# Apprentissage Statistiques

Fatoumata Badji

2 février 2021

## Exercice 1 :

**1. Déterminer n, la moyenne des xi,2, le coefficient de corrélation des xi,1 et des xi,2 :**

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

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

$$X'Y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i,1}y_i \\ \sum_{i=1}^{n} x_{i,2}y_i \end{pmatrix} = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$$

$Y'Y = \sum_{i=1}^{n} y_i^2 = 59.5$

— Par identification, on obtient 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 $X_2$ :

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

On remarque que nos deux variables ne sont pas corrélées, elles sont indépendantes.

**2. Estimer $\beta_0$, $\beta_1$, $\beta_2$, $\sigma_2$ par la méthode des moindres carrés ordinaires :**

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

$$(X'X)^{-1} = \begin{pmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix}^{-1}$$

$$Adj(a) = \frac{1}{20}\begin{pmatrix} 2 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}$$

$$\hat{\beta} = \frac{1}{20}\begin{pmatrix} 2 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} \times \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 3/2 \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

$\hat{\sigma}^2 = \frac{SCR}{n-p-1}$;

$$\begin{aligned} SCR &= \sum_{i=1}^{n}(y_i - \hat{y_i})^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \hat{\beta}_2 x_{i,2})^2 \\ &= \sum_{i=1}^{n} y_i^2 + n\hat{\beta_0}^2 + \hat{\beta_1}^2 \sum_{i=1}^{n} x_{i,1}^2 + \hat{\beta_2}^2 \sum_{i=1}^{n} x_{i,2}^2 - 2\hat{\beta}_0 \sum_{i=1}^{n} y_i - 2\hat{\beta}_1 \sum_{i=1}^{n} x_{i,1} y_i \\ &- 2\hat{\beta}_2 \sum_{i=1}^{n} x_{i,2} y_i + 2\hat{\beta}_0\hat{\beta}_1 \sum_{i=1}^{n} x_{i,1} + 2\hat{\beta}_0\hat{\beta}_2 \sum_{i=1}^{n} x_{i,2} + 2\hat{\beta}_1\hat{\beta}_2 \sum_{i=1}^{n} x_{i,1}x_{i,2} \end{aligned}$$

$SCR = 27$
$\hat{\sigma}^2 = 1$

**3. Calculer pour $\beta_1$ un intervalle de confiance à 95% et tester $\beta = 0.8$ a niveau 10% :**

— Intervalle de confiance pour $\beta_1$ à 95% :

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

$= [0.705, 2.295]$

— Testons $\beta_2 = 0.8$ à un niveau 10% : On pose l'hypothese nulle H0 : $\beta_2 = 0.8$

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

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

**4. Tester $\beta_0 + \beta_1 = 3$ Vs $\beta_0 + \beta_1 \neq 3$ au risque $\alpha = 5\%$ :**

On pose l'hypothese nulle H0 : $\beta_0 + \beta_1 = 3$

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

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

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

$\hat{\sigma}_{\hat{\beta}_0+\hat{\beta}_1} = 0.225$

$$\begin{aligned} IC^{\alpha}_{\beta_0+\beta_1} &= [(-0.5 + 1.5) - t_{27}(0.975) \times 0.224; (-0.5 + 1.5) + t_{27}(0.975) \times 0.224] \\ &= [1 - 2.052 \times 0.224; 1 + 2.052 \times 0.224] \\ &\approx [0.54; 1.46] \end{aligned}$$

L'hypothèse $H_0$ est cependant rejetée au risque $\alpha = 5\%$

**5. Calculer $\bar{y}$ et déduire le coefficient de détermination ajusté $R^2$ :**

— $\bar{y}$

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

— $R_a^2$ ajusté :

$R_a^2 = 1 - \frac{n-1}{n-p-1}\frac{SCR}{SCT} = 1 - \frac{n-1}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2}$

$R_a^2 = 1 - \frac{n-1}{\sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2} = 0.44$

Le modèle explique 44% de la variabilité de notre modèle.

**6. Construire un intervalle de prévision à 95% de $y_{n+1}$ si $x_{n+1,1} = 3$ et $x_{n+1,2} = 0.5$ :**

$IC_{y_{n+1}}^{\alpha} = [x_{n+1}^{'}\hat{\beta} \pm t_{27}(0.975)\sigma\sqrt{1 + x_{n+1}^{'}(X'X)^{-1}x_{n+1}}]$

$\sqrt{1 + x_{n+1}^{'}(X'X)^{-1}x_{n+1}} = \sqrt{1.875}$

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

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

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