

Exercices pour le cours Économétrie et évaluations statistiques (Ch. IV).

Propriétés statistiques des MCO I: Espérance des estimateurs des MCO – Variance des estimateurs des MCO

Exercice 1. Soit le modèle de régression simple:

$$y = \beta_0 + \beta_1 x + u.$$

Dans ce cas, l'estimateur des MCO du paramètre de la pente est:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x}) y_i}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

En précisant toutes les propriétés utilisées, prouvez que

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

[Indice: La solution se trouve dans: Wooldridge, Jeffrey M., Introductory Econometrics: A Modern Approach, 5th Edition. South-Western College Pub (2013)].

Exercice 2. Supposez que, dans une firme, la productivité moyenne des travailleurs (avgprod) dépend de deux facteurs, à savoir, les heures moyennes de formation de ces travailleurs (avgtrain) et les habiletés moyennes de ces travailleurs (avgabil):

$$\text{avgprod} = \beta_0 + \beta_1 \cdot \text{avgtrain} + \beta_2 \cdot \text{avgabil} + u$$

Supposez que cette équation satisfait les hypothèses de Gauss-Markov. Si les firmes qui engagent des travailleurs dont les habiletés sont inférieures à la moyenne reçoivent une subvention pour la formation des travailleurs, si bien que avgtrain et avgabil sont négativement corrélés, quel est le biais probable dans l'estimateur de β_1 si l'on estime:

$$\text{avgprod} = \beta_0 + \beta_1 \cdot \text{avgtrain} + u$$

Exercice 3. Soit la base de données HTV.xls qui comprend des informations sur le salaire, l'éducation (educ) en années, l'éducation des parents (motheduc pour la mère et fatheduc pour le père) en années et plusieurs autres variables pour 1230 hommes qui travaillent en 1991.

1. En utilisant la fonction DROITEREG de Excel, estimez le modèle de régression suivant:

$$\text{educ} = \beta_0 + \beta_1 \text{motheduc} + \beta_2 \text{fatheduc} + u$$

par les MCO et rapportez les résultats sous forme d'équation. Quelle est la proportion de la variation de l'éducation qui s'explique par la variation l'éducation des parents? Interprétez le coefficient associé à l'éducation de la mère.

2. Ajoutez la variable abil (une mesure de la capacité cognitive) à la régression de (a), et rapportez les résultats sous forme d'équation. L'habilité aide-t-elle à expliquer l'éducation, même après avoir contrôlé pour l'éducation des parents? Expliquez.

Exercice 4. Dans une étude portant sur le lien entre la moyenne (GPA or Grade Point Average) obtenue à l'université et le temps consacré à diverses activités, des étudiants ont été interrogés sur le nombre d'heures qu'ils passent chaque semaine dans quatre activités: étudier (study), dormir (sleep), travailler (work) et prendre des loisirs (leisure). Chaque heure passée est classée dans l'une de ces quatre catégories, de sorte que pour chaque étudiant, la somme des heures consacrées aux quatre activités est de 168 heures.

1. Dans le modèle suivant

$$GPA = \beta_0 + \beta_1 \text{study} + \beta_2 \text{sleep} + \beta_3 \text{work} + \beta_4 \text{leisure} + u,$$

est-ce que cela a du sens de maintenir a-t-il les temps de sommeil, de travail et de loisir constants tout en modifiant le temps d'étude?

2. Expliquez pourquoi ce modèle viole l'hypothèse H4 (absence de colinéarité parfaite)
3. Comment pourriez-vous reformuler le modèle afin que ses paramètres aient une interprétation utile et qu'ils satisfassent l'hypothèse H4?

Exercice 5. Un économètre a calculé les statistiques suivantes à l'aide d'un échantillon de 25 observations:

$$\begin{aligned} \sum_{i=1}^N x_i &= 50, & \sum_{i=1}^N y_i &= 100, \\ \sum_{i=1}^N x_i^2 &= 200, & \sum_{i=1}^N y_i^2 &= 500, & \sum_{i=1}^N x_i y_i &= 150 \end{aligned}$$

1. Calculez les variances empiriques de x , de y , et la covariance entre x et y . Quelle interprétation peut-on donner à la variance et à la covariance?
2. Calculez les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$.
3. Sachant que la variance du terme d'erreur est donné par σ^2 , calculez la variance de $\hat{\beta}_0$ et celle de $\hat{\beta}_1$.
4. Calculez le R^2 de la régression.
5. Supposez que les unités de la variable explicative soient changées de la manière suivante:

$$x^* = a + bx$$

où a et b sont des constantes différentes de 0, et x^* est la variable explicative dans les nouvelles unités. Comment se modifient les estimations a et b ? Comment se modifie le R^2 de la régression.

Exercice 6. Un chercheur utilise un échantillon de 10 observations pour étudier la relation entre une variable dépendante y_i et une variable indépendante x_i . Une analyse préliminaire de l'échantillon fournit les statistiques suivantes:

$$\begin{aligned} \sum_{i=1}^N y_i &= 100 & \sum_{i=1}^N x_i &= 50 & \sum_{i=1}^N y_i^2 &= 5000 & \sum_{i=1}^N x_i^2 &= 1000 \\ \sum_{i=1}^N x_i y_i &= 2000, \end{aligned}$$

avec $N = 10$. Utilisez ces informations pour:

1. Calculer les estimateurs des MCO de la constante β_0 et de la pente β_1 .
2. Calculer la valeur du R-carré, le coefficient de détermination, de la droite de régression de l'échantillon.
3. Sachant que l'estimateur $\hat{\sigma}^2$ de la variance du terme d'erreur est égale à 125, calculer l'estimateur de la variance de $\hat{\beta}_0$ et de $\hat{\beta}_1$.
4. La valeur donnée en (c) pour l'estimateur $\hat{\sigma}^2$ de la variance du terme d'erreur est-elle compatible avec les données de l'exercice? Prouvez-le.

Rappel (Wooldridge, Appendix A.1):

$$\begin{aligned} \sum_{i=1}^N (x_i - \bar{x})^2 &= \sum_{i=1}^N x_i^2 - N\bar{x}^2 \\ \sum_{i=1}^N (y_i - \bar{y})^2 &= \sum_{i=1}^N y_i^2 - N\bar{y}^2 \\ \sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x}) &= \sum_{i=1}^N y_i x_i - N\bar{y}\bar{x}. \end{aligned}$$