

# Université d'Orléans - Licence Economie et Gestion

## Introduction à la Statistique et à l'Econométrie

C. Hurlin. Examen Janvier 2005

**Exercice 1** Barème : 12 points.

Une entreprise spécialisée dans la production de confiseries s'engage à respecter certaines normes de fabrication concernant le pourcentage de colorant contenu dans ses produits. Ce pourcentage de colorant est représenté par une variable aléatoire  $X$  supposée suivre une loi normale  $N(m, \sigma^2)$  d'écart type  $\sigma$  connu et égal à 1. A la sortie de la chaîne de fabrication on tire un échantillon aléatoire et indépendant de  $N = 9$  mesures en pourcentage :

$$2.5 \quad 3.0 \quad 3.2 \quad 3.6 \quad 2.0 \quad 2.1 \quad 1.5 \quad 1.2 \quad 1.0 \quad (1)$$

Si la moyenne  $m$  du pourcentage de colorant est supérieure ou égal à 3, la marchandise est déclarée non conforme et détruite. Le fabricant souhaite effectuer le test suivant :

$$H_0 : m = 2$$

$$H_1 : m = 3$$

### Partie I : Test d'Hypothèses simples (8.5 points)

**Question 1 (1.5 points)** Calculez une estimation ponctuelle et une estimation par intervalle de confiance de niveau 95% de la moyenne  $m$ . Vous détaillerez la construction de l'intervalle de confiance.

**Question 2 (1 point)** Commentez le choix du fabricant de l'hypothèse nulle  $H_0$ .

**Question 3 (1 point)** Montrez que le rapport des vraisemblances de l'échantillon sous les hypothèses nulle et alternative peut s'écrire sous la forme suivante :

$$\frac{L(X_N, m_0)}{L(X_N, m_1)} = \frac{L(X_N, 2)}{L(X_N, 3)} = \exp \left[ \frac{1}{2} \left( \sum_{i=1}^N (X_i - 3)^2 - \sum_{i=1}^N (X_i - 2)^2 \right) \right] \quad (2)$$

**Question 4 (1 point)** A partir du résultat de la question 3, démontrez que la région critique  $W$  du test optimal au sens de Neyman Pearson est de la forme :

$$W = \{ X_N \mid \bar{X} > A \} \quad (3)$$

avec  $\bar{X} = (1/N) \sum_{i=1}^N X_i$  et où  $A$  est une constante déterminée par le niveau de risque de première espèce.

**Question 5 (1 point)** En admettant le résultat de la question 4, (i) construire la région critique  $W$  du test optimal pour un niveau de risque de première espèce  $\alpha = 1\%$ . (ii) Quelle décision doit prendre le fabricant dans ce cas ?

**Question 4 (1.5 points)** Pour un seuil de risque de première espèce  $\alpha = 1\%$ , calculez le risque de deuxième espèce et la puissance du test. Interprétez économiquement ces deux grandeurs.

**Question 5 (1.5 points)** Si le seuil reste fixé à 1%, quelle doit être la taille minimale de l'échantillon  $N$  pour que la puissance soit supérieure à 95%. En déduire une solution pour que le test du fabricant soit plus acceptable pour une association de consommateur.

**Partie II : Test Hypothèse simple contre hypothèse multiple (3.5 points)**

On considère à présent le test suivant :

$$H_0 : m = 2$$

$$H_1 : m > 2$$

**Question 1 (1 point)** En reprenant vos résultats concernant la région critique du test  $H_0 : m = 2$  contre  $H_1 : m = 3$ , justifiez l'existence d'un test UPP pour  $H_0 : m = 2$  contre  $H_1 : m > 2$ . Montrez que la région critique du test UPP pour un niveau de risque de première espèce  $\alpha = 1\%$  est :

$$W = \{X_N | \bar{X} > 2.7754\} \quad (4)$$

**Question 2 (2.5 point)** (i) Calculez les valeurs prises par la courbe de puissance ou courbe d'efficacité du test pour les valeurs suivantes de la moyenne  $m$

$$m = 2 \quad m = 2.5 \quad m = 3 \quad m = 5 \quad m = 10 \quad m = 50 \quad (5)$$

(ii) Commentez ces valeurs et (iii) représentez graphiquement la courbe d'efficacité du test.

