TP n° 11 : Première étude et intervalles de confiance

Exercice 1. On considère le jeu de données `airquality`, disponible dans R.

1. Charger les données et comprendre d'où elles émanent.
2. Afficher les noms des variables considérées.
3. Afficher le nombre de lignes et de colonnes.
4. Calculer les paramètres statistiques de base à l'aide de la commande `summary`.
5. Calculer séparément la moyenne, la médiane, l'étendue et les quantiles des valeurs de la variable `Temp`.
6. Calculer la variance et l'écart-type pour les valeurs de la variable `Temp`.
7. Faire un graphique tige-feuille avec la commande `stem` pour les valeurs de `Wind`.
8. Tracer l'histogramme des fréquences pour les valeurs de `Temp` avec 20 classes.
9. Tracer l'histogramme des fréquences pour les valeurs de `Ozone` avec 18 classes.
10. Extraire du jeu de données :
    (a) la 2-ème ligne ligne,
    (b) la 3-ème colonne,
    (c) les lignes 1, 2 et 4 avec une seule commande `c()`,
    (d) les lignes 3 à 6 avec la commande `:`,
    (e) tout sauf les colonnes 1 et 2,
    (f) toutes les lignes ayant une température supérieure à 90.
11. Insérer une nouvelle variable `TooWindy` dans le jeu de données `airquality` qui affiche `TRUE` pour chaque valeur de `Wind` supérieure à 10, et `FALSE` sinon. Puis supprimer cette nouvelle variable.
12. Créer un nouveau jeu de données correspondant à `airquality` privée des lignes de `Ozone` contenant des valeurs manquantes.
13. Installer le package `ggplot2`. Reproduire et analyser les commandes suivantes :

```
library(ggplot2)
qplot(Temp, Ozone, data = airquality, colour = Month)
```

Exercice 2. On considère un échantillon de 10 bouteilles de bière blanche d'une certaine marque et on mesure le volume en litres de chacune d'entre elle. Les résultats sont :

0.499	0.509	0.501	0.494	0.498	0.497	0.504	0.506	0.502	0.496

On suppose que le volume en litres d'une telle bouteille peut être modélisée par une *var* $X \sim \mathcal{N}(\mu, \sigma^2)$ avec μ et σ inconnus.

1. Calculer les paramètres statistiques de base.
2. Tracer un histogramme des fréquences.
3. Déterminer un intervalle de confiance pour μ au niveau 95% avec la commande `t.test`.
4. Déterminer un intervalle de confiance pour μ au niveau 99%.
5. Reproduire et analyser les commandes suivantes :
```
x = c(0.499, 0.509, 0.501, 0.494, 0.498, 0.497, 0.504, 0.506, 0.502, 0.496)
n = 10
alpha = 0.01
mean(x) - sd(x) * qt(1 - alpha / 2, n - 1) / sqrt(n)
mean(x) + sd(x) * qt(1 - alpha / 2, n - 1) / sqrt(n)
```
6. Reproduire les commandes suivantes :
```
(n - 1) * var(x) / qchisq(1 - alpha / 2, n - 1)
(n - 1) * var(x) / qchisq(alpha / 2, n - 1)
```
À quoi correspondent les valeurs renvoyées en matière d'intervalle de confiance ?

Exercice 3. Une ferme de Bay of Plenty en Nouvelle-Zélande produit des kiwis. On a mesuré les masses de 16 kiwis provenant de cette ferme. Les résultats, en grammes, sont :

65.06	71.44	67.93	69.02	67.28	62.34	66.23	64.16
68.56	70.45	64.91	69.90	65.52	66.75	68.54	67.90

On suppose que la masse en grammes d'un kiwi de cette ferme peut être modélisée par une *var* $X \sim \mathcal{N}(\mu, \sigma^2)$, avec μ et σ inconnus.

1. Déterminer un intervalle de confiance pour μ au niveau 95%.
2. Déterminer un intervalle de confiance pour σ^2 au niveau 95% en calculant "à la main" les bornes de celui-ci.

Exercice 4. Un radar situé sur une route départementale enregistre chaque jour la vitesse des véhicules qui passent entre deux sorties. On a regroupé les résultats en classes de 5 kilomètres heure. Le tableau suivant indique les effectifs observés le 25 janvier 2011. Tous les véhicules dont la vitesse dépasse strictement 90 kilomètres heure sont en infraction.

Classe	$]0, 75]$	$]75, 80]$	$]80, 85]$	$]85, 90]$	$]90, 95]$	$]95, 100]$	$]100, 105]$	$]105, \infty[$
Effectif	150	450	1000	2000	50	20	10	3

Déterminer un intervalle de confiance pour la proportion inconnue de véhicules en infraction un jour donné au niveau 95% *(on utilisera la commande* `binom.test`*)*.