

Université abderrahmane mira Bejaia



Faculté des Sciences Economiques, des Sciences Commerciales et de Gestion

Département des sciences économiques

Solution de la Série n°02

Econométrie des variables qualitatives

Master II Economie Quantitative

Année 2023/2024

Série n° 02 : *Estimation des modèles Probit et Logit binaires explicatifs des facteurs de la réussite en Licence (Extrait du livre : Bourbonnais, R. (2018)).¹*

Nous avons relevé sur un échantillon de 60 étudiants inscrits en dernière année de Licence d'Économie, les variables suivantes susceptibles d'expliquer la réussite ou l'échec à l'examen de Licence (variable **REUSSITE** = 0 si échec, 1 sinon) :

NENFANTS = variable discrète représentant le nombre de frères et sœurs de l'étudiant,

NECONO = la note d'économétrie sur 20 obtenue en Licence,

NMICRO = la note de micro-économie sur 20 obtenue en Licence,

GENRE = variable muette, (1 = masculin, 0 = féminin).

Un extrait des données est présenté dans le tableau 1.

Tableau 01 – Extrait de données

OBS	REUSSITE	NENFANTS	NECONO	NMICRO	GENRE
1	0	2	3,6	0	1
2	0	5	3,8	0	1
...
59	1	0	16,2	12	0
60	1	2	17	4	0

On demande :

¹ Bourbonnais, R. (2018), « Econométrie », édition DUNOD, p352.

- 1) d'estimer un modèle de type Logit permettant de prévoir la probabilité de réussite d'un étudiant en Licence,
- 2) de comparer les résultats avec un modèle de type Probit.
- 3) de donner la probabilité de réussite, à l'aide du modèle Logit estimé, pour un étudiant dont les caractéristiques sont les suivantes : NENFANTS = 1 ; NECONO = 12 ; NMICRO = 13,5 ; GENRE = masculin.

Tableau 02 : Une estimation d'un modèle Logit

Dependent Variable : REUSSITE

Method: ML – Binary Logit

Included observations: 60

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NENFANTS	– 0.682523	0.378870	– 1.801470	0.0716
NECONO	0.632062	0.239564	2.638382	0.0083
GENRE	– 3.761718	1.437068	– 2.617633	0.0089
NMICRO	0.155322	0.188916	0.822173	0.4110
C	– 3.265634	2.020060	–1.616602	0.1060

Tableau 03 : Une estimation d'un nouvel modèle Logit

Dependent Variable : REUSSITE
Method: ML – Binary Logit
Included observations: 60

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NENFANTS	– 0.746742	0.378942	– 1.970596	0.0488
NECONO	0.695857	0.231789	3.002112	0.0027
GENRE	– 3.634605	1.410945	– 2.576008	0.0100
C	– 2.859277	1.910377	– 1.496708	0.1345
Mean dependent var	0.516667	S.D. dependent var		0.503939
S.E. of regression	0.287086	Akaike info criterion		0.645890
Sum squared resid	4.615432	Schwarz criterion		0.785512
Log likelihood	– 15.37669	Hannan-Quinn criter.		0.700504
Restr. log likelihood	– 41.55549	Avg. log likelihood		– 0.25627
LR statistic (3 df)	52.35761	McFadden R-squared		0.629972
Probability(LR stat)	2.51E – 11			
Obs with Dep=0	29	Total obs		60
Obs with Dep=1	31			

Avec:

$$L_u = \text{Log likelihood}; L_R = \text{Restr. log likelihood}; LR = \text{LR statistic}; L_u/n \\ = \text{Avg. log likelihood}$$

Le critère d'information de Hannan-Quinn permet des comparaisons entre les modèles (comme les critères d'Akaike ou Schwarz) en termes d'arbitrage: apport d'information lié à l'ajout de variables explicatives et perte de degrés de liberté. En cas de modèle concurrent, celui ayant le plus faible critère d'information sera retenu.

