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

Exercices "Modèle linéaire", Enseignant : Mokhtar KOUKI

(AU : 2023/2024)

Exercice 1

On considère le modèle linéare suivant :

$$Y_i = \alpha + \beta X_i + \epsilon_i, i = 1, 2, \cdots, N$$

avec Y une variable endogène, X une variable exogène et ϵ un terme d'erreur. α et β sont des paramètres à estimer.

1. Rappeler les hypothèses de la méthode des moindres carrés ordinaires
2. Déterminer l'expression des estimateurs, par la méthode des moindres carrés ordinaires, des paramètres α et β
3. Montrer que ces deux estimteurs sont sans biais et calculer leurs variances

On considère, à présent, la relation entre la production (Q) et la quantité de travail (L) :

$$\ln(Q_i) = \alpha + \ \beta \ln(L_i) + \epsilon_i, \ i = 1, 2, \cdots, 50$$

où ϵ_i, vérifie les hypothèses de la méthode des moindres carrés ordinaires. Avec :

$$\begin{aligned} \sum_{t=1}^{50} \ln(Q_i) &= 700; \sum_{t=1}^{50} \ln(L_i) = 500; \sum_{t=1}^{50} \ln(Q_i)^2 = 15768; \\ \sum_{t=1}^{50} \ln(L_i)^2 &= 13125 \text{ et } \sum_{t=1}^{50} \ln(Q_i) \ln(L_i) = 13500 \end{aligned}$$

4. Donner l'interprétation économique de la relation définie précédemment
5. Calculer les valeurs numériques des estimateurs, par la méthode des moindres carrés ordinaires, de α et β, notés respectivement $\widehat{\alpha}$ et $\widehat{\beta}$. Commenter.
6. Dresser le tableau d'analyse de la variance. En déduire la valeur du coefficient de détermination linéaire. Commenter.
7. Donner la valeur numérique de l'estimateur sans biais de la variance du terme d'erreur, notée $\widehat{\sigma}^2$.

8. En déduire l'estimateur de la variance de $\widehat{\beta}$
9. Sous l'hypothèse de la normalité du terme d'erreur, tester, au seuil de 5%, l'hypothèse : $H_0 : \beta < 1$.
10. On considère une observation supplémentaire de L, $L_{51} = 2500$. Construire un intervalle de confiance, au niveau 95%, de $\ln(Q_{51})$. Commenter

Exercice 2 :

On considère la relation entre le logarithme de l'investissement (Y), le taux d'intérêt (X_1), le taux d'inflation (X_2) et le logarithme du produit intérieur brut (X_3) :

$$Y_i = \alpha + \beta_1\ X_{1i} + \beta_2\ X_{2i} + \beta_3\ X_{3i} + \epsilon_i,\ i = 1, 2, \cdots, 44$$

où ϵ_i, sont des termes aléatoires identiquement et indépendemment distribués selon la loi normale centrée et de variance σ^2, α, β_1, β_2 et β_3 sont des paramètres à estimer. On dispose des statistiques suivantes :

$$\begin{aligned}
\sum_{t=1}^{44}(Y_i - \overline{Y})^2 &= 135; \sum_{t=1}^{44}(X_{1i} - \overline{X_1})^2 = 206; \sum_{t=1}^{44}(X_{2i} - \overline{X_2})^2 = 442, \\
\sum_{t=1}^{44}(X_{3i} - \overline{X_3})^2 &= 1619, \sum_{t=1}^{44}(Y_i - \overline{Y})(X_{1i} - \overline{X_1}) = -112\ , \\
\sum_{t=1}^{44}(Y_i - \overline{Y})(X_{2i} - \overline{X_2}) &= 70\ , \sum_{t=1}^{44}(Y_i - \overline{Y})(X_{3i} - \overline{X_3}) = 135 \\
\sum_{t=1}^{44}(X_{1i} - \overline{X_1})(X_{2i} - \overline{X_2}) &= 65\ \sum_{t=1}^{44}(X_{1i} - \overline{X_1})(X_{3i} - \overline{X_3}) = -309 \\
\text{et } \sum_{t=1}^{44}(X_{2i} - \overline{X_2})(X_{3i} - \overline{X_3}) &= -624
\end{aligned}$$

1. **Déterminer** l'expression de l'estimateur par la méthode des moindres carrés ordinaires (mco),du vecteur $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ et calculer son espérance mathématique et sa variance.
2. Quels sont les signes attendus des paramètres β_1, β_2 et β_3.
3. Calculer les valeurs des estimateurs par la méthode mco des paramètres β_1, β_2 et β_3. sachant que

$$\begin{pmatrix} 206 & 65 & -309 \\ 65 & 442 & -624 \\ -309 & -624 & 1619 \end{pmatrix}^{-1} = 10^{-3} \begin{pmatrix} 7.547 & 2.026 & 2.221 \\ 2.026 & 5.507 & 2.509 \\ 2.221 & 2.509 & 2.009 \end{pmatrix}$$

commenter les valeurs obtenues.

