



Module 3 - Analyse exploratoire des données

Exercices - Consignes

Cette première série d'exercices a deux objectifs : Le premier est de se familiariser avec l'utilisation des données avec le logiciel R et le deuxième est de pratiquer les notions d'exploration de données.

Soit la base de données `iris` prédéfinie dans l'environnement R. Dans la suite, il vous est demandé un ensemble de questions qui vous permettront de manipuler le contenu de cette base de données et d'interpréter les résultats des analyses. Vous devez fournir, pour chaque question, les lignes de codes qui permettent de répondre à la question, s'il y a lieu.

Exercices 1 : Lecture des données

1. Téléchargez le contenu de la base de données `iris`.
2. Énumérez les différentes variables.
3. Déterminez le nombre d'échantillons.
4. Affichez le contenu des dix premières lignes de la base de données.

Exercices 2 : Caractéristiques de tendances centrales et de dispersion

1. Calculez la moyenne et les trois quartiles de chaque variable. Présentez les résultats sous la forme d'un tableau récapitulatif.
2. Calculez la moyenne et les trois quartiles de la variable `Sepal.Length` selon le type d'espèce, c'est-à-dire selon `species`. Présentez les résultats sous la forme d'un tableau récapitulatif.

Exercices 3 : Caractéristiques de formes

1. Calculez le coefficient d'asymétrie de chaque variable. Présentez les résultats sous la forme d'un tableau récapitulatif.
2. Calculez le coefficient d'aplatissement de chaque variable. Présentez les résultats sous la forme d'un tableau récapitulatif.
3. Calculez le coefficient d'asymétrie de la variable `Petal.Length` pour l'espèce `versicolor`. Que peut-on dire de la distribution de cette variable ?