Econométrie: Exercices

Institut Supérieur d'Economie et de Management

Licence 3ème année EG

Catherine Laffineur, Maître de conférences en sciences économiques

catherine.laffineur@unice.fr

Exercice pour la séance

- Le tableau ci-dessous présente le revenu moyen par habitant en dollars pour un pays

Annee	revenu
1983	8000
1984	9000
1985	9500
1986	9500
1987	9800
1988	11000
1989	12000
1990	13000
1991	15000
1992	16000

- Sachant que la propension à consommer est de 0.8 et que la consommation incompressible est de 1000, définissez la fonction de consommation

Correction

Annee	revenu	Conso théorique
1983	8000	7400
1984	9000	8200
1985	9500	8600
1986	9500	8600
1987	9800	8840
1988	11000	9800
1989	12000	10600
1990	13000	11400
1991	15000	13000
1992	16000	13800

Exercice pour la séance

- Calculez la consommation théorique de 1983 à 1992
- Supposons que les erreurs suivent une loi normale $N \sim N(0; 2000)$, déterminez la consommation observée

Correction

Annee	revenu	Conso théorique	Erreur	conso
1983	8000	7400	83	7483
1984	9000	8200	-28	8172
1985	9500	8600	206	8806
1986	9500	8600	23	8623
1987	9800	8840	110	8950
1988	11000	9800	4	9804
1989	12000	10600	-45	10555
1990	13000	11400	11	11411
1991	15000	13000	45	13045
1992	16000	13800	111	13911

Exercice pour la séance

- Réalisez un ajustement linéaire, calculez les valeurs des estimateurs et leurs variances
- $\bar{y} = 10076 \rightarrow$ moyenne conso
- $\bar{x} = 11280 \rightarrow$ moyenne revenu
- $\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$=0.798
- $\hat{a} = \bar{y} - \hat{b}\bar{x} = 1069.95$

Exercice pour la séance

- Pour déterminer la variance de a et b, il faut d'abord déterminer la variance de u_i
- $\hat{\sigma}^2 = \frac{\sum \hat{u_i^2}}{N-2}$=79246/8=9905.65
- Si on ne connaissait pas u_i, on peut passer par la formule suivante:
  $\hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{N-2}$, avec $\hat{y}_i = \hat{a} + \hat{b}x_i$
- $V(\hat{a}) = \frac{\hat{\sigma^2}}{N} + \frac{112802^2 * 9905.75}{64156000} = 20636.31$
- $V(\hat{b}) = \frac{\hat{\sigma^2}}{\sum (x_i - \bar{x})^2} = 0.00015$

Exercice pour la séance

- En prenant un seuil de confiance à 95%, peut on dire que la propension marginale à consommer est différente de zéro? Donner un intervalle de confiance pour cette propension marginale à consommer.
- Rappel de l'IC:

$$I_c = \left[\hat{b} - t_{\frac{\alpha}{2}} * \sigma_{\hat{b}}; \hat{b} + t_{\frac{\alpha}{2}} * \sigma_{\hat{b}}\right]$$

Correction

	γ										
k	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587

- $IC_{\hat{b}} = [0.798 - (2.306 * \sqrt{0.00015}); 0.798 + (2.306 * \sqrt{0.00015})]$= [0.77;0.826]
- $IC_{\hat{a}} = [1069.95 - (2.306 * \sqrt{20636.31}); 1069.95 + (2.306 * \sqrt{20636.31})]$=[801.49;1338.40]
- Conclusion: les coefficients sont significatifs
- On verra par la suite comment calculer la statistique de student
- Vérification sous logiciel

Exercice pour la séance

- On considère le modèle linéaire suivant:

$$y_i = a + bx_i + \varepsilon_i, \qquad i = 1, \ldots, 7$$

- y_i représente le rendement par hectare de blé, exprimé en quintal par hectare.
- x_i représente la quantité d'engrais utilisée exprimée en Kg/ha.
- ε_i représente le terme aléatoire qui vérifie les hypothèses classiques des MCO.
  On dispose des observations suivantes:

y_i	40	45	50	65	70	70	80
x_i	100	200	300	400	500	600	700

Exercice pour la séance

- 1) Estimer par la méthode des MCO, les paramètres du modèle (a, b et σ^2)
- 2) Etablir des intervalles de confiance, au niveau de 95%, pour a,b.
- 3) Tester, individuellement, la significativité des paramètres a et b au seuil de 5%.
- 4) Etablir le tableau d'analyse de la variance. Tester la significativité globale du modèle, au seuil de 5%.

