

TP2 : Modèle linéaire multiple

STID - IUT de Paris - Rives de Seine

2020-2021

La chenille processionnaire est un papillon nocturne de la famille des Notodontidés. Elle se développe de préférence sur des pins et peut causer des dégâts considérables. On souhaite étudier l'influence de certaines caractéristiques de peuplements forestiers sur leurs développements à partir d'un échantillon de 32 parcelles forestières de 10 hectares. Chaque parcelle est échantillonnée en placettes de 5 ares et on a calculé les moyennes (sur les placettes) des différentes mesures suivantes :

- Y : le **nombre de nids** de chenilles processionnaires par arbre,
- X_1 : l'**altitude** en mètres,
- X_2 : la **pente** en degrés,
- X_3 : le **nombre de pins** dans la placette,
- X_4 : la **hauteur** en mètres de l'arbre échantillonné au centre de la placette,
- X_5 : le **diamètre** de cet arbre,
- X_6 : la note de **densité** de peuplement,
- X_7 : l'**orientation** de la placette, allant de 1 (sud) à 2 (autre),
- X_8 : la **hauteur** en mètres des arbres dominants,
- X_9 : le **nombre** de strates de végétation,
- X_{10} : le **mélange** du peuplement, allant de 1 (non mélangé) à 2 (mélangé).

On souhaite expliquer le nombre de nids par les autres variables de la base de données.

1. Charger les données **chenilles** correspondant à l'étude :

```
chen <- read.table("chenilles.txt",header=TRUE)
attach(chen)
```

2. Calculer et interpréter la matrice des corrélations entre toutes les variables.
3. Construire un modèle linéaire gaussien multiple expliquant le nombre de nids en fonction des autres variables.
4. Comparer les valeurs du nombre de nids prédites par ce modèle avec les vrais valeurs. Que penser de la qualité d'ajustement du modèle aux données?
5. Tester au risque 5% la contribution globale des variables explicatives sur le nombre de nids.
6. Effectuer une sélection descendante des variables explicatives avec le test de Fisher. Quel est le modèle final retenu?
7. Effectuer une sélection descendante des variables explicatives avec le critère AIC à l'aide de la fonction `step()`. Quel est le modèle final retenu?
8. Comparer la qualité d'ajustement aux données de ces deux modèles.

9. Comparer les prévisions obtenues avec ces deux méthodes de sélection de modèles.
10. Calculer la valeur du nombre de nids prédite par ces deux modèles pour une parcelle d'altitude 1300m, de pente 28 degrés, de hauteur 4.4m, de diamètre 15m et de nombre de strates égal à 2.
11. Vérifier si les hypothèses du modèle linéaire multiple sont satisfaites.

On travaille désormais avec le logarithme du nombre de nids

12. Effectuer une régression linéaire multiple du log du nombre de nids en fonction des variables explicatives.
13. Analyser la qualité d'ajustement du modèle aux données.
14. Tester au risque 5% la contribution globale des variables explicatives sur le log de nombre de nids.
15. Effectuer une sélection descendante des variables explicatives avec le test de Fisher. Quel est le modèle final retenu?