**Exercice 2 (Econométrie) Barème : 6 points.**

On considère un modèle linéaire faisant dépendre le taux de croissance du PIB français (variable FRA), du taux de croissance du PIB allemand (variable DEU), du taux de croissance du PIB italien (variable ITA) et d'une constante. On dispose d'un échantillon de données semestrielles sur la période 1963:2 à 2002:2, soit un total de 79 observations.

**Question 1 (3 points)** Retrouvez alors les 6 éléments manquants de la sortie de résultats Eviews (figure 1). Vous justifierez très précisément votre réponse.

**Question 2 (2 points)** On cherche à tester si l'élasticité du PIB français au PIB italien est égale à 2 fois l'élasticité du PIB français au PIB allemand. On note  $\beta_0$  la constante,  $\beta_D$  l'élasticité au PIB allemand et  $\beta_I$  l'élasticité au PIB italien.

(i) Ecrivez cette hypothèse sous la forme d'une contrainte linéaire  $R\beta = q$ . Précisez l'écriture des vecteurs  $R$ ,  $\beta$  et  $q$ .

(ii) On admet que la réalisation de la statistique de Fisher associée au test  $H_0 : R\beta = q$  est égale à 1.046. Donnez la loi de la statistique de Fisher sous l'hypothèse nulle  $H_0$  et concluez pour un risque de première espèce de 5%.

**Question 3 (1 point)** On régresse maintenant le PIB français uniquement sur le PIB allemand (figure 2). Testez de deux façons différentes l'hypothèse nulle de nullité de l'élasticité du PIB français au PIB allemand.

Figure 1: Régression 1

Dependent Variable: FRA  
Method: Least Squares  
Date: 12/21/04 Time: 14:11  
Sample(adjusted): 1963:2 2002:2  
Included observations: 79 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEU	0.149478	0.059981	2.492108	██████
ITA	0.466263	0.075113	██████	0.0000
C	██████	0.001337	3.171449	0.0022
R-squared	0.462545	Mean dependent var		0.011514
Adjusted R-squared	██████	S.D. dependent var		0.011806
S.E. of regression	0.008769	Akaike info criterion		-6.598060
Sum squared resid	██████	Schwarz criterion		-6.508081
Log likelihood	263.6234	F-statistic		██████
Durbin-Watson stat	2.061817	Prob(F-statistic)		0.000000

Dependent Variable: FRA  
Method: Least Squares  
Date: 12/21/04 Time: 14:36  
Sample(adjusted): 1963:2 2002:2  
Included observations: 79 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEU	0.288466	0.067865	4.250600	0.0001
C	0.008769	0.001366	6.421393	0.0000
R-squared	0.190050	Mean dependent var		0.011514
Adjusted R-squared	0.179531	S.D. dependent var		0.011806
S.E. of regression	0.010694	Akaike info criterion		-6.213248
Sum squared resid	0.008806	Schwarz criterion		-6.153262
Log likelihood	247.4233	F-statistic		18.06760
Durbin-Watson stat	1.813339	Prob(F-statistic)		0.000059

Régression 2

**Exercice 3 (Estimation)** *Barème : 4 points.*

On admet qu'au cours de l'année écoulée, le chiffre d'affaires, exprimé avec une unité monétaire convenablement choisie dont il ne sera plus fait mention par la suite, des entreprises de gros dont le chiffre d'affaire est inférieur à  $c$ , d'un certain pays peut être correctement représenté par une variable aléatoire  $X$  absolument continue dont la densité de probabilité  $f$  est définie, pour tout  $x \neq 0$  et  $x \neq c$  comme suit :

$$f(x) = \frac{1}{\theta \cdot c^{\frac{1}{\theta}}} \cdot x^{\frac{1}{\theta}-1} \quad \text{pour } 0 \leq x \leq c \quad (6)$$

