

TP MISO: corrélation régression

guillemette.marot@univ-lille.fr

Dans ce TP, nous reprenons l'exemple du cours sur la liaison linéaire entre la fréquence cardiaque maximale et l'âge.

```
age=c(40,36,51,49,47, 51, 32, 55, 55, 23, 49,52,35)
fcm=c(187,195,180,190,185,183,195,185,189,201,189,185,195)
```

1 Corrélation

1. Tracer le nuage de points représentant la fréquence cardiaque maximale avec la fonction `plot`.
2. Calculer les corrélations de Pearson et de Spearman à l'aide de la fonction `cor`.

2 Régression linéaire

1. Retrouver l'équation de la droite de régression linéaire calculée en cours en utilisant la fonction `lm`:

```
model<-lm(fcm~age)
model
```

2. Retrouver les valeurs prédites et les résidus en regardant les noms de l'objet renvoyé précédemment. On pourra utiliser la fonction `names` et `$`.

Remarque: on peut retrouver ces valeurs en lançant les commandes suivantes:

```
pred2<-predict(model)
pred2
residuals<-fcm-pred2
residuals
```

3 Diagnostics régression

1. Commenter les lignes de code suivantes et interpréter leurs sorties:

```
mean(residuals)

qqnorm(residuals)
qqline(residuals)

res.student=rstudent(model)
plot(res.student~pred2,ylab="Résidus")
abline(h=c(-2,0,2),lty=c(2,1,2))
```

2. A l'aide de la fonction `summary`, retrouver le coefficient de détermination. On pourra vérifier qu'il est égal au carré du coefficient de corrélation de Pearson.
3. Calculer le PRESS en utilisant les fonctions `sum` et `lm.influence`:

```
PRESS <- 1/length(age)*(sum((residuals/(1 - lm.influence(model)$hat))^2))
PRESS
```

4. Evaluer l'influence des observations sur l'estimation des coefficients grâce à la distance de Cook (fonction `cooks.distance`). Faut-il retirer certaines observations?
5. Vérifier l'indépendance entre les résidus grâce au test de Durbin Watson. L'hypothèse nulle est-elle rejetée?

```
library(car)
durbinWatsonTest(model)
```