = \overline{x_2} - a\,\overline{x_1} = 5$ (0.25 pt).

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

$$R - \lambda_i\ I = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} - \lambda_i \times \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1-\lambda_i & -0.5 \\ -0.5 & 1-\lambda_i \end{pmatrix}$$

Alors :

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

donc : $1 - \lambda_i + 0.5 = 0 \Rightarrow \lambda_i = 1.5 = \lambda_1$

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

L'inertie total $I_G = \lambda_1 + \lambda_2 = 2$ (0.25 pt).

(b) Les pourcentages d'inerties, et le diagramme en bâtons. (1 pt)

	Valeur propre	pourcentage %	pourcentage cumulé
	λ_1 =1.5	75	75
	λ_2 =0.5	25	100
Total	2	100	//////

(c) 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)

$$C_1 = X_{cr}\ \times\ u_1 = \begin{pmatrix} -1.23 & 0 \\ 0 & 1.23 \\ 1.23 & -1.23 \end{pmatrix} \times \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix} = \begin{pmatrix} 0.87 \\ 0.87 \\ -1.75 \end{pmatrix}$$

et

$$C_2 = X_{cr}\ \times\ u_2 = \begin{pmatrix} -1.23 & 0 \\ 0 & 1.23 \\ 1.23 & -1.23 \end{pmatrix} \times \begin{pmatrix} -0.71 \\ -0.71 \end{pmatrix} = \begin{pmatrix} 0.87 \\ -0.87 \\ 1.75 \end{pmatrix}$$

donc :

Individus	Axe principal 1 Composante principale 1 Facteur principal 1	Axe principal 2 Composante principale 2 Facteur principal 2
I_1	0.87	0.87
I_2	0.87	-0.87
I_3	-1.75	1.75

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

$$C_1 = u_1\ \times\ \sqrt{\lambda_1} = \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix} \times \sqrt{1.5} = \begin{pmatrix} -0.87 \\ 0.87 \end{pmatrix}$$

et

$$C_2 = u_2\ \times\ \sqrt{\lambda_2} = \begin{pmatrix} -0.71 \\ -0.71 \end{pmatrix} \times \sqrt{0.5} = \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix}$$

donc :

Variables	Composante principale 1	Composante principale 2
X_1	-0.87	-0.5
X_2	0.87	-0.5

(d) La représentation des individus. (0.75 pt)

(e) La représentation des variables. (0.5 pt)

5. L'inertie de l'axe principal 1 est : $I_1 = \lambda_1 = 1.5$. L'inertie de l'axe 1 est : $I_2 = \lambda_2 = 0.5$. (0.5 pt)

6. La qualité globale des représentations $Q_g = 1$. (0.25)

Exercice 02 (08 points) : deux variables qualitatives

1. La population étudiée : L'ensemble d'étudiants. (0.5 pt)
   L'effectif total $n = 100$. (0.25 pt)
   Les caractères : X le type de BAC au lycée (0.5pt) et Y le type de diplôme à l'université (0.5 pt), et Les deux variables sont qualitatives (0.5 pt).

	Licence	Master	$n_{i.}$
Lettre	20	08	28
Sc.Ex	40	20	60
Gestion	10	02	12
$n_{.j}$	70	30	$n = 100$

   (1.25 pts)

2. Les coefficients de la matrice X_{th} s'écrivent comme suit : $n^*_{ij} = \dfrac{n_{i.} \times n_{.j}}{n}$. (1.5 pts)

$$X_{th} = \begin{pmatrix} 19.6 & 8.4 \\ 42 & 18 \\ 8.4 & 3.6 \end{pmatrix}$$

3. On a : $n^*_{ij} = \dfrac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}}$. (1.5 pts)

$$X^*_{th} = \begin{pmatrix} 0.01 & 0.02 \\ 0.1 & 0.22 \\ 0.3 & 0.71 \end{pmatrix}$$

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

$$\mathcal{X}^2 = \sum_i \sum_j \frac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}} = 0.01 + 0.02 + 0.1 + 0.22 + 0.3 + 0.71 = 1.36. \ (0.5\ pt)$$

4. Le coefficient de Cramer C est :

$$C = \sqrt{\frac{\mathcal{X}^2}{n\ (d-1)}} = \sqrt{\frac{1.36}{100\ (2-1)}} = 0.12 \ (0.5\ pt)$$

   On remarque que C proche de 0, donc les deux variables X_1 et X_2 ne sont pas liées c'est-à-dire indépendantes (0.5 pt).