

**ECONOMÉTRIE**  
**EXAMEN FINAL - SESSION 1**

Note : Aucun document n'est autorisé. Tout type de calculatrice est autorisé.

Pour tous les exercices, vous effectuerez les tests à un seuil de risque de 5%.

**Exercice 1 : Evolution des déterminants des distances domicile-travail**

On souhaite analyser l'évolution des déterminants des distances domicile-travail des salariés de l'aire urbaine lyonnaise entre 1995 et 2006. Pour ce faire, on a interrogé 10 739 individus en emploi en 1995 et 5 763 en 2006 sur la distance (en km) entre leur domicile et leur lieu de travail (*DistanceDT*) et sur un certain nombre de caractéristiques sociodémographiques :

- *femme* : variable qui vaut 1 si le salarié est une femme et 0 sinon ;
- *Age* et *Age2* : l'âge et l'âge au carré de l'individu ;
- *Nbreenfants* : le nombre d'enfants de l'individu ;
- la catégorie socioprofessionnelle de l'individu en 4 modalités : i) cadres et professions supérieures (*profsup*), ii) professions intermédiaires (*PI*), iii) employés (*employe*), iv) ouvriers.
- *temps\_partiel* : qui vaut 1 si l'individu travaille à temps partiel et 0 si l'individu travaille à temps plein ;
- le lieu de résidence en 3 modalités : i) en centre-ville (*centre*), ii) en 1ère couronne/proche banlieue (*couronne1*), iii) plus loin que la 1ère couronne.
- *an06* : une variable qui vaut 1 pour les individus qui ont répondu en 2006 et 0 pour ceux qui ont répondu en 1995.

En outre, des variables croisées ont été construites. Toutes les variables commençant par *v06\_* sont des termes d'interaction entre la variable *an06* et la variable explicative correspondante. Par exemple, *v06\_femme* = *an06 \* femme* et *v06\_Nbreenfants* = *an06 \* Nbreenfants*.

Une régression linéaire multiple sur *DistanceDT* a été estimée par moindres carrés ordinaires à partir du logiciel Stata. Les résultats de cette régression sont présentés en page suivante (les ??? correspondent à des résultats qui ont été volontairement enlevés).

Tableau 1

Source	SS	df	MS	Number of obs	=	16,502
				F(16, 16485)	=	????_(1)
Model	45067.2718	16	2816.70449	Prob > F	=	0.0000
Residual	385801.962	16,485	???(2)	R-squared	=	0.1046
Total	430869.233	16,501	26.1117043	Adj R-squared	=	???(3)
				Root MSE	=	4.8377

DistanceDT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
femme	-.9423776	.1026989	-9.18	0.000	-.143679 -.7410768
Age	.043301	.0273176	1.59	0.113	-.0102444 .0968464
Age2	-.000554	.0003311	-1.67	0.094	-.001203 .0000949
Nbreenfants	.5002659	.0560227	8.93	0.000	.3904554 .6100764
profsup	1.402584	.1416772	9.90	0.000	1.124881 1.680286
PI	.7255484	.1391128	5.22	0.000	.4528723 .9982245
employe	.4821767	.1409012	3.42	0.001	.2059952 .7583582
temps_partiel	-1.152966	.1114269	-10.35	0.000	???????
centre	-3.469238	.0943284	-36.78	0.000	-3.654132 -3.284344
couronne1	-2.076777	.0984805	-21.09	0.000	-2.26981 -1.883745
an06	-.065158	.1823469	-0.36	0.721	-.4225775 .2922616
v06_femme	-.0733011	.1715333	-0.43	0.669	-.409525 .2629227
v06_Nbreenfants	-.4740966	.0808937	-5.86	0.000	-.6326569 -.3155362
v06_profsup	-.4852013	.2392066	-2.03	0.043	-.9540721 -.0163305
v06_PI	.1943327	.2343251	0.83	0.407	-.2649698 .6536352
v06_employe	-.5218006	.2377741	-2.19	0.028	-.9878635 -.0557378
_cons	5.984743	.5410748	11.06	0.000	4.924178 7.045308

1. Trouver la valeur de  $F(16,16485)$  qui a été remplacée par  $????_(1)$ . Que pouvez-vous en conclure ?
2. Trouver la valeur de  $????_(2)$ ? Que représente-t-elle ?
3. Calculer le coefficient de détermination ajusté ( $????_(3)$ ). Que pouvez-vous en conclure ?
4. Déterminer l'intervalle de confiance (à 95%) pour le coefficient associé à la variable *temps\_partiel* et donner en une interprétation.
5. Quel est l'intérêt d'introduire une variable « âge au carré »? Selon les résultats de la régression, que peut-on conclure sur l'effet de l'âge sur les distances domicile-travail ?
6. Quelles conclusions pouvez-vous tirer à partir des résultats sur les 10 premières variables explicatives (*femme* → *couronne1*)? Soyez précis et quantifier les effets de chacune des variables ?
7. Que peut-on conclure à partir des variables commençant par *v06\_-*? Quantifier précisément les effets et discuter des changements de déterminants des distances domicile-travail entre 1995 et 2006 ?
8. Donner une estimation de la distance domicile-travail moyenne d'une femme cadre de 20 ans, sans enfants, travaillant à temps plein et habitant en centre ville pour 1995 et pour 2006 (vous tiendrez compte des variables non significatives dans vos calculs) ?
9. On cherche à tester :  $H_0 : \beta_{PI} = \beta_{employe}$ . Grâce à la commande appropriée sous Stata, on obtient :

```
( 1) PI - employe = 0
      F(  1, 16485) =    4.45
      Prob > F = 0.0349
```

Donner la formule qui a permis de calculer  $F(1,16485)$  (en précisant à quoi correspondent les éléments de cette formule) et conclure sur la pertinence de l'hypothèse préconisée et interpréter littérairement.

