

Solution de l'exercice 3 de la Série N°2 : ACP

Exercice 1 On s'intéresse à l'ACP sur le nuage de points défini par la matrice

$$\mathbf{X}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Donc il s'agit de 4 lignes-individus et 4 colonnes-variables.

1. Donner les moyennes et les variances des quatre variables puis déterminer la matrice de variances-covariances \mathbf{V} associée à \mathbf{X}^* .
2. Donner les valeurs propres et les vecteurs propres de \mathbf{V} .
3. Donner les coordonnées des lignes sur le deuxième axe principal de l'ACP de \mathbf{X}^* .
4. Donner les coordonnées des colonnes sur le deuxième axe principal de l'ACP de \mathbf{X}^* .

Solution

- 1) Les moyennes des variables sont respectivement: $1, 3/4, 1/2, 1/4$.
- 2) La matrice centrée des données est:

$$\mathbf{X} = \begin{pmatrix} 1-1 & 0-\frac{3}{4} & 0-\frac{1}{2} & 0-\frac{1}{4} \\ 1-1 & 1-\frac{3}{4} & 0-\frac{1}{2} & 0-\frac{1}{4} \\ 1-1 & 1-\frac{3}{4} & 1-\frac{1}{2} & 0-\frac{1}{4} \\ 1-1 & 1-\frac{3}{4} & 1-\frac{1}{2} & 1-\frac{1}{4} \end{pmatrix} = \begin{pmatrix} 0 & -0.75 & -0.5 & -0.25 \\ 0 & 0.25 & -0.5 & -0.25 \\ 0 & 0.25 & 0.5 & -0.25 \\ 0 & 0.25 & 0.5 & 0.75 \end{pmatrix}.$$

La matrice de variance covariance est

$$\begin{aligned} V &= \frac{1}{4} \begin{pmatrix} 0 & -0.75 & -0.5 & -0.25 \\ 0 & 0.25 & -0.5 & -0.25 \\ 0 & 0.25 & 0.5 & -0.25 \\ 0 & 0.25 & 0.5 & 0.75 \end{pmatrix}^T \begin{pmatrix} 0 & -0.75 & -0.5 & -0.25 \\ 0 & 0.25 & -0.5 & -0.25 \\ 0 & 0.25 & 0.5 & -0.25 \\ 0 & 0.25 & 0.5 & 0.75 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.1875 & 0.125 & 0.0625 \\ 0 & 0.125 & 0.25 & 0.125 \\ 0 & 0.0625 & 0.125 & 0.1875 \end{pmatrix} \end{aligned}$$

Les valeurs et les vecteurs propres de \mathbf{V} sont

$$u_1 = \begin{pmatrix} 0 \\ 0.5 \\ 0.7 \\ 0.5 \end{pmatrix} \leftrightarrow 0.42, \quad u_2 = \begin{pmatrix} 0 \\ 0.7 \\ 0 \\ -0.7 \end{pmatrix} \leftrightarrow 0.125,$$

$$u_3 = \begin{pmatrix} 0 \\ 0.5 \\ -0.7 \\ 0.5 \end{pmatrix} \leftrightarrow 0.07, \quad u_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \leftrightarrow 0.$$

Nous avons $\|u_1\| = \|u_2\| = \|u_3\| = \|u_4\| = 1$. Donc $\{u_1, u_2, u_3, u_4\}$ est une base orthonormée.

3) Les coordonnées des lignes sur le deuxième axe principal est la deuxième composante principale

$$c_2 = Xu_2 = \begin{pmatrix} 0 & -0.75 & -0.5 & -0.25 \\ 0 & 0.25 & -0.5 & -0.25 \\ 0 & 0.25 & 0.5 & -0.25 \\ 0 & 0.25 & 0.5 & 0.75 \end{pmatrix} \begin{pmatrix} 0 \\ 0.7 \\ 0 \\ -0.7 \end{pmatrix} = \begin{pmatrix} -0.35 \\ 0.35 \\ 0.35 \\ -0.35 \end{pmatrix}.$$

4) Les coordonnées des colonnes sur le deuxième axe principal sont les corrélations $\mathbf{cor}(X_j, c_2)$, $j = 1, 2, 3, 4$. Comme X_1 est nulle donc $\mathbf{cor}(X_1, c_2) = 0$. Nous avons

$$\begin{aligned} \mathbf{cor}(X_2, c_2) &= \frac{1}{4} \frac{X_2^t c_2}{\sqrt{\text{var}(X_2)} \sqrt{\lambda_2}} \\ &= \frac{1}{4} \frac{\begin{pmatrix} -0.75 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}^T \begin{pmatrix} -0.35 \\ 0.35 \\ 0.35 \\ -0.35 \end{pmatrix}}{\sqrt{0.1875} \times \sqrt{0.125}} = 0.57, \end{aligned}$$

$$\begin{aligned} \mathbf{cor}(X_3, c_2) &= \frac{1}{4} \frac{X_3^t c_2}{\sqrt{\text{var}(X_3)} \sqrt{\lambda_2}} \\ &= \frac{1}{4} \frac{\begin{pmatrix} -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \end{pmatrix}^T \begin{pmatrix} -0.35 \\ 0.35 \\ 0.35 \\ -0.35 \end{pmatrix}}{\sqrt{0.25} \times \sqrt{0.125}} = 0, \end{aligned}$$

$$\begin{aligned} \mathbf{cor}(X_4, c_2) &= \frac{1}{4} \frac{X_4^t c_2}{\sqrt{\text{var}(X_4)} \sqrt{\lambda_2}} \\ &= \frac{1}{4} \frac{\begin{pmatrix} -0.25 \\ -0.25 \\ -0.25 \\ 0.75 \end{pmatrix}^T \begin{pmatrix} -0.35 \\ 0.35 \\ 0.35 \\ -0.35 \end{pmatrix}}{\sqrt{0.1875} \times \sqrt{0.125}} = -0.57. \end{aligned}$$

Les coordonnées des colonnes sur le deuxième axe principal sont

$$\begin{pmatrix} 0 \\ 0.57 \\ 0 \\ -0.57 \end{pmatrix}.$$