

Exercices pour le cours Econométrie et évaluations statistiques (Ch. II).

Interprétation des MCO – Modèle de régression multiple

Exercice 1. Soit un modèle de régression simple estimé sur un échantillon de couples et permettant d'expliquer le nombre d'enfants (nbchild) par femme en fonction du niveau d'éducation de la mère (motheduc) mesuré en années:

$$\text{nbchild} = \beta_0 + \beta_1 \cdot \text{motheduc} + u.$$

1. Selon vous, quelles sont les variables qui sont contenues dans le terme d'erreur? Pensez-vous que ces variables sont corrélées avec la variable motheduc?
2. Si ces variables sont corrélées avec motheduc, quelles seront les conséquences en termes de biais des estimateurs? Soyez précis: caractérisez le signe plausible du biais pour une ou deux des variables que vous avez considérées.
3. Si les estimateurs sont biaisés, que suggérez comme solution pour corriger ce biais?

Exercice 2. Soit l'équation de consommation suivante:

$$\text{totexp} = \beta_0 + \beta_1 \cdot \text{income} + u.$$

où totexp est la dépense totale d'un ménage sur une année et income son revenu total annuel. La propensité marginale à consommer est égale à la pente β_1 . En utilisant des observations pour 100 ménages sur la dépense annuelle totale et sur le revenu total annuel (les deux mesurés en euros), l'équation suivante est obtenue:

$$\widehat{\text{totexp}} = -124.84 + 0.853 \cdot \text{income}$$

1. Interprétez la constante de cette équation et commentez son signe et son amplitude.
2. Quel est le niveau de dépense totale prédit si le revenu du ménage est de 30.000 euros?

Exercice 3. En se basant sur des données de 1988 portant sur la vente de maisons à Andover, Massachusetts, de Kiel and McClain (1995), on obtient l'équation suivante qui explique le prix des maisons (price) par la distance de celles-ci à un incinérateur à ordures récemment installé (dist):

$$\widehat{\log(\text{price})} = 9.40 + 0,312 \cdot \log(\text{dist})$$

$$N = 135, \quad R^2 = 0,162$$

1. Est-ce que le signe de β_1 est celui qui était attendu?
2. Pensez-vous que cette simple régression fournit un effet *ceteris paribus* (toutes autres choses étant égales) de l'effet de la distance sur le prix? [Indice: Pensez au choix de la ville de placer l'incinérateur.] Quels sont les autres facteurs qui peuvent expliquer le prix des maisons? Ceux-ci peuvent-ils être corrélés avec la distance à l'incinérateur?

Exercice 4. Considérez les données GPA2.xls sur 4137 étudiants en première cycle universitaire ("college"). Soit le modèle suivant:

$$\text{colGPA} = \beta_0 + \beta_1 \text{hsperc} + \beta_2 \text{sat} + u$$

où colGPA est le résultat moyen aux examens de l'étudiant, mesuré sur une échelle de 4 points, hsperc est le percentile de classement de l'étudiant en classe de terminale ("high school percentile class"), si bien que perc= 5 signifie que l'étudiant figurait parmi les 5% les meilleurs, et sat est le résultat à un test standardisé, appelé SAT, de l'étudiant avant l'entrée à l'université.

1. En utilisant la fonction DROITEREG de Excel, estimez les paramètres de ce modèle de régression par les MCO. Estimez également le coefficient de détermination?
2. Est-ce que le signe de l'estimateur de β_1 est celui attendu?
3. Quelle est la valeur prédite de colGPA quand perc= 20 et sat= 1050?

4. Supposons que deux étudiants A et B sont classés dans le même percentile en terminale, mais que le résultat au test SAT de l'étudiant A était 140 points plus élevés que celui de l'étudiant B. Quelle est la différence prédite pour les examens à l'université pour ces deux étudiants?
5. En maintenant `hsperc` fixé, quelle différence dans la variable `sat` mène à une différence de 0,50 dans `colGPA`?

Exercice 5. Un problème qui intéresse les responsables de la santé est de déterminer les effets du tabagisme pendant la grossesse sur la santé du nourrisson. Une mesure de la santé infantile est le poids à la naissance; un poids de naissance trop faible peut exposer un bébé au risque de contracter diverses maladies. Étant donné que les facteurs autres que le tabagisme qui affectent le poids à la naissance sont susceptibles d'être corrélés avec le tabagisme, nous devons tenir compte de ces facteurs. Par exemple, un revenu plus élevé donne généralement accès à de meilleurs soins prénataux, ainsi qu'à une meilleure alimentation de la mère. Soit le modèle suivant

$$\text{bwght} = \beta_0 + \beta_1 \cdot \text{cigs} + \beta_2 \cdot \text{faminc} + u$$

où `bwght` est le poids du bébé à la naissance en onces, `cigs` le nombre de cigarettes fumées par jour par la mère durant la grossesse, et `faminc` est le revenu du ménage en milliers de dollars.

