CHAPITRE 5 : LES MODELES A CHOIX ORDONNES

SECTION 1 : Exemple : le choix d'éducation post-secondaire (suite du chapitre 3)

Base de données : EDUCATION.XLS

N=1000 individus

Variable endogène :

$$\mathbf{y} = \begin{cases} \mathbf{1} & \textbf{pas au collège} \\ \mathbf{2} & \textbf{2 ans au collège} \\ \mathbf{3} & \textbf{4 ans au collège} \end{cases}$$

Variables exogènes :

ETCATHO = 1 si études secondaires catholiques

NIVEAU = index moyen en mathématiques, anglais et études sociales sur échelle de 13 points avec 1 le plus élevé et 13 le plus faible

REVENU = revenu familial brut en $

PERSON = Nombre de personnes dans la famille

DIPARENT = 1 si la plupart des parents formés étaient diplômés du collège ou avaient un diplôme plus élevé

FEMME = 1 si femme, 0 si homme

NOIR = 1 si noir, 0 si autres

On va transformer l'énoncé précédent de la façon suivante :

<u>Base de données</u> : EDUCATION.XLS

N=1000 individus

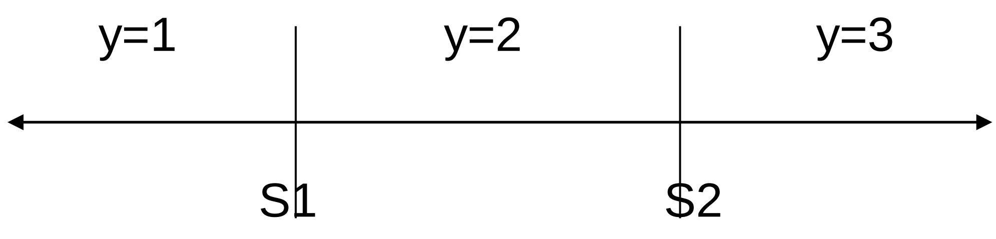
<u>Variable endogène</u> :

$$\mathbf{y} = \begin{cases} \mathbf{1} & \textbf{pas au collège} \quad \textbf{mauvaise performance} \\ \mathbf{2} & \textbf{2 ans au collège} \quad \textbf{moyen} \\ \mathbf{3} & \textbf{4 ans au collège} \quad \textbf{super performance} \end{cases}$$

Nous avons donc ordonné nos réponses de la forme :
3 TB, 2 M et 1 N (N pour Nul). C'est notre ressentiment.

$$\mathbf{y} = \begin{cases} \mathbf{1} & \mathbf{N} \\ \mathbf{2} & \mathbf{M} \\ \mathbf{3} & \mathbf{TB} \end{cases}$$

Représentons cette variable i.e notre ressentiment vis-à-vis des études dans le temps sous forme de seuil :



Nous devons rechercher ces 2 seuils :

$$\mathbf{y} = \begin{cases} \mathbf{1} & \textbf{si } \boldsymbol{y^* \le S1} \\ \mathbf{2} & \textbf{si } \mathbf{S1 \le y^* \le S2} \\ \mathbf{3} & \textbf{si } \mathbf{y^*} \boldsymbol{\ge S2} \end{cases}$$

$\boldsymbol{y^*}$ représente une variable latente, notre ressentiment vis-à-vis de notre jugement, de notre classement. C'est une variable inobservable.

Variables exogènes :

NIVEAU = index moyen en mathématiques, anglais et études sociales sur échelle de 13 points avec 1 le plus élevé et 13 le plus faible

NB : on pourrait aussi garder d'autres variables exogènes.

SECTION 2 : Ecriture du modèle

Nous devons estimer le modèle suivant :

$$\boldsymbol{y^* = \beta\, niveau + \epsilon}$$

Comme $\boldsymbol{y^*}$ est inconnu, ce n'est pas un modèle de régression standard. On appelle ce modèle le modèle à index.

