

# TP1 : Régression linéaire simple

STID - IUT Paris Descartes - 2AFA

2020-2021

## Exercice 1

Dans cet exercice, on cherche à expliquer le poids du coeur de chats adultes (**Hwt**) par le poids de leur corps (**Bwt**). Pour cela, on utilise le jeu de données **cats** de la librairie **MASS** du logiciel **R**.

1. Charger le jeu de données **cats** dans une nouvelle page **R**.

```
library(MASS)
cats
```

2. Réaliser une analyse descriptive du jeu de données.
3. Justifier la modélisation des données par un modèle linéaire gaussien.
4. Donner une estimation des paramètres du modèle ainsi qu'un intervalle de confiance des paramètres d'espérance de niveau de confiance 95%.
5. Tester au risque 5% le caractère significatif de la liaison linéaire entre le poids du coeur des chats et le poids de leur corps de deux manières différentes.
6. Etudier la qualité d'ajustement du modèle aux données.
7. Donner une prévision et un intervalle de prédiction du poids du coeur d'un chat dont le poids du corps est de 2.5kg.
8. Tracer la droite de régression estimée par le modèle linéaire sur le nuage de points ainsi que l'intervalle de confiance de la droite de régression et l'intervalle de pari des prédictions (à 95%).
9. Procéder à la validation du modèle.

## Exercice 2

Un agronome cherche à étudier la relation entre le taux de DDT d'un brochet et l'âge du brochet. Il dispose pour cela d'un échantillon de  $n = 15$  brochets pour lesquels on a relevé l'âge  $x$  et la mesure de son taux de DDT  $y$ . Les données sont disponibles dans le fichier **brochets.txt**.

1. Ouvrir un nouveau script et charger les données de cette étude en tapant :

```
Brochets = read.table("/chemin_vers_le_fichier/brochets.txt", header=TRUE)
attach(Brochets)
```

2. Décrire les données et tracer le nuage de points du taux de DDT en fonction de l'âge des brochets.
3. Calculer les moyennes et variances des taux de DDT des brochets par âge. Refaire le même calcul sur le log du taux de DDT. Qu'en pensez-vous?

4. Introduire la modélisation des données par un modèle linéaire gaussien. Donner les estimations des paramètres mis en jeu dans le modèle.
5. Tester au risque 5% le caractère significatif de la liaison linéaire entre le taux de DDT et l'âge des brochets en comparant le modèle constant au modèle linéaire gaussien.
6. Que pensez-vous de la qualité d'ajustement du modèle aux données? Comparer les prévisions obtenues avec la modèle aux observations.
7. L'analyse des résidus semble-t-elle valider les hypothèses du modèle?
8. Expliquer pourquoi il est judicieux de travailler avec le log des taux de DDT.
9. On cherche désormais à expliquer le log du taux de DDT en fonction de l'âge des brochets. Donner les estimateurs des paramètres du modèle linéaire simple gaussien correspondant.
10. Tester au risque 5% le caractère significatif du lien linéaire entre le log du taux de DDT et l'âge des brochets.
11. Discuter de la qualité d'ajustement du modèle aux données et procéder à la validation du modèle.
12. Proposer une prédiction du log du taux de DDT pour un brochet âgé de 7 ans, 10 ans ainsi qu'un intervalle de prévision de niveau de confiance 95%.