

```

### Exo1 ###

Basket = read.table(file = "NBA_data.csv", sep = ",", header = TRUE)
summary(Basket)
head(Basket)

attach(Basket)

plot(Points, Salary)

basket.lm <- lm(Salary ~ Points)
# On recupere une liste

summary(basket.lm) # synthese de la regression lineaire

names(basket.lm) # noms des elements de la liste

beta=basket.lm$coefficients # pour acceder aux coeff estimes

SCR=sum((Salary-(beta[1]+beta[2]*Points))^2) # somme des carrés des
résidus

n=nrow(Basket)
sigmahat=sqrt(SCR/(n-2))

abline(beta[1],beta[2],col="red")
# ou sinon
abline(basket.lm,col="red")

points(mean(Points),mean(Salary),pch=3,col="blue")
text(mean(Points)+5,mean(Salary),"centre de gravite",col="blue")

plot(basket.lm$residuals,ylab="residus")

Salaire15 =beta[1]+beta[2]*15

# Prediction du salaire pour 15 points
Data15=data.frame(15)
colnames(Data15)="Points"
predict.lm(basket.lm,newdata=Data15)
predict.lm(basket.lm,newdata=Data15,interval="prediction")
predict.lm(basket.lm,newdata=Data15,interval="confidence")
# on peut aussi utiliser predict() plus simplement

Points.seq <- data.frame(seq(0,30,0.1))
colnames(Points.seq)="Points"

Pred.seq <- predict.lm(basket.lm,newdata=Points.seq,interval =

```

```

"prediction")
basket.pred.max <- Pred.seq[,3]
basket.pred.min <- Pred.seq[,2]
Pred.conf <- predict.lm(basket.lm,newdata=Points.seq,interval =
"confidence")
basket.conf.max <- Pred.conf[,3]
basket.conf.min <- Pred.conf[,2]

```

```

plot(Points,Salary)
abline(basket.lm,col="red")

```

```

lines(seq(0,30,0.1),basket.pred.max,col = "blue",lty = 1)
lines(seq(0,30,0.1),basket.pred.min,col = "blue",lty = 1)
lines(seq(0,30,0.1),basket.conf.max,col = "blue",lty = 2)
lines(seq(0,30,0.1),basket.conf.min,col = "blue",lty = 2)

```

```
#####
```

```

# Name: Forced Expiratory Volume (FEV) Data
#Size: 654 observations, 5 variables

```

```

#Descriptive Abstract:
#Sample of 654 youths, aged 3 to 19, in the area of East #Boston
#during middle to late 1970's. Interest concerns the relationship
#between smoking and FEV.

```

```

DataFEV= read.table(file = "fev_data.csv",
                    col.names = c("age","fev","ht","sex","smoke"), # noms
de variables
                    colClasses=c(rep("numeric",3),"factor","factor") #
type des variables
)

```

```

summary(DataFEV)
head(DataFEV)
attach(DataFEV)

```

```

boxplot(fev ~ smoke)

```

```

# attention : pas d'observations pour age < 10 pour les fumeurs ...

```

```

mod1 <- lm(fev ~ age) # dans ce modele on ne tient pas compte de Smoke

```

(une seule droite donc)

```
summary(mod1)
```

```
plot(age, fev, cex = 0.1, xlab = "age", ylab = "fev")
```

```
# cex controle la taille des points
```

```
abline(mod1)
```

```
mod2 <- lm(fev ~ age + smoke) # dans ce modele les deux droites (pour  
smoke = 0 et Smoke = 1) sont //.
```

```
summary(mod2)
```

```
# smoke1 : indique l'ecart avec les deux droites.
```

```
# Pour Smoke = 0 : le coeff constant de la droite = Intercept
```

```
# Pour smoke = 1 : le coeff constant de la droite = Intercept + Smoke1
```

```
plot(age[smoke == "0"], fev[smoke == "0"], cex = 0.1, xlab = "age", ylab =  
"fev")
```

```
abline(mod2$coefficients[1], mod2$coefficients[2] )
```

```
points(age[smoke == "1"], fev[smoke == "1"], cex = 0.1, col = "red")
```

```
abline(mod2$coefficients[1] +
```

```
mod2$coefficients[3], mod2$coefficients[2], col = "red")
```

```
mod3 <- lm(fev ~ age + smoke + age:smoke) # Ce modele autorise deux  
droites (pour smoke = 0 et Smoke = 1) non necessairement //
```

```
# age:smoke variable d interaction entre smoke et age
```

```
summary(mod3)
```

```
plot(age[smoke == "0"], fev[smoke == "0"], cex = 0.1, xlab = "age", ylab =  
"fev")
```

```
abline(mod3$coefficients[1], mod3$coefficients[2] )
```

```
points(age[smoke == "1"], fev[smoke == "1"], cex = 0.1, col = "red")
```

```
abline(mod3$coefficients[1] + mod3$coefficients[3], mod3$coefficients[2] +  
mod3$coefficients[4], col = "red" )
```

```
# F test de mod2 contre mod3 :
```

```
anova(mod2, mod3)
```

```
# p -value tres faible, on a interet a garder mod3, il y a un effet  
significatif du tabagisme sur la capacité respiratoire.
```

```
# Pour préciser cet effet, on considère les valeurs coefficients estimés  
dans mod3.
```

```
# Dans le summary(mod3) on voit que (pente pour smoke = 1) - (pente  
pour smoke = 0) = "age:smoke1 -0.162703"
```

```
# Clairement il y a un effet nefaste du tabac sur la capacité  
respiratoire (pente + faible) car la fev progresse moins rapidement avec  
l'age lorsque Smoke =1.
```

```
# on peut aussi évaluer l'effet du tabagisme sur Fev sans tenir compte de  
l'interaction avec Age : cette fois on compare les deux modeles plus
```

```

simples mod1 mod2.
# F test de mod1 contre mod2 :
anova(mod1,mod2)
# encore une p value tres faible, il y a bien un effet du tabac sur FEV
# Pour préciser cet effet, on considère les valeurs des coefficients
  estimés dans mod2.
# dans summary(mod2) on voit que la droite pour smoke =0 est au dessus de
  la droite pour smoke = 1 "smoke1      -0.208995".
# La tabac diminue donc la capacité pulmonaire (sans tenir compte ici de
  l'interaction avec l'age)

```

```

# on pourrait complexifier encore les modèles en prenant en compte les
  autres variables disponibles dans la table.
# On pourrait refaire l'analyse pour les ages > 10.

```

```
#####
```

```

DataPolish = read.table(file = "Polish.csv",
                        col.names =
  c("Bol", "cass", "plat", "plateau", "diam", "temps", "prix"),
                        colClasses=c(rep("factor",4),rep("numeric",3)))

```

```

summary(DataPolish)
head(DataPolish)
attach(DataPolish)
par(mfrow=c(2,2))
plot(temps ~ diam)

```

```
Polish.lm <- lm(temps ~ diam)
```

```

plot(Polish.lm)
quartz()
Polish.lm <- lm(log(temps) ~ log(diam))

```

```
plot(Polish.lm)
```