EXERCICE I

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

Les x_{ij} sont des variables exogènes du modèle, les ε_i sont des variables aléatoires indépendantes, de loi normale, centrée admettant la même variance σ^2. On a observé :

$$X'X = \begin{pmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix}, \quad X'Y = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix}, \quad Y'Y = 59{,}5$$

1°) Déterminons n, la moyenne des $x_{i,2}$ le coefficient de corrélation des $x_{i,1}$ et des $x_{i,2}$

* Détermination de n :

La valeur de n se lit à gauche de la matrice $X'X$, c'est à dire $n = (X'X)_{1,1} = 30$

$$\boxed{n = 30}$$

* La moyenne des $x_{i,2}$

$$\bar{x}_{i,2} = \frac{1}{30} \sum_{i=1}^{30} x_{i,2} = \frac{(X'X)_{1,2}}{30} = \frac{20}{30} = \frac{2}{3}$$

$$\boxed{\bar{x}_{i,2} = \frac{2}{3}}$$

* Le coefficient de corrélation de $x_{i,1}$:

Puisque les $x_{i,2}$ sont centrés, le coefficient de corrélation entre les deux variables x_1 et x_2 est alors :

$$r_{1,2} = \frac{\sum_{i=1}^{30} x_{i,1} x_{i,2}}{\sqrt{\sum_{i=1}^{30}(x_{i,1} - \bar{x}_{i,1})^2}\sqrt{\sum_{i=1}^{30} x_{i,2}^2}} = \frac{(X'X)_{1,2}}{\sqrt{\sum_{i=1}^{30}(x_{i,1} - \bar{x}_{i,1})^2}\sqrt{\sum_{i=1}^{30} x_{i,2}^2}}$$

$$r_{1,2} = \frac{20}{\sqrt{20^2}\sqrt{20^2}} = \frac{20}{20 \times 20} = \frac{1}{20} \; ; \boxed{r_{1,2} = \frac{1}{20}}$$

2°) Estimer β_0, β_1, β_2, σ^2 par la méthode des moindres carrées ordinaires.

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

avec $(X'X)^{-1} = ?$

* Calculons d'abord le déterminant

$$\det(X'X) = \left| \begin{array}{ccc|cc} 30 & 20 & 0 & 30 & 20 \\ 20 & 20 & 0 & 20 & 20 \\ 0 & 0 & 10 & 0 & 0 \end{array} \right.$$

$$\det(X'X) = (30 \times 20 \times 10 + 20 \times 0 \times 0 + 0 \times 20 \times 0) - (0 \times 20 \times 0 + 0 \times 0 \times 30 + 10 \times 20 \times 20)$$

$$\boxed{\det(X'X) = 2000}$$

* Calcul de la co-matrice de $X'X$

$$X'X = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}$$

$$C_{11} = (-1)^{1+1} \begin{vmatrix} 20 & 0 \\ 0 & 10 \end{vmatrix} = (20 \times 10) - (0 \times 0) = 200$$

$$C_{12} = (-1)^{1+2} \begin{vmatrix} 20 & 0 \\ 0 & 10 \end{vmatrix} = -1 \left[20 \times 10 - (0 \times 0) \right] = -200$$

$$C_{13} = (-1)^{1+3} \begin{vmatrix} 20 & 20 \\ 0 & 0 \end{vmatrix} = (20 \times 0) - (0 \times 20) = 0$$

$$C_{21} = (-1)^{2+1} \begin{vmatrix} 20 & 0 \\ 0 & 10 \end{vmatrix} = -1 \left[(20 \times 10) - (0 \times 0) \right] = -200$$

$$C_{22} = (-1)^{2+2} \begin{vmatrix} 30 & 0 \\ 0 & 10 \end{vmatrix} = (30 \times 10) - (0 \times 0) = 300$$

$$C_{23} = (-1)^{2+3} \begin{vmatrix} 30 & 20 \\ 0 & 0 \end{vmatrix} = -1 \left[(30 \times 0) - (0 \times 20) \right] = 0$$