Correction: question 1

- $\hat{a} = \bar{y} - \hat{b}\bar{x} = 32.8$
- $\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = 0.068$
- $\hat{\sigma^2} = \frac{\sum(y_i - \hat{y}_i)^2}{(N-2)}$ où $\sum(y_i - \hat{y}_i)^2$:

$(40 - (32,8 + 0,068 * 100))^2$	=0,16
$(45 - (32,8 + 0,068 * 200))^2$	=1,96
$(50 - (32,8 + 0,068 * 300))^2$	=10,24
$(65 - (32,8 + 0,068 * 400))^2$	=25
$(70 - (32,8 + 0,068 * 500))^2$	=10,24
$(70 - (32,8 + 0,068 * 600))^2$	=12,96
$(80 - (32,8 + 0,068 * 700))^2$	=0,16
Somme	60,72

- $\sigma^2 = \frac{60.72}{5} = 12,144$

Intervalle de confiance

- Intervalle de confiance de $\hat{a}$:
- $IC = [\hat{a} - t_{\alpha/2}\sigma_{\hat{a}}; \hat{a} + t_{\alpha/2}\sigma_{\hat{a}}$
- Avec $\sigma_{\hat{a}} = \sqrt{V(\hat{a}} = \frac{\hat{\sigma^2}}{N} + \frac{\bar{x}^2 * \hat{\sigma^2}}{\sum(x_i - \bar{x})^2} = \frac{12,1444}{7} + \frac{400^2 * 12,144}{280000} = 8.674$
- $IC = [32.8 - 2.5706 * \sqrt{8.674}; 32.8 + 2.5706 * \sqrt{8.674}] = [25.23; 40.37]$
- intervalle de confiance de $\hat{b}$
- $IC = [\hat{b} - t_{\alpha/2}\sigma_{\hat{b}}; \hat{b} + t_{\alpha/2}\sigma_{\hat{b}}]$
- Avec $\sigma_{\hat{b}} = \sqrt{V(\hat{b})} = \frac{\hat{\sigma^2}}{\sum(x_i - \bar{x})^2}$
- $IC = [0.051; 0.085]$

Correction question 3

- On teste la significativité des paramètres à l'aide d'un test de student:
  - $H_0 : a = 0$
  - $H_1 : a \neq 0$
- La statistique calculée est donc: $t_c = \frac{\hat{a} - a_{H_0}}{\sigma_{\hat{a}}} = \frac{32.8 - 0}{\sqrt{8.67}} = 11.13$
- La statistique de la table pour $\alpha = 0.05$ avec $N - 2 = 7 - 2 = 5$ ddl est 2.5706
- Comme $|t_c| > t_{tab}$ on conclue que a est statistiquement différent de zéro
- On réalise le même calcul pour $t_{\hat{b}} = \frac{0.067 - 0}{\sqrt{0.00043}} = 3.28 > 2.5706 \rightarrow$ b est statistiquement différent de zéro

Correction question 4

- Tableau d'analyse de la variance et test de significativité du modèle

	Somme des carrés	
Regression	SCE	1294.72
Résidus	SCR	60.72
Total	SCT	1355.44

- Les SCE sont obtenus: $\sum(\hat{y}_i - \bar{y})^2$

$[(32.8 + 0.068 * 100) - 60]^2$	416.16
$[(32.8 + 0.068 * 200) - 60]^2$	184.96
$[(32.8 + 0.068 * 300) - 60]^2$	46.24
$[(32.8 + 0.068 * 400) - 60]^2$	0
$[(32.8 + 0.068 * 500) - 60]^2$	46.24
$[(32.8 + 0.068 * 600) - 60]^2$	184.96
$[(32.8 + 0.068 * 700) - 60]^2$	416.16
Somme	1294.72

