

Exercice 1

Airquality est un dataset intégré dans R

1. Donner la nature du dataset « Airquality » ?
2. Donner les dimensions du dataset « Airquality » ?
3. Combien de valeurs manquantes NA se trouvent dans la colonne "Ozone" du dataset « Airquality » ?
4. Donner les lignes qui contiennent au moins une valeur manquante NA ?
5. Quelle est la moyenne de la colonne "Ozone" dans cet ensemble de données ? Exclure les valeurs manquantes (codées comme NA) de ce calcul.
6. Donner un sous-ensemble du dataset « airquality » qui ne contient aucune NA en supprimant les lignes qui contiennent au moins une NA ?

Exercice 2

1. Extrayez le sous-ensemble de lignes du dataset où les valeurs d'ozone sont supérieures à 31 et les valeurs de température sont supérieures à 90.
2. Utilisez une boucle for pour créer un vecteur de longueur 6 contenant la moyenne de chaque colonne du dataset (**à l'exclusion de toutes les valeurs manquantes**).
3. Utilisez la fonction **Apply** pour calculer l'écart type de chaque colonne du dataset (**à l'exclusion de toutes les valeurs manquantes**).
4. Calculez la moyenne de « Ozone » pour chaque mois dans le dataset et créez un vecteur contenant les moyennes mensuelles (**excluez toutes les valeurs manquantes**).
5. Donner un échantillon aléatoire de 5 lignes du dataset ?

Exercice 3

1. Créé un objet qui contient les mêmes colonnes que Airquality mais la température et la vitesse du vent sont exprimées en degrés Celsius et en kilomètres par heure.
2. Calculer la corrélation entre niveau d'ozone et température.
3. Donner le nuage de points (scatter plot) du niveau d'ozone en fonction de la température.

Exercice 4

state.x77 est un dataset intégré dans R

1. Donner le revenu par habitant (per capita income) des États qui ont moins de 40 jours avec une température minimale en dessous de zéro.
2. Les états qui ont moins de 40 jours avec une température minimale en dessous de zéro et une espérance de vie supérieure à 71 ans.
3. La moyenne pour chaque variable dans cet dataset.
4. L'écart type pour chaque variable dans cet dataset.

Exercice 5

On va utiliser le package **DAAG** (Data Analysis and Graphics Data and Functions)

Installation du package

```
install.packages("DAAG")
```

1. Donner les valeurs non manquantes pour la colonne **root** pour chaque espèce.
2. Donner l'histogramme de l'espèce *Acmena smithii* en fonction du diamètre à hauteur de base (dhp).

Exercice 6

Considérez les données sur la longueur des œufs dans le dataframe **cuckoos** du package **DAAG**.

1. Tracez un histogramme de ces données.
2. Calculez la moyenne et l'écart type de ces données.
3. Les données semblent-elles être distribuées normalement ?
4. Construire un QQ-plot et des QQ-plots de référence pour ces données, en utilisant la fonction `qreference`.
5. Construisez un graphique de l'estimation de la densité pour ces données.
6. Répétez (a), (d) et (e) pour la longueur des œufs de troglodyte seulement.