
Spécification des variables

Mars 2020

Application

Dans cet exercice, nous cherchons à expliquer le salaire de ménages américains par leur âge, leur niveau d'éducation et l'origine ethnique.

Importez le fichier **salaire.xls**.

1. Estimez le modèle suivant :

$$wage = \beta_0 + \beta_1 age + \beta_2 education + \beta_3 race + \varepsilon$$

2. Analysez la régression à l'aide des éléments de synthèse proposés dans le summary. Interprétez les coefficients obtenus. Qu'observez-vous en particulier pour les variables catégoriques (race et education) ? Pourquoi toutes les catégories n'ont-elles pas été incluses ?
3. On voudrait voir si on ne peut pas en plus observer une discrimination dans la valorisation des diplômes (variable education) et/ou de la catégorie socio-professionnelle (variable jobclass). Comment pouvez-vous procéder pour vérifier cela ? Interprétez les coefficients obtenus.
4. On se focalise maintenant sur la relation entre l'âge et le salaire. Pensez-vous que cette relation est linéaire ? Pour répondre, construisez un nuage de point.
5. On part du principe que la relation n'est en effet pas linéaire. On se propose d'estimer le modèle suivant :

$$\ln(wage) = \beta_0 + \beta_1 \ln(age) + \beta_2 education + \beta_3 race + \varepsilon$$

Ce modèle est-il estimable à l'aide des MCO ? Si oui, interprétez les coefficients. Quelle est l'élasticité de la variable $\ln(wage)$ à $\ln(age)$?

6. Estimez et interprétez les coefficients des modèles :

$$(i) \ln(wage) = \beta_0 + \beta_1 age + \beta_2 education + \beta_3 race + \varepsilon$$

ou

$$(ii) wage = \beta_0 + \beta_1 \ln(age) + \beta_2 education + \beta_3 race + \varepsilon$$

7. Une autre manière d'estimer une relation non linéaire est de mettre la variable explicative au carré. Estimez ce modèle et interprétez les coefficients. Définissez la notion de point de retournement.
8. Quelle est selon vous la meilleure façon de représenter la relation entre l'âge et le salaire ? Utiliser un **Boxcox** pour répondre.