$$f(x) = 0 \quad \text{pour } x < 0 \text{ et pour } x > c \quad (7)$$

où  $c$  et  $\theta$  désignent deux paramètres réels strictement positifs. La valeur du paramètre  $c$  est connue, mais non celle du paramètre  $\theta$ . On a constitué au hasard et avec remise un échantillon de  $n$  entreprises de commerce de gros, parmi celles dont le chiffre d'affaires est inférieur à  $c$ , et déterminé leur chiffre d'affaires au cours de l'année écoulée. Soit  $x_1, x_2, \dots, x_n$  ces chiffres d'affaires. Ainsi pour tout  $i$  de 1 à  $n$ ,  $x_i$  est la réalisation de la variable aléatoire  $X_i$  représentant le chiffre d'affaire de la  $i^{\text{ème}}$  entreprise choisie lors de la constitution de l'échantillon.

**Question 1 (1 point)** Déterminez la fonction de vraisemblance  $L(X_1, X_2, \dots, X_n ; \theta)$ .

**Question 2 (1 point)** Montrez que l'estimateur  $\hat{\Theta}_n$  de  $\theta$  obtenu par la méthode du maximum de vraisemblance est :

$$\hat{\Theta}_n = \ln(c) - \frac{1}{n} \sum_{i=1}^n \ln(X_i) \quad (8)$$

**Question 3 (2 points)** On pose  $Y = \ln(c) - \ln(X)$  et l'on admet que  $E(Y) = \theta$  et  $V(Y) = \theta^2$ . L'estimateur  $\hat{\Theta}_n$  est-il sans biais ? Est-il convergent ?

# Université d'Orléans - Licence Economie et Gestion

## Introduction à la Statistique et à l'Econométrie

C. Hurlin. Janvier 2005. Eléments de Correction

**Exercice 1 (A partir de l'Examen 2003, Université Paris IX Dauphine) Barème : 7 points.**

### Partie I : Test d'Hypothèses simples (8 points)

**Question 1 (1.5 points)** On a :

$$\bar{X} = (2.5 + 3.0 + 3.2 + 3.6 + 2.0 + 2.1 + 1.5 + 1.2 + 1.0)/9 = 2.2333 \quad 0.5 \text{ points} \quad (1)$$

Justification sur la loi de  $\bar{X}$  (cf cours) :

$$\bar{X} \rightarrow N\left(m, \frac{1}{N}\right) \quad 0.5 \text{ points}$$

D'où l'on tire (à justifier 0.5 points):

$$IC = \left[2.2333 - 1.96 * \frac{1}{\sqrt{9}}, 2.2333 + 1.96 * \frac{1}{\sqrt{9}}\right] = [1.58, 2.8866]$$

**Question 2 (1 point)** On choisit toujours l'hypothèse à laquelle on souhaite le moins renoncer (0.5 point) : c'est à dire que le lot soit acceptable. On teste le lot est acceptable (on aurait pu tester  $m < 3$ ) contre lot non acceptable (0.5 point). Le choix de deux confère une certaine "marge de sécurité"

**Question 3 (1 point)** vraisemblance loi normale associé à une observation

$$L(x_i, m) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_i - m)^2\right] \quad \text{car } \sigma = 1$$

D'où l'on tire que :

$$\begin{aligned} L(X, m) &= \prod_{i=1}^N L(x_i, m) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \prod_{i=1}^N \exp\left[-\frac{1}{2}(x_i - m)^2\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2} \sum_{i=1}^N (x_i - m)^2\right] \quad 0.5 \text{ point} \end{aligned}$$

D'où l'on tire que :

$$\begin{aligned} \frac{L(X, 2)}{L(X, 3)} &= \frac{\exp\left[-\frac{1}{2} \sum_{i=1}^N (x_i - 2)^2\right]}{\exp\left[-\frac{1}{2} \sum_{i=1}^N (x_i - 3)^2\right]} \\ &= \exp\left[\frac{1}{2} \left(\sum_{i=1}^N (x_i - 3)^2 - \sum_{i=1}^N (x_i - 2)^2\right)\right] \quad 0.5 \text{ point} \end{aligned}$$

