

TP3: Modèle linéaire et généralisations

Données ozone

On considère le jeu de données “ozone.txt” qui mesure la quantité maximale d’Ozone par jour à Rennes (maxO3), à expliquer par les covariables suivantes :

- T9, T12, T15: température à 9h, 12h, 15h
- Ne9, Ne12, Ne15: nébulosité à 9h, 12h, 15h
- Vx9, Vx12, Vx15: vitesse du vent à 9h, 12h, 15h
- maxO3v: pollution maximale à l’ozone la veille
- vent: direction principale du vent
- temps: sec ou pluvieux

Chargement des données

Charger les données dans R et les visualiser rapidement. Quelle est la nature des colonnes vent et temps?

Régression linéaire simple

Effectuer et valider une régression linéaire simple de maxO3 contre T12. Interpréter les sorties.

ANOVA

Effectuer une ANOVA à 1 facteur de maxO3 contre temps, puis contre vent. Qu’en pensez-vous ?

Effectuer une ANOVA de maxO3 contre ces deux facteurs simultanément.

Régression linéaire multiple

Effectuer une régression linéaire multiple de maxO3 contre toutes les covariables. Qu’en pensez-vous ?

Proposer des interactions éventuelles entre covariables.

Choix de modèle

Effectuer une sélection de covariables, et interpréter le modèle final.

Prédiction

Pour une nouvelle observation, on a T9=15, T12=18, T15=20, Ne9=3, Ne12=4, Ne15=4, Vx9=0.7, Vx12=1.3, Vx15=1.4, maxO3v=85, vent=Ouest.

Donner une prédiction ponctuelle et un intervalle de prédiction à 95%.