

TP5

Régression linéaire multiple (2) : variables quantitatives

Objectifs du TP

- Maîtriser la régression linéaire multiple avec des covariables quantitatives : estimation et tests.
- Interpréter les sorties de R associées.
- Savoir donner une réponse statistique précise à une question formulée en français.

Pollution de l'air : de nouvelles variables explicatives

La modélisation des pics de pollution à l'ozone à Rennes proposée lors des premières séances peut sembler simpliste : la température à 12h ne suffit vraisemblablement pas à expliquer la concentration maximale journalière en ozone.

Le jeu de données '`ozone3.txt`' contient, pour chaque jour d'observation :

- **Date** : la date de l'observation.
- **O3** : le maximum journalier de la concentration en O_3 , exprimé en $\mu\text{g}/\text{m}^3$.
- **T12**, **T15** : les températures observées à 12h et à 15h, exprimées en $^{\circ}\text{C}$.
- **Vx** : la composante est-ouest du vent à 12h.
- **O3v** : le maximum journalier de la concentration en O_3 observé la veille, exprimé en $\mu\text{g}/\text{m}^3$.
- **vent** : l'orientation du vent à 12h (nord, est, sud ou ouest).

On va construire des modèles incluant certaines des covariables, en vue d'expliquer au mieux la valeur de la variable réponse **O3**.

Préliminaires

1. Importer les données ‘`ozone3.txt`’.
2. Faire une analyse descriptive des données (en se limitant aux variables quantitatives). Sélectionner un petit nombre de résultats permettant de dire quelque chose d’intéressant.

Régression multiple

On considère un modèle de régression linéaire qui explique O_3 à l’aide de O_3v , $T12$ et Vx . On appelle ce modèle « le modèle complet ».

3. Écrire ce que signifie (mathématiquement) le modèle complet.
Que vaut p , le nombre de covariables, pour le modèle complet ?
4. Ajuster les données au modèle complet.

Tests

5. Comparer la qualité du modèle complet aux modèles de régression linéaire simple expliquant O_3 à l’aide d’une seule covariable (respectivement : O_3v , $T12$ et Vx).
6. Dans le modèle complet, tester l’influence de chacune des covariables seules.
7. Y a-t-il un effet « concentration en ozone la veille ou vitesse du vent » ?
8. Une augmentation de la température induit-elle une augmentation du taux de pollution à l’ozone ? (Faire un test.)
9. Quel modèle proposer avec O_3v , $T12$ et Vx ? Interpréter.
10. Ajouter la variable $T15$ au modèle complet. La température a-t-elle un effet sur la pollution ?

Diagnostics

11. Diagnostiquer les résidus du modèle proposé à la question 9.
11. Reprendre le modèle de la question précédente et proposer une transformation de $T12$ qui améliore le modèle. Diagnostiquer les résidus.