Apprentissage Statistiques

Alioune CISSE

2 février 2021

Exercice 1 :

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

On a : $X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} \end{pmatrix} \Rightarrow X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{2,1} & \dots & x_{n,1} \\ x_{1,2} & x_{2,2} & \dots & x_{n,2} \end{pmatrix} \times \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} \end{pmatrix}$

$$X'X = \begin{pmatrix} n & \sum_{i=1}^n (x_{i,1}) & \sum_{i=1}^n (x_{i,2}) \\ \sum_{i=1}^n (x_{i,1}) & \sum_{i=1}^n (x_{i,1})^2 & \sum_{i=1}^n (x_{i,1} x_{i,2}) \\ \sum_{i=1}^n (x_{i,2}) & \sum_{i=1}^n (x_{i,1} x_{i,2}) & \sum_{i=1}^n (x_{i,2})^2 \end{pmatrix} = \begin{pmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{2,1} & \dots & x_{n,1} \\ x_{1,2} & x_{2,2} & \dots & x_{n,2} \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i,1} y_i \\ \sum_{i=1}^n x_{i,2} y_i \end{pmatrix} = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$$

$$Y'Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n y_i^2 = 59.5$$

Question 1 :

- Détermination de n :
  Par identification n = 30.
- La moyenne des $x_{i,2}$:
  $\bar{x}_2 = \frac{\sum_{i=1}^n (x_{i,2})}{n} = \frac{0}{30} = 0$
- Le coefficient de corrélation entre X_1 et X_2 :

$$\begin{aligned} \cos(\alpha) &= \frac{Cov(X_1, X_2)}{\sigma(X_1)\sigma(X_2)} = \frac{\frac{1}{n}\sum_{i=1}^n (x_{i,1} x_{i,2}) - \bar{x_1}\bar{x_2}}{\sqrt{\frac{1}{n}\sum_{i=1}^n (x_{i,1})^2 - \bar{x_1}^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (x_{i,2})^2 - \bar{x_2}^2}} \\ &= \frac{0}{\sqrt{\frac{1}{n}\sum_{i=1}^n (x_{i,1})^2 - \bar{x_1}^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (x_{i,2})^2 - \bar{x_2}^2}} = 0 \end{aligned}$$

  Interprétation : On a une absence totale de corrélation, les deux variables sont linéairement indépendantes.

Question 2 :

- Estimation de $\beta = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \end{pmatrix}'$ par la MCO :
  $\hat{\beta} = (X'X)^{-1} X'Y = a^{-1} b$ avec $a = X'X$ et $b = X'Y$

$a^{-1} = \begin{pmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix}^{-1}$
$det(a) = 30 \times 20 \times 10 - 20 \times 20 \times 10 = 2000$
$a^{'} = a$ et $Adj(a) = \begin{pmatrix} 20 \times 10 & 20 \times 10 & 0 \\ 20 \times 10 & 30 \times 10 & 0 \\ 0 & 0 & 30 \times 20 - 20 \times 20 \end{pmatrix} \times \begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$
$Adj(a) = \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix} \Rightarrow a^{-1} = \frac{1}{20} \begin{pmatrix} 2 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$
$b = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$
$\hat{\beta} = \frac{1}{20} \begin{pmatrix} 2 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} \times \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 3/2 \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$

— Estimation de σ^2 par la MCO :
$\hat{\sigma}^2 = \frac{SCR}{n-p-1}$;

$$\begin{aligned} SCR &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \hat{\beta}_2 x_{i,2})^2 \\ &= \sum_{i=1}^{n} y_i^2 + n\hat{\beta}_0^{\ 2} + \hat{\beta}_1^{\ 2} \sum_{i=1}^{n} x_{i,1}^2 + \hat{\beta}_2^{\ 2} \sum_{i=1}^{n} x_{i,2}^2 - 2\hat{\beta}_0 \sum_{i=1}^{n} y_i - 2\hat{\beta}_1 \sum_{i=1}^{n} x_{i,1} y_i \\ &- 2\hat{\beta}_2 \sum_{i=1}^{n} x_{i,2} y_i + 2\hat{\beta}_0\hat{\beta}_1 \sum_{i=1}^{n} x_{i,1} + 2\hat{\beta}_0\hat{\beta}_2 \sum_{i=1}^{n} x_{i,2} + 2\hat{\beta}_1\hat{\beta}_2 \sum_{i=1}^{n} x_{i,1} x_{i,2} \end{aligned}$$

$SCR = 59.5 + 7.5 + 45 + 10 + 15 - 60 - 20 - 30 + 0 + 0 = 27$
$SCR = 27 \Rightarrow \hat{\sigma}^2 = \frac{27}{27} = 1$

Question 3 :

— Intervalle de confiance pour β_1 à 95% :

$$\begin{aligned} IC_{\beta_1}^{\alpha} &= [\hat{\beta}_1 - t_{n-p-1}(1-\alpha/2)\hat{\sigma}\sqrt{(X'X)_{22}^{-1}}, \hat{\beta}_1 + t_{n-p-1}(1-\alpha/2)\hat{\sigma}\sqrt{(X'X)_{22}^{-1}}] \\ &= [1.5 - t_{27}(0.975)\sqrt{0.15}, 1.5 + t_{27}(0.975)\hat{\sigma}\sqrt{0.15}] \\ &= [1.5 - 2.052\sqrt{0.15}, 1.5 + 2.052\sqrt{0.15}] \\ &\approx [0.705, 2.295] \end{aligned}$$