4. Calculer la somme des carrés expliqués par le modèle (**SCE**), puis dresser le tableau d'analyse de la variance
5. Tester, au seuil de 5%, la significativité globale du modèle.
6. Calculer le coefficient de détermination linéaire. Commenter
7. Calculer l'estimateur de la matrice de variance-covariance des estimateurs des paramètres β_1, β_2 et β_3.et tester, au seuil de 5%, la significativité individuelle de ces paramètres.
8. Tester, au seuil de 5%, l'hypothèse :

$$H_0 : \begin{cases} \beta_1 = -\beta_2 \\ \beta_3 = 0.25 \end{cases}$$

9. On suppose que les investisseurs ne tiennent compte **que du taux d'intérêt réel**.
   (a) Donner le modèle correspondent à cette stratégie.
   (b) Estimer les paramètres de ce modèle (autre que la constante).

Exercice 3 :

Pour N clients de la banque BBB, le défault de paiement d'un client (D_i) obeït à la relation suivante :

$$\begin{aligned} D_i^* &= \beta_0 + \beta_1 R_i + \beta_2 A_i + \beta_3 G_i + \beta_4 MB_i + \epsilon_i \text{ pour } i = 1, ..., N \\ D_i &= \begin{cases} 1 \text{ si } D_i^* > 0 \\ 0 \text{ sinon} \end{cases} \end{aligned}$$

avec

- D : une variable indicatrice qui vaut 1 si le client est en défault de paiement et 0 sinon
- R : le revenu mensuel du client (en millier de TND),
- A_i : l'ancienneté du client avec la banque (en nombre d'années),
- G : une variable indicatrice qui vaut 1 si le client est une femme et 0 si le client est un homme.
- MB : une variable indicatrice qui vaut 1 si le client a des engagements avec d'autres banques et 0 sinon.
- ϵ_i : des termes aléatoires identiquement et indépendamment distribués selon la loi logistique.
- et $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)'$ le vecteur des paramètres.

1. Pour un client i, montrer que D_i suit une loi de Bernoulli dont on précisera le paramètre.
2. En déduire la vraisemblance d'un échantillon de taille N, $(D_1, D_2, \cdots, D_N)$.
3. L'estimation par maximum de vraisemblance du modèle défini précédemment a fourni les résultats suivant :

$$\begin{aligned} N &= 10000,\ \widehat{\beta} = \begin{pmatrix} 1.20 \\ -1.10 \\ -0.60 \\ -0.09 \\ 0.15 \end{pmatrix} \\ \text{et } \widehat{V}(\widehat{\beta}) &= 10^{-4} \begin{pmatrix} 2500 & 400 & 1000 & 10 & 10 \\ 400 & 400 & -100 & -100 & 10 \\ 1000 & -100 & 900 & 0 & 30 \\ 10 & -100 & 0 & 9 & -20 \\ 10 & 10 & 30 & -20 & 4 \end{pmatrix} \end{aligned}$$

   (a) Tester la significativité individuelle des paramètres
   (b) Commenter les résultats
   (c) Calculer :
      i. Le différentiel de probabilité de défaut de paiement entre les clients hommes et les clients femmes
      ii. Le différentiel de probabilité de défaut de paiement entre les clients qui ont des engagements avec d'autres banques comparés à ceux qui n'en ont pas.
      iii. Commenter.
4. Calculer une estimation de la probabilité **de solvabilité** d'un client homme possédant des engagements avec d'autres banques dont le salaire mensuel est de 1100 TND et ayant 3 années d'ancienneté avec la banque.

Exercice 4
On considère un modèle à choix discrets défini par :

$$\begin{aligned} S_i^* &= \beta_0 + \beta_1 R_i + \beta_2 Anc_i + \beta_3 MB_i + \epsilon_i, i = 1, \cdots, 500 \\ \text{avec } S_i &= \begin{cases} 1 \text{ si } S_i^* > 0 \\ 0 \text{ sinon} \end{cases} \end{aligned}$$

avec ϵ_i un terme d'erreur qui suit la loi logistque, et S **désigne la solvabilité du client** (1 si l'individu est solvable et 0 sinon) et R désigne le revenu annuel du client (en millier de dinar) Anc_i l'ancienneté du client avec la banque, MB une variable indicatrice qui vaut 1 si l'individu est client d'une autre baqnue et 0 sinon.

