Université Paris I, Panthéon - Sorbonne

Deuxième Année Master T.I.D.E. 2014 – 2015

Econométrie des modèles linéaires

Examen final, janvier 2015

Examen de 2h00. Tout document ou calculatrice est interdit.

1. **(10 points)**

   (a) On a tapé les commandes suivantes avec le logiciel SAS:

```
data sasuser.TableExam;
do i=1 to 100;
x1=1; x2=-3*rannor(-1); x3=x2**2; eps=rannor(-1);
y=5*x1-3*x2+2*x3+20*eps;
output; end;
keep x1 x2 x3 y;
run;
```

   On rappelle ici que la commande `rannor(-1)` produit une réalisation d'une variable aléatoire indépendante de tout le reste et de loi $\mathcal{N}(0,1)$. Les réalisations des variables y, $x1$, $x2$ et $x3$ seront supposées connues, alors que celle de la variables *eps* seront inconnues.
   Voici le début de la table produite:

x1	x2	x3	y
1	-0.38369	0.1472	22.494
1	4.48510	20.1161	35.675
1	6.00487	36.0585	42.549
1	-0.29442	0.0867	26.384
1	1.40338	1.9695	-9.390
⋮	⋮	⋮	⋮

   Questions I.1: quelles sont les lois des variables $x2$, $x3$ et y? Ecriver le modèle vérifié par y sous forme matricielle $Y = X\,\theta + \varepsilon$, en précisant la matrice X et le vecteur θ.

   (b) On fait maintenant comme si θ était inconnu. On tape ensuite les commandes:

```
proc reg data=sasuser.TableExam;
model y=x1-x3;
run;
```

   Sur la page 4, on peut voir une sélection des résultats obtenus.

   Questions I.2: Qu'a-t-on fait en tapant ces commandes? Que représentent précisément les valeurs numériques 36.95, -3.38, 0.0046. *Pourquoi a-t-on un* 0 *à la suite de $x1$? Quelles sont les valeurs estimées de θ? Est-ce conforme à ce que l'on attendait?*

   Questions I.3: Calculer uniquement en fonction de X la valeur de la matrice de covariance Σ de $\widehat{\theta}$. Si on note $x1 = (x1_i)$, $x2 = (x2_i)$ et $x3 = (x3_i)$, écrire Σ en fonction des xj_i. Déterminer $\mathbb{E}(xj_i \times xk_i)$ pour $1 \leq j \leq k \leq 3$. En déduire en justifiant que

$$\frac{1}{100}\,{}^t X\,X \simeq \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

   En déduire si la valeur numérique 0.22550 *était prévisible. Par le même raisonnement, peut-on en déduire que $\widehat{\theta}$ est un estimateur convergent de θ?*

   Questions I.4: Que pensez-vous de la valeur du R^2, des tests de Fisher et de student? Le modèle linéaire obtenu est-il acceptable?

2. **(7 points)** On s'intéresse à des données socio-économiques concernant le taux de criminalité dans les différents états des USA en 1960. On désire connaître l'influence de certaines variables sur ce taux de criminalité. La variable à expliquer, Y_i, est le nombre moyen de crimes pour 1 million d'habitants dans l'état i. Les autres variables potentiellement influentes sont :

   - la moyenne de la durée de scolarité ($\times 10$) : $Z1$;
   - le budget de la police par habitant de l'état : $Z2$;
   - le nombre d'actifs pour 1000 hommes âgés de 14 à 24 ans : $Z3$;
   - le nombre d'hommes pour 1000 femmes : $Z4$;
   - la population de l'état (unité: 100000) : $Z5$;
   - le nombre de non blancs pour 1000 habitants : $Z6$;
   - le nombre de chômeurs pour 1000 habitants de 14 à 24 ans : $Z7$;
   - le nombre de chômeurs pour 1000 habitants de 35 à 39 ans : $Z8$;
   - le revenu médian des familles en dizaines de dollars : $Z9$;
   - le taux (/1000) de familles en dessous du niveau de pauvreté : $Z10$.

   Voici un aperçu de ce tableau de données:

```
Crimes
       Y  Z1  Z2  Z3   Z4  Z5  Z6  Z7 Z8  Z9 Z10
1   79.1  91  58 510  950  33 301 108 41 394 261
2  163.5 113 103 583 1012  13 102  96 36 557 194
3   57.8  89  45 533  969  18 219  94 33 318 250
.............................................
```

   (a) On tape alors les commandes suivantes:

```
proc reg data=sasuser.Crimes;
model Y=Z1-Z10;
run; quit;
```

   Les résultats de ces commandes sont présentes en pages 5 et 6.

