Université Paris 1, UFR 02, Licence de Sciences Economiques

STATISTIQUE, cours de Mme PRADEL

Partiel 13 juin 2000

Sont autorisées, 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 jugée utiles.

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

Pour la loi de Student de plus de 30 degrés de liberté, on utilisera la loi Normale

Les exercices sont indépendants les uns des autres

Exercice 1

Nous pouvons considérer que la note finale sur 20 obtenue au baccalauréat est, pour chaque élève, une variable aléatoire, Y , de loi Normale. Les notes obtenues au baccalauréat en 1999 par les 36 élèves de terminale S du lycée Walras sont supposées être des variables $(Y_1, ..., Y_{36})$ indépendantes et identiquement distribuées selon une loi Normale, d'espérance m et de variance $\sigma^2 = 6.25 = (2.5)^2$.

1. Construire un intervalle de confiance à 95% pour m .
2. Application numérique : calculer l'intervalle obtenu précédemment sachant que la valeur moyenne observée est égale à 11.26 :

$$\bar{Y} = \frac{1}{36} \sum_{i=1}^{36} Y_i = 11.26$$

3. Vous désirez prévoir la moyenne qu'obtiendra votre petit cousin Fred (élève en terminale S au lycée Walras) au bac 2000, en faisant l'hypothèse que sa note, notée Y_{Fred} est indépendante des 36 variables déjà observées, et suit la même loi qu'elles. Calculer l'intervalle de prévision à 98% pour Y_{Fred} , et les valeurs numériques correspondant aux observations faites en 1999.

Exercice 2

Nous considérons ici que la moyenne annuelle obtenue par un élève est une variable aléatoire X de loi Normale, et que les moyennes obtenues pour l'année 1999 par les 36 élèves de terminale S au lycée Walras sont indépendantes. De plus, ces 36 élèves sont issus de deux sections distinctes, S1 et S2, et nous supposons que l'espérance et la variance de X sont les mêmes pour tous les élèves d'une même section. Les résultats observés en fin d'année sont les suivants :

$$\begin{cases} \text{section S1 :} \quad i = 1,...16 \quad \bar{x}_1 = \frac{1}{16}\sum_{i=1}^{16} x_i = 9.222 \;, \text{ et } s_1^2 = \frac{1}{15}\sum_{i=1}^{16}(x_i - \bar{x}_1)^2 = 5.811 \\ \text{section S2 :} \quad i = 17,...36 \quad \bar{x}_2 = \frac{1}{20}\sum_{i=17}^{36} x_i = 11.247 \;, \text{ et } s_2^2 = \frac{1}{19}\sum_{i=17}^{36}(x_i - \bar{x}_2)^2 = 12.083 \end{cases}$$

1. Peut-on, au seuil de 10%, considérer que les notes $(X_1, ..., X_{36})$ ont toutes la même variance ?
2. Peut-on, au seuil de 10%, considérer que les notes dans la section S2 ont une espérance mathématique supérieure à celle des notes dans la section S1 ?

Exercice 3

Dans cet exercice, nous supposons que la note finale sur 20 obtenue au baccalauréat est, pour chaque élève, une variable aléatoire, Y , de loi Normale et dont l'espérance dépend du niveau atteint par cet élève au moment de passer l'examen. Ce niveau est représenté par la moyenne annuelle x de l'élève, considérée comme certaine et connue sans erreur, et nous postulons une relation affine entre $E(Y)$ et x , de la forme : $E(Y) = ax + b$.

Les 36 élèves de terminale S du lycée Walras en 1999 sont issus de deux sections distinctes, S1 et S2, et leurs notes de baccalauréat en 1999 sont $(Y_1, ..., Y_{36})$, vérifiant les conditions suivantes :

$$(I)\begin{cases} (Y_1, ..., Y_{36}) \text{ sont indépendantes et Normales} \\ E(Y_i) = a_1 x_i + b_1 \text{ et } V(Y_i) = \sigma_1^2 \quad \text{si} \quad i = 1, ..., 16 \text{ (section S1)} \\ E(Y_i) = a_2 x_i + b_2 \text{ et } V(Y_i) = \sigma_2^2 \quad \text{si} \quad i = 17, ..., 36 \text{ (section S2)} \end{cases}$$

1. Le modèle (I) ainsi défini est-il un modèle statistique linéaire ?
A quelle condition ce modèle statistique serait-il linéaire standard ?

