

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

Ex2.r

Admin

2020-05-25

```
EX2<-read.csv(file.choose())
Y<-EX2$Azote
X1<-EX2$Agriculture
X2<-EX2$Forêts
X3<-EX2$Habitations
X4<-EX2$Com..et.ind.

##Exercice1
##Q1
library(leaps)

recherche.ex<-regsubsets(Y~X1+X2+X3+X4,int=T,nbest=1,
                          nvmax=10,method="exhaustive",really.big=T, data=EX2)
res_summary <-summary(recherche.ex)
res_summary

## Subset selection object
## Call: regsubsets.formula(Y ~ X1 + X2 + X3 + X4, int = T, nbest = 1,
##      nvmax = 10, method = "exhaustive", really.big = T, data = EX2)
## 4 Variables (and intercept)
##      Forced in Forced out
## X1      FALSE      FALSE
## X2      FALSE      FALSE
## X3      FALSE      FALSE
## X4      FALSE      FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##           X1  X2  X3  X4
## 1  ( 1 ) " " "*" " " " "
## 2  ( 1 ) " " "*" " " "*"
## 3  ( 1 ) "*" "*" " " "*"
## 4  ( 1 ) "*" "*" "*" "*"

```

Modèle groupe 1 : $B_0 + B_2.X_2 + \text{Epsilon}$

Modèle groupe 2 : $B_0 + B_2.X_2 + B_4.X_4 + \text{Epsilon}$

Modèle groupe 3 : $B_0 + B_1.X_1 + B_2.X_2 + B_4.X_4 + \text{Epsilon}$

Modèle groupe 4 : $B_0 + B_1.X_1 + B_2.X_2 + B_3.X_3 + B_4.X_4 + \text{Epsilon}$

On utilise `library(leaps)` pour appeler méthode "exhaustive" pour trouver le meilleurs modèle

```
res_summary$cp #Mallows Cp
```

```
## [1] 4.739604 1.818783 3.045633 5.000000
```

On surveille les valeur critère du Cp de Mallows

=> Choisir le modèle 3 : $B_0 + B_1.X_1 + B_2.X_2 + B_4.X_4 + \text{Epsilon}$ parce que $C_p = 3.045633$ est proche de $p = 4$

```
##Q2
```

```
step(lm(Y~X1+X2+X3+X4))
```

```
## Start: AIC=-48.89
```

```
## Y ~ X1 + X2 + X3 + X4
```

```
##
```

	Df	Sum of Sq	RSS	AIC
--	----	-----------	-----	-----

## - X3	1	0.003203	1.0559	-50.826
---------	---	----------	--------	---------

## - X1	1	0.010479	1.0632	-50.689
---------	---	----------	--------	---------

## - X2	1	0.060809	1.1135	-49.764
---------	---	----------	--------	---------

## <none>			1.0527	-48.887
-----------	--	--	--------	---------

## - X4	1	0.243326	1.2960	-46.728
---------	---	----------	--------	---------

```
##
```

```
## Step: AIC=-50.83
```

```
## Y ~ X1 + X2 + X4
```

```
##
```

	Df	Sum of Sq	RSS	AIC
--	----	-----------	-----	-----

## - X1	1	0.054261	1.1102	-51.824
---------	---	----------	--------	---------

## <none>			1.0559	-50.826
-----------	--	--	--------	---------

## - X2	1	0.130368	1.1863	-50.498
---------	---	----------	--------	---------

## - X4	1	0.289535	1.3455	-47.980
---------	---	----------	--------	---------

```
##
```

```
## Step: AIC=-51.82
```

```
## Y ~ X2 + X4
```

```
##
```

```
##           Df Sum of Sq    RSS    AIC
## <none>                1.1102 -51.824
## - X4      1    0.34535 1.4555 -48.407
## - X2      1    1.48521 2.5954 -36.840
##
## Call:
## lm(formula = Y ~ X2 + X4)
##
## Coefficients:
## (Intercept)          X2          X4
##      2.09621      -0.01648      0.18767
```

On utilise fonction step() pour trouver le meilleur modèle par critère du AIC

On voit le résultat :

Start : le plus faible c'est : -50.826 (rejette X3)

Step(X1, X2, X4) : le plus faible c'est : -51.824 (rejette X1)

Step(X2, X4) : le plus faible c'est : -51.824 (rejette none)

=> Le meilleur modele : $Y_{\text{chapeau}} = 2.09621 - 0.0164 \cdot X2_{\text{chapeau}} - 0.18767 \cdot X4_{\text{chapeau}}$