M2 Stat de la SD
TD1

~~10 janvier 2024~~
14 octobre 2024

Le travail est à réaliser en trinômes et chaque réponse doit être consciencieusement justifiée.
Un fichier compilé (pdf) est à envoyer en fin de cours à l'adresse :
hadrien.lorenzo@univ-amu.fr
Bonne chance !

Nous allons nous concentrer sur des variatoins d'un je de données, ozone, décrit par les 13 varaibles suivantes Les 13 variables recueillies sont :

- MaxO3 : Maximum de concentration d'ozone observé sur la journée
- T9, T12, T15 : Température observée à 9, 12 et 15h
- Ne9, Ne12, Ne15 : Nébulosité observée à 9,12 et 15h
- Vx9, Vx12, Vx15 : Composante E-O 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

Dans chacune des questions, proposer des visualisations, via le package **VIM** par exemple, afin d'étayer votre propos. Regarder notamment les fonctions **aggr**, **marginplot** et **matrixplot**.

Exercice 1

Ouvrir le fichier **ozone_1.csv**, conjecturer une source éventuelle de données manquantes et la classer parmi MCAR, MAR ou MNAR. Faire de même avec les fichiers **ozone_2.csv**, **ozone_3.csv** et **ozone_4.csv**.

Exercice 2

Utiliser le package **missMDA**, et l'implémentation de l'ACP qui est proposée, afin de réaliser l'imputation simple des 4 jeux de donnée précédents. Proposer des visualisations afin de rendre compte de la qualité des imputations. Mettre à défaut, ou pas, les conclusions de l'exercice précédent.

Exercice 3

Réaliser l'imputation simple de chacun des 4 jeux de données avec le package **missForest**. Comparer, dans chaque cas, les imputations réalisées en utilisant les vraies valeurs présentes dans le fichier **ozone.csv**. Discuter les hypothèses de linéarité de l'ACP dans chaque cas.