

Exercices pour le cours Économétrie et évaluations statistiques (Ch. VIII).

Prédiction – R-carré ajusté – Quelques exemples

Exercice 1. Les équations suivantes ont été estimées à l'aide des données de BWGHT.xls:

$$\begin{aligned}\widehat{\log(\text{bwght})} &= 4,66 - 0,004 \text{ cigs} + 0,009 \log(\text{faminc}) \\ &\quad (0,22) \quad (0,001) \quad (0,006) \\ &\quad + 0,016 \text{ parity} + 0,027 \text{ male} + 0,055 \text{ white} \\ &\quad (0,006) \quad (0,010) \quad (0,013) \\ N &= 1,388 \quad R^2 = 0,047\end{aligned}$$

et

$$\begin{aligned}\widehat{\log(\text{bwght})} &= 4,65 - 0,005 \text{ cigs} + 0,011 \log(\text{faminc}) \\ &\quad (0,38) \quad (0,001) \quad (0,008) \\ &\quad + 0,017 \text{ parity} + 0,034 \text{ male} + 0,045 \text{ white} \\ &\quad (0,006) \quad (0,011) \quad (0,015) \\ &\quad - 0,003 \text{ motheduc} + 0,003 \text{ fatheduc} \\ &\quad (0,003) \quad (0,003) \\ N &= 1,191 \quad R^2 = 0,049\end{aligned}$$

où bwght désigne le poids à la naissance en livres d'un enfant, cigs le nombre moyen de cigarettes que la mère a fumées par jour pendant la grossesse, parity l'ordre de naissance de l'enfant, faminc le revenu familial annuel, motheduc l'éducation de la mère en années, fatheduc l'éducation du père en années, male une variable binaire qui est égale à 1 si l'enfant est un garçon et 0 sinon, white une variable binaire qui est égale à 1 si l'enfant est blanc et 0 sinon.

1. Dans la première équation, interprétez le coefficient de la variable cigs. En particulier, quel est l'effet sur le poids de naissance de fumer 10 cigarettes de plus par jour?

2. Combien un enfant blanc devrait-il peser de plus qu'un enfant non blanc, tenant les autres facteurs de la première équation fixés? La différence est-elle statistiquement significative?
3. Commentez l'effet estimé et la signification statistique de la variable *motheduc* ?
4. À partir des informations fournies, pourquoi n'est-il pas possible de calculer la statistique de Fischer pour le test de l'hypothèse $H_0 : \beta_6 = \beta_7 = 0$ (l'éducation des parents n'a pas d'effet sur le poids de l'enfant) ? Que devriez-vous faire pour calculer la statistique de Fischer ?

Exercice 2. Utilisez les données sur les prix des logements dans *HPRICE1.xls* pour cet exercice. Supposez que les hypothèses H_1 - H_6 sont satisfaites.

1. En utilisant DROITEREG sur Excel, estimez le modèle suivant:

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{lotsize}) + \beta_2 \log(\text{sqrft}) + \beta_3 \text{bdrms} + u$$

et rapportez les résultats sous la forme habituelle.

2. Calculez la valeur prédite de $\log(\text{price})$ pour $\text{lotsize} = 20000$, $\text{sqrft} = 2500$ et $\text{bdrms} = 4$.
3. En utilisant DROITEREG sur Excel, calculez la variance l'estimateur de l'espérance conditionnelle de $\log(\text{price})$ [Une transformation des variables sur Excell est nécessaire].
4. Calculez l'intervalle de confiance au seuil de 95% de l'espérance conditionnelle de $\log(\text{price})$.
5. Calculez l'intervalle de confiance au seuil de 95% de la réalisation de $\log(\text{price})$.

6. En utilisant l'approximation suggérée dans le cours, calculez la valeur prédite de price (et non de $\log(\text{price})$)
7. Quel serait, selon-vous, l'intervalle de confiance au seuil de 95% de la réalisation de l'espérance conditionnelle de price? Attention, cette question nécessite réflexion. [Astuce: L'intervalle de confiance n'utilise pas la valeur prédite de price calculées en (6)].
8. En utilisant DROITEREG sur Excel, estimez le modèle suivant:

$$\text{price} = \beta_0 + \beta_1 \text{lotsize} + \beta_2 \text{sqft} + \beta_3 \text{bdrms} + \beta_4 \text{colonial} + u.$$

9. Décidez si vous préférez le modèle en (1) ou le modèle en (8). Justifiez.

Exercice 3. Considérez les 4 modèles alternatifs suivants qui diffèrent dans leur forme fonctionnelle :