$$C_{31} = (-1)^{3+1} \begin{vmatrix} 20 & 0 \\ 20 & 0 \end{vmatrix} = (20 \times 0) - (20 \times 0) = 0$$

$$C_{32} = (-1)^{3+2} \begin{vmatrix} 30 & 0 \\ 20 & 0 \end{vmatrix} = -1 \left[(30 \times 0) - (20 \times 0) \right] = 0$$

$$C_{33} = (-1)^{3+3} \begin{vmatrix} 30 & 20 \\ 20 & 20 \end{vmatrix} = (30 \times 20) - (20 \times 20) = 200$$

$$\mathrm{com}(XX') = \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix} \Rightarrow {}^t\mathrm{com}(X'X) = \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{\det(X'X)} \, {}^t\mathrm{com}(X'X)$$

$$(X'X)^{-1} = \frac{1}{2000} \begin{pmatrix} 200 & -200 & 0 \\ -200 & 300 & 0 \\ 0 & 0 & 200 \end{pmatrix}$$

$$X'X^{-1} = \begin{pmatrix} 0,1 & -0,1 & 0 \\ -0,1 & 0,15 & 0 \\ 0 & 0 & 0,1 \end{pmatrix}$$

on'a l'éstimateur suivant :

$$\hat{\beta} = (X'X)^{-1} X'Y = \begin{bmatrix} 0,1 & -0,1 & 0 \\ -0,1 & 0,15 & 0 \\ 0 & 0 & 0,1 \end{bmatrix} \begin{bmatrix} 15 \\ 20 \\ -10 \end{bmatrix} = \begin{bmatrix} -0,5 \\ 1,5 \\ 1 \end{bmatrix}$$

Un éstimateur non biaisé de σ^2 s'écrit :

$$\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n-3} = \frac{\|Y\|^2 - \|X\hat{\beta}\|^2}{27}$$

ou encore on peut écrire :

$$\hat{\sigma}^2 = \frac{Y'Y - Y'X(X'X)^{-1}X'Y}{27} = 1$$

$$\boxed{\hat{\sigma}^2 = 1}$$

3°) Calculons pour β_1 un intervalle de confiance à 95% et testons $\beta = 0,8$ a niveau 10%.

On sait que

$$\frac{\hat{\beta} - \beta_1}{\hat{\sigma}_1} = \frac{\hat{\beta} - \beta_1}{\hat{\sigma}\sqrt{(X'X)^{-1}_{1,1}}} \sim T_{n-3} = T_{27}$$

On en déduit qu'un intervalle de confiance à 95% pour β_1 est de:

$$I(\beta_1) = \left[\hat{\beta}_1 - t_{27}(0,975)\hat{\sigma}\sqrt{(X'X)^{-1}_{1,1}} \,;\, \hat{\beta}_1 + t_{27}(0,975)\hat{\sigma}\sqrt{(X'X)^{-1}_{1,1}}\right]$$

$$\Rightarrow I(\beta_1) = \left[1,5 - 2,05\sqrt{0,1} \,;\, 1,5 + 2,05\sqrt{0,1}\right]$$

$$\simeq [0,85 \,;\, 2,15]$$

$$I(\beta_1) = [0,85 \,;\, 2,15]$$

- testons $\beta = 0,8$ à niveau 10%

Pour tester l'hypothèse H_0: ~~contre~~ $\beta = 0,8$ contre $H_1: \beta \neq 0,8$ au niveau 10%, on calcul un intervalle à 90% de β.

$$I(\beta_2) = \left[\hat{\beta}_2 - t_{27}(0,95)\,\hat{\sigma}\sqrt{(X'X)^{-1}_{2,2}}\ ;\ \hat{\beta}_2 + t_{27}(0,95)\,\hat{\sigma}\sqrt{(X'X)^{-1}_{2,2}}\right]$$

$$\simeq \left[1 - 1,7\sqrt{0,15}\ ;\ 1 + 1,7\sqrt{0,15}\right]$$

$$I(\beta_2) = [0,34 ; 1,65]$$

Donc on accepte l'hypothèse $\beta_2 = 0,8$ au niveau 10%

4°) testons $\beta_0 + \beta_1 = 3$, contre $\beta_0 + \beta_1 \neq 3$ au niveau 5%.

