République du Sénégal
Un Peuple – Un But – Une Foi

-*-*-*-*-*-*-*-

Ministère de l'Enseignement Supérieur et de la Recherche et de l'Innovation

UNIVERSITE DE THIES

UFR SET

Master Science des Données et Applications
Option Econométrie-Statistiques

Devoir de modèle linéaire

Auteurs :

Ousmane DIA

Abdoulaye Bara DIAW

Professeur :

Dr GNINGUE

Exercice 1

On considère le modèle linéaire $y_i = \beta + \varepsilon_i$, i = 1, ……, n
Où ε_i idd de loi N (0, σ²)
Proposons deux estimateurs différents du paramètre β

- Estimation par MCO
  On appelle estimateur des moindres carrés ordinaires de β, l'estimateur $\hat{\beta}$ obtenu par minimisation du risque quadratique
  $$\hat{\beta} = \underset{\beta \in \mathbb{R}}{argmin}\, J(\beta)$$
  Avec $J(\beta) = \frac{1}{n}\Sigma(y_i - \beta)^2$ la fonction coût quadratique
  $$J(\beta) = \frac{-2}{n}\Sigma(y_i - \beta) = 0 => \Sigma y_i - \Sigma\beta = 0 => n\beta = \Sigma y_i$$
  $$=> \quad \hat{\beta} = \frac{1}{n}\Sigma y_i$$
  $$\Leftrightarrow \quad \boxed{\boldsymbol{\hat{\beta} = \bar{y}}}$$
- Pour pouvoir avoir la valeur prédite $\hat{y}_i$ de y_i, il faut trouver un estimateur $\hat{\beta}$ de β c'est-à-dire $\hat{y}_i = \hat{\beta}$
  $$\Leftrightarrow \quad \boxed{\boldsymbol{\hat{\beta} = \hat{y}_i}}$$

Exercice 2

On considère le modèle de régression linéaire $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i$
i = 1, ……, n, que l'on écrit sous la forme $Y = \beta + \varepsilon_i$.

X'X = $\begin{pmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix}$, X'Y = $\begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$, Y'Y = 59.5

1. **Déterminons n , la moyenne $\bar{x}_{i,2}$ des $x_{i,2}$ et le coefficient de corrélation (r) des $x_{i,1}$ et $x_{i,2}$**

   On sait que par définition X'X = $\begin{pmatrix} n & \Sigma x_{i,1} & \Sigma x_{i,2} \\ \Sigma x_{i,1} & \Sigma {x^2}_{i,1} & \Sigma x_{i,1}x_{i,2} \\ \Sigma x_{i,2} & \Sigma x_{i,1}x_{i,2} & \Sigma {x^2}_{i,2} \end{pmatrix}$

Par identification : $n = 30$, $\Sigma x_{i,1} = \Sigma {x^2}_{i,1} = 20$, $\Sigma x_{i,2} = \Sigma x_{i,1}x_{i,2} = 0$, $\Sigma {x^2}_{i,2} = 10$

Or $\bar{x}_{i,2} = \frac{1}{n}\,\Sigma x_{i,2}$ et $r = \frac{cov(x_{i,1}, x_{i,2})}{\sigma(x_{i,1}) * \sigma(x_{i,2})}$

Comme $\Sigma x_{i,2} = \Sigma x_{i,1}x_{i,2} = 0$,

$$=> \bar{x}_{i,2} = \frac{1}{n}\,\Sigma x_{i,2} = 0\ ;\ cov(x_{i,1}, x_{i,2}) = \frac{1}{n}\Sigma x_{i,1}x_{i,2} - \bar{x}_{i,1}\bar{x}_{i,2} = 0$$

et $\sigma(x_{i,1}) = \sqrt{\frac{1}{n} * \sum {x^2}_{i,1} - \overline{x}^2_{i,1}} \neq 0$; $\sigma(x_{i,2}) = \sqrt{\frac{1}{n} * \sum {x^2}_{i,2} - \overline{x}^2_{i,2}} \neq 0$

