

T.D. n° IV - Quelques applications

Exercice 1

Soit le tableau de données

$$\mathbf{X} = \sqrt{10} \begin{bmatrix} 2 & 2 & 3 \\ 3 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 1 & 4 \\ 2 & 1 & 3 \end{bmatrix}$$

correspondant à des observations effectuées sur 5 individus de poids statistiques égaux pour les trois variables T^1 , T^2 et T^3 . On veut effectuer une ACP centrée-réduite sur ce tableau.

1. Calculer l'individu moyen, le vecteur $(\sigma_1, \sigma_2, \sigma_3)'$ des écart-types des variables et la matrice \mathbf{X}_{cr} des données centrées-réduites.
2. Calculer la matrice des corrélations \mathbf{R} .
3. Effectuer la décomposition aux valeurs propres de \mathbf{R} . En déduire les composantes principales.
4. Représenter les plans factoriels (1, 2) pour les variables et les individus. Donner une interprétation de l'ACP.

Exercice 2

Soit la matrice $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3]$ dont les variables ont pour corrélation

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & -\rho \\ \rho & 1 & \rho \\ -\rho & \rho & 1 \end{bmatrix}$$

avec $-1 \leq \rho \leq 1$. On va effectuer l'ACP centrée réduite de \mathbf{X} .

1. Vérifier que \mathbf{R} admet pour vecteur propre $\boldsymbol{\xi}_1 = \frac{1}{\sqrt{3}}(1, -1, 1)'$.
2. Déterminer les autres éléments propres de cette matrice.
3. Quelles sont les valeurs possibles de ρ ? On se placera ensuite dans le cas où $-1 \leq \rho < 0$.
4. Calculer les pourcentages de variance expliquée et tracer l'éboulis des valeurs propres.
5. Comment s'interprète en fonction des variables, l'unique composante principale à retenir dans cet exemple?

Corrections

Exercice 1

1. L'individu moyen est obtenu en faisant la moyenne des colonnes du tableau \mathbf{X} , soit $\bar{\mathbf{x}} = \sqrt{10}(2, 1, 3)'$. On obtient, en soustrayant ce vecteur à chaque ligne du tableau, le tableau centré

$$\mathbf{X}_c = \sqrt{10} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Le vecteur $(\sigma_1, \sigma_2, \sigma_3)'$ est donné par la racine carrée de trace de la matrice de variance covariance

$$\mathbf{V} = \mathbf{X}'_c \mathbf{D}_p \mathbf{X}_c = \frac{10}{5} \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

soit $(\sigma_1, \sigma_2, \sigma_3)' = (2, 2, 2)'$, où la matrice \mathbf{D}_p correspond à la métrique diagonale des poids, soit $\frac{1}{5}$ sur la diagonale. On en déduit la matrice des données centrées réduites de terme général $x_{ij}^{cr} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$

$$\mathbf{X}_{cr} = \frac{\sqrt{10}}{2} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

2. La matrice de corrélation est donnée par

$$\mathbf{R} = \mathbf{X}'_{cr} \mathbf{D}_p \mathbf{X}_{cr} = \frac{10}{20} \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix}.$$

Cette matrice correspond à une matrice de variance particulière calculée sur un tableau centré réduit.

3. Faire l'ACP du tableau centré réduit correspond à diagonaliser la matrice \mathbf{R} . On cherche donc à résoudre le problème aux valeurs propres

$$\mathbf{R}\mathbf{A} = \mathbf{A}\mathbf{\Lambda}$$

où \mathbf{A} est la matrice des vecteurs propres et $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ est la matrice diagonale des valeurs propres. On commence par résoudre $\det(\mathbf{R} - \lambda\mathbf{I}) = 0$ et on obtient l'équation

$$(1 - \lambda) \left(1 - \lambda + \frac{1}{\sqrt{2}}\right) \left(1 - \lambda - \frac{1}{\sqrt{2}}\right) = 0.$$

On obtient donc 3 valeurs propres $\lambda_1 = 1 + \frac{\sqrt{2}}{2}$, $\lambda_2 = 1$, $\lambda_3 = 1 - \frac{\sqrt{2}}{2}$. Il faut maintenant trouver les vecteurs propres correspondants à chaque valeur propre en résolvant

$$\mathbf{R}\mathbf{a} = \lambda\mathbf{a}$$

et on trouve $\mathbf{a}_1 = \frac{1}{2}(\sqrt{2}, 1, -1)'$, $\mathbf{a}_2 = \frac{1}{\sqrt{2}}(0, 1, 1)$ et $\mathbf{a}_3 = \frac{1}{2}(\sqrt{2}, -1, 1)'$. Attention, ces vecteurs propres ont été normalisés pour être de norme unité pour la métrique \mathbf{M} , c'est à dire $\|\mathbf{a}_j\|_{\mathbf{M}} = \mathbf{a}'_j \mathbf{M} \mathbf{a}_j = 1$. Ici, la métrique $\mathbf{M} = \mathbf{I}$, mais attention de travailler avec le tableau centré réduit. On vérifiera également qu'ils sont orthogonaux deux à deux $\mathbf{a}'_k \mathbf{M} \mathbf{a}_j = 0$, $k \neq j$. On en déduit les composantes principale

$$\begin{aligned} \mathbf{c}^1 &= \mathbf{X}_{cr} \mathbf{a}_1 = \frac{\sqrt{10}}{4} (1, 1 + \sqrt{2}, -1 - \sqrt{2}, -1, 0)' \\ \mathbf{c}^2 &= \mathbf{X}_{cr} \mathbf{a}_2 = \frac{\sqrt{20}}{4} (1, -1, -1, 1, 0)' \\ \mathbf{c}^3 &= \mathbf{X}_{cr} \mathbf{a}_3 = \frac{\sqrt{10}}{4} (-1, \sqrt{2} - 1, 1 - \sqrt{2}, 1, 0)' \end{aligned}.$$

On vérifie que ces composantes principales sont orthogonales deux à deux pour la métrique \mathbf{D}_p , c'est à dire que $\mathbf{c}^{k'} \mathbf{D}_p \mathbf{c}^j = 0$, $k \neq j$ et que leur \mathbf{D}_p -norme est telle que $\|\mathbf{c}^j\|_{\mathbf{D}_p}^2 = \lambda_j$.

4. Les coordonnées des points sont données par les composantes principales. Les coordonnées des variables sur le $j^{\text{ème}}$ axe sont données par $\mathbf{v}_j = \sqrt{\lambda_j} \mathbf{a}_j$ dans ce cas. Les variables \mathbf{v}_j sont de \mathbf{M} -norme telle que $\|\mathbf{v}_j\|_{\mathbf{M}}^2 = \lambda_j$ et orthogonales deux à deux également. Dans ce cas particulier (travail avec la matrice de corrélation) les coordonnées des variables sont directement les corrélations entre les axes principaux et les différentes variables. On peut également tracer le cercle des corrélation puisque les variables sont de norme unité (elles sont donc situées dans un cercle de rayon 1 lorsqu'elles sont projetées). Plus une variable sera proche de ce cercle, mieux elle sera représentée dans le plan de où l'on observe sa projection. Dans ce cas, la variable v_1 est inféodée à l'axe 1, la variable v_2 est également bien représentée dans le plan.