10. Expliquer en détails comment vous auriez mis en œuvre le test de Chow pour savoir si les déterminants des distances domicile-travail étaient identiques entre 1995 et 2006 (régression(s) effectuée(s), variables explicatives, la/les statistique(s) de test, la/les règle(s) de décision). Au vu des résultats de la régression de la page précédente, peut-on savoir quelle aurait été la conclusion de ce test de Chow ? Justifiez votre réponse.

11. De manière générale, qu'est-ce qui pourrait remettre en cause la robustesse des résultats de la régression du Tableau 1 ? Quelle(s) solution(s) préconiseriez-vous ?

## Exercice 2 : Santé mentale

On cherche à connaître les facteurs associés à la survenue d'un trouble mental (troubles anxieux et épisodes dépressifs). Pour ce faire, on dispose d'une base de données américaines contenant la variable *trouble* qui vaut 1 si l'individu a souffert d'un trouble mental au cours des 12 derniers mois et 0 sinon, une variable d'âge, l'origine ethnique de l'individu (Asiatique, Latino, Afro-américain, Blanc), son statut matrimonial (Célibataire, Marié, Divorcé), le nombre d'années d'éducation en classes (moins de 12 années d'études, 12 années d'études, 13 à 15 années d'études, 16 années ou plus d'études), une variable indiquant si l'individu souffre de maladies chroniques (hors santé mentale : oui/non), une variable indiquant le nombre de fois où des problèmes d'anxiété ont été traités dans l'enfance de l'individu et sa région de résidence (Midwest, South, West, North). En supposant que les termes d'erreur suivent une loi normale de moyenne nulle et de variance unitaire, les résultats de l'estimation du modèle approprié sont reportés ci-dessous.

	Coefficient	Ecart-type
Age	0.030	(0.01)
Age au carré	-0.01	(0.001)
Origine ethnique :		
Asiatique	-0.037	(0.01)
Latino	-0.023	(0.02)
Afro-américain	-0.055	(0.02)
Statut matrimonial :		
Marié	-0.068	(0.03)
Divorcé	0.001	(0.03)
Années d'éducation :		
12 années	0.017	(0.02)
13-15 années	0.003	(0.02)
16 années ou plus	-0.001	(0.02)
A une maladie chronique	0.031	(0.01)
Région de résidence :		
Midwest	-0.016	(0.02)
South	-0.005	(0.02)
West	0.003	(0.02)
Nbre de pbs d'anxiété dans l'enfance	0.076	(0.01)
Nbre d'obs.	2 779	

1. Ecrire le modèle qui a été estimé et quelle a été la technique d'estimation ?
2. Interpréter "littérairement" l'ensemble des résultats de la régression ?
3. Donner la formule qui permettrait de calculer l'effet marginal de l'âge pour un individu avec des caractéristiques  $X$  données.

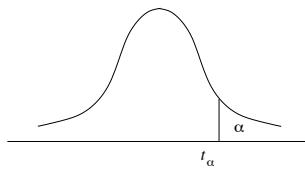
4. Le tableau ci-dessous confronte les prédictions du modèle aux réponses réelles.

- (a) Expliquez comment a été construit ce tableau ?
- (b) Quelle(s) conclusion(s) pouvez-vous tirer de ce tableau ?

	Observé	Total
Prédit	313	65
	1	2400
Total	314	2 779

5. Expliquer littérairement en quoi l'hypothèse de normalité des résidus est nécessaire pour ce modèle de régression.

### Loi de Student



d.d.l.	$\alpha$	0.10	0.05	0.025	0.01	0.005
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807
24		1.318	1.711	2.064	2.492	2.797
25		1.316	1.708	2.060	2.485	2.787
26		1.315	1.706	2.056	2.479	2.779
27		1.314	1.703	2.052	2.473	2.771
28		1.313	1.701	2.048	2.467	2.763
29		1.311	1.699	2.045	2.462	2.756
$\infty$		1.282	1.645	1.960	2.326	2.576

Exercice 1 :

$$1) F(16, 16485) = \frac{R^2}{1-R^2} \times \frac{16485}{16} \quad R^2 = 0.1046$$

$$= 120,3604$$

$\Rightarrow$  Le modèle est globalement significatif.

$$2) \text{Variane résiduelle} = 385801.962 / 16485 = 23,403$$

Elle représente la variance résiduelle

$$3) \bar{R}^2 = 1 - \frac{SCR/(F-k)}{SCT/(F-1)} = 1 - \frac{385801.962 / 16485}{430863.233 / 16501} = 0.1037$$

Le modèle peut d'expliquer 10,37% de variance de Distanc DT.

$$4) IC_{\text{tapispart}} = [\hat{\beta} \pm t_{F-k}^{0.025} \times \hat{\sigma}_{\beta}]$$

$$= [-1,15286 \pm 1,96 \times 0,111663]$$

$$= [-3,0015; 0,9184]$$

Dès à 95% de chance le coeff associé à la variable partiel sera compris dans IC

5) L'intérêt est de tenir compte de l'effet non linéaire de l'âge.

Car selon la va étudiée l'effet de l'âge peut être grand au début puis pas après etc.

6) Toutes ces va sont significatives sauf Age et Age2.

effets...

7)