Correction question 4

	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

- $R^2 = \frac{SCE}{SCT} = 0.95$
- $H_0 : b = 0$ et $H_1 : b \neq 0$
- $F_c = \frac{SCE}{SCR}(N-2) = \frac{1294.72}{60.72} * 5 = 106.61$
- La statistique de la table de Fisher $F_{1;5}^{0.05}$=6.61
- Comme $F_c > F_{tab}$ nous rejetons H_0 le modèle est significatif

Exercice

- Soit une équation économétrique $Salaire_i = a + bExprience + u_i$

Individu	Salaire	Expérience
1	2500	12
2	2000	8
3	3000	14

Questions

- Calculez $\hat{a}$ et $\hat{b}$
- Comment s'interprète le coefficient a et b?
- L'effet marginal de l'expérience est-il statistiquement différent de zéro?
- Déterminez un intervalle de confiance pour $\hat{b}$
- Calculez le coefficient de détermination et effectuer le test de Fisher permettant de déterminer si la régression est globalement significative
- Quelle est la conséquence sur le salaire d'une augmentation de 3 ans d'expérience?
- Quel sera le salaire de l'individu 3 dans 1 et 2 ans?

Exercice modèle linéaire multiple

- On considère le modèle linéaire suivant: $y_i = b_0 + b_1 x_1 + b_2 x_2 + u_i$. On dispose des observations suivantes:

$$y_i : 0, 24, 12, 8, 12, 16$$

$$x_{i1} : -2, -1, 0, 0, 1, 2$$

$$x_{i2} : -5, 4, 0, -2, 2, 1$$

- Ecrire ce modèle sous forme matricielle
- Calculez $\hat{B}$ et sa matrice de variance covariance
- Testez la significativité de chaque coefficient
- Calculez le R^2 et interprétez le résultat

Exercice sur l'interprétation des résultats

Dans cet exercice nous illustrons l'utilisation des variables indicatrices par une série de régressions examinant les déterminants du poids d'un bébé à la naissance.

1. Dans la première régression, le poids à la naissance, mesuré en grammes, est régressé sur le nombre moyen de cigarettes fumées quotidiennement par la mère durant sa grossesse (x). Le résultat obtenu est le suivant :

$$\text{Poids} = \underset{(14)}{3418} - \underset{(2,1)}{7{,}2}\,x \quad R^2 = 0{,}012 \quad n = 964 \quad SCR = 158{,}6$$

   où n est le nombre d'observations et où l'écart-type des estimateurs est donné entre parenthèses.

   - Quel est le poids moyen à la naissance d'un enfant dont la mère ne fume pas?
   - Quel est le poids moyen d'un enfant dont la mère fume 10 cigarettes par jour?
   - Que pensez-vous du R^2 de cette régression? Qu'en concluez-vous sur la relation mise en évidence?
   - Comment aurait-on pu modifier la régression pour autoriser l'effet du nombre de cigarettes fumées quotidiennement à varier avec celui-ci (par ex. on peut imaginer que fumer une cigarette par jour n'ait aucun effet et que les effets négatifs commencent à partir de 5 cigarettes. Comment pourrait-on faire apparaître ceci dans la régression?)

Exercice sur l'interprétation des résultats

2. En pratique on sait que le poids moyen d'un bébé à la naissance varie selon que celui-ci est ou n'est pas le premier bébé auquel sa mère donne naissance: les aînés tendent à avoir un poids plus faible. Pour examiner les déterminants du poids à la naissance il est donc utile de tenir compte de ce paramètre. On considère donc la variable indicatrice d telle que d=1 si la mère a déjà eu un enfant et 0 sinon. Les résultats obtenus lorsque l'on ajoute d à la liste des régresseurs sont :

$$\text{Poids} = \underset{(17)}{3373} + \underset{(26)}{119}\,d - \underset{(2{,}1)}{7{,}8}\,x \qquad R^2 = 0{,}032 \quad (*)$$

Quel est le poids moyen à la naissance d'un enfant dont la mère ne fume pas :

- s'il est le premier né ?
- S'il n'est pas le premier né ?

