Université Paris 1, UFR 02, Licence de Sciences Economiques
STATISTIQUE, cours de Mme PRADEL
Partiel 14 juin 2001
Elements de Corrigé

Exercice 1

$Y_1, Y_2, .., Y_{25} \approx i.i.d.\boldsymbol{N}(m, \sigma^2)$

1. Le meilleur estimateur de m est la moyenne empirique $\bar{y} = \frac{1}{25}\sum_{i=1}^{25} y_i$

2. $\bar{y} \approx \boldsymbol{N}\left(m, \frac{\sigma^2}{25}\right)$

3. $U = \frac{\bar{y}-m}{\sigma/5} \approx \boldsymbol{N}(0,1)$ et $P(|U| \leq 1.96) = 0.95$. On en déduit

$$P\left(-1.96 \leq \frac{\bar{y}-m}{\sigma/5} \leq 1.96\right) = 0.95$$

$$P\left(\bar{y} - \frac{1.96}{5}\sigma \leq m \leq \bar{y} + \frac{1.96}{5}\sigma\right) = 0.95$$

Si $\sigma^2 = 2.25 = (1.5)^2$, $\frac{1.96}{5}\sigma = 0.588$. L'intervalle de confiance 95% pour m.est donc

$$\bar{y} - 0.588 \leq m \leq \bar{y} + 0.588$$

4. Lorsque la variance est inconnue, son estimation est $s^2 = \frac{1}{24}\sum_{i=1}^{25}(y_i - \bar{y})^2$, et $T = \frac{\bar{y}-m}{s/5} \approx STUDENT(24)$. Lecture dans la table de Student :

$$P(|student(24)| \leq 2.064) = 0.95$$

L'intervalle de confiance 95% précédent devient $\bar{y} - \frac{2.064}{5}s \leq m \leq \bar{y} + \frac{2.064}{5}s$:

$$P(\bar{y} - 0.4128s \leq m \leq \bar{y} + 0.4128s) = 0.95$$

5. Application numérique : pour $\bar{y} = 13.5$ et $s^2 = 1.69 = (1.3)^2$:
le premier intervalle est $12.912 \leq m \leq 14.088$
le second intervalle est $12.963 \leq m \leq 14.037$
Remarque : l'intervalle où la variance est inconnue est plus petit que le premier, bien que 2.064 soit plus grand que 1.96. En effet, la variance estimée est plus petite que la variance théorique. Ce sont des choses qui arrivent..

Exercice 2

On désire tester l'efficacité d'un nouveau médicament censé corriger l'hypertension avec moins d'effets secondaires génants que ceux existant sur le marché. Pour les médicaments déjà sur le marché, la probabilité d'apparition de ces effets génants, notée p_a, est connue : $p_a = 0.60$. Le gain apporté par le nouveau médicament ne sera considéré comme décisif que si la probabilité d'apparition qui lui correspond est inférieure à $p_e = 0.54$. Pour simplifier, nous supposerons que la probabilité d'apparition des effets secondaires avec le nouveau médicament ne peut prendre que l'une des deux valeurs p_a ou p_e.

1. Le modèle : soit X la variable qui vaut 1 si le patient souffre des effets secondaires et 0 sinon : nous observons 80 réalisations indépendantes de cette variable de Bernoulli :

$$X_1,..,X_{80} \approx i.i.d.\boldsymbol{B}(1,p)$$
$$\text{avec } p = p_a = 0.60 \ (\text{médicament sans intérêt thérapeutique})$$
$$\text{ou } p = p_e = 0.54 \ (\text{médicament efficace})$$

Le problème de test qui se pose est de choisir entre les deux valeurs de p. L'hypothèse de base du laboratoire est que son médicament apporte un gain décisif : $H_o = \{p = p_e = 0.54\}$. C'est en effet l'hypothèse dont le rejet à tord a les conséquences les plus facheuses pour le labo.

2. Le test de Neyman a pour région critique $F = \frac{1}{80}\sum X_i \geq A$. Sous H_o, $80p_e(1-p_e) = 19,9 > 15$, et nous pouvons utiliser l'approximation de la loi de F par une loi Normale $N\left(0.54; \frac{0.54*0.46}{80}\right)$. Pour un seuil de 5%, nous lisons : $P\{N(0;1) \geq 1.645\} = 0.05$.La région critique du test de seuil 5% est donc
$$\frac{F-0.54}{\sqrt{0.54*0.46/80}} \geq 1.645$$
$$F \geq 0.54 + 1.645\sqrt{0.54*0.46/80} = 0.63166$$
Le médicament sera considéré comme insuffisamment efficace si $F \geq 0.63166$.

3. La puissance de ce test est $\eta = P\{F \geq 0.63166 \mid p = p_a = 0.60\}$. Sous $H_1 = \{p = p_a = 0.60\}, F\#N\left(0.60; \frac{0.60*0.40}{80}\right)$
$P\left\{N(0;1) \geq \frac{0.63166-0.60}{\sqrt{0.6*0.4/80}}\right\} = P\{N(0;1) \geq 0.57803\} = 0.2816$
La puissance du test est donc $\eta = 28.16\%$

4. Si 47 d'entre les patients ont éprouvé des effets secondaires génants , la fréquence observée est $f_{obs} = \frac{47}{80} = 0.5875 < 0.6317$: l'hypothèse de base n'est pas rejetée, le médicament est considéré (par son fabricant) comme apportant un gain décisif.

