

Exercices

Chapitre 4 : Analyse descriptive et visualisation des données

Hiver 2025

L'objectif de ces exercices est de réaliser un prétraitement sur un ensemble de données.

Les questions suivis de * ne sont pas corrigés. Ils sont ajoutés pour des fins de pratiques.

Soit l'ensemble de données `heart.txt` disponible sur Léa.

Étape 1 : Lecture des données

1. Lisez la base de données `heart.txt`
2. Déterminez le nombre d'individu (`df`)
3. Déterminez le nombre de variables (`df`)
4. Affichez les valeurs des variables des six premiers individus (`df`)
5. Énumérez les différentes variables et leur type (`df`)
6. Considérez un sous-ensemble de variables (les 8 premières) et vérifiez sa dimension (nommez le `df1`)
7. Énumérez les différentes variables de (`df1`)
8. Considérez un sous-ensemble d'individus (les 10 premiers) et vérifiez sa dimension (nommez le `df2`)

Étape 2 : Représentation graphique des données qualitative

1. Vérifiez les modalités de la variable (`type_douleur`) de `df`
2. Calculez l'effectif de chaque modalité de la variable (`type_douleur`) de l'ensemble `df`

3. Calculez la fréquence (en %) de la variable (`type_douleur`) de l'ensemble `df`
4. Représentez le diagramme en bâtons de la variable (`type_douleur`) de l'ensemble `df`
5. Représentez le diagramme circulaire de la variable (`type_douleur`) de l'ensemble `df`
6. * Refaire les questions 1 à 5 pour la variable (`vaisseau`)

Étape 3 : Description de variables quantitatives

1. Déterminez les caractéristiques de tendances centrales de la variable `age` : la moyenne, la médiane, les quartiles Q1, Q2, Q3.
2. Déterminez les caractéristiques de dispersion de la variable `age` : Étendue, la variance et l'écart type
3. Représentez l'histogramme de la variable `age`
4. Représentez l'histogramme de la variable `age` selon le `sexe`
5. Représentez le diagramme en boîte de la variable `age`
6. Calculez le coefficient d'asymétrie et interpréter le résultat
7. Calculez le coefficient d'aplatissement et interpréter le résultat
8. * Refaire les questions 1 à 7 pour la variable (`pression`)