

Solutions Exercices Régression linéaires, Série 2 (Multiple)

A2

Modèle: $y_i = \beta \ln x_i + \varepsilon_i$, $i=1,2$

* Par application de MCO,

$$\min_{\beta} \sum \varepsilon_i^2 = \min_{\beta} \sum (y_i - \beta \ln x_i)^2$$

* On applique les conditions du 1^{er} ordre:

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta} = 0 \Rightarrow 2 \sum \ln x_i (y_i - \beta \ln x_i) = 0$$

$$\Rightarrow \sum_{i=1}^2 y_i \ln x_i - \beta \sum (\ln x_i)^2 = 0$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{i=1}^2 y_i \ln x_i}{\sum (\ln x_i)^2} = \frac{y_1 \ln x_1 + y_2 \ln x_2}{(\ln x_1)^2 + (\ln x_2)^2}$$

* Si $x_1 = e$ et $x_2 = e^2$, il vient

$$\hat{\beta} = \frac{y_1 \cdot \ln e + y_2 \cdot \ln e^2}{(\ln e)^2 + (\ln e^2)^2} = \frac{y_1 + 2y_2}{1 + 2^2} = \frac{y_1 + 2y_2}{5}$$

\Rightarrow Réponse c

A4 . Modèle: $Y_i = \alpha \beta^{x_i} u_i$, $i = 1, \dots, 8$

* Linéarisation du modèle par passage aux logarithmes:

$$\ln Y_i = \ln \alpha + x_i \ln \beta + \ln u_i$$

i.e: $Y_i^* = \alpha^* + x_i \beta^* + \varepsilon_i$

$$\text{on } \left\{ \begin{array}{l} Y_i^* = \ln Y_i \\ \alpha^* = \ln \alpha \\ \beta^* = \ln \beta \end{array} \right.$$

* Application de MCO:

$$\min_{(\alpha, \beta)} \sum (\ln u_i)^2 = \min_{(\alpha, \beta)} \sum (\ln Y_i - \ln \alpha - x_i \ln \beta)^2$$

* Conditions du 1^{er} ordre:

$$\frac{\partial \sum (\ln u_i)^2}{\partial \beta} = 0 \Rightarrow \ln \beta = \frac{\sum x_i \ln Y_i - n \cdot \frac{\sum \ln x_i \cdot \bar{x}}{n}}{\sum x_i^2 - n \bar{x}^2}$$

$$\Rightarrow \ln \beta = \frac{\sum x_i \ln Y_i - (\sum \ln x_i) \cdot \bar{x}}{\sum x_i^2 - n \cdot \bar{x}^2}$$

* On a:

$$\sum x_i \ln Y_i = 52,36 \quad ; \quad \bar{x} = \frac{\sum x_i}{8} = \frac{36}{8} = 4,5$$

$$\sum x_i^2 = 204 \quad ; \quad \sum \ln Y_i = 8,61$$

$$A0: \ln \beta = \frac{52,36 - 8,61 \times 4,5}{204 - 8 \times (4,5)^2} = \frac{13,615}{42} = 0,324$$

* On prend l'exponentielle,

$$\hat{\beta} = e^{0,324} \approx e^{0,32}$$

\Rightarrow Réponse c

A5

$$\text{Modèle: } \sqrt{y_i} = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i)^{-1/2}$$

$$i = 1, \dots, 5$$

$$\text{ou } X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 6 & 3 & 4 \\ 2 & 3 & 4 & 5 & 6 \end{bmatrix} \text{ et } (X'X)^{-1} = \frac{1}{1263} \begin{bmatrix} 2141 & -299 & -274 \\ -299 & 86 & 7 \\ -274 & 7 & 74 \end{bmatrix}$$

$$* \text{ On pose: } y_i^* = (\sqrt{y_i})^{-2}$$

$$\Rightarrow y_i^* = \left[(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i)^{-1/2} \right]^2$$

$$\Rightarrow y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (\text{linéarisation})$$

$$* \text{ On a: } y_1^* = (\sqrt{y_1})^{-2} = (\sqrt{4})^{-2} = \frac{1}{4} = \frac{36}{144}$$

