

Deuxième Année Master T.I.D.E. 2016 – 2017

Econométrie des modèles linéaires

Examen final, janvier 2017

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

On étudie l'espérance de vie annuelle en France de 1816 à 2014 en fonction de l'année, du sexe (femme ou homme) et du fait qu'il y ait ou non la guerre sur le territoire français. On travaille ainsi avec la table `sasuser.Espvie`, dont voici un aperçu

An	Esp	Sexe	Guerre
1816	41.07	1	0
1817	40.19	1	0
1818	39.06	1	0
1819	37.90	1	0
⋮	⋮	⋮	⋮
1917	52.01	1	1
⋮	⋮	⋮	⋮
2014	85.40	1	0
1816	39.03	0	0
1817	38.21	0	0
1818	38.08	0	0
⋮	⋮	⋮	⋮

et on essaye de mettre en place un modèle permettant de modéliser au mieux cette espérance de vie. Par la suite on notera $E_{k,i}$ l'espérance de vie pour $k = 1$ (femme) ou 2 (homme) et $i = 1, \dots, n$, et on notera $F = {}^t(E_{1,1}, \dots, E_{1,n}, E_{2,1}, \dots, E_{2,n})$

1. On a tout d'abord tapé les commandes suivantes avec le logiciel SAS:

```
proc glm data=sasuser.Espvie;
class Sexe;
model Esp=Sexe;
means Sexe;
run;
```

Questions I.1: Expliquer ce qui a été réalisé (voir aussi les résultats proposés page 3). Ecrivez à l'aide de F le modèle vectoriel qui a été mis en place. Que vaut n et expliquer pourquoi? Des paramètres et une variance ont été estimés dans ce modèle (donner leurs valeurs numériques). Donner leur écriture formelle (en fonction des $E_{k,i}$). Expliquer pourquoi si le modèle est vrai alors les coefficients estimés convergent et pourquoi ils sont approximativement gaussiens. Quelles conclusions obtenez vous?

2. On effectue alors une autre procédure:

```
proc reg data=sasuser.Espvie;
model Esp=An;
run;
```

Sur les pages 4 et 5, on peut voir une sélection des résultats obtenus.

Questions I.2: Qu'a-t-on réalisé avec ces commandes? Donner le modèle théorique sous-jacent en fonction de F . Si on considère la variable `An2` telle que `An2=An-1815` à la place de la variable `An2`, que valent approximativement les coefficients estimés de ce nouveau modèle? Pour ce nouveau modèle, préciser la matrice X du modèle et montrer que $({}^tX X)^{-1} = \frac{1}{n(n^2-1)} \begin{pmatrix} (n+1)(2n+1) & -3(n+1) \\ -3(n+1) & 6 \end{pmatrix}$ pour tout $n \in \mathbf{N}^$ (on rappelle que $\sum_{i=1}^k i^2 = \frac{1}{6}(2k+1)(k+1)k$). En déduire que les deux paramètres estimés sont convergents. Que pensez-vous des résultats numériques et graphes obtenus? A la page 5 se trouve également le résultat d'une transformation de Box-Cox. Conclusion?*

3. On introduit désormais la variable `Guerre` dans le modèle:

```
proc glm data=sasuser.Espvie;  
class Sexe Guerre;  
model Esp=Guerre|Sexe;  
means Sexe*Guerre;  
run; quit;
```

Questions I.3: Ecrire formellement le modèle proposé. Que pensez vous des résultats obtenus page 6 et 7? Avait-on besoin de demander des tests de comparaison entre modalités?

4. Après plusieurs essais, on en arrive aux commandes suivantes:

```
proc glm data=sasuser.Espvie;  
class Sexe Guerre;  
model Esp=Guerre*Sexe Sexe*An;  
run; quit;
```

Questions I.4: Ecrire formellement le modèle proposé. Combien de paramètres doit-on estimer? Que pensez vous des résultats obtenus page 8? A votre avis, quel serait à peu près l'espérance de vie d'une française en 2100?