   Questions II.1: Quelles conclusions pouvez vous tirer de ces résultats?

   (b) On tape ensuite les commandes:

```
proc reg data=sasuser.Crimes;
model Y=Z1-Z10/selection = rsquare bic best=1;
run; quit;
proc reg data=sasuser.Crimes;
model Y=Z1 Z2 Z10;
run; quit;
```

   Les résultats de ces commandes sont présentés en page 7.

   Questions II.2: Quelles conclusions pouvez vous tirer de ces résultats?

   (c) On tape enfin les commandes:

```
proc transreg details data=sasuser.Crimes ss2
              plots=(transformation(dependent) obp);
   model BoxCox(Y / convenient lambda=-3 to 3 by 0.05) =
         qpoint(Z1 Z2 Z10);
run;
data sasuser.Crimes;
set sasuser.Crimes;
YY=log(Y);
Z12=exp(log(Z1)+log(Z2)); Z110=exp(log(Z1)+log(Z10)); Z210=exp(log(Z2)+log(Z10));
Z1Z1=Z1**2; Z2Z2=Z2**2; Z10Z10=Z10**2;
run;
proc reg data=sasuser.Crimes;
model YY=Z1 Z2 Z10 Z1Z1 Z2Z2 Z10Z10 Z12 Z210 Z110/selection = backward;
run; quit;
```

Une partie des résultats de ces commandes est présentée en page 8 et 9.

Questions II.3: Qu'a-t-on fait? Quelles conclusions pouvez vous tirer de ces résultats? Est-on satisfait?

3. **(5 points)** On dispose d'une table des données présentant les revenus mensuels (variable Y) de 80 adultes en fonction du sexe (variable $X1$, valant 1: homme, ou 0: femme), du lieu de travail (variable $X2$, valant 1: Ile de France, 2: agglomération de plus de 100000 habitants, ou 3: agglomération de moins de 100000 habitants) et de l'âge (variable $X3$, de 25 à 65 ans).
On a tapé les commandes suivantes:

```
proc glmselect data=sasuser.Revenu15;
class X1 X2;
model Y=X1| X2| X3 /select = bic;
run; quit;
```

Une partie des résultats apparaît en page 10.

Questions III.1: Qu'a-t-on fait avec ces commandes? A quel modèle aboutit-on? Si on considère un homme de 50 ans, dans une agglomération de plus de 100000 habitants, quel sera approximativement la prédiction de ses revenus mensuels?

Le Système SAS

Procédure REG
Modèle : MODEL1
Variable dépendante : y

Nb d'observations lues	100
Nb d'obs. utilisées	100

Analysis of Variance

Source	DDL	Sum of Squares	Mean Square	Valeur F	Pr > F
Modèle	2	30481	15240	36.95	<.0001
Erreur	97	40014	412.51337		
Total sommes corrigées	99	70495			

Root MSE	20.31043	**R carré**	0.4324
Moyenne dépendante	20.62843	**R car. ajust.**	0.4207
Coeff Var	98.45841		

Note: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

Note: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

x1 = Intercept

Parameter Estimates

Variable	DDL	Parameter Estimate	Standard Error	Valeur du test t	Pr > |t|
Intercept	**B**	7.59486	2.61875	2.90	0.0046
x1	**0**	0	.	.	.
x2	**1**	-2.53539	0.75067	-3.38	0.0011
x3	**1**	1.75223	0.22550	7.77	<.0001

