

Contrôle Continu d'Économétrie Appliquée - 1H30

Olivier Donni & Ulysse Lawogni & Anderson Vil
CY Cergy Paris Université
2021 – 2022

Exercice

On considère le modèle linéaire suivant :

$$y_i = \beta_0 + \beta_1 x_i + u_i \text{ avec } u_i \sim \mathcal{N}(0, \sigma^2)$$

où u_i est le terme d'erreur non observable sur l'individu i . L'économètre observe $n = 10$ individus et recueille les informations suivantes :

$$\sum_{i=1}^n x_i = 207, \quad \sum_{i=1}^n y_i = 25.7, \quad \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 5.813$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 56.875, \quad \sum_{i=1}^n \hat{u}_i^2 = 0.4347, \quad \sum_{i=1}^n (y_i - \bar{y})^2 = 1.0288$$

1. Interprétez β_0 et β_1 .
2. Donnez les expressions des estimateurs $\hat{\beta}_1$ et $\hat{\beta}_0$ par moindres carrés ordinaires de β_1 et β_0 . Calculez leurs valeurs.
3. Soit i un individu tel que $x_i = 28$.
 - (a) Donnez une prédiction pour y_i .
 - (b) Calculez le résidu $\hat{u}_i$ sur cet individu sachant $y_i = 3.4$.
4. Considérons un second j tel que $y_j = 3.7$ et $\hat{u}_j = 0.02$. Que vaut x_j ?
5. Quelle variation en x mène à une baisse moyenne en y de 0.7 ?
6. Soient c_1 et c_2 deux constantes réelles telles que $c_2 \neq 0$.
 - (a) Soient $\tilde{\beta}_1$ et $\tilde{\beta}_0$ les estimateurs par les moindres carrés ordinaires de la régression linéaire de $c_1 y$ sur $c_2 x$. Exprimez $\tilde{\beta}_1$ et $\tilde{\beta}_0$ en fonction de $\hat{\beta}_1$ et $\hat{\beta}_0$ et de c_1 et c_2 .
 - (b) Soient $\tilde{\beta}_1$ et $\tilde{\beta}_0$ les estimateurs par les moindres carrés ordinaires de la régression linéaire de $c_1 + y$ sur $c_2 + x$. Exprimez $\tilde{\beta}_1$ et $\tilde{\beta}_0$ en fonction de $\hat{\beta}_1$ et $\hat{\beta}_0$ et de c_1 et c_2 .
7. Soient i et j deux individus tels que $x_i - x_j = 0.5$. Donnez une estimation de la différence moyenne de $y_i - y_j$.

Problème

Dans un échantillon de 1126 personnes de 25 ans à 55 ans, on estime le modèle linéaire liant le logarithme du taux de salaire $ltsal$ en fonction du niveau d'éducation $educ$, de la variable d'âge age ainsi que d'une variable $paris$ indiquant si l'individu est Francilien ($paris = 1$) ou s'il est non Francilien ($paris = 0$). Notez que le terme Francilien désigne ici tout individu habitant dans la région Ile-de-France.

$$ltsal = 0.86 + 0.054 \text{ } educ + 0.0017 \text{ } age + 0.068 \text{ } paris + \hat{u}$$

(0.20) (0.0033) (0.010) (0.025)

$$N = 1126, \text{ } SCR = 91.72, \text{ } SCT = 150.$$

où les écarts-types des coefficients estimés figurent entre parenthèses en dessous des valeurs estimées.

1. Donner une expression du R^2 et calculez sa valeur.
2. Donner une estimation de la différence moyenne absolue de salaire entre deux jumeaux ayant tous deux le même niveau d'éducation, l'un habitant à Paris et l'autre à Bordeaux ?
3. Donner une estimation de $ltsal$ pour un individu de 30 ans, ayant 10 ans d'années d'éducation et résidant à Cergy-Pontoise.
4. Donner une estimation de la différence en $ltsal$ entre un Parisien de 37 ans et un Marseillais de 42 ans, tous deux ayant le même niveau d'éducation.
5. Donner une estimation de la différence moyenne en niveau d'éducation entre un Francilien et un non Francilien ayant le même taux de salaire et le même âge. Précisez celui qui a le plus grand niveau d'éducation.
6. Quelle est la différence moyenne en niveau d'éducation faut-il y avoir entre un Francilien et un Provincial pour espérer avoir le même taux de salaire ?
7. Danaerys est une jeune Lyonnaise très ambitieuse. Elle a 27 ans et elle détient un Master (Bac +5) en Finance. Elle bosse dans une start-up dans le centre de Lyon. N'étant pas très satisfaite de son salaire, elle décide d'effectuer un Mastère spécialisé (Bac +6) en Data Science d'un an dans une Grande École parisienne. Elle est recrutée directement à la fin de sa formation dans une grande entreprise implantée à la Défense. Quelle augmentation moyenne de taux de salaire peut-elle espérer par rapport à l'ancien taux de salaire qu'elle percevait à Lyon un an plus tôt ?
8. On définit la variable $province$ qui vaut 1 si l'individu est en non Francilien c'est-à-dire si $paris = 0$ et qui vaut 0 si l'individu est Francilien, c'est-à-dire si $paris = 1$.
 - (a) Quelles seraient les conséquences de l'introduction de cette variable dans le modèle ?
 - (b) On remplace la variable $paris$ par la variable $province$ dans le modèle. Estimez ce nouveau modèle.