Licence 3 - SOE / EG 15 Mars 2022
Nom, Prénom : ____________ Université Côte d'Azur

Partiel - Économétrie

Tout outil électronique est prohibé, *e.g.* calculatrice, téléphone, montre connectée et *cie.*

Le barème des points est donné à titre indicatif et est destiné à évoluer à votre avantage

La qualité de la rédaction sera prise en compte.

Exercice 1 : Calcul de matrice (TD1)

Considérons la matrice

$$A = \begin{pmatrix} 2 & 3 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

(1) 1. Calculer la trace $\text{tr}(A)$ de A.

Solution:

$$\text{tr}(A) = 2 + 2 + 8 = 12$$

(1) 2. Quelle est la transposée de A ?

Solution:

$$A' = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

(2) 3. Calculer le déterminant de A.

Solution: On le fait par bloc.

$$\det(A) = 8 \times (2 \times 2 - 3) = 8$$

(1) 4. A est-elle idempotente ?

Solution: Non, son déterminant est $8 \neq 1$. Sinon, $(AA)_{33} = 8 \times 8 \neq 8$.

(2) 5. Calculer l'inverse A^{-1} de A.

Solution: Astuce : le faire par bloc *i.e.* on inverse la matrice 2×2 en haut à droite et on inverse le 8.

$$A^{-1} = \begin{pmatrix} 2 & -3 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1/8 \end{pmatrix}$$

Exercice 2 : Équation de Mincer (TD2)
On considère l'équation donnée par une régression linéaire classique, représentant la quantité moyenne de donnée qu'une entreprise peut collecter via une application F sur ses utilisateurs,

$$Y = 40 + 60 \times \frac{x}{12} - 20 \times d_1 - 5 \times d_2;$$ [1]

où Y est la quantité de donnée collectée en Mégaoctet par jour, x est le temps passé en heure sur l'application F, $d_1 = 1$ si l'utilisateur vit en Europe, et 0 sinon ; et $d_2 = 1$ si une sanction a récemment été donnée à l'entreprise pour violation des droits des citoyens, et 0 sinon.

(½) 1. (a) Quelle est la quantité moyenne de donnée prélevée à un Niçois (*i.e.* personne vivant à Nice), n'utilisant pas F, si F n'a jamais été sanctionnée ?
○ 50 ○ 40 √ **20**

(½) (b) Et si ce citoyen utilise l'application 6h par jour ?
○ 60 ○ 40 √ **50**

(½) (c) Et si ce même citoyen voyage en Amérique ?
○ 65 √ **70** ○ 50

(2) 2. On sait aussi que le nombre de donnée prélevées dépend de l'activité, dénotée par «act», exprimé en *logarithme*, de l'utilisateur sur F.

$$Y = 40 + 30 \times \frac{x}{12} - 20 \times d_1 - 5 \times d_2 + 30 \times \text{act}.$$

Quelle est l'interprétation du coefficient de l'activité «act» ?

○ Si l'activité augmente de 1 alors la collecte de données de F augmente de $0,3$Mo/J.

√ **Si l'activité augmente de 1% alors la collecte de données de F augmente de** $0,3$Mo/J.

○ Si l'activité augmente de 1 alors la collecte de données de F augmente de 30Mo/J.

○ Si l'activité augmente de 1% alors la collecte de données de F augmente de 30Mo/J.

(2) 3. Interprétez le coefficient de d_2.

Solution: Si F a eu une sanction récemment, alors elle réduit son apport journalier en données de 5Mo par utilisateur.

Exercice 3 : Moindres Carrés (TD3)
Note : Dans cet exercice, certains calculs peuvent être pénibles ; quand c'est raisonnable (*i.e.* produit / division de nombres non entiers) vous pouvez ne pas finir ces calculs.
On considère le modèle linéaire suivant

$$y_i = a + bx_i + \varepsilon_i,$$

avec ε_i indépendantes de loi vérifiant les hypothèses classique des MCO, et $i \in \{1, \ldots, 6\}$.
On dispose des observations suivantes.

y_i	2,1	3,8	11,9	8,2	10,3	-0,1
x_i	1	2	6	4	5	0

On donne :

$$\text{var}(x) = 4 + \frac{2}{3}, \quad \sum_i (y_i - \bar{y})^2 \simeq 115,4, \quad \sum_i (\hat{y}_i - \bar{y})^2 \simeq 115,2, \quad \sum_i (x_i - \bar{x})(y_i - \bar{y}) = 56,8.$$

1. Ces données sont fictives.

(2) 1. Estimer a et b. Vous pouvez donner les résultats sous forme de fraction de deux nombres.

Solution: Par le cours, on a

$$b \approx \hat{b} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n\mathrm{var}(x)} = \frac{56,8}{28} \approx 2,03$$

et

$$a \approx \hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{36,2}{6} - \frac{56,8}{28} * 3 \approx -0,05.$$

Dans la suite vous pouvez utiliser $\hat{a} = -0,05$ et $\hat{b} = 2,03$.

(2) 2. Établir le tableau d'analyse de la variance, calculer R^2, et interpréter.

Solution: L'énoncé donne SCT = 115,4, SCE = 115,2. De plus, par le cours SCT = SCR + SCE, donc SCR = 0,2. Finalement

$$R^2 = \frac{\mathrm{SCE}}{\mathrm{SCT}} = \frac{115,2}{115,4} \approx 0,999 \approx 1.$$

La variable y est donc expliquée à 99,9% par le modèle et x.

Source de variation	$\sum \cdot^2$	d.d.l.	carrés moyens
Régression	SCE = 115.2	1	SCE/1 = 115.2
Résidu	SCR = 0.2	$6 - 2 = 4$	SCR/(6 − 2) = 0.05
Total	SCT = 115.4	$6 - 1 = 5$	

(1) 3. Quelle est la meilleure approximation de σ^2 ? ○ 0,5 √ **0,05** ○ 0,01 ○ 0,001

Solution: Par le cours,

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{\mathrm{SCR}}{n-2} = \frac{\mathrm{SCT} - \mathrm{SCE}}{n-2} = \frac{115,4 - 115,2}{4} = \frac{0,2}{4} = 0,05$$

(1) 4. Quelle est la variance estimée de $\hat{a}$?

Solution: Par le cours,

$$\mathrm{var}(\hat{a}) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2} \approx \frac{\hat{\sigma}^2}{n} + \bar{x}^2 \frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2} \approx \frac{0,05}{6} + 3^2 * \frac{0,05}{28}$$

(1) 5. (a) Quelle est la t-stat du test de significativité de $\hat{a}$? On pourra utiliser $\sqrt{\mathrm{var}(\hat{a})} \approx 0.156 \approx 0.15$.
○ $-0,15$ ○ $-0,05$ ○ -1 √ $-1/3$ ○ $0,3$

(1) (b) La variable a est-elle significative à 5% ?
√ **non**
○ oui

Solution: On regarde la t-stat :

$$T = \frac{\hat{a}}{\sqrt{\mathrm{var}(\hat{a})}} = \frac{-0,05}{0,156} = \frac{-5}{15,6} \approx \frac{-5}{15} \approx -\frac{1}{3}$$

Le quantile (bilatéral) à 5% de la loi de Student à 6-2=4 degrés de liberté est 2,776 donc on doit garder l'hypothèse nulle : la variable n'est pas significative.