— Test de $\beta_2 = 0.8$ à un niveau 10% :
Pour cela on calcule l'intervalle de confiance à un risque de 10% et on regarde si β_2 est dans cet intervalle.

$$\begin{aligned} IC^{\alpha}_{\beta_2} &= [\hat{\beta}_2 - t_{n-p-1}(1-\alpha/2)\hat{\sigma}\sqrt{(X'X)^{-1}_{33}}, \hat{\beta}_2 + t_{n-p-1}(1-\alpha/2)\hat{\sigma}\sqrt{(X'X)^{-1}_{33}}] \\ &= [1 - t_{27}(0.95)\sqrt{0.1}, 1 + t_{27}(0.95)\sqrt{0.1}] = [1 - 1.703\sqrt{0.1}, 1 + 1.703\sqrt{0.1}] \\ &\approx [0.461, 1.539] \end{aligned}$$

$\beta_2 = 0.8$ appartient à cet intervalle donc l'hypothèse que $\beta_2 = 0.8$ est accepté au risque $\alpha = 10\%$

Question 4 :

— Test $\beta_0 + \beta_1 = 3$ Vs $\beta_0 + \beta_1 \neq 3$ au risque $\alpha = 5\%$

$IC^{\alpha}_{\beta_0+\beta_1} \quad = \quad [(\hat{\beta}_0 + \hat{\beta}_1) - t_{n-p-1}(1-\alpha/2)\hat{\sigma}_{\hat{\beta}_0+\hat{\beta}_1}, (\hat{\beta}_0 + \hat{\beta}_1) + t_{n-p-1}(1-\alpha/2)\hat{\sigma}_{\hat{\beta}_0+\hat{\beta}_1}]$

$Var(A+B) = Var(A)+Var(B)+2Cov(A,B) \Rightarrow \hat{\sigma}^2_{\hat{\beta}_0+\hat{\beta}_1} = \hat{\sigma}^2_{\hat{\beta}_0}+\hat{\sigma}^2_{\hat{\beta}_1}+2Cov(\hat{\beta}_0, \hat{\beta}_1)$

$\hat{\sigma}^2_{\hat{\beta}_0+\hat{\beta}_1} = \hat{\sigma}^2[(X'X)^{-1}]_{11} + \hat{\sigma}^2[(X'X)^{-1}]_{22} + 2\hat{\sigma}^2[(X'X)^{-1}]_{12}$

$\hat{\sigma}_{\hat{\beta}_0+\hat{\beta}_1} = \hat{\sigma}\sqrt{[(X'X)^{-1}]_{11} + [(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{12}} = \sqrt{\frac{2}{20} + \frac{3}{20} + 2 \times \frac{-2}{20}}$

$\hat{\sigma}_{\hat{\beta}_0+\hat{\beta}_1} \approx 0.224$

$$\begin{aligned} IC^{\alpha}_{\beta_0+\beta_1} &= [(-0.5+1.5) - t_{27}(0.975) \times 0.224; (-0.5+1.5) + t_{27}(0.975) \times 0.224] \\ &= [1 - 2.052 \times 0.224; 1 + 2.052 \times 0.224] \\ &\approx [0.540; 1.460] \end{aligned}$$

Donc on rejette l'hypothèse H_0 selon laquelle $\beta_0 + \beta_1 = 3$ au risque $\alpha = 5\%$

Question 5 :

— Calcul de $\bar{y}$

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{15}{30} = 0.5$

— Le coefficient de détermination ajusté R^2_a :

$R^2_a = 1 - \frac{n-1}{n-p-1}\frac{SCR}{SCT} = 1 - \frac{n-1}{n-p-1}\frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2} = 1 - \frac{(n-1)\times 27}{27 \times (\sum_{i=1}^{n}(y_i - \bar{y_i})^2)}$

$R^2_a = 1 - \frac{n-1}{\sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2} = 1 - \frac{29}{59.5 - (2\times 0.5 \times 15) + (30 \times 0.5^2)} \approx 0.442$

Interprétation : Le modèle explique 44.2% de la variabilité de notre modèle.

Question 6 :

— Intervalle de prévision à 95% de y_{n+1} si $x_{n+1,1} = 3$ et $x_{n+1,2} = 0.5$
$IC^{\alpha}_{y_{n+1}} = [x'_{n+1}\hat{\beta} \pm t_{27}(0.975)\sigma\sqrt{1 + x'_{n+1}(X'X)^{-1}x_{n+1}}]$

$$\sqrt{1+x'_{n+1}(X'X)^{-1}x_{n+1}} = \sqrt{1+\begin{pmatrix}1 & 3 & 0.5\end{pmatrix}\begin{pmatrix}0.1 & -0.1 & 0\\ -0.1 & 0.15 & 0\\ 0 & 0 & 0.1\end{pmatrix}\begin{pmatrix}1\\3\\0.5\end{pmatrix}} = \sqrt{1.875}$$

$$x'_{n+1}\hat{\beta} = \begin{pmatrix}1 & 3 & 0.5\end{pmatrix}\begin{pmatrix}-0.5\\1.5\\1\end{pmatrix} = 4.5$$

$$IC^{\alpha}_{y_{n+1}} = [4.5 - 2.052\times\sqrt{1.875}; 4.5 + 2.052\times\sqrt{1.875}]$$

$$IC^{\alpha}_{y_{n+1}} = [1.690; 7.310]$$