

Analyse des données

Série 1

Analyse en composantes principales

Exercice 1 Considérons la matrice de données X suivante :

$$X = \begin{pmatrix} 0 & 1 & 1 & 2 & 3 & 3 & 4 \\ 1 & 2 & 4 & 3 & 2 & 4 & 5 \end{pmatrix}.$$

- 1) Déterminer la matrice centrée Y .
- 2) Déterminer la matrice V des variances-covariances. En déduire l'inertie de ce nuage autour du centre de gravité.
- 3) Déterminer les valeurs propres de V ainsi que les axes principaux associés à ces valeurs propres.
- 4) Calculer les valeurs propres de la matrice $Y^t \cdot Y$.
- 5) En déduire les valeurs propres de la matrice $Y \cdot Y^t$.

Exercice 2 Soit la matrice de données suivante :

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

- 1) Donner le nuage des points $N(I)$.
- 2) Calculer le centre de gravité de ce nuage. Que pouvons-nous déduire ?
- 3) Déterminer la matrice des variances-covariances V .
- 4) Posons $S = X^t \cdot X$. Montrer que S possède une valeur propre nulle (sans faire de calculs).
- 5) Calculer les valeurs propres de S et en déduire celles de V .
- 6) Déterminer le meilleur plan qui ajuste $N(I)$. Indication : Il s'agit de réaliser une ACP non normée.

Exercice 3 On considère la matrice de données X de type (11, 3) :

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 6 & 1 & 4 & 5 & 3 & 2 & 9 & 7 & 8 & 10 & 11 \\ 2 & 6 & 5 & 3 & 4 & 1 & 8 & 9 & 7 & 10 & 11 \end{pmatrix}$$

- 1) Déterminer la matrice des variances-covariances.
- 2) Déterminer la matrice centrée réduite.
- 3) Déterminer la matrice des corrélations (en gardant les coefficients en fractions). Interpréter et analyser cette matrice.

Analyse des données

Série 1

- 4) Déterminer la matrice des corrélations (en prenant deux chiffres après la virgule). Donner les valeurs propres. Interpréter ces valeurs en termes d'inertie. En déduire le taux d'inertie projeté sur chaque axe.
- 5) Déterminer le meilleur plan ajustant le nuage des individus. Vérifier que ses axes sont orthogonaux.
- 6) Poumons-nous se contenter d'une seule composante principale ?

Exercice 4 Considérons la matrice des données X suivante :

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

- 1) Calculer l'individu moyen ainsi que les écarts types.
- 2) Déterminer la matrice des données centrées réduites.
- 3) Déterminer la matrice des corrélations.
- 4) Déterminer le meilleur sous espace ajustant le nuage des points tout en précisant le taux d'inertie exprimé par chaque axe. Justifier votre réponse.
- 5) Représenter les individus dans le nouveau plan.
- 6) Déterminer les individus les mieux représentés dans le nouveau plan. Calculer leurs contributions à l'inertie de chaque axe.
- 7) Donner les coordonnées des variables sur le sous espace factoriel. Qu'expriment ces coordonnées.
- 8) Représenter graphiquement le nuage des variables dans le nouveau plan. Que pouvez-vous déduire de cette représentation ? Expliquer.