- (I) : $\text{intensrd} = \beta_0 + \beta_1 \log(\text{sales}) + u,$
- (II) : $\text{intensrd} = \beta_0 + \beta_1 \text{sales} + \beta_2 \text{sales}^2 + u$
- (III) : $\text{intensrd} = \beta_0 + \beta_1 \text{sales} + u$
- (IV) : $\text{intensrd} = \beta_0 + \beta_1 \log(\text{sales}) + \beta_1 (\log(\text{sales}))^2 + u,$

où "intensrd" représente l'intensité de la R&D (le rapport des dépenses de R&D sur les ventes) et "sales" représente les ventes de l'entreprise. En utilisant la base de données RDCHEM.xls qui contient 32 observations sur des entreprises de l'industrie chimique,

les résultats suivants, résumés dans un tableau, sont obtenus :

	I	II	III	IV
constant	1,103 (1,578)	2,612 (0,429)	-3,063 (0,369)	6,933 (5,210)
sales $\div$ 1000	- -	0,300 (0,139)	0,053 ($< 0,001$)	- -
log(sales)	0,302 (0,256)	- -	- -	-1,420 (1,483)
sales ² $\div$ 1000000	- -	0,007 (0,004)	- -	- -
log(sales) ²	- -	- -	- -	0,121 (0,103)
R ²	0,061	0,148	0,104	0,047
$\bar{R}^2$	0,030	0,090	0,042	0,015

1. (a) Si vous comparez les modèles I et IV, lequel pensez-vous est meilleur ?
Précisez le critère utilisé pour la comparaison.
- (b) Si vous comparez les modèles I et II, lequel pensez-vous est meilleur ?
Précisez le critère utilisé pour la comparaison.
- (c) Si vous comparez les modèles II et III, lequel pensez-vous est meilleur ?
Précisez le critère utilisé pour la comparaison.

Exercice 4. Soit le modèle suivant:

$$\widehat{\log(\text{wage})} = \beta_0 + \beta_1 \text{male} \cdot \text{married} + \beta_2 \text{female} \cdot \text{married} + \beta_3 \text{female} \cdot \text{single} \\ + \beta_4 \text{educ} + \beta_5 \text{exper} + \beta_6 \text{exper}^2 + \beta_7 \text{tenure} + \beta_8 \text{tenure}^2$$

En utilisant la base de données WAGE1.xls, ce modèle suivant est estimé :

$$\begin{aligned} \widehat{\log(\text{wage})} = & 0,321 + 0,213 \text{ male} \cdot \text{married} - 0,198 \text{ female} \cdot \text{married} \\ & (0,100) \quad (0,055) \quad (0,058) \\ & - 0,110 \text{ female} \cdot \text{single} + 0,079 \text{ educ} + 0,027 \text{ exper} \\ & (0,056) \quad (0,007) \quad (0,005) \\ & - 0,0005 \text{ exper}^2 + 0,029 \text{ tenure} - 0,0005 \text{ tenure}^2 \\ & (0,0001) \quad (0,007) \quad (0,0002) \end{aligned}$$

$$N = 526, \quad R^2 = 0,461$$

où "male", "female", "married" et "single" sont des variables binaires égales à un pour les hommes, les femmes, les individus mariés et les individus célibataires, respectivement, et zéro sinon, "educ" représente l'éducation en années, "exper" l'expérience en années et "tenure" l'ancienneté en années.

1. Parmi les variables explicatives, interprétez l'effet des trois variables croisées: "male·married", "female·married" and "female·single". En utilisant un t-test, vérifiez si ces variables ont un effet significativement différent de zéro au seuil de 5%.
2. Considérez le modèle simplifié suivant:

$$\begin{aligned} \widehat{\log(\text{wage})} = & 0,417 + 0,297 \text{ female} + 0,080 \text{ educ} + 0,029 \text{ exper} \\ & (0,099) \quad (0,036) \quad (0,007) \quad (0,005) \\ & - 0,0006 \text{ exper}^2 + 0,032 \text{ tenure} - 0,0006 \text{ tenure}^2 \\ & (0,0001) \quad (0,007) \quad (0,0002) \end{aligned}$$

$$N = 526, \quad R^2 = 0,441$$

Est-ce que ce modèle simplifié peut être considéré comme un modèle restreint du modèle précédent, plus général. Si oui, quelles sont les contraintes qui doivent être imposées aux paramètres du modèle général pour obtenir le modèle restreint?

3. En utilisant un test de Fisher au seuil de signification de 5%, pouvez-vous déterminer si les trois variables "male-married", "female-married" and "female-single", sont nécessaires ou si elles peuvent être remplacées par une seule variable binaire "female"?
4. Utilisant le deuxième modèle, calculez l'intervalle de confiance au niveau de signification de 95% pour le paramètre β_1 ("female").

Exercice 5. On estime successivement :

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

et

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + u \tag{2}$$

Les sommes des carrés résiduels sont : $SCR_1 = 220$ et $SCR_2 = 190$. L'échantillon compte $N = 100$.

1. On teste l'hypothèse nulle $H_0 : \beta_2 = \beta_3 = 0$ à l'aide d'un test de Fisher, en retenant un seuil de signification de 5%.
2. Comparez la décision du test de Fisher à celle du R-carré ajusté.
3. Expliquez pourquoi les deux critères peuvent diverger (d'un point de vue économétrique).