

TP3

Régression linéaire simple (3) : intervalles de confiance, tests et diagnostics

Objectifs du TP

- Savoir déterminer un intervalle de confiance pour β_1 , β_2 et σ^2 .
- Savoir déterminer un intervalle de confiance pour une valeur moyenne, un intervalle de prévision, et comprendre la différence entre les deux.
- Savoir effectuer des tests sur β_1 et β_2 .
- Interpréter les sorties de la fonction `lm`, en particulier les tests et le coefficient de détermination.
- Faire un diagnostic sur les résidus, et utiliser le résultat pour construire un meilleur modèle.

Intervalles de confiance, intervalle de prévision

1. Charger le jeu de données ‘`ozone_simple.txt`’ et ajuster un modèle linéaire expliquant O3 à l’aide de T12.
2. En utilisant les formules du cours, déterminer des intervalles de confiance de probabilité de couverture 0.95 pour β_1 , puis pour β_2 , et enfin pour σ^2 .
Faire de même pour β_1 et β_2 avec la fonction `confint`.
3. Sur un même graphe, représenter les données, tracer la droite de régression, les intervalles de confiance à 95% pour $\mathbb{E}[y]$ et les intervalles de prévision à 95%. On pourra utiliser la fonction `predict`. Commenter.

Tests, analyse de la variance

4. Tester les paramètres du modèle (en précisant bien H_0 et H_1 pour chaque test). Commenter.
5. Calculer à la main la valeur du coefficient de détermination et la valeur de la statistique de Fisher. Commenter.

Diagnostics

6. Diagnostiquer les résidus. On pourra notamment utiliser la commande `plot` appliquée à la sortie de la fonction `lm`.
7. Proposer une transformation des données susceptible de mieux expliquer la relation entre concentration maximale O3 et température à midi T12. Interpréter le modèle final (*i.e.* refaire les questions précédentes avec ce nouveau modèle).

Questions subsidiaires

8. Il semble intuitif que la pente soit strictement positive. Tester la réalité de cette intuition (en précisant bien H_0 et H_1 , et en expliquant pourquoi le test est valide).
9. À la question 3, les intervalles de confiance obtenus pour les $\mathbb{E}[y_i]$ sont-ils valables *simultanément* pour tous les $i \in \{1, \dots, n\}$?