$$y_2^* = (\sqrt{y_2})^{-2} = (\sqrt{9})^{-2} = \frac{1}{9} = \frac{16}{144}$$

$$y_3^* = (\sqrt{y_3})^{-2} = (\sqrt{36})^{-2} = \frac{1}{36} = \frac{4}{144}$$

$$y_4^* = (\sqrt{y_4})^{-2} = (\sqrt{64})^{-2} = \frac{1}{64} = \frac{2,25}{144}$$

$$y_5^* = (\sqrt{y_5})^{-2} = (\sqrt{144})^{-2} = \frac{1}{144} = \frac{1}{144}$$

Donc,

$$X'Y^* = \begin{pmatrix} \sum y_i^* \\ \sum x_{1i} y_i^* \\ \sum x_{2i} y_i^* \end{pmatrix} = \begin{pmatrix} \frac{59,25}{144} \\ \frac{102,75}{144} \\ \frac{141,25}{144} \end{pmatrix}$$

$$\Rightarrow \hat{\beta}_1 = \frac{1}{1263} \left[-299 \left(\frac{59,25}{144} \right) + 86 \left(\frac{102,75}{144} \right) + 7 \left(\frac{141,25}{144} \right) \right]$$

$$= \frac{1}{1263} \left[-\frac{7890,5}{144} \right] = -0,043384908$$

Aucune réponse est juste, il avait une erreur!

A7 Modèle $y_i = \alpha e^{\beta x_i} u_i$, $i = 1, \dots, 7$

* On linéarise le modèle en travaillant avec le logarithme

$$\ln y_i = \ln \alpha + \beta x_i + \ln u_i$$

Puis $SCE = \sum (\ln u_i)^2$

En utilisant le nco ;

$$\hat{\beta} = \frac{\sum x_i \ln y_i - 7 \bar{x} \left(\frac{\sum \ln y_i}{7} \right)}{\sum x_i^2 - 7 \bar{x}^2}$$

On a :

$$\sum x_i \ln y_i = 105,24 \quad , \quad \bar{x} = 3$$

$$\sum x_i^2 = 91 \quad ; \quad \sum \ln y_i = 31,76$$

$$\Rightarrow \hat{\beta} = \frac{105,24 - 7 \times 3 \times \frac{31,76}{7}}{91 - 7 \times 3^2} = \frac{9,96}{28} = 0,355714285$$

$$\begin{aligned} \Rightarrow \hat{\alpha} &= e^{\left(\frac{\sum \ln y_i}{7} - \hat{\beta} \bar{x} \right)} \\ &= e^{\left(\frac{31,76}{7} - \frac{(9,96)}{28} \times 3 \right)} \\ &= e^{3,47} \approx 32,136 \approx 32 \end{aligned}$$

Réponse E.

A8 | modèle $y_i = \alpha e^{2x_i} + \varepsilon_i$,

* Par application de MCO:

$$\text{Min}_{\alpha} \sum \varepsilon_i^2 = \text{Min}_{\alpha} \sum (y_i - \alpha e^{2x_i})^2$$

* Conditions du 1^{er} ordre:

$$\frac{\partial \sum \varepsilon_i^2}{\partial \alpha} = 0 \Rightarrow 2 \sum e^{2x_i} (y_i - \alpha e^{2x_i}) = 0$$

$$\Rightarrow \sum y_i e^{2x_i} - \alpha \sum e^{4x_i} = 0$$

$$\Rightarrow \hat{\alpha} = \frac{\sum y_i e^{2x_i}}{\sum e^{4x_i}}$$

→ réponse B

A12 | Modèle $y_t = \beta_0 + \beta_1 (X_t + X_t^2) + \varepsilon_t, t=1, \dots, 6$