2. Les observations $(y_1, x_1), ..., (y_{16}, x_{16})$ dans la section S1 conduisent à l'ajustement MCO suivant :

$$(S1) \quad y_i = 3.252 + 0.772x_i + \hat{u}_i \qquad SCR1 = 0.4394 \quad \text{si} \quad i = 1, ..., 16$$

tandis que les observations $(y_{17}, x_{17}), ..., (y_{36}, x_{36})$ dans la section S2 conduisent à l'ajustement MCO suivant :

$$(S2) \quad y_i = 3.039 + 0.794x_i + \hat{u}_i \qquad SCR2 = 0.7008 \quad \text{si} \quad i = 17, ..., 36$$

Peut-on considérer, au seuil de 10%, que le modèle statistique est linéaire standard ?

3. Les observations $(y_1, x_1), ..., (y_{36}, x_{36})$ prises toutes ensemble conduisent à l'ajustement MCO suivant :

$$y_i = 3.105 + 0.788x_i + \hat{u}_i \quad SCR = 1.1707$$

a. A quelles contraintes sur le paramètre (a_1, b_1, a_2, b_2) correspond cet ajustement ?

b. Ces contraintes sont-elles acceptables, au seuil de 5%?

4. On suppose maintenant que $(II)\begin{cases} (Y_1, ..., Y_{36}) \text{ sont indépendantes et Normales} \\ E(Y_i) = ax_i + b \text{ et } V(Y_i) = \sigma^2 \quad \text{si} \quad i = 1, ..., 36 \end{cases}$

L'ajustement MCO obtenu à partir des 36 observations est donné à la question précédente, et nous complétons les résultats par la donnée de :

$$(X'X)^{-1} = \begin{pmatrix} 0,331 & -0,0293 \\ -0,0293 & 0,00283 \end{pmatrix}$$

a. Donner en fonction de σ^2 la valeur de $Var(\hat{a}), Var(\hat{b})$ et $Cov(\hat{a}, \hat{b})$.

b. Calculer l'estimation de la variance σ^2 .

c. En déduire les estimations de $Var(\hat{a}), Var(\hat{b})$ et $Cov(\hat{a}, \hat{b})$.

5. Dans le cadre du modèle (II), la moyenne annuelle est-elle une variable explicative de la note du baccalauréat ?

6. Votre petit cousin Fred, qui est élève en terminale S (section S2) à Walras en 2000, a obtenu la note annuelle $x_{Fred} = 11$.
Calculer l'intervalle de prévision de sa note de Bac à 98%.

Exercice 4

On dispose d'un échantillon de taille 4 d'une variable aléatoire de Poisson, dont le paramètre peut prendre les valeurs réelles supérieures ou égales à 1. On désire tester $\{\lambda = 1\}$ contre $\{\lambda > 1\}$, au seuil de 10%.

1. Dans un premier temps, on décide de tester $\{\lambda = 1\}$ contre $\{\lambda = 2\}$.

a. Existe-t-il un test le plus puissant parmi les tests de seuil exactement égal à 10% ?

b. Déterminer le test le plus puissant parmi les tests de seuil inférieur ou égal à 10%.

c. Quel est son seuil exact, noté α ?

d. Quelle est sa puissance ?

2. Montrer que le test défini en 1°) est uniformément le plus puissant parmi les tests de seuil inférieur ou égal à 10% , pour tester $\{\lambda = 1\}$ contre $\{\lambda > 1\}$.

3. Quel seuil α devrait-on choisir pour que la puissance en $\lambda = 2$ soit supérieure ou égale à $1 - \alpha$?

Table (partielle) de la loi de Poisson : $Z \sim Poisson(\lambda)$

	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$	$\lambda = 11$	$\lambda = 12$
$P(Z \geq 5)$	0,1848	0,3711	0,5595	0,7149	0,8271	0,9004	0,9451	0,9707	0,9849	0,9924
$P(Z \geq 6)$	0,0840	0,2148	0,3840	0,5543	0,6994	0,8088	0,8844	0,9329	0,9625	0,9797
$P(Z \geq 7)$	0,0336	0,1106	0,2378	0,3937	0,5504	0,6867	0,7933	0,8698	0,9214	0,9542
$P(Z \geq 8)$	0,0120	0,0511	0,1334	0,2560	0,4014	0,5471	0,6762	0,7797	0,8568	0,9105
$P(Z \geq 9)$	0,0039	0,0213	0,0681	0,1527	0,2710	0,4075	0,5444	0,6671	0,7680	0,8450
$P(Z \geq 10)$	0,0012	0,0081	0,0318	0,0839	0,1696	0,2834	0,4126	0,5420	0,6595	0,7576
$P(Z \geq 11)$	0,0004	0,0028	0,0137	0,0426	0,0986	0,1841	0,2940	0,4169	0,5401	0,6528
$P(Z \geq 12)$	0,0002	0,0009	0,0055	0,0201	0,0534	0,1119	0,1970	0,3032	0,4207	0,5384