Université Paris 1, UFR 02, Licence de Sciences Economiques
STATISTIQUE, cours de Mme PRADEL
Partiel 21 janvier 2004

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 cohérence et de la pertinence des arguments avancés autant que de l'exactitude des résultats.

Les exercices sont indépendants les uns des autres

Exercice 1 (5 pts)

Nous disposons d'un échantillon de $n = 343$ entreprises nouvellement créées dont l'activité est hautement technologique et pour lesquelles nous observons sur chacune d'elles si elle a pu, ou non, bénéficier d'une aide publique au moment de sa création. Nous notons p la probabilité qu'un projet de création dans ce secteur obtienne une aide publique.

1. Quel estimateur de p proposez-vous ?
2. Construire un intervalle bilatéral de confiance 66,8% pour p.
3. application numérique : nous avons constaté que 126 entreprises ont bénéficié d'une aide publique au moment de leur création. Calculer l'estimation de p et l'intervalle correspondant à cette observation.

Exercice 2 (8 pts)

Nous avons observé le capital initialement investi par des entreprises du secteur industriel, dont certaines ont une activité faiblement technologique (désignées par "entreprises FT") et les autres une activité hautement technologique (désignées par "entreprises HT"). Nous pouvons supposer que le capital initialement investi par une entreprise est une variable aléatoire suivant une loi Normale, les capitaux initiaux des entreprises FT étant indépendamment et identiquement distribués ainsi que ceux des entreprises HT. Nous nous demandons si les créations d'entreprises HT nécessitent en moyenne le même capital que les entreprises FT. Les statistiques observées sur chaque groupe d'entreprises sont les suivantes :

$$\begin{cases} \text{entreprises FT} & \overline{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i = 141{,}46 \quad s_x^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \overline{x})^2 = 1973 \\ \text{entreprises HT} & \overline{y} = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j = 160{,}86 \quad s_y^2 = \frac{1}{n_2}\sum_{j=1}^{n_2}(y_j - \overline{y})^2 = 9085 \end{cases}$$

1. Ecrire précisément le modèle statistique qui est ici postulé et le problème que nous désirons résoudre.
2. Nous avons observé $n_1 = 14$ entreprises FT et $n_2 = 11$ entreprises HT.
   (a) Quelles sont les hypothèses sous lesquelles il est possible de tester l'égalité des espérances mathématiques entre les deux populations ?
   (b) Les données sont-elles (précisez le seuil utilisé) compatibles avec ces hypothèses ?
   (c) Avec un seuil de 10%, quelle est votre conclusion au vu des résultats obtenus ?
3. Comment sont changées les réponses à la question précédente si les mêmes statistiques ont été obtenues avec des échantillons de tailles respectives $n_1 = 140$ et $n_2 = 110$?

Exercice 3 (7 pts)

Nous disposons d'un échantillon de 145 entreprises du secteur industriel, pour lesquelles nous avons observé les variables suivantes :

- DUR_i = la durée d'activité de l'entreprise
- INV_i = le capital investi à la création
- $PRET_i$ = le montant du financement extérieur (prêts, bancaires ou non)

La régression des Moindres Carrés Ordinaires de la durée d'activité sur INV et $PRET$ fournit les résultats suivants :

AJUSTEMENT	1	DEPENDENT VARIABLE	DUR
SMPL 1 145		TOTAL OBSERVATIONS	145
FISHER global	17.91116	Prob(F-statistic)	0.000000
DURBIN-WATSON	1.765096	SCR	523.4665

VARIABLE	*COEFFICIENT*	*STAND.ERROR*	*T − STATISTIC*	*PROB*
INV	0.029428	0.021458	1.371453	0.1724
PRET	-0.004397	0.028404	-0.154811	0.8772
constante	59.17879	0.938442	63.06065	0.0000

1. Rappeler les hypothèses du modèle statistique linéaire standard. Les résultats obtenus ici sont-ils compatibles avec ces hypothèses ?
2. Les deux variables considérées sont-elles globalement explicatives ?
3. Quelles sont celles qui sont, individuellement, explicatives au seuil de 10% ?
4. Comparez vos réponses aux questions 2°) et 3°) : comment expliquez-vous cette apparente contradiction ?

Extraits de tables statistiques

LOI NORMALE : $U \approx N(0,1)$: table de $F(u) = P(U \leq u)$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633

LOI DE FISHER : Table du quantile $F_{0,95}$en fonction de r_1 et r_2 : $P(F(r_1, r_2) > F_{0,95}) = 5\%$

r2 \ r1	9	10	11	12	13	14	15	100	110	120	130	140
9	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,76	2,75	2,75	2,74	2,74
10	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,59	2,58	2,58	2,58	2,57
11	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,46	2,45	2,45	2,44	2,44
12	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,35	2,34	2,34	2,34	2,33
13	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,26	2,26	2,25	2,25	2,25
14	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,19	2,18	2,18	2,17	2,17
15	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,12	2,12	2,11	2,11	2,11
100	1,97	1,93	1,89	1,85	1,82	1,79	1,77	1,39	1,38	1,38	1,37	1,36
110	1,97	1,92	1,88	1,84	1,81	1,78	1,76	1,38	1,37	1,36	1,36	1,35
120	1,96	1,91	1,87	1,83	1,80	1,78	1,75	1,37	1,36	1,35	1,35	1,34
130	1,95	1,90	1,86	1,83	1,80	1,77	1,74	1,36	1,35	1,34	1,34	1,33
140	1,95	1,90	1,86	1,82	1,79	1,76	1,74	1,35	1,34	1,33	1,33	1,32

TABLE de DURBIN-WATSON :
Test unilatéral de $\rho = 0$ contre $\rho > 0$, au seuil de 5% (test bilatéral : seuil de 10%)

	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
n	dL	du	dL	du	dL	du	dL	du	dL	du
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78
150	1,72	1,75	1,71	1,76	1,69	1,77	1,68	1,79	1,66	1,80
200	1,73	1,78	1,75	1,79	1,73	1,80	1,73	1,81	1,72	1,82