L3 2006/2007
Statistique et Économétrie
Interrogation 2, Éléments de correction

Exercice 1

Soit un échantillon de femmes âgées de 26 à 54 ans[1]. On dispose de trois variables : *lnsalaire* = le logarithme du taux de salaire horaire en euros, *lnage* = le logarithme de l'âge de l'individu, *lnetudes* = le logarithme du nombre d'années d'études. En régressant la variable *lnsalaire* sur *lnage* et *lnetudes*, on obtient l'équation suivante :

$$\widehat{lnsalaire} = -0,31 + 0,49 \cdot lnage + 0,31 \cdot lnetudes$$

1. Écrire le modèle économétrique qu'on a estimé.

$$lnsalaire = \beta_0 + \beta_1 \cdot lnage + \beta_2 \cdot lnetudes + \varepsilon$$

2. Rappeler les propriétés des estimateurs des MCO.
   Les estimateurs des MCO des coefficients β sont les meilleurs (les plus efficaces) estimateurs linéaires sans biais.
3. Interpréter la régression estimée.
   Comme on estime un modèle logarithmique, les coefficients β_1 et β_2 sont des élasticités. L'élasticité du taux de salaire horaire par rapport à l'âge est 0,49 (le taux de salaire horaire augmente de 0,49% si l'âge augmente de 1%) et l'élasticité du taux de salaire horaire par rapport au nombre d'années d'études est 0,31 (le taux de salaire horaire augmente de 0,31% si le nombre d'années d'études augmente de 1%).

Exercice 2

On dispose d'un échantillon $(x_1, \ldots, x_n)$ de taille n d'une variable suivant une loi de Bernoulli de paramètre p.

1. Écrire la vraisemblance de cet échantillon.

$$L = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{N-\sum_{i=1}^{n} x_i}$$

2. Déterminez l'estimateur du maximum de vraisemblance de p.

$$\ln(L) = \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)(N - \sum_{i=1}^{n} x_i)$$

$$\frac{\partial \ln(L)}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{N - \sum_{i=1}^{n} x_i}{1-p} = 0 \Rightarrow \hat{p}_{MV} = \frac{\sum_{i=1}^{n} x_i}{N} = \bar{x}.$$

[1]Source : l'Enquête « Emploi » 2002 de l'INSEE.

3. Montrez que la méthode des moments donne le même estimateur.
   $E(X) = (1-p)\cdot 0 + p \cdot 1 = p$, alors $p = E(X)$ le premier moment théorique, $\hat{p}_{MM} = \bar{x}$ le premier moment empirique.

4. Est-ce un estimateur sans biais ? convergent ? Expliquez.
   $E(\hat{p}) = E(\bar{x}) = E(\frac{\sum_{i=1}^n x_i}{n}) = \frac{\sum_{i=1}^n E(x_i)}{n} = \frac{\sum_{i=1}^n p}{n} = \frac{np}{n} = p$ donc $\hat{p}$ est un estimateur sans biais de p.
   $V(\hat{p}) = V(\bar{x}) = V(\frac{\sum_{i=1}^n x_i}{n}) = \frac{\sum_{i=1}^n V(x_i)}{n^2} = \frac{\sum_{i=1}^n p(1-p)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$. Alors $\hat{p}$ est un estimateur sans biais et sa variance tend vers 0 quand n tend vers l'infini. Donc, $\hat{p}$ est un estimateur convergent de p.

Exercice 3

Soit X une variable aléatoire de loi normale d'espérance m et de variance σ^2. On dispose d'un échantillon de 23 tirages indépendants de X. La moyenne de cet échantillon est égale à 6, la variance empirique (l'estimateur sans biais) est égale à 6,25. Construire un intervalle de confiance à 99% pour m.

Donc, on dispose d'un échantillon $x_1, \ldots, x_{23}$ de $X \rightsquigarrow N(m, \sigma^2)$.

$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 6$, $s^2 = \frac{\sum_{i=1}^n (x_i-\bar{x})^2}{23-1} = 6,25$.

$\bar{x} \rightsquigarrow N(m, \frac{\sigma^2}{23}) \Rightarrow \frac{\bar{x}-m}{\sqrt{\sigma^2/23}} \rightsquigarrow N(0,1)$.

Mais la variance théorique σ^2 est inconnue donc on la remplace par son estimateur s^2.

$\frac{\bar{x}-m}{\sqrt{s^2/23}} \rightsquigarrow t_{23-1}$.

$\Pr(-2,819 < \frac{\bar{x}-m}{\sqrt{s^2/23}} < 2,819) = 0,99$

où 2,819 est la valeur critique d'une distribution t à 22 degrés de liberté.

$\Pr(\bar{x} - 2,819\frac{s}{\sqrt{23}} < m < \bar{x} + 2,819\frac{s}{\sqrt{23}}) = 0,99$

$\Pr(6 - 2,819\frac{2,5}{\sqrt{23}} < m < 6 + 2,819\frac{2,5}{\sqrt{23}}) = 0,99$

L'intervalle de confiance à 99% est $]4,53; 7,47[$.

Exercice 4

Une compagnie d'assurance a envoyé n propositions à des clients potentiels tirés dans une très large population. En retour elle a obtenu 12% de réponses favorables. Elle construit un intervalle de confiance à 95% pour la valeur de la probabilité p de succès pour chaque envoi. La longueur de cet intervalle est inférieure à 5%. Combien de propositions ont été envoyées ?

On dispose de n tirages d'une v.a. de Bernoulli : $X_1, \ldots, X_n \rightsquigarrow B(1,p)$.

$\hat{p} = \sum X_i/n = 0,12$.

$n\hat{p} \rightsquigarrow B(n,p)$ (loi binomiale) mais pour les calculs numériques on utilise l'approximation normale : $\hat{p} \rightsquigarrow N(p, \frac{p(1-p)}{n})$.

$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \rightsquigarrow N(0,1) \Rightarrow P(\hat{p} - 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = 0,95 \Rightarrow$

$P(0,12 - 1,96\sqrt{\frac{0,12(1-0,12)}{n}} < p < 0,12 + 1,96\sqrt{\frac{0,12(1-0,12)}{n}}) = 0,95 \Rightarrow$ la longueur de l'intervalle de confiance est $2 \cdot 1,96\sqrt{\frac{0,12(1-0,12)}{n}}$ et elle est inférieure à 5% :

$2 \cdot 1,96\sqrt{\frac{0,12(1-0,12)}{n}} < 0,05 \Rightarrow n > (2 \cdot 1,96\sqrt{0,12(1-0,12)}/0,05)^2 \Rightarrow n > 649$.

Barème :

Exercice 1 :
1. 1 point
2. 1 point
3. 1 point

Exercice 2 :
1. 1 point
2. 4 points
3. 2 points
4. 2 points

Exercice 3 : 4 points

Exercice 4 : 4 points