UNIVERSITÉ TOULOUSE 1 CAPITOLE

TSE
Toulouse School of Economics
École d'économie de Toulouse
Année Universitaire 2012-2013
SESSION 1

Semestre 6
LICENCE 3 mention ÉCONOMIE
LICENCE 3 mention ÉCONOMIE et MATHEMATIQUES

ECONOMETRIE
(durée 1h30)

F. GASMI
T. MAGNAC

Mardi 7 mai 2013 ~ 12h30 – 14h00

-=-=-=-=-

PARTIE EMPIRIQUE (12 points)

Problème : On considère un échantillon de 1000 personnes, extrait de l'Enquête Emploi effectuée en France en 1999. On classe ces personnes en fonction de leur education 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 *educ_col*, *educ_sec* et *educ_sup* (par exemple. *educ_col* = 1 si la personne appartient au premier groupe et *educ_col* = 0 sinon).

(*Rappel* : le quantile d'ordre 0.995 d'une loi normale centrée et réduite est approximativement égal à 2.58).

Voici maintenant les résultats de la régression du logarithme du taux de salaire (*ltsal*) sur les variables indicatrices *educ_sec* et *educ_sup*, l'âge (*agd*), l'âge au carré (*agd2*) et le sexe ($s \in \{Homme = 0, Femme = 1\}$).

Source	SS	df	MS
Model	46.9788694	5	9.39577389
Residual	103.461621	994	.104086138
Total	150.44049	999	.150591082

Number of obs	=	1000
F(5, 994)	=	90.27
Prob > F	=	0.0000
R-squared	=	0.3123
Adj R-squared	=	0.3088
Root MSE	=	.32262

ltsal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ_sec	.2098154	.0300892	6.97	0.000	.1507698	.268861
educ_sup	.5725776	.0329807	17.36	0.000	.5078578	.6372975
agd	.0255319	.0125094	2.04	0.042	.000984	.0500798
agd2	-.0001281	.0001585	-0.81	0.419	-.0004392	.0001829
s	-.1469443	.0206648	-7.11	0.000	-.1874959	-.1063926
_cons	2.987725	.2422177	12.33	0.000	2.512409	3.463042

1. Pourquoi n'a-t-on pas introduit la variable indicatrice *educ_col* dans cette régression?
2. Quelle est l'interprétation du coefficient de *educ_sup* dans cette régression? Ecrivez de façon brève une phrase en français qui résume ce résultat.
3. Est-ce que ce dernier coefficient est significatif à un niveau de 1%? Pourquoi?
4. Quelle est la probabilité que le coefficient estimé de s soit supérieur à -.1063?
5. Comment écrivez-vous l'hypothèse nulle que l'éducation n'a pas d'influence sur le taux de salaire?

On régresse maintenant le logarithme du taux de salaire (*ltsal*) sur les variables indicatrices *educ_col* et *educ_sup*, l'âge (*agd*), l'âge au carré (*agd2*) et le sexe ($s \in \{Homme = 1, Femme = 2\}$). Vous vous servirez de la Table ci-dessus pour répondre aux questions suivantes:

6. Quelle est la valeur du coefficient estimé de *educ_col* dans cette nouvelle régression?
7. Quelle est la valeur de la statistique de Student qui est associée à ce dernier coefficient?
8. Quelle est la nouvelle valeur du coefficient estimé de *educ_sup*?

On crée maintenant une nouvelle variable qui est l'interaction entre le fait d'avoir été à l'université avec la variable indicatrice homme/femme (*s*). Cette nouvelle variable est notée *s_sup* = *s*educ_sup*.

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

On ajoute maintenant dans la première régression cette variable et on obtient les résultats suivants:

Source	SS	df	MS
Model	47.023717	6	7.83728616
Residual	103.416774	993	.104145794
Total	150.44049	999	.150591082

Number of obs	=	1000
F(6, 993)	=	75.25
Prob > F	=	0.0000
R-squared	=	0.3126
Adj R-squared	=	0.3084
Root MSE	=	.32272

ltsal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ_sec	.2086618	.0301491	6.92	0.000	.1494986	.2678251
educ_sup	.5278817	.0756803	6.98	0.000	.3793699	.6763935
agd	.0258963	.0125253	2.07	0.039	.0013171	.0504754
agd2	-.0001325	.0001587	-0.83	0.404	-.0004439	.000179
s	-.1560756	.024918	-6.26	0.000	-.2049737	-.1071775
s_sup	.0292775	.0446154	0.66	0.512	-.0582739	.1168288
_cons	2.994337	.2424965	12.35	0.000	2.518473	3.470202

10. Calculer la statistique de Fisher correspondant au test de $H_0 : \beta_{s\_\sup} = 0$ contre $H_a : \beta_{s\_\sup} \neq 0$ (en utilisant les R^2 ou les sommes des carrés des résidus que vous tirerez des 2 tableaux de résultats). Quelle conclusion en tirez-vous ?

On considère un homme d'éducation secondaire et d'âge égal à 30 ans dont le logarithme du taux de salaire est égal à 4.04.

11. Quelle serait la meilleure prédiction de la valeur du logarithme du taux de salaire pour une femme de même âge et de même éducation?
12. Calculer l'intervalle de confiance à 95% pour cette valeur.

PARTIE THEORIQUE (8 points)

1. Considérez le modèle de régression simple

$$y = \beta_0 + \beta_1 x + u$$

Sous les 4 hypothèses standards (Linéarité, échantillonnage aléatoire, espérance conditionnelle du terme d'erreur nulle et absence de colinéarité parfaite), l'estimateur des MCO de β_1, $\hat{\beta}_1$, est non-biaisé et convergent. Montrez que l'estimateur des MCO de β_0, $\hat{\beta}_0$, satisfait également ces deux propriétés (*Suggestion : Vous pourriez vous servir de l'expression de* $\hat{\beta}_0$ *en fonction de* $\hat{\beta}_1$).

2. Le "R-carré" d'une régression représente la part de la somme des carrés totaux expliquée par le modèle, i.e., $R^2 \equiv \frac{\text{ESS}}{\text{TSS}}$ où $\text{ESS} \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ est la somme des carrés expliqués et $\text{TSS} \equiv \sum_{i=1}^{n}(y_i - \bar{y})^2$ est la somme des carrés totaux. Le "Rbar-carré" est donné par $\bar{R}^2 \equiv 1 - \left(\frac{n-1}{n-k-1}\right)\left(\frac{\text{SSR}}{\text{TSS}}\right)$ où $\text{SSR} \equiv \sum_{i=1}^{n}\hat{u}_i^2$ est la somme des carrés des résidus, n le nombre d'observations et k le nombre de variables indépendantes. Montrez que :

   - $0 \leq R^2 \leq 1$
   - $\bar{R}^2 < R^2$
   - Quand k augmente R^2 ne peut qu'augmenter ou rester le même alors que $\bar{R}^2$ peut diminuer.
   - Alors que R^2 ne peut jamais être négatif, montez que $\bar{R}^2$ peut l'être.

3. Répondez par "Oui" ou "Non" (sans donner d'explications): Lorsque les hypothèses standards du modèle de régression multiple sont satisfaites mais les erreurs sont hétéroscédastiques.

   - L'estimateur des MCO est biaisé
   - L'estimateur des MCO est convergent
   - Le "ratio t" usuel ne suit plus une loi de student
   - L'estimateur des MCO est BLUE