Le Système SAS

The GLM Procedure

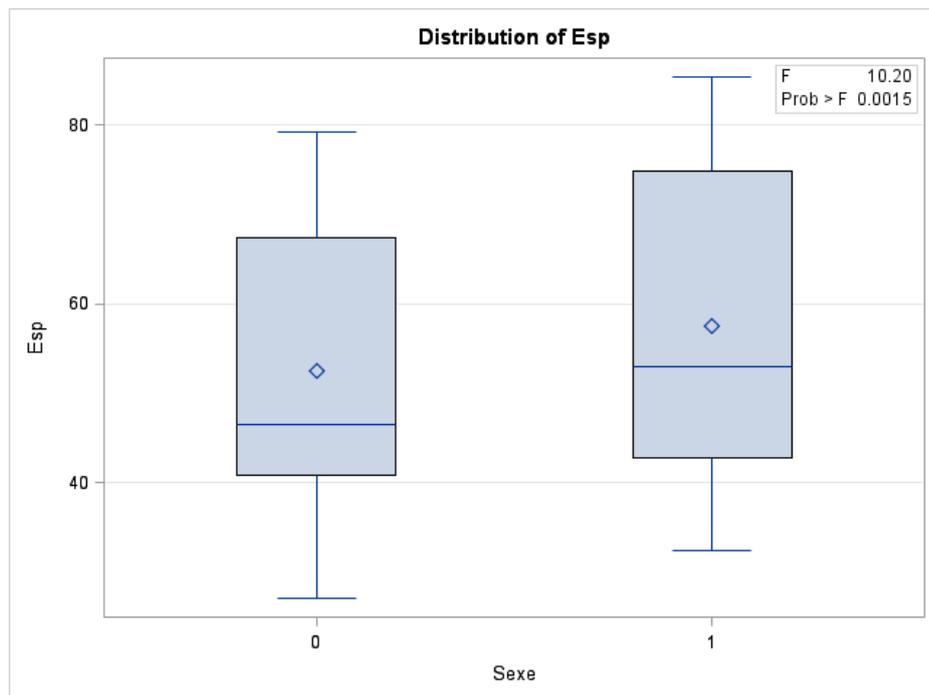
Dependent Variable: Esp

Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Model	1	2406.91347	2406.91347	10.20	0.0015
Error	396	93443.79084	235.96917		
Corrected Total	397	95850.70431			

R-Square Coeff Var Root MSE Esp Mean
 0.025111 27.91071 15.36129 55.03726

Source	DDL	Type I SS	Moyenne quadratique	Valeur F	Pr > F
Sexe	1	2406.913474	2406.913474	10.20	0.0015

Source	DDL	Type III SS	Moyenne quadratique	Valeur F	Pr > F
Sexe	1	2406.913474	2406.913474	10.20	0.0015



Level of Sexe	N	Esp Mean	Std Dev
0	199	52.5780905	14.3933585
1	199	57.4964322	16.2717414

Le Système SAS

Procédure REG
 Modèle : MODEL1
 Variable dépendante : Esp
Nb d'observations lues 398
Nb d'obs. utilisées 398