* Posons: $X_t^* = X_t + X_t^2$, Il vient

$$\text{MCO: } \hat{\beta}_1 = \frac{\sum X_t^* y_t - 6 \left(\frac{\sum X_t^*}{6} \right) \bar{y}}{\sum (X_t^*)^2 - 6 \left(\frac{\sum X_t^*}{6} \right)^2}$$

$$\hat{\beta}_1 = \frac{\sum (x_t + x_t^2) y_t - 6 \bar{y} \left(\frac{\sum (x_t + x_t^2)}{6} \right)}{\sum (x_t + x_t^2)^2 - 6 \left(\frac{\sum (x_t + x_t^2)}{6} \right)^2}$$

où $\sum (x_t + x_t^2) y_t = 110$; $\sum (x_t + x_t^2)^2 = 80$

$\sum (x_t + x_t^2) = 16$; $\bar{y} = \frac{14}{6}$

Donc, $\hat{\beta}_1 = \frac{110 - 6 \left(\frac{14}{6} \right) \left(\frac{16}{6} \right)}{80 - 6 \left(\frac{16}{6} \right)^2} = \frac{72,666\dots}{37,33} \approx 1,946$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \frac{\sum (x_t + x_t^2)}{6}$$

$$= \frac{14}{6} - 1,946428572 \times \frac{16}{6}$$

$$= -2,857142859$$

$$\hat{\beta}_0 \approx -2,9$$

Réponse A

A14

Modèle estimé :

$$\hat{Y}_i = 8 + 5X_{1i} + 2(X_{1i} - 4)X_{2i} + 9X_{3i} - 2X_{1i}X_{3i}$$

où X_{1i} : distance en km. du poste de pompier le plus près

$$X_{2i} = \begin{cases} 1 & \text{si } X_{1i} \geq 4 \\ 0 & \text{sinon} \end{cases}$$

$$X_{3i} = \begin{cases} 1 & \text{si ville A} \\ 0 & \text{si ville B} \end{cases}$$

Y_i : dommage causé par un feu en millier de dollars

- Si un feu a lieu à au moins 4 km du poste, on cherche la distance (X_{1i}) de la caserne la plus proche pour laquelle le montant de dommage moyen est le même à la ville A et à la ville B.

- Posons : \bar{Y}_i^A : montant de dommage, si ville est A

\bar{Y}_i^B : montant de dommage, si ville est B

- On cherche que :

$$\bar{Y}_i^A = \bar{Y}_i^B$$

l.e.

$$\begin{aligned} & 8 + 5x_{1i} + 2(x_{1i} - 4)(x_{2i} = 1) + 9(x_{3i} = 1) - 2x_{1i}(x_{3i} = 1) \\ & = 8 + 5x_{1i} + 2(x_{1i} - 4)(x_{2i} = 1) + 9(x_{3i} = 0) - 2x_{1i}(x_{3i} = 0) \end{aligned}$$

$$\Rightarrow 8 + 5x_{1i} + 2(x_{1i} - 4) + 9 - 2x_{1i} = 8 + 5x_{1i} + 2(x_{1i} - 4)$$

il vient,

$$8 + 9 - 8 = 5x_{1i} + 2(x_{1i} - 4) - 5x_{1i} - 2(x_{1i} - 4) + 2x_{1i}$$

$$9 = 2x_{1i} \Rightarrow x_{1i} = \frac{9}{2} = 4,5$$

\Rightarrow Réponse B

A15

Modèle: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + \varepsilon_i$

$$i = 1, \dots, 6$$

* On cherche $\hat{\beta}_2 + t_{(n-k, \alpha)} \sqrt{\frac{\hat{\sigma}^2}{\sum (X'X)^{-1}_{2,2}}}$

ou $t_{(n-k, \alpha)} = t_{(n-k, \alpha/2)} = t_{(6-4; 0,025)} = 4,303$

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix}$$

$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$
 $x_{2i} \quad x_{3i} \quad x_{2i}x_{3i}$

$$X'X = \begin{bmatrix} n & \sum x_{2i} & \sum x_{3i} & \sum x_{2i} x_{3i} \\ \sum x_{2i}^2 & \sum x_{2i} x_{3i} & \sum x_{2i} (x_{2i} x_{3i}) & \\ \sum x_{3i}^2 & \sum x_{3i} (x_{2i} x_{3i}) & \sum (x_{2i} x_{3i})^2 & \\ -1 & & & \end{bmatrix}$$

