Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Carthage

Ecole Supérieure de la Statistique et de l'Analyse de l'Information

Examen

Module : Modèle linéaire	Enseignant : Mokhtar KOUKI
Classe : $2^{ème}$ Année	Nombre de pages : 02
Date : 6 juin 2018	Durée : 1h30min
Session : Contrôle	Documents : non autorisés

Exercice 1 (12 points : 2+3+1+2+2+2)

On considère la régression entre le logarithme de la production (q), le logarithme du stock de capital (k) et le logarithme de l'emploi (l) pour 63 entreprises :

$$q_i = \alpha + \beta_1 \, k_i + \beta_2 \, l_i + \epsilon_i, \; i = 1, 2, \cdots, 63$$

où ϵ_i, sont des termes aléatoires identiquement et indépendamment distribués d'espérance mathématique zéro et de variance σ^2, α, β_1 et β_2 sont des paramètres à estimer. On dispose des statistiques suivantes :

$$\begin{aligned}
\sum_{i=1}^{63} q_i &= 126; \sum_{i=1}^{63} k_i = 252; \sum_{i=1}^{63} l_i = 189, \sum_{i=1}^{63} q_i^2 = 230; \sum_{i=1}^{63} k_i^2 = 1120; \\
\sum_{i=1}^{63} l_i^2 &= 300, \; \sum_{i=1}^{63} q_i k_i = 600, \; \sum_{i=1}^{63} q_i l_i = 422 \text{ et } \sum_{i=1}^{63} k_i l_i = 800
\end{aligned}$$

1. Donner l'interprétation économique de la relation définie précédemment ainsi que celle des paramètres β_1 et β_2
2. Donner l'expression des estimateurs par la méthode des moindres carrés ordinaires des paramètres α, β_1 et β_2, que l'on note $\widehat{\alpha}$, $\widehat{\beta}_1$ et $\widehat{\beta}_2$ et calculer leurs valeurs numériques. Commenter.
3. Donner le valeur numérique de l'estimateur sans biais de la variance des résidus, notée $\widehat{\sigma}^2$.
4. ~~Calculer la variance estimée du vecteur~~ $\cancel{\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix}}$.
5. Dresser le tableau d'analyse de la variance.
6. ~~Sous l'hypothèse de la normalité des résidus, tester, au seuil de 5%,~~ l'hypothèse $H_0 : \beta_1 + \beta_2 = 1$ contre $H_1 : \beta \neq 1$. Interpréter ce résultat.

Exercice 2 (8 points : 1+1+2+2+2)

On considère la relation entre la teneur en huile des olives (Y), mesurée par la quantité d'huile (en litre) par 100 kg d'olives, et 4 variétés d'oliviers. Cette relation est définie par :

$$Y_{ij} = \theta_i + \epsilon_{ij}, i = 1, 2, 3, 4 \text{ et } j = 1, \cdots, n_i$$

avec ϵ_{ij} identiquement et indépendamment distribués selon la loi normale centrée et de variance σ^2 et j la plantation. On suppose que les autres facteurs qui influent sur la teneur en huile sont identiques.

1. Déterminer l'expression des estimateurs des paramètres θ_i, notée $\widehat{\theta}_i$.
2. L'observation des récoltes a permis d'obtenir les mesures suivantes :

Variétés	n_i	$\sum_{j=1}^{n_i} Y_{ij}$	$\sum_{j=1}^{n_i} Y_{ij}^2$
1	21	525	13200
2	30	600	12200
3	25	251	3300
4	40	400	4200

i. Calculer les valeurs numériques des estimateurs $\widehat{\theta}_i, i = 1, ..4$.

ii. Calculer la somme des carrés des résidus du modèle défini précédemment.

iii. Calculer la somme des carrés des résidus sous l'hypothèse que les olives ont la même teneur en huile quelque-soit la variété.

iv. Y-a-t-il une différence significative de la teneur en huile entre les différentes variétés ? Justifier.

NB : Arrondir à 4 chiffres après la virgule, $P(F(2,60) > 3.15) = 0.05$, $P(|N(0,1)| > 1.96) = 0.95$, $P(F(3,111) > 2,686) = 0.05$.