Donc $\boxed{\boldsymbol{n = 30}}$ $\boxed{\boldsymbol{\bar{x}_{i,2} = 0}}$ et $\boxed{\boldsymbol{r = 0}}$

Le coefficient de corrélation linéaire des $x_{i,1}$ et $x_{i,2}$ nul montre qu'il y a absence totale de colinéarité entre les variables $x_{i,1}$ et $x_{i,2}$

2. Calcul des estimateurs

- Estimateurs $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ de β_0, β_1 et β_2

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

L'estimateur $\hat{\beta}$ des moindres carrés ordinaires a pour expression

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Posons $(X'X) = A$ et C la matrice des cofacteurs de A

$$A^{-1} = \frac{1}{detA} * adjA$$

$$detA = \begin{vmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{vmatrix} = 2000 \qquad C = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$$

$$c_{11} = +\begin{vmatrix} 20 & 0 \\ 0 & 10 \end{vmatrix} = 200\ ;\ c_{12} = -\begin{vmatrix} 20 & 0 \\ 0 & 10 \end{vmatrix} = -200\ ;$$

$$c_{13} = +\begin{vmatrix} 20 & 20 \\ 0 & 0 \end{vmatrix} = 0\ ;\ c_{21} = -\begin{vmatrix} 20 & 0 \\ 0 & 10 \end{vmatrix} = -200\ ;$$

$$c_{22} = +\begin{vmatrix} 30 & 0 \\ 0 & 10 \end{vmatrix} = 300\ ;\ c_{23} = -\begin{vmatrix} 30 & 20 \\ 0 & 0 \end{vmatrix} = 0\ ;$$

$$c_{31} = +\begin{vmatrix} 20 & 0 \\ 20 & 0 \end{vmatrix} = 0; c_{32} = -\begin{vmatrix} 30 & 0 \\ 20 & 0 \end{vmatrix} = 0; c_{32} = +\begin{vmatrix} 30 & 20 \\ 20 & 20 \end{vmatrix} = 200$$

$$C = \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix}$$

$$adjA = C^t = \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix}$$

$$A^{-1} = \frac{1}{detA} * adjA = \frac{1}{2000} * \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.15 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}$$

D'où $\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.15 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} * \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$

$$= \begin{pmatrix} (0.1 * 15) + (-0.1 * 20) + (0 * 10) \\ (-0.1 * 15) + (0.15 * 20) + (0 * 10) \\ (0 * 15) + (0 * 20) + (0.1 * 10) \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} -0.5 \\ 1.5 \\ 1 \end{pmatrix} \Leftrightarrow \boxed{\boldsymbol{\hat{\beta}_0 = -0.5}} \ ; \boxed{\boldsymbol{\hat{\beta}_1 = 1.5}} \ ; \ \boxed{\boldsymbol{\hat{\beta}_2 = 1}}$$

- Estimateur $\hat{\sigma}$ de σ

  L'estimateur $\widehat{\sigma^2}$ de σ^2 des moindres carrés ordinaires a pour expression

$$\widehat{\sigma^2} = \frac{\|\hat{\varepsilon}\|^2}{n-p-1} = \frac{\|y_i - \hat{y}\|^2}{n-p-1} = \frac{\|y_i\|^2 - \|\hat{y}\|^2}{n-p-1}$$

Or $\|y_i\|^2 = \sum {y_i}^2$ et $\|\hat{y}\|^2 = \hat{\beta}'(X'Y)$

De plus on sait que $Y'Y = \sum {y_i}^2$ donc $\boxed{\widehat{\sigma^2} = \frac{Y'Y - \hat{\beta}'(X'Y)}{n-p-1}}$

Avec $\hat{\beta}'(X'Y) = 32.5$; $Y'Y = 59.5$; $p = 2$; $n = 20$

On a donc $\widehat{\sigma^2} = \frac{59.5-32.5}{30-2-1}$ $\Leftrightarrow$ $\boxed{\boldsymbol{\widehat{\sigma^2} = 1}}$

3. Calcul d'intervalle et test

- Intervalle de confiance à 95% pour β_1
  Par définition : $IC_{95}(\beta_1) = \left[\hat{\beta}_1 \mp t_{n-p-1}\left(1-\frac{\alpha}{2}\right)\hat{\sigma}\sqrt{(X'X)^{-1}_{2,2}}\right]$