$$X'X = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$$\det(X'X) = 6 \times 4 \times 6 \times 4 \neq 0$$

$$\Rightarrow (X'X)^{-1} = \begin{bmatrix} 1/6 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix}$$

$$\Rightarrow X'y = \begin{bmatrix} \sum y_i \\ \sum x_{2i} y_i \\ \sum x_{3i} y_i \\ \sum (x_{2i} x_{3i}) y_i \end{bmatrix} = \begin{bmatrix} 150 \\ 12 \\ -6 \\ 44 \end{bmatrix}$$

done

$$\hat{\beta} = (X'X)^{-1} X'y = \begin{bmatrix} 1/6 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 150 \\ 12 \\ -6 \\ 44 \end{bmatrix} = \begin{bmatrix} 150/6 \\ 12/4 \\ -6/6 \\ 44/4 \end{bmatrix} = \begin{bmatrix} 25 \\ 3 \\ -1 \\ 11 \end{bmatrix}$$

$$\Rightarrow \hat{\beta}_2 = 3$$

Toujours $\hat{\sigma}_\varepsilon^2$

$$\text{On sait: } \hat{\sigma}_\varepsilon^2 = \frac{SES}{n-k} = \frac{\sum (y_i - \hat{y}_i)^2}{6-4}$$

où $k=4$: nombre de paramètres ($\beta_1, \beta_2, \beta_3$ et β_4)

$$\hat{y}_i = 25 + 3X_{2i} - 1 \cdot X_{3i} + 11 \cdot X_{2i} \cdot X_{3i}$$

$$\hat{i}=1, \quad \hat{y}_1 = 25 + 3 - 1 + 11 = 38$$

$$\hat{i}=2, \quad \hat{y}_2 = 25 + 3 + 1 - 11 = 18$$

$$\hat{i}=3, \quad \hat{y}_3 = 25 - 3 - 1 - 11 = 10$$

$$\hat{i}=4, \quad \hat{y}_4 = 25 - 3 + 1 + 11 = 34$$

$$\hat{i}=5, \quad \hat{y}_5 = 25 + 0 - 1 + 0 = 24$$

$$\hat{i}=6, \quad \hat{y}_6 = 25 + 0 + 1 + 0 = 26$$

$$\begin{aligned} \Rightarrow \sum (y_i - \hat{y}_i)^2 &= (28 - 38)^2 + (13 - 18)^2 + (0 - 10)^2 + (29 - 34)^2 \\ &\quad + (24 - 44)^2 + (26 - 36)^2 \\ &= 100 + 25 + 100 + 25 + 400 + 100 \\ &= 750 \end{aligned}$$

$$\Rightarrow \hat{\sigma}_\varepsilon^2 = \frac{750}{2} = 375$$

il vient, la limite supérieure de l'IC (β_2) est:

$$3 + 4,303 \cdot \sqrt{375 \cdot (1/4)} = 44,6636 \approx 45$$

\Rightarrow réponse E

(11)

B1

Y = salaire

X_2 : expérience

$X_3 = \begin{cases} 1 & \text{si homme} \\ 0 & \text{si femme} \end{cases}$

- Sous le modèle complet

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + \epsilon_i$$

On a : $SCE_c = 12,8156$ ("somme des carrés des erreurs")

avec $n - k = 7$ degré de liberté

- Sous le modèle réduit

$$Y_i = \beta_1^* + \beta_2^* X_{2i} + \epsilon_i^*$$

On a : $SCE_R = 27,7281$

$$\beta_1^* \neq \beta_1 \text{ et } \beta_2^* \neq \beta_2$$

- Test d'hypothèse : $H_0 : \beta_3 = \beta_4 = 0$

$H_1 : \beta_3 \neq 0$ ou $\beta_4 \neq 0$
(i.e. : teste si le sexe n'influence pas le salaire)

⇒ la statistique associée est celle de Fisher :

$$F_{obs} = \frac{(SCE_R - SCE_c)}{(n - k_R) - (n - k_c)} = \frac{SCE_R - SCE_c}{k_c - k_R} = \frac{SCE_c}{n - k_c}$$

k_R : nombre de paramètres du modèle réduit

k_c : nombre de paramètres dans le modèle complet. (12)

$$AN: \quad \overline{F}_{obs} = \frac{(27,7281 - 12,8156) / (4-2)}{12,8156 / 7} \approx 4,072673148$$

$$\overline{F}_{obs} \approx 4,1$$

⇒ Réponse D.

