Université Paris 1, UFR 02, Licence de Sciences Economiques
STATISTIQUE, cours de Mme PRADEL
Examen de septembre 2005

Eléments de corrigé

Exercice 1:

$(X_1, ..., X_{20}) \overset{i.i.d.}{\approx} loiUniforme\,[0; a]$, avec $E(X) = \frac{a}{2}$ et $V(X) = \frac{a^2}{12}$.

1. Méthode des moments : une seule équation suffit :

$$E(X) = \frac{a}{2} \Longleftrightarrow a = 2E(X)$$

$E(X)$ est estimé par la moyenne empirique $\overline{X} = \frac{1}{20}\sum X_i$. L'estimateur des moments de a est donc

$$\widehat{A} = 2\overline{X}$$

2. L'estimation correspondant aux données est

$$\widehat{a} = 2 * 12 = 24$$

Exercice 2 :

Les variables observées sont les indicatrices de la propriété "utiliser un progiciel de Gestion Intégrée".

$$X_1, ..., X_N \text{ sont indépendantes, } X_i = \begin{cases} 1 \text{ si } i \text{ utilise un progiciel de Gestion Intégrée} \\ 0 \text{ sinon} \end{cases} \approx B(1, p_i)$$

1. L'hypothèse faite ici est que $p_i = p_1$si i est un établissement industriel et $p_i = p_2$si i est un établissement commercial. Nous testons $H_o : \{p_1 = p_2\}$ contre $H_1 : \{p_1 \neq p_2\}$.

   (a) Les fréquences observées sont

$$f_1 = \tfrac{839}{3019} = 0.2779 \qquad f_2 = \tfrac{402}{1505} = 0.2671$$

   (b) Les fréquences observées sont évidemment différentes : $f_1 \neq f_2$, mais la question est de savoir si $p_1 \neq p_2$.Nous effectuons le test de comparaison de fréquences, fondé sur la statistique

$$\begin{aligned} Z &= \frac{F_1 - F_2}{\sqrt{F(1-F)(1/3019 + 1/1505)}} = 31,691\frac{F_1 - F_2}{\sqrt{F(1-F)}} \\ \textit{où} &: F = \frac{3019F_1 + 1505F_2}{4524} \end{aligned}$$

   - Construction du test :Sous H_o, $Z \approx N[0; 1]$ car $3019f_1(1 - f_1) = 605$ et $1505f_2(1 - f_2) = 295$ sont largement supérieurs à 15 : l'approximation normale de la loi de Z est excellente.
     Nous déciderons que $\{p_1 \neq p_2\}$ si $|Z| > A$. Pour que ce test soit de seuil 10%, il faut utiliser la valeur A déterminée par

$$\begin{aligned} P(|N[0; 1]| > A) &= 0,10 \\ P(N[0; 1] < A) &= 1 - \frac{0,10}{2} = 0,95 \end{aligned}$$

     La lecture de la table de la loi $N[0; 1]$ fournit $P(N[0; 1] < 1,645) = 0,95$. Nous en déduisons que $A = 1,645$. Le test de seuil 10% est donc de décider que la probabilité d'utiliser un progiciel de Gestion intégrée dépend du type d'activité de l'établissement si:

$$31,691\frac{|f_1 - f_2|}{\sqrt{f(1-f)}} > 1,645$$

- Application du test : $f = \frac{839+402}{4524} = 0,27431$ et $z = 31,691\frac{0,2779-0,2671}{\sqrt{0,27431(1-0,27431)}} = 0,7671 < 1,645$. Nous ne sommes pas dans la région de rejet de H_o : au seuil de 10%, les proportions d'utilisateurs d'un progiciel de Gestion Intégrée parmi les établissements industriels et les établissements commerciaux ne sont pas significativement différentes.

2. Ici, nous avons 4 sous populations constituées par les différents types de biens : la variable qualitative Y correspondante prend 4 valeurs (4 niveaux). L'hypothèse testée est toujours que la loi de X ne dépend pas du type de bien Y. Le test est un test d'indépendance entre les variables "utilisation d'un progiciel" et "type de biens concernés".

   (a) les effectifs théoriques sont calculés sous l'hypothèse d'indépendance : dans la case correspondant à la ligne i et à la colonne j, l'effectif théorique est $Np_{i,j} = NP[Y=i, X=j] = NP[Y=i]P[X=j]$. Il est estimé par les fréquences marginales observées

$$N\widehat{p_{i,j}} = N\frac{N_{i.}}{N}\frac{N_{.j}}{N} = \frac{N_{i.}N_{.j}}{N}$$

	utilisateurs	non-utilisateurs	Total
Biens de Consommation	173,367	458,633	632
Biens Agro-Alimentaires	340,150	899,850	1240
Total	1241	3283	4524

   (b) La statistique utilisée est

$$\begin{aligned} \Delta &= \sum_{i,j} \frac{\left(N_{ij} - N\widehat{p_{i,j}}\right)^2}{N\widehat{p_{i,j}}} \underset{H_o}{\approx} CHI-DEUX\,(r) \\ r &= (4-1)(2-1) = 3 \end{aligned}$$

   (c) La règle est de refuser l'indépendance des facteurs si $\Delta_{observé} > A$.
   Plutôt que de rechercher A qui correspond au seuil de 10% (lecture de table : on lirait $A = 6,251$), nous pouvons utiliser la p-value associée.

$$PROB = P\left(\chi^2(3) > \Delta_{observé}\right)$$

   La règle est alors de refuser l'indépendance des facteurs si $PROB < 0,10$.
   Ici, $PROB = 0,0674 < 0,10$. En conclusion, nous décidons, avec un risque d'erreur de 10%, que l'utilisation d'un progiciel de Gestion Intégré dépend du type de biens concernés par l'activité de l'établissement.