Tableau 04 : Table de prédiction des résultats

Dependent Variable : REUSSITE
Method: ML – Binary Logit
Included observations: 60
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep = 0	Dep = 1	Total	Dep = 0	Dep = 1	Total
P(Dep =1) <= C	26	4	30	0	0	0
P(Dep =1) > C	3	27	30	29	31	60
Total	29	31	60	29	31	60
Correct	26	27	53	0	31	31
% Correct	89.66	87.10	88.33	0.00	100.00	51.67
% Incorrect	10.34	12.90	11.67	100.00	0.00	48.33
Total Gain*	89.66	– 12.90	36.67			
Percent Gain**	89.66	NA	75.86			

Tableau5 : Estimation d'un modèle Probit :

Included observations: 60

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NENFANTS	– 0.428197	0.219223	– 1.953247	0.0508
NECONO	0.363148	0.110230	3.294454	0.0010
GENRE	– 1.824203	0.650426	– 2.804629	0.0050
C	– 1.491466	1.108767	– 1.345157	0.1786

➤ **Solution**

1- Validation statistique des résultats :

À la lecture des résultats du **tableau 02**, nous constatons que :

- la variable NMICRO à une probabilité critique de 0,41, elle n'est donc pas significative,
- la variable NENFANTS à une probabilité critique de 0,07, elle est donc faiblement significative.

Nous procédons à une nouvelle estimation en retirant la variable NMICRO dont le coefficient n'est pas significativement différent de 0. A cet effet, Les résultats complets fournis par Eviews sont reportés dans le **tableau 03**. Les coefficients sont tous significativement différents de 0,

hormis le terme constant. La statistique de la Log vraisemblance est égale à $LR = 52,35$ que l'on compare à un χ^2 lu dans la table à un seuil de 0,95 % et à 3 degrés de liberté,

$$\chi_3^{2\ 0,95} = 9,28 < 52,35 \rightarrow \text{rejet de } H_0$$

Le pseudo- R^2 est donné par :

$$1 - \frac{\text{Log}(L_u)}{\text{Log}(L_R)} = 1 - \frac{-15,38}{-41,56} = 1 - 0,37 = 0,63$$

Le logiciel Eviews propose une table permettant d'appréhender les qualités prévisionnelles du modèle sur l'échantillon en comparant la probabilité estimée pour un individu i d'être $y_i = 1(P(\text{Dep} = 1))$ au seuil arbitraire de 50 % à la valeur observée des $y_i = 0$ ou 1.

D'après le **tableau 05**, les individus (29) pour lesquels $y_i = 0$, le modèle indique que 26 individus ont une probabilité estimée de réussite inférieure à 50 %. Dans 89,66 % des cas, les échecs sont donc correctement prévus. Pour les individus (31) pour lesquels $y_i = 1$, le modèle indique que 27 individus ont une probabilité estimée supérieure à 50 %. Dans 87,10 % des cas, les réussites sont correctement prévues. Le taux d'erreur est donc faible.

Le modèle est validé sur le plan statistique.

- Le modèle s'écrit :

$$\ln\left(\frac{P_i}{1 - P_i}\right) = -0,75NENFANTS_{(1,97)} + 0,70NECONO_{(3,00)} - 3,63GENRE_{(2,57)} - 2,86 + e_i$$

(.) = \mathcal{Z} -Statistique

e_i = Résidu d'estimation

– Le nombre de frères et sœurs du foyer agit négativement, les étudiants issus de familles nombreuses ont un taux de réussite plus faible.

– La note d'économétrie est un facteur positif de réussite.

– Enfin, les étudiants de genre masculin réussissent en général moins bien (signe négatif) que les étudiants de genre féminin.

2- D'après le **tableau 07**, les valeurs des coefficients sont de même signe mais différentes par rapport au modèle Logit car la spécification n'est pas la même. Cependant, nous pouvons retrouver, approximativement, les valeurs estimées du modèle Logit en multipliant chacun des coefficients des variables explicatives par la constante $\frac{\pi}{\sqrt{3}} \approx 1.81$.

3- Soit les caractéristiques de l'étudiant : NENFANTS = 1 ; NECONO = 12 ; NMICRO = 13,5 ; GENRE = masculin. Le modèle Logit estimé (la note de micro-économie ne figurant pas dans le modèle final, elle n'est pas intégrée dans le calcul, *cf.* question 1) est le suivant :

$$\ln\left(\frac{P_i}{1-P_i}\right) = -0,75NENFANTS_{(1,97)} + 0,70NECONO_{(3,00)} - 3,63GENRE_{(2,57)} - 2,86 + e_i$$

$$\ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = -0,75 \times 1 + 0,70 \times 12 - 3,63 \times 1 - 2,86 - 2,86 = 1,109$$

$$\ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = e^{1,109} = 3,033 \rightarrow \hat{P}_i = \frac{3,033}{1+3,033} = 0,75$$

La probabilité de réussite de cet étudiant de licence est donc de 75 %.