B2

- Modèle A : modèle complet avec 8 variables

$$On a \quad SCE_c = 76,893$$

$$n - k = 18 \text{ . degré de liberté'}$$

- Modèle B : modèle réduit

$$\left. \begin{array}{l} H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0 \\ H_1: \exists \beta_k \neq 0, k=4,5,6,7,8,9 \end{array} \right\}$$

Le modèle réduit, ~~complet~~ contient 3 paramètres ($\beta_1, \beta_2, \beta_3$)

$$On a \quad SCE_R = 126,471$$

Il vient,

$$\overline{F}_{obs} = \frac{(SCE_R - SCE_c) / (k_c - k_r)}{SCE_c / (n - k_c)}$$

$$= \frac{(126,471 - 76,893) / (9-3)}{76,893 / 18} \approx 1,934298311$$

$$\approx 1,9$$

⇒ Réponse D

B3

- Modèle : $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i, i=1, \dots, 10$

On a : $SCR_c = 61,3$ et $SCT = 128$

- On ajoute une 4^{ème} variable explicatives, soit X_4
on donne : $SCR_R = 65,6$ et $SCT = 128$

- On teste $\left. \begin{array}{l} H_0 : \beta_4 = 0 \\ H_1 : \beta_4 \neq 0 \end{array} \right\}$

* Sous H_0 : modèle réduit

on a $SCE_R = SCT - SCR_R = 128 - 61,3 = 66,7$

* Sous H_1 : modèle complet

$$SCE_c = SCT - SCR_c = 128 - 65,6 = 62,4$$

On a :
$$F_{obs} = \frac{(SCE_R - SCE_c) / (k_c - k_r)}{SCE_c / (n - k_c)}$$

$$F_{obs} = \frac{(66,7 - 62,4) / (4 - 3)}{62,4 / (10 - 4)} = 0,41346138 \approx 0,41$$

\Rightarrow Réponse B

B8

3 modèles et 20 observations

Modèle 1: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$, avec $SCE_1 = 484$

Modèle 2: $Y_i = \beta_1 + \beta_2 (X_{2i} + X_{3i}) + \varepsilon_i$, avec $SCE_2 = 925$

Modèle 3: $Y_i - X_{3i} = \beta_1 + \beta_2 (X_{2i} - X_{3i}) + \varepsilon_i$, avec $SCE_3 = 982$

- On veut tester: $H_0: \beta_2 + \beta_3 = 1 \Leftrightarrow \beta_3 = 1 - \beta_2$

$H_1: \beta_3 \neq 1 - \beta_2$

* Sous H_1 :

On utilise le modèle 1, puisque c'est le seul qui contient β_3 .

$\Leftrightarrow SCE_C = SCE_1$

* Sous H_0 :

On utilise le modèle 3, puisque: En partant du modèle 1, on a:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

soi $\beta_3 = 1 - \beta_2$ (H_0 :)

Alors: $Y_i = \beta_1 + \beta_2 X_{2i} + (1 - \beta_2) X_{3i} + \varepsilon_i$

$$Y_i = \beta_1 + \beta_2 X_{2i} + X_{3i} - \beta_2 X_{3i} + \varepsilon_i$$

$$Y_i - X_{3i} = \beta_1 + \beta_2 (X_{2i} - X_{3i}) + \varepsilon_i \quad (\text{modèle 3})$$

$$\Leftrightarrow SCE_R = SCE_3$$

$$F_{obs} = \frac{(SCE_3 - SCE_4) / (3 - 2)}{SCE_4 / (20 - 3)} = 17,49173554$$

\Rightarrow Réponse D

B9

Modèle $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \varepsilon_i$

$$i = 1, \dots, 20$$

- SCT = 15000

Test: $\left. \begin{array}{l} H_0: \beta_4 = \beta_5 = \beta_6 = 0 \\ H_1: \exists \beta_k \neq 0, k = 4, 5, 6 \end{array} \right\}$

