Université Paris 1, UFR 02, Licence de Sciences Economiques
STATISTIQUE, cours de Mme PRADEL
Partiel 14 juin 2002

Sont autorisés, outre une calculette et les tables statistiques : deux feuilles (recto verso) manuscrites, remplies par l'étudiant (à sa convenance et de sa main) des formules qu'il a choisies.

Il sera tenu compte de la justesse et de la pertinence des arguments justifiant les réponses

Les exercices sont indépendants les uns des autres

Exercice 1

Nous avons envoyé 600 propositions d'assurance à des clients potentiels tirés dans une très large population et nous avons en retour obtenu 78 réponses favorables. Nous nous intéressons à la valeur de la probabilité p de succès pour chaque envoi et nous pouvons supposer que les comportements des individus sont indépendants les uns des autres.

1. Décrire avec précision le modèle statistique correspondant à ces observations : quelle est la variable aléatoire observée, quel est le nombre d'observations, quelle est la loi de ces observations ?
2. Quel estimateur proposez vous pour p ? Quelle est l'estimation correspondant aux observations faites ?
3. Quelle est la loi de cet estimateur ? Pouvez vous utiliser l'approximation normale pour les calculs numériques ?
4. Construisez un intervalle bilatéral de confiance proche de 95% pour p, et calculez l'intervalle ici observé.

Exercice 2

Nous désirons proposer, par courrier, une assurance vie : notre fichier contient un grand nombre de clients potentiels dont nous connaissons l'âge, et nous nous demandons si les individus agés de “35 ou moins de 35 ans” ont moins de chance d'être intéressés par notre offre que ceux agés de “plus de 35 ans”.

Nous envoyons $n_1 = 250$ propositions à des jeunes et $n_2 = 350$ propositions à des plus de 35 ans. Nous observons 29 retours favorables provenant des premiers et 49 retours favorables provenant des seconds.

1. Test d'égalité des probabilités de retour favorable provenant des deux classes d'âge.
   (a) Précisez l'alternative (H_o contre H_1) que nous nous posons.
   (b) Construire le test de seuil 10%. Précisez notamment la statistique utilisée, sa loi sous l'hypothèse de base, la forme de la région de rejet de cette hypothèse et enfin la règle de décision correspondant au seuil de 10% choisi.
   (c) A quelle décision les observations faites nous conduisent-elles ?
2. Nous remarquons en fait que, parmi les plus de 35 ans, il peut être intéressant de distinguer les plus de 70 ans. Nos envois et les retours favorables sont répartis de la façon suivante :

	$\hat{a}ge \leq 35$	$35 < \hat{a}ge \leq 70$	$70 < \hat{a}ge$
nombre de retours favorables	29	35	14
nombre total d'envois	250	200	150

   (a) Construire le tableau de contingence associé aux deux variables “âge” et “intérêt pour l'assurance proposée”
   (b) Sous l'hypothèse d'indépendance entre l'âge et l'intérêt pour l'assurance proposée, quel est l'effectif estimé de retours favorables que nous aurions du observer dans la classe des individus d'âge supérieur à 70 ans ? Pourra-t-on utiliser le test du chi-deux pour tester cette hypothèse ?
   (c) Si vous effectuez un test du chi-deux d'indépendance, quelle sera la contribution de la classe des retours favorables parmi les plus de 70 ans dans le calcul de la statistique du chi-deux ?

(d) Quel est le degré de liberté de la statistique du chi-deux ?

(e) Un logiciel vous fournit la valeur $chi - deux = 5,797$, et la p-value associée : $PROB = 0,05510$. Finalement, au seuil de 10%, l'intérêt des personnes pour le type d'assurance que nous proposons dépend-il de l'âge ?

Exercice 3

Nous disposons d'un échantillon de taille 9 d'une variable X suivant une loi Normale de variance connue $\sigma^2 = 4$. et dont l'espérance peut prendre des valeurs supérieures ou égales à 10. Nous désirons tester l'hypothèse $\{m = 10\}$ contre $\{m > 10\}$.

1. Dans un premier temps, nous nous fixons l'alternative $\{m = 10\}$ contre $\{m = 11\}$.

   (a) Construire le test le plus puissant de seuil 10% pour tester $\{m = 10\}$ contre $\{m = 11\}$.

   (b) Calculer sa puissance

2. Nous revenons à notre problème initial, qui est de tester $\{m = 10\}$ contre $\{m > 10\}$

   (a) Montrer que le test construit à la première question est uniformément le plus puissant pour tester $\{m = 10\}$ contre $\{m > 10\}$.

   (b) Pour quelles valeurs de m sa puissance sera-t-elle supérieure ou égale à 0,90 ?

