

Thème 1 : Analyse en composantes principales (ACP)

Méthode & Calculs

Enoncé du problème

A l'aide d'une méthode appropriée, nous avons collecté des informations relatives à 3 variables quantitatives pour 5 individus. Les données collectées sont fournies dans le tableau ci-dessous :

Individus\Variables	X_1	X_2	X_3
1	50	4	15
2	40	2	25
3	30	1	35
4	20	5	45
5	10	3	55

Nous appelons $X(n,p)$ la matrice des données brutes associée au tableau précédent. Nous souhaitons réaliser une analyse en composantes principales sur la matrice des données brutes $X(5,3)$.

Question 1 :

- a- Justifier le choix de réaliser une ACP plutôt qu'un autre type de traitement.
- b- Présenter les différentes étapes de réalisation d'une ACP.

Question2 :

- a- Transformer la matrice des données brutes X en matrice des données centrées-réduites Z .
- b- Quel est l'objectif de cette opération de centrage et réduction (ou de standardisation) ?
- c- En quoi consiste-t-elle au plan géométrique ?

Question 3 :

- a- Calculer la matrice R des coefficients de corrélation linéaire entre les variables.
- b- Que contient-elle ? Quelles propriétés en découlent ?
- c- Quel est l'objectif de cette opération ? L'ACP est-elle réalisable ?

Rappel : $R = \frac{1}{n} Z'Z$

Question 4 : On cherche à présent à diagonaliser la matrice R .

- a- Calculer les valeurs propres λ_j associées à la matrice R . Que représentent-elles dans le contexte de l'ACP ? Quelles propriétés doivent-elles vérifier ?

b- Déterminer les vecteurs propres normés associés aux différentes valeurs propres. Que représentent-ils ?

Question 5 :

a- Calculer la matrice de dispersion des individus V .

b- A quoi correspond-elle ? Quelles sont ses propriétés ?

c- Sans calcul, déterminer les valeurs propres de la matrice V .

Rappel : $V = \frac{1}{n}ZZ'$

Question 6 : On travaille désormais à partir des résultats obtenus à la question 4, c'est-à-dire sur ceux portant sur la diagonalisation de la matrice des coefficients de corrélation linéaire entre les variables R .

a- Déterminer les coordonnées des individus dans le nouvel espace constitué par les deux premiers axes factoriels (F1 et F2).

b- Représenter graphiquement le résultat.

c- Calculer les contributions absolues et contributions relatives des individus (CTA et CTR). Que représentent-elles ? En quoi sont-elles utiles ?

d- Qu'en concluez-vous ?

Question 7 :

a- Déterminer les coordonnées des variables dans le nouveau repère constitué par les deux premiers axes factoriels (F1 et F2).

b- Représenter graphiquement le résultat.

c- Calculer les contributions absolues et contributions relatives des variables (CTA et CTR). Que représentent-elles ? En quoi sont-elles utiles ?

d- Pour les variables, le calcul de ces aides à l'interprétation est-il indispensable ?

e- Qu'en concluez-vous ?

Question 8 :

a- Retrouver les différents éléments calculés lors des questions précédentes sur les sorties du logiciel SAS fournies en annexe.

b- Que remarquez-vous ?