1. Ecrire la vraisemblance relative à un échantillon $(S_i, X_i), i = 1, \cdots, N$ (avec $X_i = (R_i, Anc_i, MB_i)$).
2. Quels sont les signes attendus des paramètres β_1, β_2, et β_3?
3. Sous l'hypothèse que $\beta_i = 0 \ i = 1, 2, 3$, montrer que $F(\widehat{\beta}_0) = \overline{S}$. En déduire l'expression du logarithme de la vraisemblance sous cette hypothèse.
4. L'estimation par maximum de vraisemblance des paramètres a donné le résultat suivant : $\widehat{\beta}_1 = 0.10$, $\widehat{\beta}_2 = 0.05$, $\widehat{\beta}_3 = -0.05, \widehat{V(\widehat{\beta}_1)} = 0.0016$, $\widehat{V(\widehat{\beta}_2)} = 0.0004$, et $\widehat{V(\widehat{\beta}_3)} = 0.0004$. Commenter.
5. Test, au seuil de 5%, la significativité globale du modèle sachant que $\widehat{Cov}(\ \widehat{\beta}_1,\ \widehat{\beta}_2) = -0.0002$, $\widehat{Cov}(\ \widehat{\beta}_1,\ \widehat{\beta}_2) = 0.0002$ et $\widehat{Cov}(\ \widehat{\beta}_2,\ \widehat{\beta}_3) = -0.0002$.

Eexercice 5 :

On considère l'observation du logarithme du PIB (Y), le logarithme de la demande locale (X) et le logarithme du taux de coouverture (Z). Les variables sont mesurées en volumes. Et on adopte le modèle linéaire suivant :

$$Y_t = \alpha + \beta_1 X_t + \beta_2 Z_t + \epsilon_t \text{ pour } t = 1, \cdots, 40$$

le terme d'erreur vérifie les hyothèses de la méthode des moindres carrés ordinaires. On suppose que ϵ_t suit la loi normale.

1. Rappeler les hypothèses de la méthodes des moindres carrés ordinaires.
2. Déterminer l'estimateur par mco des paramètres α, β_1 et β_2.
3. Calculer les valeurs des estimateurs sachant que la matrice de variance covariace entre les variables X, Z et Y est fournie comme suit

$$\begin{pmatrix} & X & Z & Y \\ X & 50 & 30 & 35.5 \\ Z & & 80 & 30.6 \\ Y & & & 29.375 \end{pmatrix}$$

avec $\sum_{t=1}^{40} Y_t = 600$, $\sum_{t=1}^{40} X_t = 480$ et $\sum_{t=1}^{40} Z_t = 400$.

4. Dresser le tableau d'analyse de la variance du modèle défini précédemment.
5. Tester, au seuil de 5%, la significativité globale du modèle.
6. Calculer le coefficient de détermination linéaire. Commenter
7. Test au seuil de 5%, l'hypothèse $H_0 : \beta_1 - \beta_2 = 0.5$

Exercice 6

On considère une expérience de plantation de quatre variétés d'un arbre fruitier plantées sur la même parcelle. Et on s'intéresse au rendement annuel par arbre. Le modèle linéaire correspondant est défini comme suit :

$$R_{ij} = \beta_j + \varepsilon_{ij} \text{ avec } j = 1, 2, 3, 4 \text{ et } i = 1, \cdots, N_j$$

R_{ij} le rendement de l'arbre i de la variété j et N_j le nombre d'arbres de la variété j. ε_{ij} un terme d'erreur identiquement et indépendemmentt distribué selon la loi $N(0, \sigma^2)$.

1. Déterminer les estimateurs par mco des paramètres β_j, noté $\widehat{\beta}_j$ pour $j = 1, 2, 3, 4$.
2. Construire un estimateur sans biais de σ^2, noté $\widehat{\sigma}^2_w$.
3. On fournit les informations suivantes :

Variété	N_j	$\overline{R}_j = \sum_{i=1}^{N_j} R_{ij}$	$S_j^2 = \frac{\sum_{i=1}^{N_j}\left(R_{ij}-\overline{R}_j\right)^2}{N_j}$
1	30	53.838	2.152
2	20	65.996	1.502
3	40	183.651	2.260
4	35	34.471	1.441

– Calculer les valeurs estimés $\widehat{\beta}_j$, $j = 1, 2, 3, 4$.
– Dresser le tableau ANOVA
– Calculer la valeur de l'estimateur $\widehat{\sigma}^2_w$
– Tester, au seuil de 5%, l'hypothèse selon laquelle les quatre variétés fournissent le même rendement.