

TP – Régression linéaire simple**Exercice 1.**

Récupérer la base `anscombe`.

```
> data(anscombe)
> attach(anscombe)
```

1. Réaliser l'ajustement d'un modèle linéaire simple pour chacun des 4 jeux de données (x_1, y_1) , (x_2, y_2) , (x_3, y_3) et (x_4, y_4) à l'aide de la fonction `lm`. On ne demande pas de réaliser l'étude des résidus.
2. Comparer les principales statistiques de l'ajustement, données par la fonction `summary`.
3. Faire les représentations graphiques de (x_1, y_1) , (x_2, y_2) , (x_3, y_3) et (x_4, y_4) ainsi que des droites de régression linéaire obtenues. Commenter.

Exercice 2.

On utilise la base `cats` de la library `MASS`.

```
> library(MASS)
> attach(cats)
```

On souhaite expliquer le poids du coeur (Hwt) par le poids du corps (Bwt)

1. Tracer un nuage de points représentant le poids du coeur en fonction du poids du corps.
2. On cherche à étudier la relation linéaire entre ces deux variables. On utilise pour cela la fonction `lm` de R.

```
> cats.lm <- lm(Hwt ~ Bwt)
> summary(cats.lm)
```

Commenter les résultats obtenus.

3. Tracer la droite de régression estimée par le modèle linéaire sur le nuage de points.
4. Faire l'étude des résidus.

Exercice 3.

On utilise la base `cars` de R. Ce paquet contient deux variables : la vitesse et la distance de freinage de 50 voitures, mesurées en 1920. On souhaite expliquer la distance de freinage `dist` par la vitesse de la voiture `speed`.

1. Tracer un nuage de points représentant la distance de freinage en fonction de la vitesse de la voiture.

2. On cherche à étudier la relation linéaire entre ces deux variables à l'aide de la fonction `lm` de R. Commenter les résultats obtenus.
3. Tracer la droite de régression estimée par le modèle linéaire sur le nuage de points.
4. Faire les représentations graphiques des résidus studentisés et tester les hypothèses du modèle.
5. Représenter les intervalles de confiance et de prédiction.
6. On souhaite maintenant expliquer la distance de freinage par le carré de la vitesse. On introduit `speed2 <- speed^2`. Faire la représentation graphique et ajuster le modèle avec une contrainte de constante nulle : $dist_i = a \cdot speed2_i + \varepsilon_i$. Pour cela on utilisera la commande `cars.lm2 <- lm(dist~0+speed2)`. Comparer les résultats obtenus avec le premier modèle.

Exercice 4.

On utilise la base `p9.10` du paquet `MPV`. L'objectif est d'expliquer l'usure de l'asphalte `y` en fonction des caractéristiques du revêtement. Nous nous intéresserons plus spécifiquement à l'influence de la viscosité, donnée par `x1` en échelle logarithmique.

1. Faire l'ajustement du modèle linéaire de `y` par rapport à `x`. Vérifier les hypothèses du modèle. Qu'en concluez-vous ?
2. Au vu des représentations graphiques (`pairs(p9.10)`) on souhaite faire intervenir dans le modèle la variable `x4`. Cette variable est un indicateur prenant les valeurs `-1` et `+1`.
 - (a) Ajuster un modèle de régression linéaire sur le sous-ensemble `x4==-1`. On pourra utiliser l'option `subset` de `lm`. Faire l'étude des résidus.
 - (b) Ajuster un modèle de régression linéaire sur le sous-ensemble `x4==+1`. Faire l'étude des résidus.