

Contrôle Continu d'Économétrie Appliquée - 1H30

Olivier Donni & Ulysse Lawogni
Université de Cergy-Pontoise
2019 – 2020

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 $n = 8$ 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) Comment estimez-vous l'erreur commise par le modèle sur cet individu sachant que $y_i = 3.4$?
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

On considère un échantillon de 1000 personnes, extrait l'Enquête Emploi effectuée en France en 1999. On classe ces personnes en fonction de leur niveau d'éducation dans trois sous-échantillons : le premier sous-échantillon comprend 166 personnes qui sont allées à l'école au plus jusqu'au collège (moins de 8 ans d'éducation depuis l'âge de 6 ans) ; le deuxième sous-échantillon comprend 528 personnes qui sont allées à l'école après le collège mais au plus jusqu'à la fin du secondaire (entre 9 et 11 ans d'éducation depuis l'âge de 6 ans) ; enfin, le troisième sous-échantillon comprend 306 personnes qui sont allées à l'université. Les variables indicatrices de ces trois groupes sont respectivement COL , SEC , et SUP (par exemple : $COL = 1$ si la personne appartient au premier groupe et $COL = 0$ sinon).

On décrit le modèle donnant le logarithme du taux de salaire ($ltsal$) en fonction des variables indicatrices du niveau d'éducation SEC et SUP , en fonction de l'âge age exprimé en années et du carré de l'âge $age2$ et en fonction du sexe sex ($sex = 1$ si l'individu est une femme et 0 sinon).

$$ltsal = \beta_0 + \beta_1 SEC + \beta_2 SUP + \beta_3 age + \beta_4 age2 + \beta_5 sex + u$$

On suppose :

$$u \sim \mathcal{N}(0, \sigma^2)$$

L'estimation de ce modèle sur le logiciel SAS par les moindres carrés ordinaires donne :

$$\hat{ltsal} = 2.988 + 0.210SEC + 0.573SUP + 0.026age - 0.0001281age2 - 0.147sex$$

$$n = 1000 \quad SCT = 150.44 \quad SCR = 103.46$$

1. Interprétez les paramètres β_1 , β_2 et β_5 .
2. Pourquoi n'a-t-on pas introduit la variable indicatrice du collège COL dans ce modèle ?
3. Donnez une expression du R^2 et calculez sa valeur.
4. On fixe le niveau d'éducation et le sexe. Donnez l'allure graphique moyenne de $ltsal$ en fonction de age et déterminez l'âge où le taux de salaire est maximal.
5. Considérons un couple formé par une femme j de 27 ans ayant 10 ans d'éducation et un homme i de 30 ans détenteur d'un Master d'Économie.
 - (a) Estimez la différence en taux de salaire entre les deux conjoints.
 - (b) Quel âge devrait avoir la femme j pour espérer avoir un salaire égal à celui de son conjoint i ?
6. On décide d'exprimer l'âge en mois. Et on estime de nouveau le modèle. Que deviennent les estimateurs $\hat{\beta}_3$ et $\hat{\beta}_4$?
7. On définit la variable $S = 1 - sex$. On régresse à présent le taux de salaire sur les variables COL , SUP , age , $age2$ et S . L'âge est exprimé en années.
 - (a) Interprétez les coefficients de COL , SUP et S .
 - (b) Écrivez le modèle estimé.