



Année universitaire 2022-2023

Feuille de TD N° 2

Analyses de données : Analyse en Composantes Principales

Exercice : Supposons qu'on a observé le jeu de données suivant :

Sujet/Descripteur	D_1	D_2	D_3	D_4
S_1	-11	-60	110	40
S_2	-12	-62	93	25
S_3	-15	-80	113	39
S_4	-14	-75	94	25
S_5	-14,5	-82	100	30
S_6	-13	-72	102	32
Moyenne :
Écart-type :

On note les variables $X_1 = D_1$, $X_2 = D_2$, $X_3 = D_3$ et $X_4 = D_4$.

1. Compléter le tableau de données et donner le centre de gravité et l'inertie totale.
2. Calculer la distance entre les sujets (D_1 et D_2), (D_1 et D_3) et (D_1 et D_4). Interpréter le résultat.
3. Donner la formule de centrage et réduction d'une variable, et donner la matrice centrée réduite Z .
4. Calculer la matrice de corrélation.
5. interpréter les corrélation entre la variable X_3 et les autres variables.
6. On donne les valeurs propres et les vecteurs propres de R :

$$\lambda_1 = 2.001, \quad \lambda_2 = 1.967, \quad \lambda_3 = 0.032, \quad \lambda_4 = 0.0003$$

et

$$v_1 = \begin{pmatrix} 0,441933 \\ 0,467633 \\ -0,57415 \\ -0,50633 \end{pmatrix}, v_2 = \begin{pmatrix} -0,55021 \\ -0,52745 \\ -0,41487 \\ -0,49694 \end{pmatrix},$$

$$v_3 = \begin{pmatrix} 0,655642 \\ -0,68925 \\ -0,23389 \\ 0,200895 \end{pmatrix}, v_4 = \begin{pmatrix} -0,2685 \\ 0,167476 \\ -0,66598 \\ 0,675519 \end{pmatrix}$$

7. Compléter le tableau donnant les composantes principales :

Sujet/Descripteur	Y_1	Y_2	Y_3	Y_4
S_1				
S_2				
S_3				
S_4				
S_5				
S_6				

8. Calculer l'inertie expliquée par chaque composante principale et les inerties cumulées.
9. Combien de composantes principales on doit choisir pour expliquer au moins 80% d'informations.
10. Calculer les corrélations entre les variables initiales et les composantes principales choisies. Et dessiner le cercle de corrélations.