

TP 1 : Analyse des résidus

Exercice 1. Données ozone

Nous expliquons le pic d'ozone "max03" par 6 variables : la teneur en ozone maximale la veille "max03v", la température prévu par Météo France à 9h "T9", à midi "T12", une variable synthétique (la projection du vent sur l'axe est-ouest notée "Vx12"), et enfin la nébulosité prévue à midi "Ne12" et à 15h "Ne15". Nous voulons valider le modèle de régression linéaire multiple suivant :

$$\text{max03} = \beta_0 + \beta_1 \text{max03v} + \beta_2 T9 + \beta_3 T12 + \beta_4 Vx12 + \beta_5 Ne12 + \beta_6 Ne15 + \varepsilon$$

1. Proposer une estimation de ce modèle de régression en utilisant la fonction `lm()`.
2. Construire le vecteur `res` des résidus calculés.
3. Représenter les différents graphes des résidus vus en cours, en précisant pour chacun ce qu'il permet de vérifier.
4. Construire la Hat matrix `H`. Que peut-on en conclure ?
5. Construire le vecteur `rstand` des résidus standardisés.
6. Calculer l'erreur de test pour chaque observation i , puis le PRESS de Allen.
7. Construire puis représenter les résidus studentisés en fonction du numéro de l'observation. Que peut-on en conclure ?
8. Calculer la distance de Cook pour chaque observation i .
9. Tracer la droite de Henry ou Normal QQ-plot.
10. Que proposez vous avant de poursuivre l'estimation de ce modèle de régression linéaire ?

Exercice 2. Espérance de vie

On étudie la relation entre l'espérance de vie et le nombre d'habitants par TV de 38 pays du monde (fichier "lifeexp.txt").

Diagnostiquer le modèle de régression linéaire simple :

$$\text{LifeExp} = \beta_0 + \beta_1 \text{People.per.TV} + \varepsilon,$$

et proposer par conséquent des améliorations significatives de ce dernier.