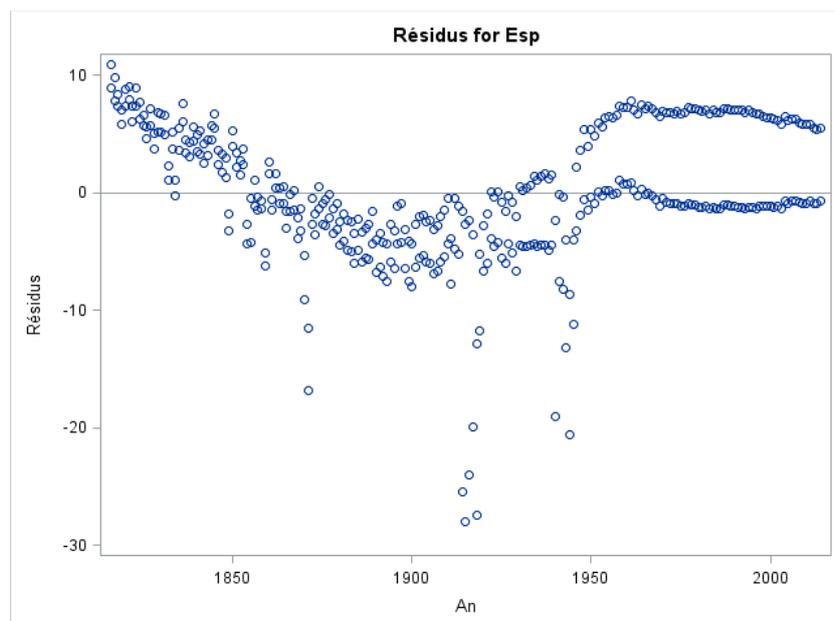
Analysis of Variance

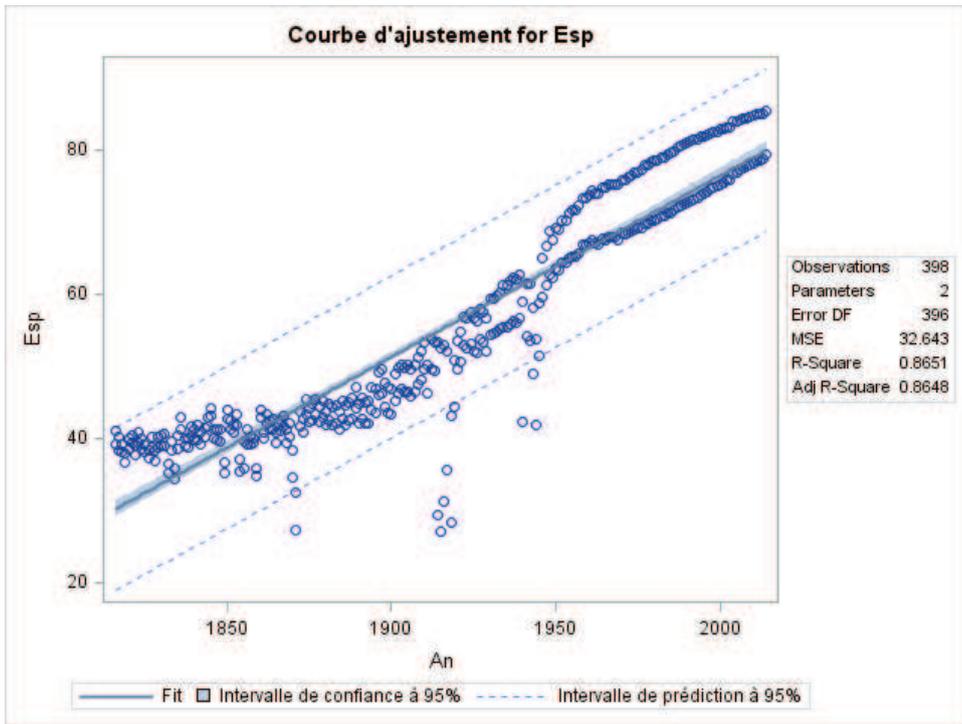
Source	DDL	Sum of Squares	Mean Square	Valeur F	Pr > F
Modèle	1	82924	82924	2540.29	<.0001
Erreur	396	12927	32.64342		
Total sommes corrigées	397	95851			

Root MSE 5.71344 **R carré** 0.8651
Moyenne dépendante 55.03726 **R car. ajust.** 0.8648
Coeff Var 10.38104

Parameter Estimates

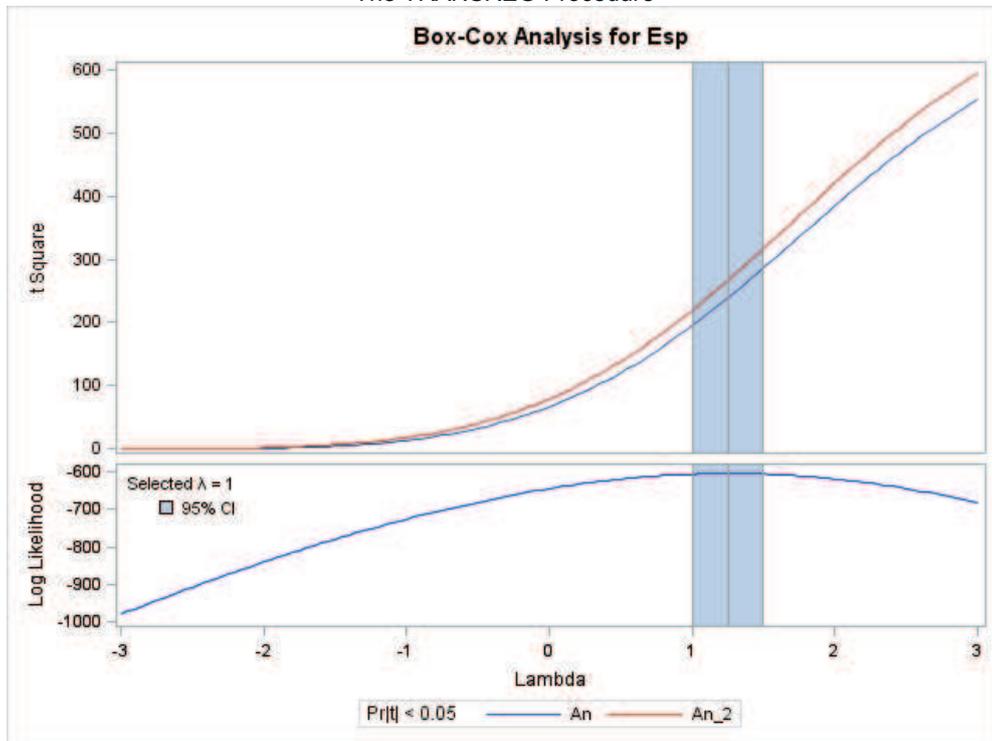
Variable	DDL	Parameter Estimate	Standard Error	Valeur du test t	Pr > t
Intercept	1	-426.14561	9.55132	-44.62	<.0001
An	1	0.25127	0.00499	50.40	<.0001