  Où $t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$ est le quantile de niveau $1-\frac{\alpha}{2}$ d'une loi de Student τ_{n-p-1}

  $\hat{\beta}_1 = 1.5$; $t_{n-p-1}\left(1-\frac{\alpha}{2}\right) = 2.052$; $\hat{\sigma} = 1$; $(X'X)^{-1}_{2,2} = 0.15$

AN: $IC_{95}(\beta_1) = \left[1.5 \mp 2.052 * 1\sqrt{0.15}\right] \Rightarrow$ $\boxed{\boldsymbol{IC_{95}(\beta_1) = [0.705; 2.295]}}$

- Test de $\beta = 0.8$ à niveau 10%
  Le test de significativité pour chaque coefficient est les suivant :

$$\begin{cases} H_0 : \beta = 0.8 \\ H_1 : \beta \neq 0.8 \end{cases}$$

On a la statistique suivante : $T_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim \tau_{n-p-1}$ avec

$j = (0,1,2)$; $\hat{\sigma}_{\hat{\beta}_j} = \hat{\sigma}\sqrt{(X'X)^{-1}_{j+1,j+1}}$

Règle de décision :

Si $|T_j| \geq t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$ alors on rejette H_0

Si $|T_j| < t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$ alors on ne peut pas rejeter H_0

Application :

$$T_0 = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{(X'X)^{-1}_{1,1}}} = \frac{-0.5 - 0.8}{1 * \sqrt{0.1}} = -4{,}111$$

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\sqrt{(X'X)^{-1}_{2,2}}} = \frac{1.5 - 0.8}{1 * \sqrt{0.15}} = 1{,}807$$

$$T_2 = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}\sqrt{(X'X)^{-1}_{3,3}}} = \frac{1 - 0.8}{1 * \sqrt{0.1}} = 0{,}663$$

$t_{n-p-1}\left(1-\frac{\alpha}{2}\right) = 1.703$

$|T_0| > t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$: On rejette H_0

$|T_1| > t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$: On rejette H_0

$|T_2| < t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$: On ne peut pas rejeter H_0

Ainsi au seuil de 10%, on rejette l'hypothèse de nullité statistique du coefficient associé à chaque variable explicative excepté celui associé à la variable $x_{i,2}$

4. **Testons $\boldsymbol{\beta_0} + \boldsymbol{\beta_1} = 3$ contre $\boldsymbol{\beta_0} + \boldsymbol{\beta_1} \neq 3$ au niveau 5%**

Hypothèses :

$$\begin{cases} H_0: \beta_0 + \beta_1 = 3 \\ H_1: \beta_0 + \beta_1 \neq 3 \end{cases}$$

On a la statistique suivante : $T = \frac{(\widehat{\beta}_0+\widehat{\beta}_1)-(\beta_0+\beta_1)}{\widehat{\sigma}_{(\widehat{\beta}_0+\widehat{\beta}_1)}} \sim \tau_{n-p-1}$ avec

$$\hat{\sigma}_{(\widehat{\beta}_0+\widehat{\beta}_1)} = \sqrt{\widehat{\sigma^2}_0 + 2cov\big((\hat{\beta}_0, \hat{\beta}_1\big) + \widehat{\sigma^2}_1}$$

$$= \hat{\sigma}\sqrt{(X'X)^{-1}_{1,1} + 2 * (X'X)^{-1}_{1,2} + (X'X)^{-1}_{2,2}}$$

Règle de décision :

Si $|T| \geq t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$ alors on rejette H_0

Si $|T| < t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$ alors on ne peut pas rejeter H_0

Application :

$t_{n-p-1}\left(1-\frac{\alpha}{2}\right) = 2.052$

$$T = \frac{(-0.5+1.5)-3}{1*\sqrt{0.1+2(-0.1)+0.15}} = -8.944$$

$|T| \geq t_{n-p-1}\left(1-\frac{\alpha}{2}\right)$ Alors au niveau 5% on rejette $H_0 : \beta_0 + \beta_1 = 3$ contre $\beta_0 + \beta_1 \neq 3$.

