

Première Année Master M.A.E.F. 2017 – 2018

Econométrie II

Contrôle continu n°2, avril 2018

Examen de 1h00. Tout document ou calculatrice est interdit.

1. On lance les commandes suivantes:

```
x1=3*rexp(30); x2=c(1:30); x3=cos(2*pi/4*x2)
x4=x2^2; x5=log(x2); x6=log(x4)
epsi=c(6*rnorm(20),2*rnorm(10))
y=-5+2*x1+7*x3-4*x5+epsi
ylm1=lm(y~x1+x2+x3+x4+x5+x6); summary(ylm1)
```

Voici les résultats numériques obtenus:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.75979	3.43984	-0.512	0.614
x1	1.94213	0.16387	11.852	1.62e-11 ***
x2	-0.62478	1.13383	-0.551	0.587
x3	6.75746	1.14496	5.902	4.34e-06 ***
x4	0.01227	0.02388	0.514	0.612
x5	-2.81998	4.84202	-0.582	0.566
x6	NA	NA	NA	NA

Residual standard error: 4.371 on 24 degrees of freedom
Multiple R-squared: 0.9097, Adjusted R-squared: 0.8909
F-statistic: 48.35 on 5 and 24 DF, p-value: 9.378e-12

Question 1: Les vecteurs y, x_1, \dots, x_6 sont supposés observés. Ecrire formellement le modèle suivi par y (modèle que l'on suppose inconnu), en précisant la loi du bruit nommé ϵ . Expliquer les NA apparaissant dans les résultats. Comment le logiciel a calculé explicitement la valeur 0.612 à partir de y, x_1, \dots, x_6 ? Que peut-on conclure quant à la régression effectuée? (6pts)

2. On poursuit avec les commandes qui suivent:

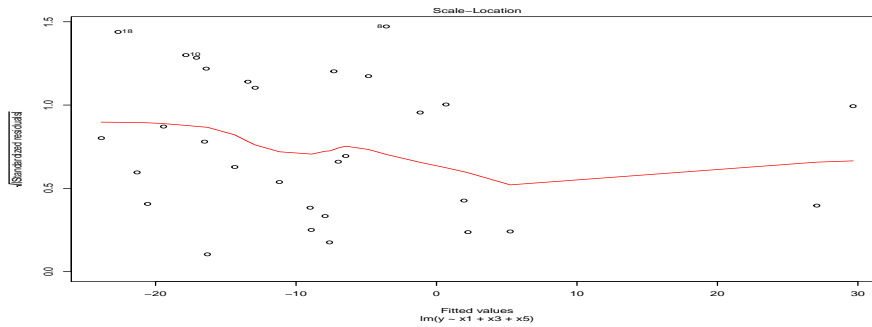
```
library(MASS)
ylm2=stepAIC(ylm1,k=log(30),direction="both", trace=FALSE)
summary(ylm2); plot(ylm2,3)
```

Ci-dessous se trouvent les résultats numériques et le graphe obtenus:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.9723	2.5227	-0.385	0.703
x1	1.9459	0.1530	12.714	1.15e-12 ***
x3	6.7564	1.0982	6.152	1.66e-06 ***
x5	-5.4796	0.9257	-5.919	3.03e-06 ***

Residual standard error: 4.228 on 26 degrees of freedom
Multiple R-squared: 0.9085, Adjusted R-squared: 0.8979
F-statistic: 86.04 on 3 and 26 DF, p-value: 1.263e-13



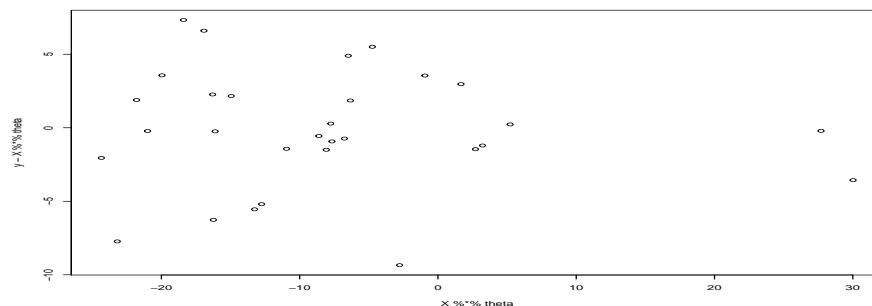
Question 2: Qu'a-t-on fait et qu'obtient-on? Est-on désormais satisfait? La valeur numérique -0.9723 vous semble-t-elle convenable? (3pts)

3. Soit maintenant les commandes suivantes:

```
sigma1=mean((y[1:20]-ylm2$fit[1:20])^2)
sigma2=mean((y[21:30]-ylm2$fit[21:30])^2)
var=c(rep(1/sigma1,20),rep(1/sigma2,10))
Sig=diag(var,30); X=cbind(rep(1,30),x1,x3,x5)
theta=solve(t(X)%*%Sig%*%X)%*%t(X)%*%Sig%*%y
plot(X%*%theta,y-X%*%theta)
```

Les résultats numériques et graphiques sont les suivants:

```
> theta
-1.057484
x1  1.964777
x3  7.445878
x5 -5.394425
```



Question 3: Expliquer ce qui a été fait et pourquoi. A-t-on gagné par rapport à la précédente régression? (5pts)

4. Enfin on continue avec les commandes:

```
ylm41=lm(I(y[1:20])~I(x1[1:20])+I(x3[1:20])+I(x5[1:20]))
ylm42=lm(I(y[21:30])~I(x1[21:30])+I(x3[21:30])+I(x5[21:30]))

BIC1=30*log(mean(ylm1$residuals^2))+30*log(30)*7; BIC1
BIC2=30*log(mean(ylm2$residuals^2))+30*log(30)*5; BIC2
BIC3=20*log(sigma1)+10*log(sigma2)+t(y-X%*%theta)%*%Sig%*(y-X%*%theta)+log(30)*6; BIC3
BIC4=20*log(mean(ylm41$residuals^2))+10*log(mean(ylm42$residuals^2))+20+10+log(30)*10; BIC4
```

Les résultats numériques sont 135.6199, 129.2109, 122.495, 134.3327.

Question 4: Expliquer d'abord ce qui est fait avec ylm41 et ylm42. Détailler formellement ce qui a conduit au calcul de BIC1, BIC3 et BIC4. Vis-à-vis des résultats obtenus, quel modèle choisiriez vous? Pourquoi cela semble logique? (7pts)