Il y a 3 alternatives, ce qui implique donc 2 seuils à chercher. La constante ne peut pas être présente car il y a multicolinéarité parfaite avec les seuils.

Comment estimer ce modèle ? Nous avons deux possibilités :

1) Par le modèle **PROBIT** ordonné,
2) Par le modèle **LOGIT** ordonné.

Choisissons le modèle **LOGIT** (totalement arbitraire).

SECTION 3 : Formules des probabilités et des effets marginaux

Les probabilités des choix sont données par :

$$\boldsymbol{p(y=1)=\varphi(S1-\beta\, niveau)}$$

$$\boldsymbol{p(y=2)=\varphi(S2-\beta\, niveau)-\varphi(S1-\beta\, niveau)}$$

$$\boldsymbol{p(y=3)=1-\varphi(S2-\beta\, niveau)}$$

Nb : calculs détaillés

$$\boldsymbol{p(y=1)=p(y^{*}\leq S1)=p(\beta\, niveau+\varepsilon\leq S1)}$$

$$\boldsymbol{=p(\varepsilon\leq S1-\beta\, niveau)=\varphi(S1-\beta\, niveau)}$$

On en déduit les effets marginaux :

$$\frac{dp(y=1)}{dniveau} = -\emptyset(S1 - \beta\, niveau) * \beta$$

$$\frac{dp(y=2)}{dniveau} = [\emptyset(S1 - \beta\, niveau) - \emptyset(S2 - \beta\, niveau)] * \beta$$

$$\frac{dp(y=3)}{dniveau} = \emptyset(S2 - \beta\, niveau) * \beta$$

SECTION 4 : Estimations

Par SAS 9.4 fichier EDUCATION2.SAS

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	Y
Number of Observations	1000
Missing Values	2
Log Likelihood	-877.29561
Maximum Absolute Gradient	1.7664E-6
Number of Iterations	10
Optimization Method	Quasi-Newton
AIC	1761
Schwarz Criterion	1775

Mesures du critère qualificatif de lissage

Mesure	Valeur	Formule
Likelihood Ratio (R)	282.72	2 * (LogL - LogL0)
Upper Bound of R (U)	2037.3	- 2 * LogL0
Aldrich-Nelson	0.2204	R / (R+N)
Cragg-Uhler 1	0.2463	1 - exp(-R/N)
Cragg-Uhler 2	0.2832	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.2624	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.2573	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.1388	R / U
Veall-Zimmermann	0.3286	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.5717	
N = # d'observations, K = # de régresseurs		

Algorithm converged.

Résultats estimés des paramètres

Paramètre	DDL	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr > |t|
NIVEAU	1	-0.510148	0.033926	-15.04	<.0001
_Limit1	1	-4.916916	0.267263	-18.40	<.0001
_Limit2	1	-3.477195	0.241669	-14.39	<.0001

Obs.	Y	NIVEAU	Meff_P1_NIVEAU	Meff_P2_NIVEAU	Meff_P3_NIVEAU	Prob1_Y	Prob2_Y	Prob3_Y
1001	.	6.64	0.074663	0.052617	-0.12728	0.17806	0.29950	0.52244
1002	.	2.635	0.013553	0.034764	-0.04832	0.02731	0.07862	0.89407

Interprétations :

Avec un niveau de 6.64 (la médiane), [élève médian], la probabilité de ne pas aller au collège est de 17.80%, 29.95% de rester 2 ans au collège et 52.24% de rester 4 ans au collège.

Pour un bon élève (niveau de 2.635), les probabilités sont 2.7% pour 1, 7.8% pour 2 et 89% pour 3.

A comparer avec (Cf. chapitre 3) :

Effets marginaux (au point median)

Obs.	p1	p2	p3	em1	em2	em3
1	0.18101	0.28558	0.53341	0.084148	0.044574	-0.12872

Effets marginaux (au point 5ième décile)

Obs.	p1	p2	p3	em1	em2	em3
1	0.017766	0.096545	0.88569	0.011642	0.033452	-0.045094