**Question 4 (1 point)**

$$\frac{L(X, m_0)}{L(X, m_1)} < K$$

$$\sum_{i=1}^N (X_i - 3)^2 - \sum_{i=1}^N (X_i - 2)^2 < 2 \ln(K)$$

$$-2 \sum_{i=1}^N X_i + 5 < 2 \ln(K) \quad (2)$$

$$\frac{1}{N} \sum_{i=1}^N X_i > A = \frac{2 \ln(K) - 5}{-2N} \quad 0.5 \text{ point}$$

$$W = \{X_N | \bar{X} > A\} \quad (0.5 \text{ point})$$

où  $A$  est une constante

**Question 5 (1 point)** Dans ce cas, on a :

$$\alpha = \Pr[W/H_0] = \Pr\left[\frac{\bar{X} - m_0}{1/\sqrt{N}} > \frac{A - m_0}{1/\sqrt{N}}\right]$$

D'où l'on tire que :

$$A = m_0 + \Phi(1 - \alpha)^{-1} \frac{1}{\sqrt{N}} = 2 + \frac{\Phi(0.99)^{-1}}{3} = 2.7754 \quad 0.5 \text{ point}$$

ici on a  $\bar{X} = 2.23$ , donc on accepte  $H_0$  marchandise conforme. (0.5 point)

**Question 6 (1.5 point)**

$$\begin{aligned} \text{RisqueII} &= \Pr[\bar{W}/H_1] = \Pr[\bar{X} < A | m = m_1] \\ &= \Pr\left[\frac{\bar{X} - m_1}{1/\sqrt{N}} < \frac{A - m_1}{1/\sqrt{N}}\right] \quad 0.5 \text{ point} \\ &= \Phi\left(\frac{A - m_1}{1/\sqrt{N}}\right) = \Phi\left(\frac{2.7754 - 3}{1/\sqrt{9}}\right) \\ &= \Phi(-0.6738) = 0.25 \quad 0.5 \text{ point} \end{aligned}$$

$$\text{Puissance} = 1 - \beta = 0.75 \quad 0.5 \text{ point}$$

**Question 7 (1.5 point)** On cherche  $N$  tel que :

$$\text{Puissance} = 1 - \Phi\left(\frac{A - m_1}{1/\sqrt{N}}\right) = 0.95$$

$$\frac{A - m_1}{1/\sqrt{N}} = \Phi^{-1}(0.05)$$

$$\sqrt{N} = \frac{\Phi^{-1}(0.05)}{A - m_1} = \frac{-1.6449}{2.7754 - 3} = 7.32377 \implies N > 53.58 \quad (1 \text{ point})$$

Il faut augmenter la taille de l'échantillon pour avoir une puissance acceptable pour un même niveau risque I (0.5 point)

### Partie II : Test Hypothèse simple contre hypothèse multiple (4 points)

On considère à présent le test suivant :

$$H_0 : m = 2$$

$$H_1 : m > 2$$

**Question 1 (1 point)** On a montré que la RC était :

$$W = \{X_N | \bar{X} > A\}$$

où  $A$  est indépendant de  $m_1$ , donc il existe un test UPP bilatéral dont la région critique correspond à celle du test hypothèse simple contre simple. Donc la RC est la même que dans le premier exo (1 point)

$$W = \{X_N | \bar{X} > 2.7754\}$$

**Question 1 (2 point)**

$$Puissance = 1 - \Phi\left(\frac{A - m_1}{1/\sqrt{N}}\right) = 1 - \Phi\left(\frac{2.7754 - m_1}{1/\sqrt{9}}\right)$$

par construction  $m = 2$ ,  $p_{ui} = \alpha = 1\%$  risque de première espèce (0.5 point si il n'y a pas de calcul)

$m = 2.5$ ,  $p_{ui} = 20.43$  0.5 point

$m = 3$ ,  $p_{ui} = 74.98$  0.5 point

$m = 5$ ,  $p_{ui} = 1$  0.5 point

$m = 10$  et  $m = 50$ , puissance=1, 0.5 point que s'il n'y a PAS DE CALCUL.