- $R^2_{adj} = R^2_c = R^2$ du modèle $\approx 0,38$

- $SCR_R = 5565$ (sous H_0)

- Sous H_0 : (modèle réduit)

$$SCE_R = SCT - SCR_R = 15000 - 5565 = 9435$$

- Sous H_1 : (modèle complet)

On sait: $R^2_c = 1 - \frac{SCE_c}{SCT}$

$$\Rightarrow SCE_c = (1 - R^2_c) SCT = (1 - 0,38) \times 15000 \approx 9300$$

Il vient,

$$F_{obs} = \frac{(SCE_R - SCE_c) / (6-3)}{SCE_c / (3120-6)} = 15,06774194$$

⇒ Réponse D

C1

Modèle estimé:

$$\hat{y}_t = 90 + 30 x_{t1} + 0,15 x_{t2} + 20 x_{t1} x_{t3} + 10 x_{t4}$$

$x_{t1} = \begin{cases} 1 & \text{si mois de Janvier, Février, Juin, juillet, Août ou Décembre} \\ 0 & \text{sinon} \end{cases}$

$x_{t2} = (\text{Année en cours} - 1987)$

$x_{t3} = \begin{cases} 1 & \text{si température} > 90^\circ \text{ F en 2 jours ou plus} \\ 0 & \text{sinon} \end{cases}$

$x_{t4} = \text{nombre de tempête de neige dans le mois}$

$y_t = \text{Facture d'électricité mensuelle.}$

⊛ - Facture pour Février 1995

pour une tempête de neige:

$$\hat{y}_t = 90 + 30 (x_{t1} = 1) + 0,15 (1995 - 1987) + 20 (x_{t1} = 1) (x_{t3} = 1) + 10 (x_{t4} = 1)$$

$$= 90 + 30 + 0,15 \times 8 + 20 + 10$$

$$\hat{y}_t = 151,2$$

(*) Facture pour mars 1995

Pour 2 tempêtes de neige:

$$\begin{aligned} \hat{y}_t &= 90 + 30 (\chi_{t1} = 0) + 0,15 (1995 - 1987) + 20 (\chi_{t1} = 0)(\chi_{t3} = 1) \\ &\quad + 10 (\chi_{t4} = 2) \\ &= 90 + 0 + 0,15 \times 8 + 0 + 20 \\ &= 111,20 \end{aligned}$$

(*) Facture moyenne pour Février - Mars 1995

$$\frac{151,2 + 111,20}{2} = 131,2 \approx 131$$

⇒ Réponse A

c2 | Modèle estimé:

$$\hat{y}_i = 12 + 3X_i + 2(X_i - 5)D_{1i} - 2(X_i - 2)D_{2i} + (X_i - 5)D_{3i}$$

y_i : vente journalier

X_i : Années d'expérience du vendeur

$D_{1i} = \begin{cases} 1 & \text{si } X_i \geq 5 \\ 0 & \text{sinon} \end{cases}$; $D_{2i} = \begin{cases} 1 & \text{si New-York} \\ 0 & \text{sinon} \end{cases}$

$D_{3i} = \begin{cases} 1 & \text{si Chicago} \\ 0 & \text{sinon} \end{cases}$

* On cherche \hat{y} pour un vendeur de Chicago, ayant 7 ans d'expérience :

$$\begin{aligned}\hat{y} &= 12 + 3(x=7) + 2(x=7-5)(D_1=1) \\ &\quad - 2(x=7-2)(D_2=0) + (x=7-5)(D_3=1) \\ &= 12 + 21 + 4 - 0 + 2\end{aligned}$$

$$\hat{y} = 39$$

\Rightarrow Réponse F.

~~Verily~~