Contrôle 2018

Ex1 :

1) $Q = \alpha K^{\beta_1} L^{\beta_2}$ avec $\beta_1 = e_{Q/K}$ et $\beta_2 = e_{Q/L}$

avec $\beta_1 + \beta_2$ = rendement d'échelle

2) $$X = \begin{pmatrix} k_1 & l_1 \\ \vdots & \vdots \\ k_{63} & l_{63} \end{pmatrix} \quad Y = \begin{pmatrix} q_1 \\ \vdots \\ q_{63} \end{pmatrix}$$ x et y données centrées

$$\hat\beta = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} : \hat\beta = (x'x)^{-1} x'y$$

$$\hat\alpha = \bar q - \hat\beta_1 \bar k - \hat\beta_2 \bar l \quad (\hat\alpha = \bar y - \bar x \hat\beta \text{ avec } \bar x = (\bar k\ \bar l))$$

$$\hat\theta = \begin{pmatrix} \hat\alpha \\ \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} = (Z'Z)^{-1} Z'Y \text{ avec } Z = \begin{pmatrix} 1 & k_1 & l_1 \\ \vdots & \vdots & \vdots \\ 1 & k_{63} & l_{63} \end{pmatrix} \quad Y = \begin{pmatrix} q_1 \\ \vdots \\ q_{63} \end{pmatrix}$$

or $$Z'Z = \begin{pmatrix} 63 & \sum k_i & \sum l_i \\ \sum k_i & \sum k_i^2 & \sum k_i l_i \\ \sum l_i & \sum k_i l_i & \sum l_i^2 \end{pmatrix} = \begin{pmatrix} 63 & 252 & 189 \\ 252 & 1120 & 800 \\ 189 & 800 & 600 \end{pmatrix}$$

et $$Z'Y = \begin{pmatrix} \sum q_i \\ \sum k_i q_i \\ \sum l_i q_i \end{pmatrix} = \begin{pmatrix} 126 \\ 600 \\ 422 \end{pmatrix}$$

calculons $(Z'Z)^{-1}$: on pose $A = Z'Z$ rq A symétrique donc il suffit de calculer (→ 3 déterminants)

$$\text{com}(A) = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \text{ avec } A_{11} = (-1)^2 \begin{vmatrix} 1120 & 800 \\ 800 & 600 \end{vmatrix} = 32\,000$$

$$A_{12} = (-1)^3 \begin{vmatrix} 252 & 800 \\ 184 & 600 \end{vmatrix} = 0 \qquad A_{22} = (-1)^4 \begin{vmatrix} 63 & 189 \\ 189 & 600 \end{vmatrix} = 2079$$

$$A_{13} = (-1)^4 \begin{vmatrix} 252 & 1120 \\ 184 & 800 \end{vmatrix} = -10080 \qquad A_{23} = (-1)^5 \begin{vmatrix} 63 & 252 \\ 189 & 800 \end{vmatrix} = -2772$$

$$A_{21} = (-1)^3 \begin{vmatrix} 252 & 184 \\ 800 & 600 \end{vmatrix} = 0 \qquad A_{31} = (-1)^4 \begin{vmatrix} 252 & 189 \\ 1120 & 800 \end{vmatrix} = -10080$$

$A_{32} = (-1)^5 \begin{vmatrix} 63 & 189 \\ 252 & 800 \end{vmatrix} = -2772 \qquad A_{33} = (-1)^6 \begin{vmatrix} 63 & 252 \\ 252 & 1120 \end{vmatrix} = 7056$

$$\Rightarrow com(A) = \begin{pmatrix} 32000 & 0 & -10080 \\ 0 & 2079 & -2772 \\ -10080 & -2772 & 7056 \end{pmatrix}$$

Or $\det(A) = \begin{vmatrix} 63 & 252 & 189 \\ 252 & 1120 & 800 \\ 189 & 800 & 600 \end{vmatrix}$ $\quad L_2 \leftarrow L_2 - 4L_1$, $\quad L_3 \leftarrow L_3 - 3L_1$

$$= \begin{vmatrix} 63 & 252 & 189 \\ 0 & 112 & 44 \\ 0 & 44 & 33 \end{vmatrix} \quad L_3 \leftarrow L_3 - \frac{44}{112} L_2$$

$$= \begin{vmatrix} 63 & 252 & 189 \\ 0 & 112 & 44 \\ 0 & 0 & \frac{1760}{112} \end{vmatrix} = 110880$$

$$\Rightarrow \boxed{A^{-1} = \frac{1}{\det A}\, {}^t com(A)} = \frac{1}{110880} \begin{pmatrix} 32000 & 0 & -10080 \\ 0 & 2079 & -2772 \\ -10080 & -2772 & 7056 \end{pmatrix} = (Z'Z)^{-1}$$