Le Système SAS

The TRANSREG Procedure



The GLM Procedure

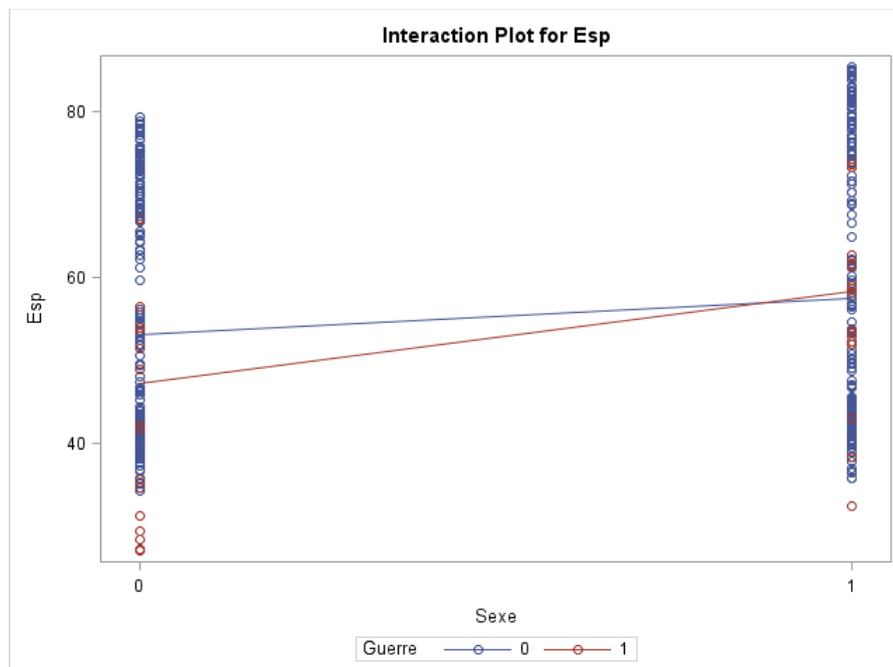
Dependent Variable: Esp

Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Model	3	3025.67030	1008.55677	4.28	0.0054
Error	394	92825.03402	235.59653		
Corrected Total	397	95850.70431			

R-Square Coeff Var Root MSE Esp Mean
 0.031566 27.88866 15.34915 55.03726

Source	DDL	Type I SS	Moyenne quadratique	Valeur F	Pr > F
Guerre	1	219.531032	219.531032	0.93	0.3350
Sexe	1	2406.913474	2406.913474	10.22	0.0015
Sexe*Guerre	1	399.225793	399.225793	1.69	0.1938

Source	DDL	Type III SS	Moyenne quadratique	Valeur F	Pr > F
Guerre	1	219.531032	219.531032	0.93	0.3350
Sexe	1	2025.022923	2025.022923	8.60	0.0036
Sexe*Guerre	1	399.225793	399.225793	1.69	0.1938



 Le Système SAS

The GLM Procedure

Tests t (LSD) pour Esp

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Degrés de liberté de l'erreur	394
Erreur quadratique moyenne	235.5965
Valeur critique de t	1.96600
Plus petite différence significative	3.0252

**Means with the same letter
are not significantly different.**

t Grouping	Mean	N	Sexe
A	57.496	199	1
B	52.578	199	0

Tests t (LSD) pour Esp

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Degrés de liberté de l'erreur	394
Erreur quadratique moyenne	235.5965
Valeur critique de t	1.96600
Plus petite différence significative	5.1472
Moyenne harmonique des tailles de cellules	68.74372

Note: Cell sizes are not equal.

**Means with the same letter
are not significantly different.**

t Grouping	Mean	N	Guerre
A	55.279	360	0
A			
A	52.751	38	1

Le Système SAS

The GLM Procedure

Dependent Variable: Esp

Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Model	5	88103.86620	17620.77324	891.63	<.0001
Error	392	7746.83812	19.76234		
Corrected Total	397	95850.70431			

R-Square Coeff Var Root MSE Esp Mean
 0.919178 8.077229 4.445486 55.03726

Source	DDL	Type I SS	Moyenne quadratique	Valeur F	Pr > F
Sexe*Guerre	3	3025.67030	1008.55677	51.03	<.0001
An*Sexe	2	85078.19590	42539.09795	2152.53	<.0001

Source	DDL	Type III SS	Moyenne quadratique	Valeur F	Pr > F
Sexe*Guerre	3	2703.61409	901.20470	45.60	<.0001
An*Sexe	2	85078.19590	42539.09795	2152.53	<.0001