Le Système SAS

Procédure REG
Modèle : MODEL1
Variable dépendante : Y

Nb d'observations lues 47

Nb d'obs. utilisées 47

Analysis of Variance

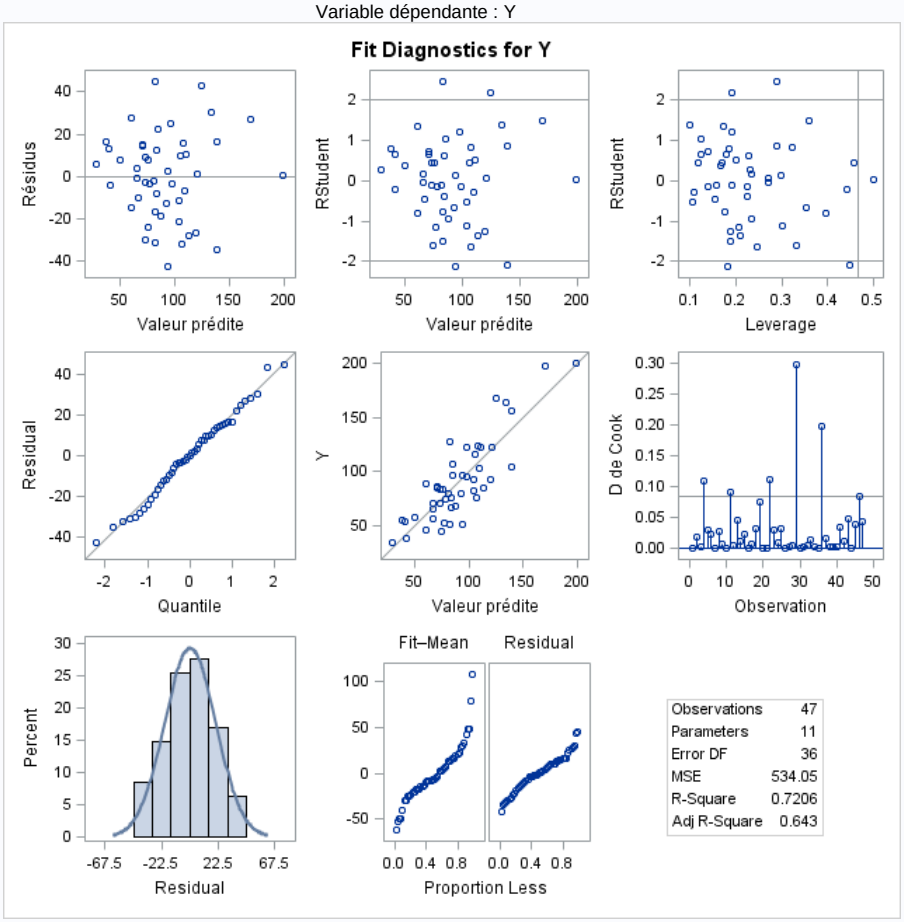
Source	DDL	Sum of Squares	Mean Square	Valeur F	Pr > F
Modèle	10	49583	4958.34360	9.28	<.0001
Erreur	36	19226	534.05113		
Total sommes corrigées	46	68809			

Root MSE	23.10955	**R carré**	0.7206
Moyenne dépendante	90.50851	**R car. ajust.**	0.6430
Coeff Var	25.53301		

Parameter Estimates

Variable	DDL	Parameter Estimate	Standard Error	Valeur du test t	Pr > |t|
Intercept	**1**	-612.43310	157.98962	-3.88	0.0004
Z1	**1**	1.51128	0.66239	2.28	0.0285
Z2	**1**	0.95895	0.25291	3.79	0.0006
Z3	**1**	-0.03201	0.13213	-0.24	0.8100
Z4	**1**	0.29730	0.20623	1.44	0.1581
Z5	**1**	-0.03726	0.13598	-0.27	0.7856
Z6	**1**	0.03618	0.05370	0.67	0.5048
Z7	**1**	-0.63264	0.42395	-1.49	0.1443
Z8	**1**	1.43268	0.87375	1.64	0.1098
Z9	**1**	0.09805	0.10913	0.90	0.3749
Z10	**1**	0.74950	0.23028	3.25	0.0025

Procédure REG
Modèle : MODEL1
Variable dépendante : Y



Procédure REG

Nb d'observations lues	47
Nb d'obs. utilisées	47

Number in Model	R-Square	BIC	Variables in Model
1	**0.4728**	317.0701	Z2
2	**0.5803**	308.9533	Z2 Z10
3	**0.6656**	301.8024	Z1 Z2 Z10
4	**0.6826**	302.0873	Z1 Z2 Z4 Z10
5	**0.6928**	303.2911	Z1 Z2 Z4 Z9 Z10
6	**0.7117**	303.6926	Z1 Z2 Z4 Z7 Z8 Z10
7	**0.7155**	305.8537	Z1 Z2 Z4 Z7 Z8 Z9 Z10
8	**0.7192**	308.0861	Z1 Z2 Z4 Z6 Z7 Z8 Z9 Z10
9	**0.7201**	310.6126	Z1 Z2 Z4 Z5 Z6 Z7 Z8 Z9 Z10
10	**0.7206**	313.1871	Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10

