Corrigé de la Série 2

Exercice 1

La matrice de données est de type $(\mathbf{4}, \mathbf{3})$. Alors :

J / *I*	***J1***	***J2***	***J3***
I1	*0*	*2*	*3*
I2	*0*	*0*	*3*
I3	*4*	*2*	*1*
I4	*4*	*0*	*1*

1) Les matrices des liaisons entre les variables sont données comme suit.
   i) La matrice des variances-covariances est définie par :

$$V = \frac{1}{4} Y^t . Y.$$

Où, Y est la matrice des données centrée dont le centre de gravite est $g = (2, 1, 2)$.

$$Y = \begin{pmatrix} -2 & 1 & 1 \\ -2 & -1 & 1 \\ 2 & 1 & -1 \\ 2 & -1 & -1 \end{pmatrix}.$$

Par conséquent,

$$V = \begin{pmatrix} 4 & 0 & -2 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}.$$

   ii) La matrice de corrélation se définit à partir de la matrice Z , matrice des données centrées réduites, par les écarts types des variables :

$$R = \frac{1}{4} Z^t . Z.$$

Où, $z_{ij} = \frac{y_{ij}}{\sigma(X^j)}$ pour tout $i = 1, 2, 3, 4$ et $j = 1, 2, 3$.

Les écarts sont :

$$\sigma(X^1) = \sqrt{\frac{16}{4}} = 2, \quad \sigma(X^2) = \sqrt{\frac{4}{4}} = 1 = \sigma(X^3).$$

Ce qui donne :

$$Z = \begin{pmatrix} -1 & 1 & 1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix}.$$

   iii) Par conséquent, la matrice de corrélation est donnée par :

$$R = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

2) Le meilleur sous espace ajustant le nuage des points individus est déterminé par les axes principaux correspondants aux vecteurs propres normés des valeurs propres de la matrice des corrélations.

Ainsi :

i) Les valeurs propres de R sont les racines de $\boldsymbol{det(R - \lambda Id) = 0}$.

$$\boldsymbol{det(R - \lambda Id) = 0}$$

Est équivalent à :

$$\boldsymbol{(1-\lambda) \times (1-\lambda)^2 - (1-\lambda) = (1-\lambda) \times [(1-\lambda)^2 - 1] = 0.}$$

Ce qui donne :

$$\boldsymbol{\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 2}$$

Ces valeurs propres nous permettent d'exprimer les taux d'inerties expliquées par les axes principaux.
Par conséquent, les taux d'inerties portés par les axes sont :

$$\boldsymbol{T_1 = \frac{1}{3} = 0.333, \quad T_2 = \frac{0}{3} = 0, \quad T_3 = \frac{2}{3} = 0.666.}$$

C'est-à-dire :

$$\boldsymbol{T_1 \sim 33.33\,\%, \quad T_2 = 0\,\%, \quad T_3 \sim 66.66\,\%.}$$

Donc, nous en déduisons que le meilleur sous espace ajustant le nuage des points individus est constitué de deux axes (un plan). Le premier axe principal associé à la plus grande valeur propre recouvre plus que $\boldsymbol{66.66}$ % de l'inertie totale et deuxième recouvre plus de $\boldsymbol{33.33}$ % de l'inertie totale. Par suite, le taux d'inertie expliqué par le plan principal n'est que $\boldsymbol{(\lambda_1 + \lambda_3)/3}$ c'est-à-dire $\boldsymbol{100}$ **%.**

ii) Les vecteurs propres associés aux valeurs propres sélectionnées sont définis comme suit :

- *Le premier vecteur propre*

Soit $u = (x, y, z)^t \in \Re^3 - \{0_{\Re^3}\}$ tel que : $R \cdot u = 2 \cdot u$, alors :

$$\begin{cases} x - z = 2 \cdot x \\ y = 2 \cdot y \\ -x + z = 2 \cdot z \end{cases} \Leftrightarrow \begin{cases} -x - z = 0 \\ y = 0 \\ -x - z = 0 \end{cases}$$. Ce qui donne : $z = -x$ et $y = 0$.

Donc,

$u = (x, 0, -x)^t = x(1, 0, -1)^t$ avec $x \in \Re^*$. En particulier, nous prenons :

$$u = (1, 0, -1)^t .$$

Par conséquent, le premier axe principal (Δu_1) est engendré par le vecteur unitaire :

$$u_1 = \frac{u}{\|u\|} = \left(\frac{\sqrt{2}}{2}, 0, -\frac{\sqrt{2}}{2}\right)^t .$$

Où,

$$\|u\| = \sqrt{2}\,.$$

- *Le deuxième vecteur propre*

Soit $v = (x, y, z)^t \in \Re^3 - \{0_{\Re^3}\}$ tel que : $R \cdot v = 1 \cdot v$, alors :

$\begin{cases} x - z = x \\ y = y \\ -x + z = z \end{cases} \Leftrightarrow \begin{cases} z = 0 \\ y \\ x = 0 \end{cases}$. Ce qui donne : $z = x = 0$ et y est quelconque.

