

Fiche de TD n° 15 et 17

Travaux pratiques en R

Régression linéaire

Avertissement

Ce TP fera l'objet d'un compte rendu déposé sur le portail avant le 22 janvier.

1° - Le problème

On désire prévoir le volume, exploitable en scierie, de merisiers en fonction de paramètres facilement mesurables sur le terrain : le diamètre mesuré à hauteur d'homme, et la hauteur - évaluée simplement à l'aide d'un piquet de hauteur donnée et du théorème de Thalès (faire un dessin si vous voulez deviner le procédé!).

On a recueilli à cet effet les données pour 31 arbres (fichier `arbres.txt`).

2° - L'étude

Il s'agit de proposer un modèle de régression permettant de prévoir le volume à l'aide des deux quantités mesurées sur le terrain, hauteur et diamètre. Pour ce faire, nous vous suggérons la démarche qui suit :

1. Lecture et examen rapide des données (fonctions `read.table`, `hist`, `plot`, `pairs`)
2. Essai d'un modèle de régression linéaire (fonction `lm`). Commentaires sur les coefficients, la fonction obtenue.
3. Représenter les résidus en fonction des deux variables hauteur et diamètre. Examiner également la normalité des résidus. Remarques ?
4. Séparer les données en deux parties selon le niveau de la variable diamètre et représenter sur un même graphique la dépendance de volume en la variable hauteur pour ces deux parties. Recommencer en échangeant les rôles de hauteur et diamètre. Les observations sont-elles compatibles avec le modèle de régression ?
5. Recommencer en considérant le log des trois variables. Examiner notamment les résidus (cf. point 3 + normalité) et expliquer en quoi le modèle résultant peut être compatible avec les observations du point 4.

6. Examiner les coefficients obtenus, et expliquez pourquoi un forestier passant par là s'esclaffe en disant que vous venez de réinventer la roue.
7. Proposer alors un troisième modèle de régression ne contenant plus qu'un seul coefficient à estimer, et discuter de la valeur de ce dernier. Conclure et répondre à la question suivante : "Quelle est la forme d'un merisier?".
8. Question bonus. Donner un intervalle de confiance sur le coefficient du modèle.