


<b>Module : Analyse et Fouille de Données</b>	
<b>Responsables du Cours :</b> Bouaziz Souhir, Abbes Amal	<b>Auditoire:</b> D-LSI-ADBD <b>A-U:</b> 2023-2024
<b>Exercice Corrigé : ACP</b>	

## Enoncé :

Soit le tableau de données suivant :

	Informatique	Gestion
Noam	4	5
Jean	6	7
Li	8	0

Où les lignes représentent les individus (noms de quelques étudiants) et les colonnes les variables (notes en informatique et gestion). Ce tableau de données peut être représenté par la matrice  $X$  de données brutes :

$$X = \begin{pmatrix} 4 & 5 \\ 6 & 7 \\ 8 & 0 \end{pmatrix}$$

1. Calculer la matrice  $Z$  de données centrées et réduites.
2. Calculer la matrice  $R(X)$  des corrélations de  $X$ .
3. Vérifier que  $\lambda_1 = 1.96$  et  $\lambda_2 = 0.31$  sont deux valeurs propres associées à  $R(X)$ .  
Puis vérifier que les vecteurs unitaires :

$$V_1 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \text{ et } V_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

sont les vecteurs propres de  $R(X)$  associés à  $\lambda_1$  et  $\lambda_2$  respectivement.

4. À partir des résultats précédents, déterminer les deux composantes principales de l'ACP normé du nuage des individus associé au tableau  $X$ . Pour chacun de ces axes, préciser le pourcentage de la quantité d'information projetée sur l'axe considéré, et le pourcentage cumulé.
5. Déterminer la matrice des scores des individus.
6. Calculer la saturation des variables et interpréter le cercle de corrélation.

## Correction :

1.

$$\overline{X_{(1)}} = \frac{18}{3} = 6 \quad \overline{X_{(2)}} = \frac{12}{3} = 4$$

$$\sigma(x_1) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{i1} - \overline{X_{(1)}})^2} = 1.633 \quad \sigma(x_2) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{i2} - \overline{X_{(2)}})^2} = 2.944$$

$$Z = \begin{pmatrix} -1.225 & 0.34 \\ 0 & 1.02 \\ 1.225 & -1.36 \end{pmatrix}$$

2. La matrice de corrélations:

$$R(X) = \begin{pmatrix} 1 & -0.69 \\ -0.69 & 1 \end{pmatrix}$$

3. On a la matrice de corrélations  $R$  de taille  $p \times p$ , carrée et symétrique, alors elle admet  $p$  valeurs propres et  $p$  vecteurs propres associés, tels que :

$$R V_j = \lambda_j V_j$$

Dans notre exemple à deux variables,  $R$  admet 2 valeurs propres et 2 vecteurs propres tels que soit vérifié les égalités suivantes:

$$\begin{pmatrix} 1 & -0.69 \\ -0.69 & 1 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} = \lambda_1 \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix}$$

$$\begin{pmatrix} 1 & -0.69 \\ -0.69 & 1 \end{pmatrix} \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix} = \lambda_2 \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$

### I. Détermination des valeurs propres:

Résolution de l'équation:  $\det(R - \lambda I) = 0$

$$R - \lambda I = \begin{pmatrix} 1 & -0.69 \\ -0.69 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$R - \lambda I = \begin{pmatrix} 1-\lambda & -0.69 \\ -0.69 & 1-\lambda \end{pmatrix}$$

$$\det(R - \lambda I) = (1-\lambda)(1-\lambda) - 0.69^2 = 0$$

$$\lambda^2 - 2\lambda + 0.524 = 0$$

$$\Delta = b^2 - 4ac = 1.904$$

$$\lambda_1 = \frac{-b + \sqrt{\Delta}}{2a} = 1.69$$

$$\lambda_2 = \frac{-b - \sqrt{\Delta}}{2a} = 0.31$$

## II. Détermination des vecteurs propres:

### 1) Détermination de $V_1$

- Résolution de système d'équations  $RV_1 - \lambda_1 V_1 = 0$

$$\begin{cases} 1 \times v_{11} - 0.69 \times v_{12} - 1.69 \times v_{11} = 0 \\ -0.69 \times v_{11} + 1 \times v_{12} - 1.69 \times v_{12} = 0 \end{cases} \quad (\text{I})$$

- Une solution possible : le vecteur unitaire :

$$v_{11}^2 + v_{12}^2 = 1$$

- Prenons l'équation:  $v_{11} - 0.69 \times v_{12} - 1.69 \times v_{11} = 0$  du système (I)

$$-0.69 \times v_{11} - 0.69 \times v_{12} = 0 \rightarrow v_{11} = -v_{12}$$

$$\text{Remplacer dans : } v_{11}^2 + v_{12}^2 = 1$$

$$(-v_{12})^2 + v_{12}^2 = 1 \rightarrow \text{Solution Possible : } v_{1,2} = \frac{1}{\sqrt{2}} \quad v_{1,1} = -\frac{1}{\sqrt{2}}$$

### 2) Détermination de $V_2$

- Résolution de système d'équations  $RV_2 - \lambda_2 V_2 = 0$

$$\begin{cases} 1 \times v_{21} - 0.69 \times v_{22} - 0.31 \times v_{21} = 0 \\ -0.69 \times v_{21} + 1 \times v_{22} - 0.31 \times v_{22} = 0 \end{cases} \quad (\text{II})$$

- Une solution possible : le vecteur unitaire :

$$v_{21}^2 + v_{22}^2 = 1$$

- Prenons l'équation:  $v_{21} - 0.69 \times v_{22} - 0.31 \times v_{21} = 0$  du système (II)

$$0.69 \times v_{21} - 0.69 \times v_{22} = 0 \rightarrow v_{21} = v_{22}$$

$$\text{Remplacer dans : } v_{21}^2 + v_{22}^2 = 1$$

$$v_{22}^2 + v_{22}^2 = 1 \rightarrow \text{Solution Possible : } v_{2,2} = \frac{1}{\sqrt{2}} \quad v_{2,1} = \frac{1}{\sqrt{2}}$$

4. En appliquant le théorème spectral pour les matrices symétriques, nous diagonalisons la matrice des corrélations :  $R(X) = QDQ'$  tel que:

$$Q = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} : \text{Matrice des deux composantes principales de l'ACP}$$

$$\text{et } D = C(Y) = \begin{bmatrix} 1.69 & 0 \\ 0 & 0.31 \end{bmatrix}$$

$$\lambda_1 + \lambda_2 = \text{Var}(x_1) + \text{Var}(x_2)$$

$$1.69 + 0.31 = 1 + 1 = 2 \text{ (nombre de variables)}$$

	Valeur propre	Pourcentage (%)	Pourcentage cumulé (%)
$CP_1$	1.69	84.5	84.5
$CP_2$	0.31	15.5	100

5.

$$Y = Z Q = \begin{bmatrix} -1.225 & 0.34 \\ 0 & 1.02 \\ 1.225 & -1.36 \end{bmatrix} \times \begin{bmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{bmatrix} = \begin{bmatrix} 1.106 & -0.647 \\ 0.721 & 0.721 \\ -1.828 & -0.095 \end{bmatrix}$$

6. Les données sont centrées-réduites donc :  $X_{(j)} \equiv Z_{(j)}$

	$Y_{(1)}$	$Y_{(2)}$
$X_{(1)}$	-0.919	0.394
$X_{(2)}$	0.919	0.394