Procédure REG

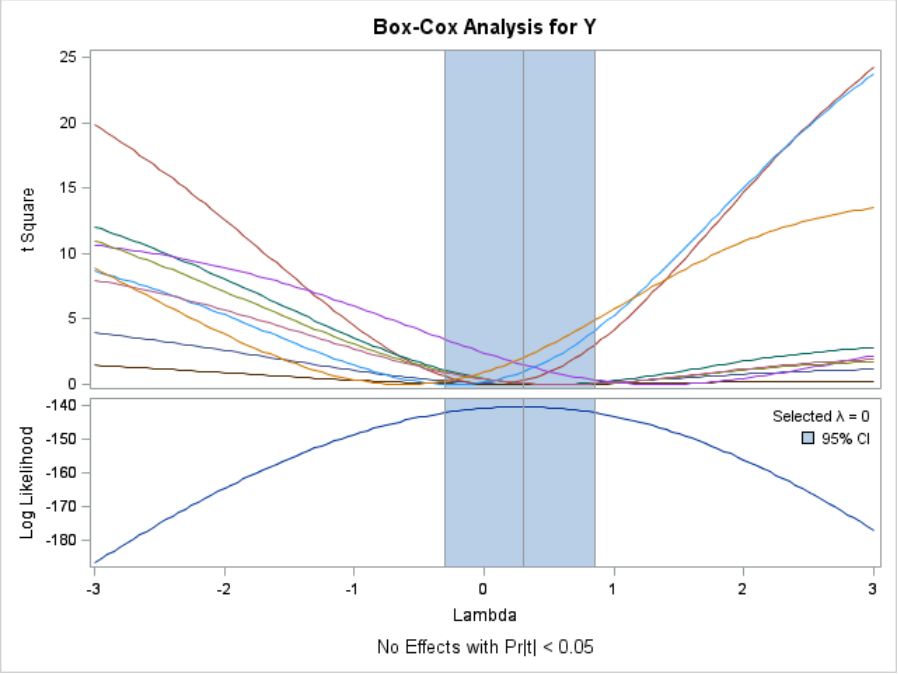
Analysis of Variance

Source	DDL	Sum of Squares	Mean Square	Valeur F	Pr > F
Modèle	3	45802	15267	28.53	<.0001
Erreur	43	23008	535.05987		
Total sommes corrigées	46	68809			

Root MSE	23.13136	**R carré**	0.6656
Moyenne dépendante	90.50851	**R car. ajust.**	0.6423
Coeff Var	25.55711		

Parameter Estimates

Variable	DDL	Parameter Estimate	Standard Error	Valeur du test t	Pr > |t|
Intercept	**1**	-327.54088	76.91367	-4.26	0.0001
Z1	**1**	1.57869	0.47661	3.31	0.0019
Z2	**1**	1.24314	0.14785	8.41	<.0001
Z10	**1**	0.75058	0.15077	4.98	<.0001


Box-Cox Analysis for Y
t Square
25
20
15
10
5
0
Log Likelihood
-140
-150
-160
-170
-180
Selected λ = 0
95% CI
-3
-2
-1
0
1
2
3
Lambda
No Effects with Pr|t| < 0.05

Le Système SAS

Sélection descendante : Etape 6

Variable Z110 supprimée : R carré = 0.7029 et C(p) = 1.0387

Analysis of Variance

Source	DDL	Sum of Squares	Mean Square	Valeur F	Pr > F
Modèle	3	5.46303	1.82101	33.90	<.0001
Erreur	43	2.30958	0.05371		
Total sommes corrigées	46	7.77261			

Variable	Parameter Estimate	Standard Error	Type II SS	Valeur F	Pr > F
Intercept	1.54497	0.48666	0.54131	10.08	0.0028
Z1	0.01266	0.00375	0.61256	11.40	0.0016
Z2Z2	-0.00003073	0.00001236	0.33219	6.18	0.0168
Z210	0.00011349	0.00001669	2.48477	46.26	<.0001

Class Level Information

Class	Levels	Values
X1	2	0 1
X2	3	1 2 3

Dimensions

Number of Effects	6
Number of Parameters	12

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 1).

Effects: Intercept X3*X1*X2

Analysis of Variance

Source	DDL	Sum of Squares	Mean Square	Valeur F
Model	6	30128451	5021408	75.54
Error	73	4852692	66475	
Corrected Total	79	34981143		

R-Square 0.8613

Parameter Estimates

Parameter	DDL	Valeur estimée	Erreur type	Valeur du test t
Intercept	1	606.834864	123.730639	4.90
X3*X1*X2 0 1	1	33.437027	3.351998	9.98
X3*X1*X2 0 2	1	23.730412	2.753387	8.62
X3*X1*X2 0 3	1	14.458537	2.954883	4.89
X3*X1*X2 1 1	1	51.066582	3.139925	16.26
X3*X1*X2 1 2	1	41.120514	3.169372	12.97
X3*X1*X2 1 3	1	31.318611	3.156557	9.92