Économétrie

Corrigé détaillé – Examen blanc 1

Licence 3 – Économie

Anderson Vil
vladerinson@gmail.com

THEMA
CY Cergy Paris Université
Année universitaire 2026-2027

Exercice 1 – Canoës dans le Marais poitevin

10 points

Un loueur de canoës du Marais poitevin relève, pendant six jours de juin, la température maximale de la journée x (en degrés Celsius) et le nombre de canoës loués dans la journée y.

Jour i	1	2	3	4	5	6
Température x_i (°C)	19	21	23	25	27	29
Canoës loués y_i	23	30	34	35	43	57

1. (2,5 pts) Écrire le modèle de régression linéaire simple de y sur x en précisant ce que désignent y_i, x_i et u_i. Écrire les formules des estimateurs des moindres carrés ordinaires (MCO) $\hat{\beta}_1$ et $\hat{\beta}_0$, puis calculer leurs valeurs.

Solution. Le modèle s'écrit $y_i = \beta_0 + \beta_1 x_i + u_i$, pour $i = 1, \dots, 6$. y_i est le nombre de canoës loués le jour i, x_i la température maximale de ce jour, et u_i regroupe les autres facteurs qui agissent sur les locations ce jour-là (jour de la semaine, vent, vacances scolaires…), qui ne sont pas observés. Les estimateurs des MCO sont

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

On a $\bar{x} = 144/6 = 24$ et $\bar{y} = 222/6 = 37$.

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	19	23	−5	−14	70	25	196
2	21	30	−3	−7	21	9	49
3	23	34	−1	−3	3	1	9
4	25	35	1	−2	−2	1	4
5	27	43	3	6	18	9	36
6	29	57	5	20	100	25	400
$\sum$	144	222	0	0	210	70	694

Donc $\hat{\beta}_1 = 210/70 = 3$ et $\hat{\beta}_0 = 37 - 3 \times 24 = -35$. La droite des MCO est $\hat{y} = -35 + 3\,x$.

Barème. 0,5 pour le modèle et le sens de y_i, x_i, u_i ; 0,5 pour les deux formules ; 0,5 pour $\bar{x}$, $\bar{y}$ et les deux sommes ; 0,5 pour $\hat{\beta}_1$; 0,5 pour $\hat{\beta}_0$.

Durée estimée 9 min.

2. (1,5 pt) Interpréter $\hat{\beta}_1$ en une phrase, avec les unités. Dire si la constante $\hat{\beta}_0$ a ici une interprétation utile, en justifiant.

Solution. Sur ces six jours, un degré de température maximale en plus est associé à 3 canoës loués en plus dans la journée. La constante, -35, est le nombre de locations prédit pour une journée à 0 °C. Un nombre négatif de locations est impossible, et 0 °C est très loin des températures observées, de 19 à 29 °C. La droite n'a aucune raison d'être valable aussi loin des données, et la constante n'a pas d'interprétation utile.

Barème. 0,75 pour la pente avec ses unités (0,25 retiré pour une lecture causale affirmée sans réserve) ; 0,25 pour la lecture de la constante comme valeur prédite en $x = 0$; 0,5 pour la conclusion justifiée par la plage des données.

Durée estimée 3 min.

3. (1,5 pt) Calculer la valeur prédite et le résidu du jour 4. Dire si la droite surestime ou sous-estime les locations de ce jour.

> **Solution.** $\hat{y}_4 = -35 + 3 \times 25 = 40$ canoës et $\hat{u}_4 = y_4 - \hat{y}_4 = 35 - 40 = -5$ canoës. Le résidu est négatif, le point est sous la droite. La droite surestime les locations du jour 4 de 5 canoës.

Barème. 0,5 pour la valeur prédite ; 0,5 pour le résidu ; 0,5 pour la conclusion tirée du signe du résidu.

Durée estimée 4 min.

4. (1 pt) Calculer le nombre de locations que la droite prévoit pour une journée à 8 °C. Dire si cette température se trouve dans la plage des données et ce que cela implique pour la prévision.

> **Solution.** $\hat{y} = -35 + 3 \times 8 = -11$ canoës, une valeur impossible. 8 °C est hors de la plage observée (19 à 29 °C). Rien ne garantit que la relation reste linéaire en dehors de cette plage, et la droite ne doit pas servir à prévoir aussi loin des données.

Barème. 0,5 pour la prévision ; 0,5 pour le constat hors plage et sa conséquence.

Durée estimée 3 min.

5. (2 pts) Calculer la somme des carrés totale SCT et la somme des carrés expliquée SCE, puis le coefficient de détermination R^2. Interpréter le R^2 en une phrase.