5. Le responsable de l'administration a pour hypothèse de base que le médicament n'est pas efficace : $K_o = \{p = p_a = 0.60\}$, qui est l'hypothèse H_1du laboratoire. Pour lui, il n'accepte le médicament que si F est suffisamment petite. La p-value associée à l'observation de $0;5875$ est
$PROB = P\{F < 0.5875 \mid p = p_a = 0.60\} = P\left\{N(0;1) < \frac{0.5875-0.6}{\sqrt{0.6*0.4/80}}\right\}$
$PROB = P\{N(0;1) < -0.22822\} = 0,4097$. Cette probabilité est pour lui trop élevée : il ne refuse pas son hypothèse de base, et il n'accepte pas le médicament.

6. La règle d'acceptation du médicament par l'administration est de la forme $F < B$. Pour un seuil de 5%, $P\{F < B \mid p = p_a = 0.60\} = 0,05$ et $P\{N(0;1) < -1,645\}$ entraînent : $\frac{B-0.6}{\sqrt{0.6*0.4/80}} = -1.645$, c'est-à-dire $B = 0,5099$. La règle de l'administration est d'accepter le médicament si moins de 50,99% de patients sur 80 ont eu des effets secondaires génants.

7. $0.51 * 80 = 40.8$. Le nombre de patients présentant des effets secondaires génants n'aurait donc pas du dépasser 40 pour que le médicament soit accepté.

8. Accepter la règle de l'administration, c'est accepter un trop grand risque de se tromper en abandonnant le médicament. Il faut donc poursuivre les observations de façon à ce que les deux règles soient compatibles : $A \leq B$ est atteint pour
$0.54 + 1.645\sqrt{0.54*0.46/n} \leq 0.60 - 1.645\sqrt{0.60*0.40/n}$
$\frac{1.645}{\sqrt{n}}\left(\sqrt{0.60*0.40} + \sqrt{0.54*0.46}\right) \leq 0.60 - 0.54 = 0.06$
$\frac{1.645}{0.06}\left(\sqrt{0.60*0.40} + \sqrt{0.54*0.46}\right) \leq \sqrt{n}$
$n \geq \left(\frac{1.645}{0.06}0.9883\right)^2 = 734.18$
Il faut donc effectuer l'essai thérapeutique sur au moins 735 patients

Exercice 3

1. Sur données temporelles, nous effectuons le test de Durbin et Watson. Nous supposons donc que les pertubations sont stationnaires au second ordre, telles que

$$u_t = \rho u_{t-1} + \varepsilon_t, \ |\rho| < 1$$
$$\varepsilon_t \approx BB(0, \sigma^2)$$

Nous testons $\rho = 0$ contre $\rho \neq 0$. Pour un seuil de 10% $\big($test bilatéral$\big)$, $n = 25$ et $k' = 2$ (nombre de variables explicatives hors la constante) , nous lisons dans la table de Durbin et Watson : $d_L = 1,21$ et $d_U = 1,55$. Ici, $DW = 2,382$ et $4 - DW = 1,618 > d_U$. Nous sommes dans la région d'acceptation de l'hypothèse de base $\{\rho = 0\}$.

L'hypothèse d'absence d'autocorrélations des observations est donc acceptable au seuil de 10%.

2. Test du Fisher global Sous $H_o = (a = b = 0), F = \frac{SCE/2}{SCR/(25-3)} \approx FISHER(2,22)$.

Lecture de table : $P\{FISHER(2,22) > 3,44\} = 0,05$.

$F_{observé} = 233,97 > 3,44$.

Au seuil de 5%, nous rejetons l'hypothèse $(a = b = 0)$: un au moins de ces coefficients est différent de zéro.

3. Tests individuels de significativité des variables du modèle : Pour tester $(a = 0)$ comme pour tester $(b = 0)$, il suffit de regarder la p-value donnée dans la table : nous refusons la nullité du coefficient si cette p-value est inférieure au seuil choisi, ici 5%.

Or pour a : $PROB = 0,3241 > 0,05$. Nous ne pouvons refuser l'hypothèse de nullité de a.

De même, pour b : $PROB = 0,5401 > 0,05$. Nous ne pouvons refuser l'hypothèse de nullité de b.

4. Les résultats ne sont pas contradictoires : nous avons admis qu'au moins l'un des coefficients est différent de zéro mais les tests individuels ne nous permettent pas de désigner lequel : les deux variables x et z sont sans doute fortement corrélées.

Exercice 4

1. L'effectif théorique correspondant à la case "$X = 1, Y = 2$" est $n_{1,2} = \frac{76*103}{808} = 9.6881 > 5$

Tous les autres effectifs théoriques seront plus grands que $n_{1,2}$ puisque cette case correspond aux effectifs marginaux minimums. Nous pourrons donc utiliser la loi de chi-deux dans le calcul du test.

2. La contribution δ_{12} de l'effectif observé dans cette case à la statistique du chi-deux est : $\delta_{12} = \frac{(13-9.6881)^2}{9.6881} = 1.1322$

3. Tous calculs faits, la statistique observée est $\Delta = 18.65$, et la p-value associée est $PROB = 0.0550$.

***Au seuil de* 10%**, $PROB = 0.0550 < 0,10$: nous sommes dans la région critique, et nous devons rejeter l'hypothèse d'indépendance entre les deux facteurs "activité innovante" de l'entreprise et le "capital humain" de son fondateur.

Variante : refus de l'indépendance des facteurs si $\Delta > A$

Le degré de liberté du chi-deux est $(3-1)(6-1) = 10$

Or sous H_o : $\Delta \approx \chi^2_{10}$et lecture sur la table du chi-deux : $P\{\chi^2_{10} > 15,987\} = 0,10$.

Ici, $18.65 > 15,987$: nous refusons l'hypothèse d'indépendance.