Exercice 4

Nous examinons la production de 25 entreprises d'un même secteur en fonction du capital et du travail mis en oeuvre : nous notons

$$\begin{cases} Q_i & \text{la production de l'entreprise} \\ W_i & \text{le nombre de salariés (équivalents temps plein)} \\ K_i & \text{le capital investi} \end{cases}$$

1. La régression de Q sur W et K fournit les résultats suivants

AJUSTEMENT	1	DEPENDENT VARIABLE	Q	
SMPL 1 25		TOTAL OBSERVATIONS	25	
FISHER global	4427,470	SSR	7,14 E+12	
DURBIN-WATSON	2.902	SE of regression	569731,7	
VARIABLE	COEFFICIENT	STAND. ERROR	T-STATISTIC	PROB
W	126,9178	4,972678	25,52302	0,0000
K	0,029528	0,049432	0,597351	0,5564
constante	65784,78	130468,2	0,504221	0,6191

.

   (a) Quelles sont les hypothèses statistiques sous lesquelles ont été calculées ces différentes statistiques, notamment l'écart type résiduel et la colonne "PROB" figurant sur le listing ci-dessus ?

   (b) La statistique de Durbin et Watson observée est-elle compatible avec ces hypothèses ?

   (c) Quel est le résultat qui conduit votre voisin a affirmer que le capital K n'est pas une variable explicative au seuil de 10%? Pourquoi cette affirmation est-elle statistiquement infondée ?

2. Nous considérons maintenant une production de type Cobb-Douglas :

$$Q = \lambda W^{\alpha} K^{\beta} \qquad \text{(mod\textbackslash\ `ele th\textbackslash'eorique)}$$

et les variables prises en logarithmes :

$$LQ_i = \ln(Q_i)\,; LW_i = \ln(W_i)\,; LK_i = \ln(K_i)$$

Nous supposons que :

$$\begin{aligned} LQ_i &= aLW_i + bLK_i + c + u_i \qquad \text{(mod\textbackslash\ `ele statistique)} \\ & \quad (a,b,c) \text{ réels non contraints} \\ u_1, u_2, ..., u_{25} &\sim i.i.d.N\left[0, \sigma^2\right] \end{aligned}$$

c'est-à-dire que $(u_1, u_2, ..., u_{25})$ sont identiquement et indépendamment distribuées selon une loi normale centrée, de variance σ^2.

Précisez les relations qui lient les paramètres (λ, α, β) du modèle théorique aux paramètres (a, b, c) du modèle statistique.

3. La régression de LQ sur LW et LK fournit l'ajustement suivant (ce sont les t-de student qui figurent sous les coefficients):

$$\begin{aligned} LQ_i &= \underset{[14{,}42]}{0.8620} LW_i + \underset{[2{,}31]}{0.1384} LK_i + \underset{[12{,}29]}{4.246} + \widehat{u}_i \\ DW &= 2,394, SCR = 0,8652 \end{aligned}$$

   (a) La statistique de Durbin et Watson est-elle compatible avec les hypothèses du modèle statistique postulé ?

   (b) Le capital est-il une variable significativement explicative au seuil de 5% ? Comparez avec l'affirmation de votre voisin, en 1c : expliquez l'apparente contradiction.

4. Remarquant que $0.8620 + 0.1384 = 1,0004$, nous désirons tester l'hypothèse $H_o = \{a + b = 1\}$.

   (a) quel est l'intérêt de cette contrainte dans l'équation de Cobb Douglas ?

   (b) Le modèle contraint par H_o est il toujours un modèle linéaire standard ?

5. Nous définissons les nouvelles variables "capital par tête" et "production par tête" en divisant respectivement K et Q par W. En logarithmes, cela nous conduit aux variables :

$$LQpt = Ln\left(\frac{Q}{W}\right) \quad \text{et} \quad LKpt = Ln\left(\frac{K}{W}\right)$$

   (a) Ecrire l'équation du modèle statistique en fonction de ces nouvelles variables, de LW et des paramètres (a, b, c) non contraints.

   (b) La régression de $LQpt$ sur LW et $LKpt$ fournit l'ajustement suivant (ce sont les t-de student qui figurent sous les coefficients):

$$\begin{aligned} LQpt_i &= \underset{[0{,}0220]}{????}\ LW_i + \underset{[2{,}31]}{????} LKpt_i + \underset{[12{,}29]}{4.246} + \widehat{v}_i \\ DW &= ????, SCR = ???? \end{aligned}$$

   Calculez les valeurs manquantes désignées par des points d'interrogation.

   (c) L'hypothèse $H_o = \{a + b = 1\}$ est-elle acceptable au seuil de 10%?