5. **Calcul de $\bar{y}$ et du coefficient de détermination ajusté**

- Calculons la moyenne empirique $\bar{y}$ des y_i

$$\bar{y} = \frac{1}{n}\sum y_i$$

On sait que $X'Y = \begin{pmatrix} \sum y_i \\ \sum x_{i,1} y_i \\ \sum x_{i,2} y_i \end{pmatrix}$; Or $X'Y = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$

Donc $\bar{y} = \frac{1}{n}\sum y_i = \frac{15}{30}$ $\Leftrightarrow$ $\boxed{\boldsymbol{\bar{y} = 0.5}}$

- Déduction du coefficient de détermination ajusté $R^2{}_a$

Par définition : $R^2{}_a = 1 - \frac{n-1}{n-p-1}\frac{SCR}{SCT} = 1 - \frac{n-1}{n-p-1}\frac{\|y_i - \hat{y}\|^2}{\|y_i - \bar{y}_i\|^2}$

Or $\widehat{\sigma^2} = \frac{SCR}{n-p-1} = \frac{\|y_i - \hat{y}\|^2}{n-p-1}$ et

$\|y_i - \bar{y}_i\|^2 = \|y_i\|^2 - \|\bar{y}_i\|^2 = \sum y_i^2 - \sum \bar{y}_i^2 = Y'Y - n\bar{y}^2$

Donc $\boxed{R^2{}_a = 1 - \frac{n-1}{Y'Y - n\bar{y}^2}\widehat{\sigma^2}}$

AN : $R^2{}_a = 1 - \frac{30-1}{59.5-(30*0.5^2)}1$ $\Leftrightarrow$ $\boxed{\boldsymbol{R^2{}_a = 0.442}}$

6. **Intervalle de prévision à 95% de $\boldsymbol{y_{n+1}}$ si $\boldsymbol{x_{n+1,1} = 3}$ et $\boldsymbol{x_{n+1,2} = 0.5}$**

$$\boxed{IP(y_{n+1}) = \left[x'_{n+1}\hat{\beta} \mp t_{n-p-1}\left(1 - \frac{\alpha}{2}\right)\hat{\sigma}\sqrt{1 + x'_{n+1}(X'X)^{-1}x_{n+1}}\right]}$$

Notons que $x'_{n+1} = (1;\ \ x_{n+1,1};\ \ x_{n+1,2}) = (1, 3, 0.5)$

et $\hat{\beta} = \begin{pmatrix} -0.5 \\ 1.5 \\ 1 \end{pmatrix}$ => $x'_{n+1}\hat{\beta} = (1, 3, 0.5) * \begin{pmatrix} -0.5 \\ 1.5 \\ 1 \end{pmatrix} = 4.5$;

$t_{n-p-1}\left(1 - \frac{\alpha}{2}\right) = 2.052$;

$$x'_{n+1}(X'X)^{-1} = (1, 3, 0.5) * \begin{pmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.15 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}$$

$$x'_{n+1}(X'X)^{-1} = (1 * 0.1 + 3 * (-0.1) + 0, 1 * (-0.1) + 3 * 0.15 + 0, 0 + \ 0 + 0.5 * 0.1)$$

$$x'_{n+1}(X'X)^{-1} = (-0.2, 0.35, 0.05)$$

$$x'_{n+1}(X'X)^{-1}x_{n+1} = (-0.2, 0.35, 0.05) * \begin{pmatrix} 1 \\ 3 \\ 0.5 \end{pmatrix}$$

$$= -0.2 * 1 + 0.35 * 3 + 0.05 * 0.5$$

$$x'_{n+1}(X'X)^{-1}x_{n+1} = 0.875$$

AN :

$$IP(y_{n+1}) = \left[4.5 \mp 2.052 * 1 * \sqrt{1 + 0.875}\right]$$

$$\boxed{\boldsymbol{IP(y_{n+1})} = [\mathbf{1.691}\ ; \mathbf{7.309}]}$$