Exercice 1 Considérons un échantillon de $n = 5$ individus où chaque individu $\boldsymbol{x}_i \in \mathbb{R}^d$ est décrit par $d = 3$ variables réelles. Cet échantillon est représenté par la matrice $\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5)^t$ suivante :

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

On va faire une ACP centrée réduite de ce jeu de données.

1. Calculer l'individu moyen (le centre de gravité du nuage de données) $\bar{\boldsymbol{x}}$
2. Calculer la matrice $\mathbf{Y}$ des données centrées
3. Calculer les écarts types σ_j de chacune des variables
4. Calculer la matrice $\mathbf{Z}$ des données centrées-réduites
5. Calculer la matrice de variance-covariance $\mathbf{\Sigma}$ de $\mathbf{Z}$ et la matrice de corrélation $\mathbf{R}$ de $\mathbf{X}$. Commenter.
6. Effectuer une décomposition spectrale de la matrice de corrélation $\mathbf{R}$: déterminer les valeurs propres λ_j associées aux vecteurs propres non-nuls $\mathbf{u}_j$ de $\mathbf{R}$.
7. Déterminer les facteurs principaux $\mathbf{f}_j$ et les axes principaux $\mathbf{a}_j$ du nuage des individus. Vérifier leurs propriétés statistiques
8. Calculer pour chacun des axes factoriels, l'inertie du jeu de données projetées sur l'axe considéré, et la part d'inertie qu'il explique.
9. Calculer les composantes principales $\mathbf{c}_j$ pour les individus. Comment s'interprètent les composantes principales en fonction des variables de départ. Vérifier leur propriétés statistiques.
10. Représenter graphiquement le nuage des individus sur le plan factoriel défini par les deux premiers axes factoriels. Commenter.
11. Représenter graphiquement le nuage des variables sur le plan factoriel défini par les deux premiers axes factoriels. Commenter.

Exercice 2 Considérons un échantillon de n individus où chaque individu $\boldsymbol{x}_i \in \mathbb{R}^d$ est décrit par $d = 3$ variables réelles qui ont pour matrice de corrélation

$$\mathbf{R} = \begin{pmatrix} 1 & \rho & -\rho \\ \rho & 1 & \rho \\ -\rho & \rho & 1 \end{pmatrix}$$

ave $-1 \leq \rho \leq 1$.

On va faire une ACP centrée-réduite de ce jeu de données.

1. Effectuer une décomposition spectrale de la matrice de corrélation $\mathbf{R}$: déterminer les valeurs propres λ_j associées aux vecteurs propres non-nuls $\mathbf{u}_j$ de $\mathbf{R}$.
2. Quelles sont les valeurs possibles pour ρ. Justifier que ρ doit vérifier $-1 \leq \rho \leq 1$.
3. Calculer pour chacun des axes factoriels, l'inertie du jeu de données projetées sur l'axe considéré, et la part d'inertie qu'il explique. Faire une représentation graphique.
4. Calculer les composantes principales $\mathbf{c}_j$ pour les individus. Comment s'interprètent les composantes principales en fonction des variables de départ. Vérifier leur propriétés statistiques.
5. Comment s'interprète en fonction des données d'origines $\boldsymbol{x}_i$ l'unique composante à retenir dans ce cas.