1. Quel est le signe le plus probable pour β_2 ?
2. Pensez-vous que `cigs` et `faminc` sont susceptibles d'être corrélés? Expliquez pourquoi la corrélation peut être positive ou négative.
3. En utilisant la fonction `DROITEREG`, et la base de données `BWGHT.xls`, estimez l'équation précédente avec et sans la variable `faminc`. Rapportez les résultats sous forme d'équation, en mentionnant la taille de l'échantillon et le R-carré.

4. Discutez vos résultats. Sont-ils conformes à votre intuition? Concentrez-vous également les conséquences de l'ajout de la variable faminc et sur comment cette variable modifie l'effet estimé des cigarettes sur le poids.
5. Quel est le poids prédit d'un bébé dont la mère fume environ 10 cigarettes par jours? Avec ces seules informations, quelle équation estimée allez-vous utiliser pour votre prédiction?
6. Sachant qu'une once vaut environ 28 grammes, comment se modifieraient les estimations si la variable bwght était exprimée en grammes?

Exercice 6. Considérez le modèle de régression simple. Est-ce qu'un changement dans les unités de mesure de la variable dépendante modifie le R-carré? Et un changement dans l'unité de mesure de la variable indépendante? Démontrez.

Exercice 7. Browning et Goertz (dans Spending Time and Money within the Household) disposent d'un échantillon de 615 couples danois, chacun composé d'un homme et d'une femme de 19 à 64 ans et travaillant à temps-plein, avec ou sans enfants. Ils modélisent ainsi la répartition des dépenses au sein du couple à travers l'équation suivante:

$$\begin{aligned} \text{relatexp} = & \beta_0 + \beta_1 \cdot \log(\text{income}) + \beta_2 \cdot \log(\text{relatwage}) \\ & + \beta_3 \cdot \text{relatage} + \beta_4 \cdot \text{femedu} + \beta_5 \cdot \text{maledu} \\ & + \beta_6 \cdot \text{ychild} + \beta_7 \cdot \text{ochild} \end{aligned}$$

où "relatexp" représente le rapport des dépenses pour la femme par rapport à celles pour l'homme, "income" le revenu brut annuel du ménage en couronnes danoises, "relatwage" le rapport du salaire horaire de la femme sur celui de l'homme, "femedu" et "maledu" le niveau d'éducation de la femme et de l'homme en années, "ychild" une variable binaire égale à 1 si le couple a au moins un enfant de moins de 6 ans, et "ochild" une variable binaire égale à 1 si le couple a au moins un enfant entre 6 ans

et 18 ans. Les dépenses pour la femme (l'homme) sont les dépenses pour des biens qui seront utilisés exclusivement par la femme (l'homme) dans le couple. Les résultats sont les suivants:

$$\begin{aligned}\text{relatexp} = & -0.706 + 0.146 \cdot \log(\text{income}) + 0.191 \cdot \log(\text{relatwage}) \\ & -0.016 \cdot \text{relatage} - 0.011 \cdot \text{femedu} - 0.003 \cdot \text{maledu} \\ & -0.021 \cdot \text{ychild} - 0.023 \cdot \text{ochild}\end{aligned}$$

1. Interprétez le paramètre β_1 . Quel est l'effet du revenu sur la dépense relative?
2. Interprétez le paramètre β_2 . Quel est l'effet du salaire relatif sur la dépense relative? Est-ce que le signe vous semble normal?
3. Interprétez les paramètres β_6 et β_7 . Comment les dépenses relatives se modifient-elles pour un couple avec un enfant de 5 ans par rapport avec un couple avec les mêmes caractéristiques mais sans enfants? Et pour couple avec un enfant 5 ans et un autre de 10 ans? Comment pouvez-vous interpréter ce résultat?
4. Supposez que $\text{income} = 700000$ (environ 90000 euros), $\text{relatwage} = 0,90$, $\text{relatage} = 0,95$, $\text{femedu} = 13$, $\text{maledu} = 13$, $\text{ychild} = 0$ et $\text{ochild} = 0$. Quelle est la valeur prédite de la dépense relative?

Exercice 8. Soient la régression: $y = \beta_0 + \beta_1 x + u$ et les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ des MCO. Soient c_1 et c_2 deux constantes réelles et $\tilde{\beta}_0$ et $\tilde{\beta}_1$ les estimateurs par les moindres carrés ordinaires de la régression linéaire de $y' = (y + c_1)$ sur $x' = (x + c_2)$. Montrez que $\tilde{\beta}_0$ et $\tilde{\beta}_1$ peuvent s'écrire en fonction de $\hat{\beta}_0$ et $\hat{\beta}_1$ et de c_1 et c_2 de la manière suivante:

$$\tilde{\beta}_0 = \hat{\beta}_0 + c_1 - \hat{\beta}_1 c_2; \quad \tilde{\beta}_1 = \hat{\beta}_1.$$