

# Université d'Orléans - Maîtrise Econométrie

## Econométrie des Variables Qualitatives

Contrôle Continu Avril 2008

### Exercice 1 (13 points)

On s'intéresse à l'efficacité d'un type de traitement analgésique employé dans le traitement de patients âgés atteints de névralgie faciale. Pour cela, on dispose d'un échantillon de 60 patients pour lesquels les variables suivantes sont renseignées : la variable d'intérêt notée  $y$  est codée 1 si la douleur persiste suite au traitement (traitement inefficace) et 0 en cas de disparition (traitement efficace), la variable  $x_1$  renseigne sur le type de traitement reçu et a trois modalités (A et B pour deux dosages de traitement différents et P s'il s'agit d'un placebo). Enfin les variables  $x_2$ ,  $x_3$  et  $x_4$  désignent respectivement le sexe (M=masculin, F=féminin), l'âge du patient et la durée (en mois) de la présence de la maladie avant le début du traitement.

L'objectif bien évidemment est de mesurer l'impact de la variable  $x_1$  (type de traitement) sur la variable  $y$ , les autres variables étant des variables de contrôle.

**Question 1** (2 points) : On décide de modéliser la probabilité de rémission de la névralgie ( $y = 0$ ) en fonction des variables explicatives avec la procédure "logistic" de SAS. (i) La variable  $x_1$  étant qualitative, quel choix vous semble le plus adéquat (au vu de l'objectif de l'étude) pour la modalité de référence en ce qui concerne cette variable. Justifiez votre réponse. (ii) Ecrivez alors le modèle de régression logistique correspondant en fonction des variables explicatives et des paramètres à estimer.

**Question 2** (2 points) : Ecrivez la log-vraisemblance du modèle associé à un échantillon de  $T$  observations. Soit  $\hat{\beta}$  l'estimateur du MV du vecteur de paramètres. Quelles sont les propriétés asymptotiques de cet estimateur ?

**Question 3** (5 points) : Les résultats de l'estimation par maximum de vraisemblance sont présentés en Annexe. Retrouvez les valeurs non reportées marquées xxxx.

**Question 4** (3 points) : Interprétez les résultats obtenus. Quel type de traitement (A, B ou P) vous semble le plus efficace ?

**Question 5** (1 point) : Ecrivez le programme SAS utilisé pour générer les résultats de l'Annexe.

### Exercice 2 (7.5 points)

Soit  $z_1$  un vecteur de variables explicatives,  $z_2$  une variable continue et  $d_1$  une variable dichotomique.

**Question 1** (1.5 points) : Dans le modèle suivant

$$P(y = 1 | z_1, z_2) = \Phi(z_1\delta_1 + \gamma_1 z_2 + \gamma_2 z_2^2) \quad (1)$$

quel est l'effet marginal de  $z_2$  sur la probabilité que  $y = 1$  ?

**Question 2** (3 points) : Dans le modèle suivant

$$P(y = 1 | z_1, z_2) = \Phi(z_1\delta_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1) \quad (2)$$

(i) trouvez l'effet marginal de  $z_2$  sur la probabilité que  $y = 1$  ?

(ii) trouvez l'effet marginal de  $d_1$  sur la probabilité que  $y = 1$  ?

On considère une modélisation pour une variable qualitative dichotomique  $y_i$ . La spécification du modèle est la suivante:

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{sinon} \end{cases} \quad (3)$$

$y_i^*$  est une variable latente fonction de deux variables exogènes  $x_{1i}$  et  $x_{2i}$ . La spécification retenue est:

$$y_i^* = \alpha_1 x_{1i} + \alpha_2 x_{2i} + \varepsilon_i \quad (4)$$

Les perturbations du modèle sont supposées *i.i.d.* et suivent une distribution logistique de moyenne nulle et de variance égale à 1.

**Question 3** (1point) : Donnez l'expression du vecteur gradient ( $G_{L_i}(\alpha)$ ) pour un individu  $i$ , en fonction des exogènes et des paramètres  $\alpha = (\alpha_1 \alpha_2)'$ .

**Question 4** (2point) : Vous ne disposez que de façon partielle des résultats de l'estimation, à savoir  $\hat{\alpha}_1 = 3$ . Mais vous savez néanmoins que pour un individu de l'échantillon d'étude,  $x_{1i}$ ,  $x_{2i}$  et la première composante de  $G_{L_i}(\alpha)$  sont respectivement égales à  $-1.5$ ,  $2$  et  $1.4938$ . On suppose que pour cet individu  $y_i = 0$ . Donnez la valeur estimée du paramètre  $\alpha_2$ .

## The LOGISTIC Procedure

Model Information	
Data Set	WORK.NEURALGIA
Response Variable	y
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	60
Number of Observations Used	60

Response Profile		
Ordered Value	y	Total Frequency
1	0	35
2	1	25

Probability modeled is y='0'.

Class Level Information			
Class	Value	Design Variables	
X1	A	1	0
	B	0	1
X2	P	-1	-1
	M	-1	

<b>Model Convergence Status</b>
Convergence criterion (GCONV=1E-8) satisfied.

<b>Model Fit Statistics</b>		
<b>Criterion</b>	<b>Intercept Only</b>	<b>Intercept and Covariates</b>
<b>AIC</b>	83.503	60.736
<b>SC</b>	85.598	73.302
<b>-2 Log L</b>	81.503	48.736

<b>Testing Global Null Hypothesis: BETA=0</b>			
<b>Test</b>	<b>Chi-Square</b>	<b>DF</b>	<b>Pr &gt; ChiSq</b>
<b>Likelihood Ratio</b>	32.7675	5	<.0001
<b>Score</b>	25.6666	5	0.0001
<b>Wald</b>	14.4512	5	xxxx (1)

<b>Type 3 Analysis of Effects</b>			
<b>Effect</b>	<b>DF</b>	<b>Wald Chi-Square</b>	<b>Pr &gt; ChiSq</b>
<b>x1</b>	2	12.5310	0.0019
<b>x2</b>	1	5.2946	0.0214
<b>x3</b>	1	7.2977	0.0069
<b>x4</b>	1	0.0315	0.8591

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	18.7872	6.9653	7.2752	0.0070
x1	A	1	0.8849	0.5291	2.7969	0.0944
x1	B	1	1.4118	0.6079	5.3933	0.0202
x2	F	1	0.9161	0.3981	5.2946	0.0214
x3		1	xxxx (3)	0.0970	xxxx (2)	0.0069
x4		1	0.00586	0.0330	0.0315	0.8591

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
x1 A vs P	24.087	3.288	176.481
x1 B vs P	40.794	4.362	381.552
x2 F vs M	xxxx (4)	1.312	29.750
x3	0.769	0.636	0.931
x4	1.006	0.943	1.073

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.5	Somers' D	0.810
Percent Discordant	9.5	Gamma	0.810
Percent Tied	0.0	Tau-a	0.401
Pairs	875	c	0.905