

Feuille-réponse 6 Régression linéaire

1 Droite de régression

Former le vecteur \mathbf{x} des *empans de la main qui écrit Wr.hnd* du fichier `survey` nettoyé des données manquantes :

```
library(MASS); data(survey); survey.cc<-survey[complete.cases(survey), ]  
x<-survey.cc$Wr.Hnd
```

et former de même les vecteurs \mathbf{y} et \mathbf{z} des deux autres variables `NW.Hnd` et `Height`.

1. Former les trois nuages de points (x, y) , (x, z) et (y, z) , soit en utilisant trois fois la commande `plot` ou une fois la commande `pairs` comme indiqué dans le cours. Parmi ces trois nuages, quel est celui qui vous paraîtrait le plus adapté à un modèle linéaire? Expliquer.

2. Ajouter au nuage (x, y) son centre de gravité au moyen de la commande `points`. Le centre de gravité est-il l'un des point du nuage? Qu'observez-vous concernant sa position?

3. Les commandes suivantes permettent de créer un *data frame* avec les trois variables \mathbf{x} , \mathbf{y} et \mathbf{z} et de réaliser une régression linéaire de \mathbf{y} sur \mathbf{x} . Quel est le modèle linéaire $y = ax + b$ obtenu (utiliser `coef`)? Le centre de gravité du nuage est-il sur la droite? Expliquer.

```
MonTableau=data.frame(x,y,z)
```

```
DteReg_y_sur_x=lm(y~x,data=MonTableau)
```

4. Avec la commande `abline(coef(DteReg_y_sur_x))` vous pouvez ajouter au nuage de point (obtenu avec `plot`) son modèle linéaire calculé avec `lm`. Expliquer comment vous pouvez vérifier sur votre figure les valeurs des coefficients de la droite? L'examen de cette figure vous permet-elle d'estimer l'ordre de grandeur des résidus?

5. Reprendre les deux questions précédente pour une régression linéaire de \mathbf{z} sur \mathbf{y} cette fois.

2 Résidus

1. Expliquer pourquoi la commande `res_y2x<-y-fitted(DteReg_y_sur_x)` permet de calculer les résidus de la régression de y sur x .
2. Tracer la boîte à moustache de `res_y2x`, indiquer combien de résidus sont exceptionnels et repérer à quels points du nuage de points correspondent ces résidus exceptionnels. Expliquer.
3. Discuter, en examinant l'histogramme de `res_y2x` puis en utilisant `qqnorm` s'il est raisonnable de considérer que ces résidus sont gaussiens.

3 Régression sur des données simulées

1. Formez un échantillon gaussien, noté `Residu` de même taille et de même moyenne et écart-type que `res_z2y`. Vérifier la longueur, la moyenne et l'écart type de l'échantillon gaussien simulé. Tracez son histogramme. Commentez.
2. Posez `zSim<-3*y+118+Residu`; expliquez le nom donné à ce nouvel échantillon.
3. Quelle est l'équation de la droite de régression des `zSim` sur les `y`?
4. Représentez le nuage des `(y,zSim)` et la droite de régression obtenue. Ajoutez la droite d'équation $z = 3y + 118$ au moyen de la commande `abline(c(118,3))`. Commentez le résultat obtenu.