

TP : Régression Multiple et Polynomiale

Partie I :

Nous voulons étudier les problèmes respiratoires d'une population de 725 individus afin de prédire les causes de ces problèmes. Nous avons à cet effet 6 variables nommées respectivement **LungCap**, **Age**, **Height**, **Smoke**, **Caesarean** représentant respectivement la capacité respiratoire, l'âge, la taille, si l'individu est fumeur ou non et enfin s'il a eu une naissance normale ou par césarienne.

Il est demandé d'exécuter le programme R suivant d'interpréter les résultats de chaque étape.

```
LungCapData <- read.table(file.choose(), header = T, sep="\t")
attach(LungCapData)
names(LungCapData)
class(Age)
class(Smoke)
levels(Smoke)
modell <- lm(LungCap ~ Age + Height)
summary(modell)
cor(Age, Height, method="pearson")
confint(modell, coef.level=0.95)
model2 <- lm(LungCap ~ Age + Height + Smoke + Gender +
Caesarean)
summary(model2)
plot(model2)
```

Partie II :

Cette fois-ci, nous espérons trouver une liaison entre une ou plusieurs variables de la base **Boston** que vous trouverez dans le package **MASS**. Cette base nous renseigne sur la valeur d'un logement dans la banlieue de Boston (USA). Elle contient enregistrement de variables à l'instar de :

- **crim** : pourcentage d'habitants criminels par ville.
- **rad** : indice d'accessibilité aux autoroutes radiales.
- **black** : proportion de noirs par 1000 individus et par ville.
- **lstat** : statut minimal de la population (pourcentage).
- **medv** : valeur médiane des habitants-propriétaires en 1000\$ par unité.
- Etc

Questions :

1. Charger la table Boston.
2. Afficher les informations.
3. Afficher le type de chaque donnée.
4. Que pouvez-vous dire de ces données.
5. Diviser les données en utilisant les 400 premières observations que les données d'entraînement et le reste en tant que données de test.
6. Vérifier s'il existe une relation linéaire entre **medv** et **age**.

7. Dessiner le nuage de points de ces deux variables.
8. Faire la même chose avec **medv** et **lstat**.
9. Dessiner la droite de régression.
10. Vérifier statiquement et graphiquement si la variable **medv** peut être expliqué par un modèle linéaire simple par **lstat**.
11. Que pouvez-vous déduire ?
12. Exécutez et commenter l'instruction **pairs(Boston)**.
13. Répéter l'instruction mais uniquement avec les première, la troisième et la septième variable.
14. utiliser l'ensemble de données de formation pour former le modèle linéaire multiple avec comme variable expliquée **medv** et comme variables explicatives la variable **lstat** et la variable **age**.
15. Que peut-on conclure ? Expliquez
16. utiliser l'ensemble de données de formation pour former le modèle linéaire multiple avec comme variable expliquée **medv** et comme variables explicatives le logarithme de la variable **lstat** et la variable **age**.
17. Que peut-on conclure ? Expliquez
18. Vérifier la relation linéaire **medv** entre et toutes les autres variables.
19. Que peut-on conclure ? Expliquez
20. Reconstruire le modèle linéaire avec toutes les variables sauf celles non significatives.
21. Que peut-on conclure ? Expliquez
22. Construire le modèle linéaire avec la variable **medv** et l'interaction entre **lstat** et **age**.
23. Vérifier si le modèle est non linéaire de degré 2, 3, 4, 5, 6 et 7.
24. Que peut-on conclure.