

# M1 SSD - UE Tests - TP2

## Exercice 1

Le jeu de données `ozone.csv` provient de l'association Air Breizh. L'une des missions de ces associations de surveillance de la qualité de l'air est de construire des modèles de prévision de la concentration en ozone, pour anticiper un dépassement légal du pic d'ozone ( $180 \mu\text{g}/\text{m}^3$ ). L'objectif des données récoltées dans ce jeu de données est de mettre en évidence l'influence de certains paramètres sur la concentration d'ozone.

Les 13 variables observées sont :

- `maxO3`: Maximum de concentration d'ozone observé sur la journée ( $\mu\text{g}/\text{m}^3$ )
- `T9`, `T12`, `T15` : Température observée à 9, 12 et 15h
- `Ne9`, `Ne12`, `Ne15` : Nébulosité observée à 9, 12 et 15h
- `Vx9`, `Vx12`, `Vx15` : Composante Est-Ouest du vent à 9, 12 et 15h
- `maxO3v` : Teneur maximum en ozone observée la veille
- `vent` : orientation du vent à 12h
- `pluie` : occurrence ou non de précipitations

La question principale est l'étude de la dépendance du maximum de concentration d'ozone observé sur la journée avec les autres variables. La suite des questions peut vous servir de guide mais ne doit pas vous restreindre. L'objectif de ce TP est de répondre à la question principale avec une méthodologie statistique rigoureuse. Pour chaque test, vous discuterez des conditions d'application. Par exemple, vous pourrez utiliser les tests de Student, Wilcoxon, ANOVA, Kruskal-Wallis, Shapiro-Wilks, Kolmogorov-Smirnov, chi-deux, corrélation.

1. Résumer graphiquement les différentes variables. Proposer aussi des résumés quantitatifs. Commenter.
2. Comparer avec un test la différence entre le maximum de concentration d'ozone des jours avec pluie et sans pluie. Est-ce que leurs moyennes sont significativement différentes au niveau 5% ? Est-ce que leurs distributions sont différentes ?
3. Qu'en est-il entre les différentes orientations du vent ? Comparer toutes les paires de groupes et résumer les résultats de cette comparaison multiple.
4. Quel est le lien entre la concentration maximale d'ozone et la température ? et la nébulosité ? et la force du vent Est-Ouest ?
5. Enfin quel est le lien entre la concentration maximale d'ozone et la concentration de la veille ?
6. On peut utiliser une analyse multivariée en appliquant un modèle linéaire pour expliquer la variable maximum de concentration d'ozone en fonction de toutes les autres variables. Qu'observez-vous ? Est-ce cohérent avec les résultats de chaque test obtenu précédemment ? Pourquoi ?
7. Conclure.

## Exercice 2

L'objectif de cet exercice est d'écrire une fonction `myt.test` qui implémente le test de Welch ou de Student, de façon équivalente à la fonction `t.test`. En particulier :

- Le format de la fonction doit être : `myt.test(x, y, alternative = "two.sided", paired = FALSE, var.equal = FALSE)`
- Les paramètres d'entrée :
  - `x` and `y` sont les vecteurs de données
  - `alternative` prend les valeurs `c("two.sided", "less", "greater")`
  - `paired` TRUE/FALSE
  - `var.equal` TRUE/FALSE
- Les sorties: une liste avec les objets
  - `method` avec une valeur dans `c("Welch two sample t-test", "Student two sample t-test")`
  - `alternative` avec une valeur dans `c("two.sided", "less", "greater")`
  - `p.value` contenant la  $p$ -valeur du test
  - `df` contenant le nombre de degré de liberté du test
  - `statistic` la valeur de la statistique de test

Tester votre fonction sur un jeu de données simulées et comparer avec les résultats de la fonction `t.test`. Vous devez obtenir exactement les mêmes résultats.