

MASS 42 – TP 1

Introduction au modèle linéaire

A. Perrut - A. Huet

2010-2011

Exercice 1 : Simuler un échantillon X de taille $n = 30$ de la loi uniforme sur $[0, 5]$. Simuler ensuite un échantillon Z de taille $n = 30$ de la loi normale centrée d'écart-type 0.5. Calculer $Y = 2X + 7 + Z$. Tracer le graphe des observations Y en fonction de X . Effectuer une régression linéaire grâce à la fonction `regression=lm(Y~X)`. Afficher le résumé de `regression` pour obtenir l'équation de la droite des moindres carrés, ainsi que le coefficient de corrélation. Faire une étude statistique des résidus obtenus avec la commande `residus=regression$residuals` : histogramme, droite de Henry, boxplot.

Exercice 2 : Ouvrir le fichier `modelelineaire.txt` disponible sur la page <http://math.univ-lyon1.fr/~perrut>. Pour chacun des couples d'observations $(X1, Y1)$, $(X2, Y2)$, $(X3, Y3)$, $(X4, Y4)$, $(X5, Y5)$, tracer le nuage de points et deviner l'ordre de grandeur du coefficient de corrélation avant de la calculer. Étudier les résidus.

Exercice 3 : Toujours à la même adresse, télécharger le fichier `QI.txt` puis importer les données dans R en utilisant la commande

```
read.table("QI.txt",header=T).
```

Deux variables ont été mesurées sur des enfants dans un laboratoire américain de psychologie : le quotient intellectuel (QI) et un score de problèmes comportementaux (PC). On cherche à étudier la relation entre ces problèmes et le QI. Réaliser une régression linéaire : tracer le nuage de points, calculer le coefficient de corrélation linéaire, calculer la droite des moindres carrés, extraire les résidus. Que pensez-vous de cette modélisation ?