

## GROUP 3

Thành viên

Nguyễn Thị Thùy Trang 1752048

Nguyễn Minh Nguyệt 1752038

Trương Thanh Trúc 1752051

Nguyễn Anh Khôi 1752024

Đặng Hữu Phước Vinh 1752052

Nguyễn Hoàng Long 1752028

Ex1.r

Admin

2020-05-25

```

EX1<-read.csv(file.choose())
Y<-EX1$FUEL.x_i5
X1<-EX1$TAX.x_i1
X2<-EX1$DLIC.x_i2
X3<-EX1$INC.x_i13
X4<-EX1$ROAD.x_i4

##Exercice1
##Q1
reg1<-lm(log(Y)~X1+X2+X3+X4)
step(reg1)

## Start:  AIC=-214.13
## log(Y) ~ X1 + X2 + X3 + X4
##
##           Df Sum of Sq      RSS      AIC
## - X4      1    0.00374 0.45387 -215.74
## <none>                0.45013 -214.13
## - X1      1    0.08004 0.53017 -208.28
## - X3      1    0.27337 0.72349 -193.35
## - X2      1    0.61391 1.06404 -174.84
##
## Step:  AIC=-215.74
## log(Y) ~ X1 + X2 + X3
##
##           Df Sum of Sq      RSS      AIC
## <none>                0.45387 -215.74
## - X1      1    0.09131 0.54518 -208.94
## - X3      1    0.28468 0.73855 -194.37
## - X2      1    0.69469 1.14855 -173.17
##
## Call:
## lm(formula = log(Y) ~ X1 + X2 + X3)
##
## Coefficients:
## (Intercept)          X1          X2          X3
##      5.97173      -0.04850      0.02322     -0.13764

```

On a des valeur AIC par fonction step()

Donc pour trouver le meilleur modèle, on choisit le moindre AIC,

c'est -215.74 (-x4)

Fin : meilleur modèle à l'aide de méthode stepwise c'est :

$y_{\text{chapeau}} = 5.97173 - 0.04850 \cdot X1_{\text{chapeau}} + 0.02322 \cdot X2_{\text{chapeau}} + 0.13764 \cdot X3_{\text{chapeau}}$

##Q2

##Etape1:

cor(Y,X1)

## [1] -0.4512803

cor(Y,X2)

## [1] 0.6989654

cor(Y,X3)

## [1] -0.2448621

cor(Y,X4)

## [1] 0.01912788

regX2<-lm(log(Y)~X2)

regX2

##

## Call:

## lm(formula = log(Y) ~ X2)

##

## Coefficients:

## (Intercept)           X2

##       5.00690       0.02338

Ici, on utilise fonction cor() de 4 variable.

On voit cor(Y, X2)= 0.6989654 est la plus grande valeur => choisissez le modèle :

$y_{\text{chapeau}} = 5.00690 + 0.02338 \cdot X2_{\text{chapeau}}$

##Etape2:

regresX2<-resid(regX2)

cor(regresX2,X1)

## [1] -0.3482288

```
cor(regresX2,X3)
```

```
## [1] -0.5920886
```

```
cor(regresX2,X4)
```

```
## [1] 0.09580514
```

```
regX2X3<-lm(log(Y)~X2+X3)
```

```
regX2X3
```

```
##
```

```
## Call:
```

```
## lm(formula = log(Y) ~ X2 + X3)
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)          X2          X3
```

```
##      5.47906      0.02569     -0.14241
```

C'est comme étape 1, utilise `resid(lm(log(Y)~X2))` avec 3 variable.

On voit `cor(resid(lm(log(Y)~X2)), X3) = -0.5920886` est la plus grande valeur => choisissez le modèle :

$\hat{Y}_{\text{chapeau}} = 5.47906 + 0.02569 \cdot X2_{\text{chapeau}} - 0.14241 \cdot X3_{\text{chapeau}}$

```
#Etape3:
```

```
regresX2X3<-resid(regX2X3)
```

```
cor(regresX2X3,X1)
```

```
## [1] -0.3911721
```

```
cor(regresX2X3,X4)
```

```
## [1] 0.165258
```

```
regX1X2X3<-lm(log(Y)~X1+X2+X3)
```

```
regX1X2X3
```

```
##
```

```
## Call:
```

```
## lm(formula = log(Y) ~ X1 + X2 + X3)
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)          X1          X2          X3
```

```
##      5.97173     -0.04850      0.02322     -0.13764
```

C'est comme étape 1, utilise `resid(lm(log(Y)~X2+X3))` avec 2 variable.

On voit `cor(resid(lm(log(Y)~X2+X3)), X1) = -0.5920886` est la plus grande valeur => choisissez le modèle :

$$Y_{\text{chapeau}} = 5.97173 - 0.04850 \cdot X1_{\text{chapeau}} + 0.02322 \cdot X2_{\text{chapeau}} - 0.13764 \cdot X3_{\text{chapeau}}$$

#Etape4:

```
regresX1X2X3<-resid(regX1X2X3)
```

```
cor(regresX1X2X3,X4) #khong chon vi cor(regresX1X2X3,X4) = -0.07419771 < 15%
```

```
## [1] -0.07419771
```

C'est comme étape 1, utilise `resid(lm(log(Y)~X1+X2+X3))` avec 1 (X4) variable.

On retrouve le résultat est -0.07419771 par fonction `cor(resid(lm(log(Y)~X1+X2+X3)), X4)`

Et  $-0.07419771 < 15\%$

DONC ON REJETTE X4

DONC, LE MEILLEUR MODELE À L'AIDE DE METHODE STAGewise, C'EST :

$$Y_{\text{chapeau}} = 5.97173 - 0.04850 \cdot X2_{\text{chapeau}} + 0.02322 \cdot X3_{\text{chapeau}} - 0.13764 \cdot X4_{\text{chapeau}}$$