> **Solution.** $\mathrm{SCT} = \sum_i (y_i - \bar{y})^2 = 694$, dernière colonne du tableau. Comme $\hat{y}_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x})$, $\mathrm{SCE} = \hat{\beta}_1^2 \sum_i (x_i - \bar{x})^2 = 9 \times 70 = 630$. Le calcul direct donne le même résultat, avec les valeurs prédites 22, 28, 34, 40, 46 et 52, d'où $\mathrm{SCE} = 225 + 81 + 9 + 9 + 81 + 225 = 630$. Donc
>
> $$R^2 = \frac{\mathrm{SCE}}{\mathrm{SCT}} = \frac{630}{694} = 0{,}908.$$
>
> La température explique 90,8 % de la variation du nombre de locations sur ces six jours. On peut vérifier $\mathrm{SCR} = 694 - 630 = 64$, somme des carrés des résidus $1, 2, 0, -5, -3, 5$.

Barème. 0,5 pour SCT ; 0,5 pour SCE (ou SCR puis SCE = SCT − SCR) ; 0,5 pour R^2 ; 0,5 pour l'interprétation, qui doit parler de la variation de y dans l'échantillon.

Durée estimée 7 min.

6. (1,5 pt) On suppose les hypothèses RLS.1 à RLS.5 satisfaites, avec $\sigma^2 = 14$. Calculer $\mathrm{Var}(\hat{\beta}_1)$ et l'écart-type de $\hat{\beta}_1$. Indiquer laquelle de ces cinq hypothèses ne sert pas à l'absence de biais mais seulement à cette formule.

> **Solution.** À $x_1, \ldots, x_6$ donnés, $\mathrm{Var}(\hat{\beta}_1) = \sigma^2 / \mathrm{SCT}_x$ avec $\mathrm{SCT}_x = \sum_i (x_i - \bar{x})^2 = 70$. Donc $\mathrm{Var}(\hat{\beta}_1) = 14/70 = 0{,}2$ et $\mathrm{sd}(\hat{\beta}_1) = \sqrt{0{,}2} = 0{,}45$ canoë par degré. L'hypothèse d'homoscédasticité RLS.5, $\mathrm{Var}(u \mid x) = \sigma^2$, ne sert qu'à cette formule. L'absence de biais repose sur RLS.1 à RLS.4.

Barème. 0,5 pour la variance ; 0,5 pour l'écart-type ; 0,5 pour RLS.5.

Durée estimée 4 min.

Erreurs fréquentes. Écrire le résidu $\hat{y}_4 - y_4$ au lieu de $y_4 - \hat{y}_4$, ce qui inverse la conclusion de la question 3. Confondre le résidu $\hat{u}_i$, calculé, et l'erreur u_i, jamais observée. Lire la pente comme un effet causal de la température, alors que les jours chauds peuvent aussi être des jours de week-end ou de vacances. Interpréter la constante ou prévoir à 8 °C sans regarder la plage des données. Diviser une seule des deux sommes par $n - 1$ dans $\hat{\beta}_1$. Donner la variance (0,2) comme écart-type, ou citer RLS.4 à la question 6.

Exercice 2 – Revenu et dépenses culturelles — 10 points

Une enquête auprès de 400 ménages niortais relève les dépenses culturelles du ménage (`depcult`, en euros par an, pour le cinéma, les spectacles, les livres et les musées), son revenu (`revenu`, en milliers d'euros par mois) et le nombre d'années d'études de la personne de référence (`etudes`). On obtient les régressions suivantes par les MCO.

	(1)	(2)	(3)
Variable dépendante	`depcult`	`etudes`	`depcult`
`revenu`	91,89***	1,55***	
	(12,11)	(0,10)	
`etudes`			41,35***
			(4,77)
Constante	315,94***	5,47***	224,06***
	(59,64)	(0,48)	(62,78)
n	400	400	400
R^2	0,126	0,387	0,159

Erreurs-types entre parenthèses. *** significatif à 1 %, ** à 5 %, * à 10 %.

1. (2 pts) Écrire l'équation estimée de la colonne (1). Interpréter la pente en une phrase, avec les unités. Calculer les dépenses culturelles prédites pour un ménage dont le revenu est de 4 000 euros par mois.

> **Solution.** $\widehat{\texttt{depcult}} = 315{,}94 + 91{,}89\,\texttt{revenu}$, avec $n = 400$ et $R^2 = 0{,}126$. Un revenu supérieur de 1 000 euros par mois est associé à des dépenses culturelles supérieures de 91,89 euros par an. Un revenu de 4 000 euros correspond à $\texttt{revenu} = 4$, d'où $315{,}94 + 91{,}89 \times 4 = 315{,}94 + 367{,}56 = 683{,}50$ euros par an.

Barème. 0,5 pour l'équation ; 0,75 pour l'interprétation avec les deux unités ; 0,75 pour la prédiction (0,25 retiré si l'on remplace `revenu` par 4 000).

Durée estimée 5 min.

2. (2 pts) On exprime désormais les dépenses culturelles en centaines d'euros par an. Donner, sans nouvelle estimation, la pente, la constante et le R^2 de la colonne (1), en justifiant chaque réponse.

> **Solution.** La nouvelle variable est $\texttt{depcult}/100$. Diviser chaque y_i par 100 divise $\bar{y}$ et chaque écart $y_i - \bar{y}$ par 100. La formule de $\hat{\beta}_1$ donne une pente divisée par 100, 0,9189 centaine d'euros par an pour 1 000 euros de revenu mensuel, et $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ est aussi divisée par 100, soit 3,1594. SCE et SCT sont toutes deux divisées par 100^2, et le R^2 reste égal à 0,126. Le changement d'unité ne modifie pas la relation, seulement la façon de l'écrire.