**Exercice 2** 0.5 point par élément

Question 2

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_D \\ \beta_I \end{pmatrix}$$

On cherche donc à tester si  $\beta_I = 2\beta_D$  ou encore  $\beta_i - 2\beta_D = 0$  (0.5 point)

Dependent Variable: FRA  
 Method: Least Squares  
 Date: 02/14/05 Time: 18:59  
 Sample(adjusted): 1963:2 2002:2  
 Included observations: 79 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEU	0.149478	0.059981	2.492108	0.0149
ITA	0.466263	0.075113	6.207479	0.0000
C	0.004239	0.001337	3.171449	0.0022
R-squared	0.462545	Mean dependent var		0.011514
Adjusted R-squared	0.448402	S.D. dependent var		0.011806
S.E. of regression	0.008769	Akaike info criterion		-6.598060
Sum squared resid	0.005843	Schwarz criterion		-6.508081
Log likelihood	263.6234	F-statistic		32.70362
Durbin-Watson stat	2.061817	Prob(F-statistic)		0.000000

Figure 1:

$$R\beta = q$$

$$R = \begin{pmatrix} 0 & -2 & 1 \end{pmatrix} \quad q = 0 \quad 0.5 \text{ point}$$

On admet que  $F = 1.046$ . Sous  $H_0$ ,  $F$  suit une  $F(N - K, 1)$  où  $N=79$ ,  $K=3$

$$F(76, 1) \quad 0.5 \text{ point}$$

Seuil à 5% = (0.5 point)? Ici pvalue=0.3 (pas à calculer), donc on accepte  $H_0$  (0.5 point)

Question 3

test avec la tstat (0.5 point)

test avec la Fstat global (0.5 point) : on teste la nullité de tous les coefs sauf la constante. Ici F stat= tstat au carré

**Exercice 3** Question 1 :

$$L = \prod_{i=1}^N f(x_i) = \left( \frac{1}{\theta \cdot c^{\frac{1}{\theta}}} \right)^N \cdot \prod_{i=1}^N x_i^{\frac{1}{\theta}-1} \quad \text{si tous les } x_i \text{ vérifient } 0 \leq x_i \leq c \quad (3)$$

$$= 0 \text{ sinon (1 point)}$$

Question 2 : supposons que tous les  $x_i$  sont tels que  $0 \leq x_i \leq c$ , alors

$$\frac{\partial \log L}{\partial \theta} = \frac{\partial}{\partial \theta} \left[ -N \log(\theta) - \frac{N}{\theta} \log(c) + \left( \frac{1}{\theta} - 1 \right) \sum \log(x_i) \right] \quad 0.5 \text{ point}$$

$$= -\frac{N}{\theta} + \frac{N}{\theta^2} \log(c) - \frac{1}{\theta^2} \sum \log(x_i) = 0 \quad 0.5 \text{ point}$$

D'où l'on tire

$$\hat{\Theta}_n = \ln(c) - \frac{1}{n} \sum_{i=1}^n \ln(X_i) \quad (4)$$

Question 3. On pose  $Y = \ln(c) - \ln(X)$  et l'on admet que  $E(Y) = \theta$  et  $V(Y) = \theta^2$ . On alors :

$$\begin{aligned} E(\hat{\Theta}_n) &= \ln(c) - \frac{1}{n} \sum_{i=1}^n E[\ln(X_i)] \\ &= \ln(c) - \frac{1}{n} \sum_{i=1}^n E[\ln(c) - \ln(Y_i)] \\ &= \frac{1}{n} \sum_{i=1}^n E[\ln(Y_i)] \\ &= \frac{1}{n} \sum_{i=1}^n \theta = \theta \quad \text{1 point sans biais} \end{aligned}$$

$$\begin{aligned} V(\hat{\Theta}_n) &= \ln(c) - \frac{1}{n^2} \sum_{i=1}^n V[\ln(X_i)] \\ &= \ln(c) - \frac{1}{n^2} \sum_{i=1}^n V[\ln(c) - \ln(Y_i)] \\ &= \frac{1}{n^2} \sum_{i=1}^n V[\ln(Y_i)] \\ &= \frac{1}{n^2} \sum_{i=1}^n \theta^2 \\ &= \frac{\theta^2}{n}, \quad \text{convergent 1point} \end{aligned}$$