3. Au vu des résultats précédents, il semble clair que le fait d'être ou non le premier né a un effet significatif sur le poids de naissance. On peut se demander si le nombre de frères et sœurs a un effet ou si cela est juste le fait que la mère ait eu des enfants auparavant qui importe. Pour vérifier cette hypothèse on peut effectuer la régression :

$$\text{Poids} = \alpha + \beta x + \gamma z + u$$

où z est le nombre de naissances ayant précédé. On peut également envisager de créer une série de variables indicatrices d_0, d_1, d_2, d_3 correspondant au nombre de naissances ayant précédé :

- $d_0 = 1$ si aucune naissance n'a précédé et 0 sinon
- $d_1 = 1$ si une naissance a précédé et 0 sinon
- $d_2 = 1$ si deux naissances ont précédé et 0 sinon
- $d_3 = 1$ si trois naissances ou plus ont précédé et 0 sinon.

Le modèle à estimer pourrait alors s'écrire :

$$\text{Poids} = \alpha + \beta x + \delta_1 d_1 + \delta_2 d_2 + \delta_3 d_3 + u$$

- En quoi cette spécification est-elle différente de celle où le nombre de naissances ayant précédé est inclus sous la forme d'une variable continue ? Quelle spécification vous semble préférable ? Pour quelle raison ?

- Pour quelle raison la variable indicatrice d_0 n'est-elle pas incluse dans la régression ? Pouvez-vous proposer une autre spécification ?

L'estimation du modèle ci-dessus donne les résultats suivants :

$$\text{Poids} = \underset{(17)}{3373} - \underset{(2,1)}{7,8}\, x + \underset{(30)}{127}\, d_1 + \underset{(49)}{102}\, d_2 + \underset{(61)}{105}\, d_3 \qquad R^2 = 0,033 \qquad SCR = 155,3$$

- Comment s'interprètent les coefficients des indicatrices d_1, d_2, d_3 ? Quel est le poids moyen à la naissance d'un enfant ayant deux frères et/ou sœurs aînés ? Et celui d'un premier né ?
- A la lecture de ces résultats, pensez-vous que le nombre d'enfants ayant précédé est important, ou que ce qui compte le plus c'est le fait d'être ou non l'aîné ?
- Testez l'hypothèse que les coefficients des trois indicatrices d_1, d_2, d_3 sont conjointement nuls.

Exercice sur l'interprétation des résultats

4. Le modèle a également été estimé en ajoutant cette fois les indicatrices d_0, d_2, d_3 au modèle de base. L'estimation obtenue est la suivante :

$$\text{Poids} = \underset{(26)}{3500} - \underset{(2,1)}{7{,}8}\,x - \underset{(30)}{127}\,d_0 - \underset{(52)}{25}\,d_2 - \underset{(64)}{22}\,d_3 \qquad R^2 = 0{,}033$$

Par rapport à la régression précédente, comment expliquez-vous les changements constatés dans la valeur et le degré de significativité des coefficients des variables indicatrices ?

5. On désire maintenant examiner si la situation de famille de la mère a un effet sur le poids du bébé. Les mères célibataires sont en effet souvent dans des situations difficiles et pour cette raison on peut imaginer qu'elles ont des bébés d'un poids inférieur à la moyenne. Soit donc *Celib* une variable indicatrice prenant la valeur 1 si la mère et seule et 0 sinon. La régression obtenue lorsque l'on ajoute cette variable parmi les régresseurs est :

$$\text{Poids} = \underset{(28)}{3386} + \underset{(27)}{109}\, d - \underset{(47)}{132}\, Celib - \underset{(2,1)}{7{,}2}\, x \qquad R^2 = 0{,}040 \quad (**)$$

- Quel est le poids moyen des enfants premiers nés dont la mère est célibataire et ne fume pas ?
- Le tableau ci-dessous présente un résumé statistique de certaines caractéristiques de l'échantillon :

	Effectif	% dont c'est la première naissance	% ayant fumé durant la grossesse
Mères célibataires	83	80,7	42,2
Autres mères	881	58,7	21,6

- En quoi ces statistiques vous permettent-elles d'expliquer les différences observées entre les régressions (*) et (**) ? Examinez en particulier les coefficients des variables *x* et *d*.