

Projet 9 : Relation entre pollution et mortalité

Dataset :

Le jeu de données contient, pour 60 zones géographiques américaines, des données socio-économique, météo, le taux de certains composés chimique dans l'atmosphère, ainsi que le taux de mortalité.

Le dataset est disponible à l'adresse suivante :

<http://lib.stat.cmu.edu/datasets/pollution>

Objectif :

L'objectif de l'étude est d'explorer en détail, à l'aide d'un modèle de régression linéaire multiple, l'impact de la pollution sur la santé, et plus spécifiquement sur la mortalité.

Il conviendra de bien comprendre les relations entre l'ensemble des variables explicatives, et leur impact sur la régression.

Ce projet est l'occasion d'approfondir les concepts vus en cours, à un niveau plus avancé, et de comprendre l'analyse de données en présence de plusieurs variables qui peuvent interagir et/ou être liées les unes aux autres, ainsi que les méthodes qui permettent de contrôler les effets qui en découlent.