D'où $\hat{\theta} = \frac{1}{110880} \begin{pmatrix} -221760 \\ 77616 \\ 44352 \end{pmatrix} = \begin{pmatrix} -2 \\ 0,7 \\ 0,4 \end{pmatrix}$

<u>Commentaire</u> :

- $\hat{\beta}_1 = 0,7$: Si le capital augmente de 1% alors la production augmentera de 0,7 %
- $\hat{\beta}_2 = 0,4$: Si le travail " " " " " " de 0,4%.
- La production est plus sensible au capital qu'au travail car $\hat{\beta}_1 > \hat{\beta}_2$
- $\hat{\beta}_1 + \hat{\beta}_2 = 1,1\% > 1\%$ $\Rightarrow$ Rendement d'échelle ↗

$\boxed{\hat{Y} = Z\hat{\theta}}$

3) $\hat{\sigma}^2 = \dfrac{SCR}{n-(k+1)} = \dfrac{SCR}{60}$ $\begin{pmatrix} n=63 \\ k=2 \end{pmatrix}$

Or $SCR = SCT - SCE = Y'Y - \hat{Y}'\hat{Y}$

$$= \sum_{i=1}^{63} y_i^2 - (Z\hat{\theta})'(Z\hat{\theta}) \quad \text{avec } \hat{\theta} = (Z'Z)^{-1}Z'Y$$

$$= \sum_{i=1}^{63} y_i^2 - \hat{\theta}'Z'Z(Z'Z)^{-1}Z'Y$$

$$= 430 - \hat{\theta}'Z'Y$$

or $\hat{\theta}'Z'Y = (-2 \quad 0{,}7 \quad 0{,}4)\begin{pmatrix} 126 \\ 600 \\ 422 \end{pmatrix} = 336{,}8$

$\Rightarrow \boxed{SCR = 93{,}2}$ D'où $\boxed{\hat{\sigma}^2 = 1{,}5533}$

4) $\hat{V}(\hat{\theta}) = \hat{\sigma}^2 (Z'Z)^{-1} = \begin{pmatrix} 0{,}4483 & 0 & -0{,}1412 \\ 0 & 0{,}0291 & -0{,}0388 \\ -0{,}1412 & -0{,}0328 & 0{,}0988 \end{pmatrix}$

$\Rightarrow \hat{V}(\hat{\beta}) = \begin{pmatrix} 0{,}0291 & -0{,}0388 \\ -0{,}0328 & 0{,}0988 \end{pmatrix}$

5) Tableau d'analyse de la variance :

	Σ carrés	ddl	carré moyen	Fisher
modèle	$SCE = 84{,}8$	$2 = k$	$\frac{SCE}{2} = 42{,}4$	$F = \dfrac{SCE/k}{SCR/n-(k+1)} = 27{,}2967$
résidu	$SCR = 93{,}2$	$60 = n-(k+1)$	$\hat{\sigma}^2 = 1{,}5533$	$P(F(2,60) > 3{,}15) = 0{,}05$
total	$SCT = 178$	62	$\frac{SCT}{n-1} = 2{,}8709$	$> 3{,}15$ on rejette que tous les coeff sont nuls

on a : $SCT = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2 = 430 - 63 \times \underbrace{4}_{=\left(\frac{126}{63}\right)^2} = 178$

$SCE = SCT - SCR = 178 - 93{,}2 = 84{,}8$

6) $H_0 : \beta_1 + \beta_2 = 1$ on pose $\theta = \beta_1 + \beta_2 \Rightarrow H_0 : \theta = 1$

