Matière : Analyse de données



**TD 3** 

Année: 2023/2024

## Exercice 1

Utiliser le PCA non normé pour combiner la pression la pression artérielle systolique (Systolic BP) et la pression artérielle diastolique (Diastolic BP) en se basant sur des données mesurées sur six individus.

| Systolic BP | Diastolic BP |
|-------------|--------------|
| 126         | 78           |
| 128         | 80           |
| 128         | 82           |
| 130         | 82           |
| 130         | 84           |
| 132         | 86           |

- 1. Calculer la première et la seconde composantes principales
- 2. Interpréter le résultat du PCA

## Exercice 2:

Soit le tableau suivant qui représente des données recueillies après de six individus, incluant trois variables : la pression artérielle systolique (SBP), le niveau de cholestérol (Chol) et l'âge.

| SBP | Chol | Age |
|-----|------|-----|
| 120 | 126  | 38  |
| 125 | 128  | 40  |
| 130 | 128  | 42  |
| 121 | 130  | 42  |
| 135 | 130  | 44  |
| 140 | 132  | 46  |

- 1. Calculer le coefficient de corrélation de Pearson entre le niveau de cholestérol (Chol) et l'âge. Quelle conclusion en tirez-vous ?
- 2. Chercher les paramètres du modèle de régression linéaire multiple.
- 3. Calculer les valeurs propres et les vecteurs propres de la matrice de covariance correspondante.
- 4. Calculer les deux composantes principales.

- 5. En utilisant le modèle de régression en composantes principales, faites la prédiction de la pression artérielle systolique d'une personne avec un niveau de cholestérol de 125 et âgée de 40 ans.
- 6. Représenter graphiquement la pression artérielle systolique avec la première composante principale.