Exercice 3

Le modèle postulé est un modèle linéaire standard, les variables explicatives étant les indicatrices des trimestres 2, 3 et 4 et la constante.

$$\begin{aligned} y_t &= c_1e_t + c_2Q2_t + c_3Q3_t + c_4Q4_t + u_t \\ u_1, ..., u_{40} &\approx i.i.d.N\left(0;\sigma^2\right) \end{aligned}$$

1. En particulier, $E(y_t) = c_1e_t + c_2Q2_t + c_3Q3_t + c_4Q4_t + 0$

   (a) Si t est un premier trimestre, $e_t = 1 \quad Q2_t = 0 \quad Q3_t = 0 \quad Q4_t = 0$ et $E(y_t) = c_1.1 + c_2.0 + c_3.0 + c_4.0 + 0 = c_1$.
   Si t est un deuxième trimestre, $e_t = 1 \quad Q2_t = 1 \quad Q3_t = 0 \quad Q4_t = 0$ et $E(y_t) = c_1.1 + c_2.1 + c_3.0 + c_4.0 + 0 = c_1 + c_2$.
   Si t est un troisième trimestre, $e_t = 1 \quad Q2_t = 0 \quad Q3_t = 1 \quad Q4_t = 0$ et $E(y_t) = c_1.1 + c_2.0 + c_3.1 + c_4.0 + 0 = c_1 + c_3$.
   Si t est un quatrième trimestre, $e_t = 1 \quad Q2_t = 0 \quad Q3_t = 0 \quad Q4_t = 1$ et $E(y_t) = c_1.1 + c_2.0 + c_3.0 + c_4.1 + 0 = c_1 + c_4$.

(b) $c_1$est donc le CA moyen des premiers trimestres, c_2, c_3 et c_4 sont les différences entre les CA moyens des trimestres suivant et le premier trimestre (on dit que le premier trimestre est pris comme trimestre de référence).

2. Le tableau donne les résultats classique de l'ajustement MCO.

(a) Le modèle postule en particulier que les y_t sont sans corrélation les uns avec les autres. Nous pouvons ici tester une des conséquences du modèle, qui est

$$cov(y_t, y_{t-1}) = 0$$

Le test mis en oeuvre est le test de Durbin et Watson. Il y a $k' = 3$ variables explicatives autres que la constante et $n = 40$ observations. La table de Durbin et Watson (test bilatéral de seuil 10%) nous fournit les deux valeurs $d_L = 1,34$ et $d_U = 1,66$. Le test de Durbin et Watson consiste à refuser l'hypothèse du modèle linéaire standard si DW est trop éloigné de 2.La loi de DW dépendant des variables explicatives, la règle de seuil 10% n'est pas fournie exactement, mais on sait que:

$$\begin{aligned} \text{si } 0 &< DW < 1,34 \text{ ou } 0 < 4 - DW < 1,34, \text{ alors il faut refuser le modèle linéaire standard} \\ \text{si } 1,66 &< DW < 2 \text{ ou } 1,66 < 4 - DW < 2, \text{ alors on peut accepter le modèle linéaire standard} \end{aligned}$$

Ici,$DW = 1,90$ vérifie la condition $1,66 < DW < 2$ Nous pouvons donc accepter l'hypothèse de modèle linéaire standard.

(b) Test global de Fisher : l'hypothèse nulle est $H_o : \{c_2 = c_3 = c_4 = 0\}$ .L'hypothèse est celle de l'absence d'effet saisonnier.

La statistique de Fisher est

$$F = \frac{(SCT - SCR)/3}{SCR/(40-4)} = \frac{36}{3}\frac{SCE}{SCR} \underset{H_o}{\approx} FISHER(3;36)$$

La règle de seuil α est refuser H_o si $F_{observé} > A$ ou $P(FISHER(3;36) > A) = \alpha$. De manière équivalente, nous refusons H_o si $P(FISHER(3;36) > F_{observé}) < \alpha$.

Ici, $P(FISHER(3;36) > F_{observé}) = \Pr ob(F - stat) = 0.00000$: nous refusons l'hypothèse d'absence d'effet saisonnier.

(c) Le test individuel de nullité d'un coefficient c_i : $\{c_i = 0\}$ contre $\{c_i \neq 0\}$ s'effectue avec la statistique de Student :

$$T_i = \frac{c_i}{s_i} \underset{\{c_i=0\}}{\approx} STUDENT(36)$$

en utilisant la p-value associée : $PROB_i = P(|STUDENT(36)| > t_i)$.

Au seuil de 10%, nous considérons que $\{c_i \neq 0\}$ si $PROB_i < 0,10$.

au seuil de 10%, nous acceptons les hypothèses $\{c_2 = 0\}$ et $\{c_4 = 0\}$, mais nous devons considérer que $\{c_3 \neq 0\}$. En considérant les valeurs des estimations, nous voyons que le CA présente une baisse systématique au cours du troisième trimestre.

3. Prévisions : nous devons faire l'hypothèse que le modèle sera encore valide au cours de l'année 2005 et que les CA seront indépendants des CA déjà observés. Les meilleures prévisions de $y_{41}, y_{42}, y_{43}, y_{44}$ seront donc l'estimation de leurs espérances mathématiques.

$$\begin{aligned} \widehat{y_{41}} &= \widehat{c_1} = 4,4837 \\ \widehat{y_{42}} &= \widehat{c_1} + \widehat{c_2} = 4,4837 + 0,0259 = 4,510 \\ \widehat{y_{43}} &= \widehat{c_1} + \widehat{c_3} = 4,4837 - 0,5427 = 3,941 \\ \widehat{y_{44}} &= \widehat{c_1} + \widehat{c_4} = 4,4837 - 0,0328 = 4,451 \end{aligned}$$