$\boxed{H_0 : R\beta = r}$ avec $R = (1 \quad 1)$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $r = 1$

$$\theta = R\beta \longrightarrow \hat{\theta} = R\hat{\beta} \Rightarrow \hat{\theta} \hookleftarrow N(R\beta , R\,V(\hat{\beta})\,R')$$

$$F = \frac{(R\hat{\beta} - r)' \left[R\,\hat{V}(\hat{\beta})R' \right]^{-1} (R\hat{\beta} - r)}{\underbrace{\text{rang}(R)}_{1}} \hookrightarrow \underbrace{F(1, 60) = t^2(60)}_{F(\text{rang}(R),\, n-(k+1))}$$

or $R\hat{\beta} - r = (1 \quad 1)\begin{pmatrix} 0,7 \\ 0,4 \end{pmatrix} - 1 = 0,1$

$$R\,V(\hat{\beta})\,R' = (1 \quad 1)\begin{pmatrix} 0.0291 & -0,0388 \\ -0,0388 & 0,0938 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0,0503$$

$(-9,7 \cdot 10^{-3} \quad 0,06)$

$\Rightarrow \boxed{F = 0,1988} < 1.96$ donc on accepte l'hyp $H_0 \Rightarrow \beta_1 + \beta_2 = 1$

$\Rightarrow$ le rendement d'échelle est cte ??

Rqs : . ddl = 60 > 30 ⟹ ddl ⟶ +∞ donc on approxime la loi de student par la loi normale

. $F(1, 60)$ lorsque n élevé $\hat{\sigma}^2 \longrightarrow \sigma^2$ donc on reste dans $\chi^2(1) = 1,96^2$

Ex2 : $Y_{ij} = \theta_i + \varepsilon_{ij}$ $\quad i = 1..4$, $j = 1..n_i$

1) $\min_{\theta_i} \sum_j \sum_i \varepsilon_{ij}^2$

$$\frac{\partial}{\partial \theta_i} \sum_j \sum_i (Y_{ij} - \hat{\theta}_i)^2 = 0 \Longrightarrow -2 \sum_i (Y_{ij} - \hat{\theta}_i) = 0$$

$$\Longrightarrow \sum_i Y_{ij} - n_i \hat{\theta}_i = 0 \Longrightarrow \boxed{\hat{\theta}_i = \bar{Y}_i}$$

2) .) $\hat{\theta}_i = \frac{1}{n_i} \sum_i Y_{ij}$

$$\Rightarrow \hat{\theta}_1 = \frac{525}{21} = 25 \qquad \hat{\theta}_3 = \frac{251}{25} = 10.04$$

$$\hat{\theta}_2 = \frac{600}{30} = 20 \qquad \hat{\theta}_4 = \frac{400}{40} = 10$$

ii) $SCE = \sum_i \sum_j (Y_{ij} - \hat{\theta}_i)^2 = \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2$

$$= \sum_j \left(\sum_i Y_{ij}^2 - n \bar{Y}_i \right)$$

$$= (13200 - 21 \cdot 25^2) + (12200 - 30 \cdot 20^2) + (3300 - 25 \cdot (10.04)^2) + (4200 - 40 \cdot 10^2)$$

$$= 75 + 200 + 779.16 + 200$$

$$\Rightarrow \boxed{SCE = 1254.96}$$

iii) $H_0 : \theta_i = 0 \ \forall i \qquad Y_{ij} = 0 + \varepsilon_{ij}$

$$SCR_0 = \sum_i \sum_j (Y_{ij} - \bar{Y})^2 = \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2 + \sum_i n_i (\bar{Y}_i - \bar{Y})^2$$

$$= \underbrace{V.\text{intra}}_{\substack{SCR \\ \parallel \\ 1254.96}} + V.\text{inter}$$

or $\bar{Y} = \dfrac{n_1 \bar{Y}_1 + n_2 \bar{Y}_2 + n_3 \bar{Y}_3 + n_4 \bar{Y}_4}{n_1 + n_2 + n_3 + n_4}$

$$= \frac{525 + 600 + 251 + 400}{21 + 30 + 25 + 40} = \frac{1776}{116} = 15.3103$$

$\Rightarrow$ $V.\text{inter} = 21(25 - 15.3103)^2 + 30(20 - 15.3103)^2$

$$+ 25(10.04 - 15.3103)^2 + 40(10 - 15.3103)^2$$

$$= 4453.8582$$

$\Rightarrow SCR_0 = 5708.8182$

iv) $F = \dfrac{V_{inter} / g-1}{V_{intra} / n-g} = \dfrac{4453.8582 / 3}{1254.96 / 112} = 132.4961$

$\sum n_i = 116$

$\hookrightarrow F(3, 112)$

interpretation :