

Econométrie Appliquée
Contrôle Continu – 1H30Min
Ulysse LAWOONI

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 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).

Rappel : Les quantiles d'ordre 0.95, 0.975 et 0.995 d'une loi normale centrée réduite est approximativement égaux à 1.63, 1.96 et 2.58. Les quantiles d'une loi de Student à un nombre de degrés supérieur à 900 sont approximativement les mêmes que ceux d'une loi normale centrée réduite. On précise également que :

$$F_{5;\infty}^{95\%} = \text{et } F_{2;\infty}^{95\%} = 3$$

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

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

Le terme d'erreur u suit une loi normale $N(0, \sigma^2)$.

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

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

(0.242) (0.030) (0.033) (0.013) (0.00016) (0.021)

$$n = 1000, \quad SCT = 150.44 \text{ et } SCR = 103.46$$

- 1- Sous quelles conditions sur l'erreur u cet estimateur est-il sans biais ?
- 2- Interpréter les paramètres du modèle.
- 3- Pourquoi n'a-t-on pas introduit la variable indicatrice du niveau collège **COL** dans cette régression ?
- 4- Donner l'expression du R^2 en fonction de SCT et SCR et calculer sa valeur.
- 5- Donner l'allure graphique de **ltsal** en fonction de **age** et justifier.

- 6- Déterminer l'âge optimal où $ltsal$ est maximal, toutes choses égales par ailleurs.
- 7- Tester individuellement la significativité de chacune des variables au seuil 5%.
- 8- Tester l'hypothèse que le carré de l'âge n'a aucune influence sur le taux de salaire contre l'alternative qu'il l'influence négativement au seuil 5%.
- 9- Tester la significativité globale du modèle au seuil 5%.
- 10- Construire un intervalle de confiance pour la variable SUP ?
- 11- Considérons un couple formé d'une femme j de 27 ans ayant 10 ans d'éducation et d'un homme i de 30 ans détenteur d'un Master d'Economie.
 - a- Estimer la différence de revenu entre les deux conjoints.
 - b- Quel âge devrait avoir la femme j pour espérer avoir un salaire au moins égal à celui de son conjoint i ?
- 12- Construire un intervalle de confiance de 99% pour β_3 .
- 13- Comment écririez-vous l'hypothèse que l'éducation n'a pas d'influence sur le taux de salaire ?
- 14- On régresse maintenant le taux de salaire $ltsal$ sur les variables COL , SUP , age , $age2$ et sex .
 - a- Interpréter les coefficients de COL et SUP .
 - b- Quelle est la valeur estimée du coefficient de COL dans cette nouvelle estimation ?
 - c- Quelle est la nouvelle valeur estimée du coefficient de SUP ?
 - d- Tester dans cette nouvelle régression ces deux coefficients.

On crée maintenant une nouvelle variable qui est l'interaction entre le fait d'avoir été à l'université avec la variable sex . Cette nouvelle variable est notée $S=sex*SUP$.

- 15- Pour quel groupe de personnes dans la population cette variable est-elle égale à 1 ? Quelles sont les autres valeurs possibles ?

On ajoute cette nouvelle variable dans le modèle et l'estimation du nouveau modèle donne les résultats suivants :

$$\widehat{ltsal} = 2.99 + 0.21 * SEC + 0.53 * SUP + 0.026 * age - 0.00013 * age2 - 0.16 * sex + 0.03 * S$$

$$(0.242) \quad (0.030) \quad (0.075) \quad (0.013) \quad (0.00016) \quad (0.025) \quad (0.0446)$$

$$n = 1000, \quad SCT = 150.44 \text{ et } SCR = 103.42$$

- 16- Calculer la statistique de Fisher correspondant au test de $H_0: \beta_S = 0$ contre l'alternative $H_1: \beta_S \neq 0$ en utilisant la somme des carrés des résidus.
- 17- On considère un homme de 30 ans d'éducation secondaire et dont le logarithme du taux de salaire est égal à 4.04.
 - a- Quelle serait la meilleure prédiction de la valeur du logarithme du taux de salaire pour une femme de même âge et de même éducation ?
 - b- Calculer l'intervalle de confiance de 95% pour cette estimation.