Donc,

$v = (0, y, 0)^t = y(0, 1,0)^t$ avec $y \in \Re^*$. En particulier, nous prenons :

$$v = (0, 1, 0)^t\,.$$

Par conséquent, le deuxième axe principal (Δu_2) est engendré par le vecteur unitaire $u_2 = v$.

Ainsi, le meilleur plan ajustant le nuage est le plan $\langle \vec{u}_1, \vec{u}_2 \rangle$.

3) Les composantes principales obtenues C^1 et C^2 sont données comme suit :

$$\boldsymbol{C^1 = Z \cdot u_1 = (1, -1, 1, -1)^t}.$$

Et

$$\boldsymbol{C^2 = Z \cdot u_2 = \left(-\sqrt{2}, -\sqrt{2}, \sqrt{2}, \sqrt{2}\right)^t}.$$

Avec :

$$\boldsymbol{Var\,(C^1) = 2} \qquad \textbf{et} \qquad \boldsymbol{Var\,(C^2) = 1}.$$

4) Les corrélations $r(X^j, C^k)$ entre les variables initiales et les composantes principales ne sont que les composantes :

$$\boldsymbol{\phi_j^k = r(X^j, C^k)} \text{ pour tout } j = 1, 2, 3 \text{ et } k = 1, 2.$$

Ces dernières s'expriment en fonction la relation suivante :

$$\boldsymbol{\phi^k = \sqrt{\lambda_k} \cdot u_k} \quad \text{pour tout } k = 1, 2.$$

Ce qui donne :

$$\boldsymbol{\phi^1 = \sqrt{2} \cdot u_1 = \sqrt{2} \cdot v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}.$$

Et

$$\boldsymbol{\phi^2 = \sqrt{1} \cdot u_2 = \sqrt{1} \cdot u = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}.$$

Par conséquent, les corrélations sont données comme suit :

$$\phi_1^1 = r(X^1, C^1) = 1 \qquad \text{et} \qquad \phi_1^2 = r(X^1, C^2) = 0$$
$$\phi_2^1 = r(X^2, C^1) = 0 \qquad \text{et} \qquad \phi_2^2 = r(X^2, C^2) = 1$$
$$\phi_3^1 = r(X^3, C^1) = -1 \qquad \text{et} \qquad \phi_3^2 = r(X^3, C^2) = 0.$$

Axes / Variables	**1^er axe**	**2^ème axe**
X^1	**1**	**0**
X^2	**0**	**1**
X^3	**-1**	**0**

Coordonnées des variables dans le plan principal.

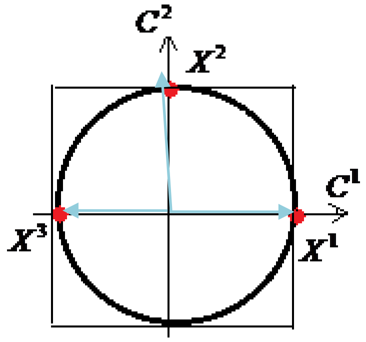
5) Ce qui nous conduit à associer à chaque point variable un point dans le plan factoriel dont les coordonnées sont les corrélations entre cette variable et les axes principaux de ce plan de l'ACP.

   Dans notre cas, nous aurons :

$$X^1 \longrightarrow \begin{pmatrix}1\\0\end{pmatrix}, \quad X^2 \rightarrow \begin{pmatrix}0\\1\end{pmatrix} \quad \text{et} \quad X^3 \rightarrow \begin{pmatrix}-1\\0\end{pmatrix}.$$

   Ces résultats peuvent être présentés dans un tableau comme suit :

6) La représentation graphique des variables dans le nouveau plan est donnée par le cercle de corrélation.



Représentation graphique des variables dans le plan factoriel.

7) De la représentation, nous en déduisons que :

   i) Le premier axe est fortement corrélé :

   - Positivement avec la variable X^1
   - Négativement avec la variable X^3.
   - Les variables X^1 et X^3 sont corrélées fortement et négativement.

   ii) Le deuxième axe principal est fortement et positivement corrélé avec la variable X^2.