Barème. 0,75 pour la pente ; 0,75 pour la constante ; 0,5 pour le R^2. Chaque valeur sans justification ne rapporte que la moitié des points.

Durée estimée 5 min.

3. (2,5 pts) Le nombre d'années d'études fait partie du terme d'erreur u de la colonne (1). À l'aide des colonnes (2) et (3), dire si l'hypothèse RLS.4 est plausible, puis dans quel sens la pente de la colonne (1) s'écarte vraisemblablement de l'effet *ceteris paribus* du revenu. Citer un autre facteur contenu dans u.

> **Solution.** La colonne (2) montre que les ménages plus aisés ont une personne de référence plus diplômée, avec 1,55 année d'études de plus par millier d'euros de revenu mensuel. La moyenne de `etudes`, donc celle de u, varie avec le revenu. $E(u \mid \texttt{revenu})$ dépend alors du revenu, et RLS.4 n'est pas plausible. La colonne (3) montre que les ménages plus diplômés dépensent davantage, 41,35 euros par an par année d'études. Les ménages à haut revenu ont donc aussi des facteurs inobservés qui poussent les dépenses vers le haut. La pente de la colonne (1) attribue au revenu une partie de l'effet des études et surestime vraisemblablement l'effet *ceteris paribus* du revenu (biais vers le haut). Autres facteurs contenus dans u, les goûts culturels, l'âge, la taille du ménage

ou la proximité des salles de spectacle.

Barème. 0,5 pour la lecture de la colonne (2) ; 0,5 pour la lecture de la colonne (3) ; 0,5 pour la conclusion sur RLS.4 ; 0,75 pour le sens du biais justifié par les deux signes ; 0,25 pour un autre facteur pertinent.

Durée estimée 8 min.

4. (2 pts) L'écart-type des résidus de la colonne (1) vaut 149, 175, 209 et 316 euros dans les quatre quarts de revenu, du plus faible au plus fort. Indiquer l'hypothèse que ces chiffres permettent d'examiner et dire si elle est plausible. Préciser la conséquence pour la pente des MCO et pour la formule $\mathrm{Var}(\hat{\beta}_1) = \sigma^2/\mathrm{SCT}_x$.

Solution. Ces chiffres examinent l'homoscédasticité RLS.5, $\mathrm{Var}(u \mid x) = \sigma^2$. La dispersion des résidus double presque entre le premier et le dernier quart. $\mathrm{Var}(u \mid \texttt{revenu})$ augmente avec le revenu, et RLS.5 n'est pas plausible (hétéroscédasticité). L'hétéroscédasticité ne crée aucun biais, puisque l'absence de biais ne repose que sur RLS.1 à RLS.4. Le problème de la pente vient ici de RLS.4 (question 3). En revanche, la formule σ^2/SCT_x n'est plus exacte et peut faire croire à une précision que la pente n'a pas.

Barème. 0,5 pour RLS.5 ; 0,5 pour la conclusion « non plausible » tirée des chiffres ; 0,5 pour l'absence de biais due à la seule hétéroscédasticité ; 0,5 pour la formule de la variance qui devient fausse.

Durée estimée 5 min.

5. (1,5 pt) Une autre enquête tire au hasard quatre fois plus de ménages dans la même population. Donner le facteur par lequel l'écart-type de la pente serait multiplié si la formule $\mathrm{Var}(\hat{\beta}_1) = \sigma^2/\mathrm{SCT}_x$ s'appliquait, en justifiant. Dire si l'écart discuté à la question 3 disparaîtrait.

Solution. $\mathrm{SCT}_x = (n-1)s_x^2$. Avec quatre fois plus de ménages et la même répartition des revenus, SCT_x est environ multiplié par 4. La variance est divisée par 4 et l'écart-type est multiplié par $1/\sqrt{4} = 1/2$. L'écart de la question 3 ne disparaîtrait pas. Il vient de la liaison entre le revenu et les études, qui ne dépend pas de la taille de l'échantillon. Un échantillon plus grand donne une pente plus précise, autour d'une mauvaise valeur.

Barème. 0,5 pour le facteur $1/2$; 0,5 pour la justification par SCT_x ; 0,5 pour le biais qui ne disparaît pas.

Durée estimée 4 min.

Erreurs fréquentes. Interpréter 91,89 comme l'effet d'un euro de revenu, ou oublier que les dépenses sont annuelles et le revenu mensuel. Remplacer `revenu` par 4 000 à la question 1. Laisser la constante inchangée, ou modifier le R^2, à la question 2. Lire la colonne (3) comme un effet causal des études, alors qu'elle ne sert qu'à indiquer le signe probable de leur effet. Inverser le sens du biais. Croire que l'hétéroscédasticité biaise la pente, ou qu'un échantillon plus grand corrige le biais, ce qui revient à confondre biais et variance.