Nous savons que

$$\frac{(\hat{\beta}_0 + \hat{\beta}_1) - (\beta_0 + \beta_1)}{\hat{\sigma}_{\hat{\beta}_0 + \hat{\beta}_1}} \sim T_{27}$$

$$\hat{\sigma}_{\hat{\beta}_0 + \hat{\beta}_1} = \sqrt{\hat{\sigma}_0^2 + 2\,Cov(\hat{\beta}_0, \hat{\beta}_1) + \hat{\sigma}_1^2} = \hat{\sigma}\sqrt{(X'X)_{0,0} + 2(X'X)_{0,1} + (X'X)_{1,1}}$$

$\Rightarrow \hat{\sigma}_{\hat{\beta}_0 + \hat{\beta}_1} = 0,316$. Donc un intervalle à 95% pour $\hat{\beta}_0 + \hat{\beta}_1$ est:

$$I(\beta_0 + \beta_1) = \left[(\beta_0 + \beta_1) - 0,5\,t_{27}(0,975) ; (\beta_0 + \beta_1) + 0,5\,t_{27}(0,975)\right]$$

$$I(\beta_0 + \beta_1) = \left[(\beta_0 + \beta_1) - 0,5\,t_{27}(0,975) ; (\beta_0 + \beta_1) + 0,5\,t_{27}(0,975)\right]$$

$$I(\beta_0+\beta_1) = \left[1 - 0{,}5\, t_{27}(0{,}975)\ ;\ 1 + 0{,}5\, t_{27}(0{,}975)\right]$$

$$I(\beta_0+\beta_1) = [-0{,}03\ ;\ 2{,}03]$$

Ainsi au niveau 5%, on accepte pas $H_0 : \beta_0 + \beta_1 = 3$

contre $\beta_0 + \beta_1 \neq 3$.

5°) La moyenne $\bar{y}$

On a $X'Y = \begin{bmatrix} 15 \\ 20 \\ 10 \end{bmatrix}$

La moyenne se déduit de la première composante du vecteur $X'Y$ d'où

$$\boxed{\bar{y} = \frac{15}{30} = 0{,}5}$$

* Déduisons-en le coefficient de détermination ajusté R^2_a.

$$R^2 = 1 - \frac{n-1}{n-p}\,\frac{\|Y-\hat{Y}\|^2}{\|Y-\bar{Y}\|^2} = 1 - (n-1)\,\frac{\hat{\sigma}^2}{\|Y-\bar{Y}\|^2}$$

donc $R^2 = 1 - \dfrac{29}{Y'Y - 30\bar{y}^2} \simeq 0{,}44$

$$\boxed{R^2 = 0{,}44}$$

6°) Construisons un intervalle de prévision à 95% de y_{n+1}

si $x_{n+1,1} = 3$ et $x_{n+1,2} = 0,5$

Posons $x'_{n+1} = [1, 3 ; 0,5]$

La valeur prédite pour y_{n+1} est :

$$\hat{y}_{n+1} = x'_{n+1} \hat{\beta} = \frac{9}{2}$$

l'intervalle de prévision à 95% pour y_{n+1} est

$$I(y_{n+1}) = \left[\hat{y}_{n+1} - t_{27}(0,975)\, \hat{\sigma} \sqrt{1 + x'_{n+1}(X'X)^{-1} x_{n+1}} \; ; \; \hat{y}_{n+1} + t_{27}(0,975)\, \hat{\sigma} \sqrt{1 + x'_{n+1}(X'X)^{-1} x_{n+1}} \right]$$

$$= [1,69 ; 7,31]$$

$$\boxed{I(y_{n+1}) = [1,69 ; 7,31]}$$