Exercice 2

Soit X de type **(96, 8)** dont la matrice des corrélations est donnée par :

$$R = \begin{array}{c} \textbf{Variables} \\ X^1 \\ X^2 \\ X^3 \\ X^4 \\ X^5 \\ X^6 \\ X^7 \\ X^8 \end{array} \begin{pmatrix} X^1 & X^2 & X^3 & X^4 & X^5 & X^6 & X^7 & X^8 \\ 1 & 0.56 & 0.28 & 0.20 & 0.43 & 0.29 & 0.31 & 0.43 \\ 0.56 & 1 & 0.34 & 0.32 & 0.48 & 0.32 & 0.47 & 0.49 \\ 0.28 & 0.34 & 1 & 0.64 & 0.72 & 0.52 & 0.70 & 0.45 \\ 0.20 & 0.32 & 0.64 & 1 & 0.67 & 0.46 & 0.58 & 0.44 \\ 0.43 & 0.48 & 0.72 & 0.67 & 1 & 0.56 & 0.68 & 0.53 \\ 0.29 & 0.32 & 0.52 & 0.46 & 0.56 & 1 & 0.64 & 0.42 \\ 0.31 & 0.47 & 0.70 & 0.58 & 0.68 & 0.64 & 1 & 0.51 \\ 0.43 & 0.49 & 0.45 & 0.44 & 0.53 & 0.42 & 0.51 & 1 \end{pmatrix}.$$

1) L'analyse de cette matrice donne :
   - Toutes les variables sont corrélées positivement c'est-à-dire :

   $$\rho(X^i, X^j) = r_{ij} > 0 \text{ pour tout } i,\ j \text{ dans } \{1, 2, \dots, 8\}.$$

   - Une forte corrélation entre les couples des variables : (X^3, X^5), (X^3, X^7), (X^5, X^7).
   Si, nous fixons le seuil de corrélation a 68 %.
   - Une faible corrélation entre les couples des variables : (X^1, X^3), (X^1, X^4), (X^1, X^6).

2) Les taux d'inertie expliqués par chaque axe sont donnés en fonction des valeurs propres déterminées de la matrice de corrélations R moyennant la formule suivante :

$$T_{\lambda_i} = \frac{\lambda_i}{n} = \frac{\lambda_i}{8} \quad \text{pour tout } \ i = 1, 2, \dots, 8.$$

Ainsi, les différents taux sont donnés dans l'ordre comme suit :

$$T_{\lambda_7} \approx 55.30\ \%,\ T_{\lambda_1} \approx 14.21\ \%,\ T_{\lambda_5} \approx 7.35\ \%,\ \ T_{\lambda_6} \approx 6.63\ \%,\ \ T_{\lambda_4} \approx 5.72\ \%,$$

$$T_{\lambda_2} \approx 4.70\ \%,\ \ T_{\lambda_8} \approx 3.23\ \% \ \text{ et } \ T_{\lambda_3} \approx 2.88\ \%.$$

3) Le meilleur plan ajustant le nuage des individus est engendré par les deux vecteurs propres normés associés aux deux premières plus grandes valeurs propres de la matrice de corrélations R :

$$\lambda_7 = 4.4242\,, \qquad \lambda_1 = 1.1366.$$

Avec un taux égal a :

$$T_{\lambda_7+\lambda_1} = \frac{\lambda_7 + \lambda_1}{8} \approx 69.51\ \%.$$

4) Les coordonnées des individus dans le plan principal ne sont que les projections des individus sur les axes principaux de ce plan. Autrement dit, elles ne sont que les coordonnées des composantes principales. Elles sont données par :

$$C^1 = Z \cdot \vec{u} \quad \text{et} \quad C^2 = Z \cdot \vec{v}.$$

Où,

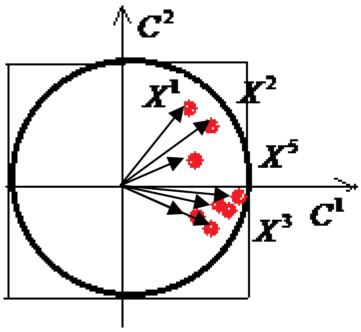
C^1 et C^2 sont les composantes principales associées aux vecteurs propres normés associés des valeurs propres λ_7 et λ_1 respectivement.

5) La réalisation de l'ACP donne :

Axes / Variables	1^{er} Axe	$2^{ème}$ Axe
X^1	*0.55*	*0.68*
X^2	*0.65*	*0.55*
X^3	*0.81*	*- 0.32*
X^4	*0.75*	*- 0.35*
X^5	*0.87*	*- 0.1*
X^6	*0.72*	*- 0.2*
X^7	*0.84*	*- 0.18*
X^8	*0.71*	*0.25*

Les valeurs du tableau représentent……………… Expliquer.

6) La visualisation graphique



Représentation des variables dans le plan factoriel.

7) Ce cercle nous permet de créer des groupes homogènes à partir de l'analyse des liaisons entre les variables, d'étudier les corrélations entre les variables et les axes principaux. Noter bien que les petites flèches correspondent aux variables faiblement corrélées aves les deux axes du plan factoriel. Par conséquent, notre intérêt se focalise sur les variables qui sont proches du périmètre du cercle.