

Examen : Econométrie des variables qualitatives

Année Universitaire : 2020-2021

Durée : 1h45min

Vous allez répondre aux questions relatives à deux exercices. Le barème est le suivant et vous pouvez allouer votre temps en fonction.

Exercice 1 (17 points) :

Q1 (0,5 point)

Q2 (1 point)

Q3 (1,5 point)

Q4 (14 points)

Exercice 2 (3 points) :

Q1 (0,5 point)

Q2 (1,5 point)

Q3 (0,75 point)

Q4 (0,25 point)

Bonne chance !

Exercice 1 : modèle logit simple et décision de confiance

Nous considérons un “jeu de confiance” dans lequel il y a deux joueurs : **envoyeur** et **receveur**.

L’envoyeur décide en premier et a deux actions possibles : “**faire confiance**” et “**ne pas faire confiance**”.

- Si l’envoyeur décide de “**ne pas faire confiance**” alors lui et le receveur gagneront chacun **10 euros**.
- Si l’envoyeur décide de “**faire confiance**” alors les gains sont déterminés par le receveur. Le receveur dans ce cas a deux actions possibles : “**bienveillant**” et “**égoïste**” :
 - Si le receveur décide “**bienveillant**”, alors lui et l’envoyeur gagneront chacun **15 euros**
 - Si le receveur décide “**égoïste**”, alors il gagnera **22 euros** et l’envoyeur gagnera **8 euros**.

La base de données “**Trust**” (en format stata, i.e. avec extension “**.dta**”) qui vous a été envoyé par email contient les données de 100 individus (indités $i = 1, 2, \dots, 100$) qui jouent le rôle d’envoyeur. La base contient trois variables :

- **y**: variable qui prend la valeur 1 si l’envoyeur a décidé de “**faire confiance**” et 0 s’il a décidé de “**ne pas faire confiance**”.
- **r**: variable indiquant l’aversion au risque de l’envoyeur
- **p**: variable indiquant la probabilité (subjective) de l’envoyeur sur le fait que le receveur est “**bienveillant**”

Nous considérons une variable latente $y_i^* = a_0 + a_1 r_i + a_2 p_i + u_i$ qui est telle que :

Examen : Econométrie des variables qualitatives

Année Universitaire : 2020-2021

Durée : 1h45min

$$y_i = \begin{cases} 1 & \text{si } y_i^* = a_0 + a_1 r_i + a_2 p_i + u_i > 0 \\ 0 & \text{si } y_i^* = a_0 + a_1 r_i + a_2 p_i + u_i \leq 0 \end{cases}$$

et u_i suit une loi continue de fonction de répartition $F(\cdot)$ et de fonction de densité $f(\cdot)$. De plus supposons que la loi est symétrique autour de 0, impliquant $F(-x) = 1 - F(x)$ pour tout $x \in R$.

Ci-après se trouvent les questions relatives à cet exercice 1. Le symbole "Feuille" signifie que la réponse doit être écrite sur votre feuille d'examen, alors que le symbole "DO" signifie que la réponse doit être incluse dans votre fichier ".do" (dofile ou code Stata). Parfois, les 2 symboles apparaissent ensemble !

Utilisez votre nom et prénom(s) pour nommer votre dofile comme suit : **“nom_prénoms.do”**. De plus, mettez votre nom et prénoms comme un commentaire dans la première ligne du dofile. A la fin de l'examen, envoyez le dofile à l'adresse email **yt.kpegli@univ-lyon2.fr**, en mettant comme objet du mail « Exam2020 » et votre nom et prénom(s) dans le corps du mail.

Q1 (Feuille) : Montrez que la probabilité que l'envoyeur i décide de **“faire confiance”** est donnée par $P(y_i = 1 | r_i, p_i) = F(a_0 + a_1 r_i + a_2 p_i)$.

Q2 (Feuille) : Estimation par maximum de vraisemblance de a_0 , a_1 et a_2

- Montrez que la log-vraisemblance sur l'échantillon est donnée par

$$l_N(a_0, a_1, a_2; y | r, p) = \sum_{i=1}^{100} y_i \log[F(a_0 + a_1 r_i + a_2 p_i)] + \sum_{i=1}^{100} (1 - y_i) \log[1 - F(a_0 + a_1 r_i + a_2 p_i)]$$

- Comment retrouver l'estimateur maximum de vraisemblance de a_0 , a_1 et a_2 ? (Expliquez ou indiquez juste le programme de maximisation qui permet d'obtenir a_0 , a_1 et a_2).

Q3 (Feuille) : Effets marginaux

- Montrez que l'effet marginal de r_i (i.e. aversion aux risques) sur la probabilité que l'envoyeur i décide de **“faire confiance”** est donnée par

$$a_1 * f(a_0 + a_1 r_i + a_2 p_i)$$

- Expliquez la notion d'« effet marginal évalué à la moyenne de l'échantillon » (MEM)
- Expliquez la notion de « moyenne des effets marginaux sur l'échantillon » (AME)

Dans la question 4, nous supposons que u_i suit une loi logistique.

Q4 (DO + Feuille) :

- Importez la base de données **“Trust”** dans stata. **(DO)**

Examen : Econométrie des variables qualitatives

Année Universitaire : 2020-2021

Durée : 1h45min

- Faites une statistique descriptive de la **variable binaire y** et interprétez la moyenne de cette variable. **(DO + Feuille)**
- Utilisez la commande « logit » de STATA pour donner les valeurs estimées de a_0 , a_1 et a_2 . **(DO + Feuille)**
- Interprétez a_1 et a_2 . **(Feuille)**
- Utilisez la commande « margin » de STATA pour déterminer les effets marginaux évalués à la moyenne de l'échantillon (MEM). Interprétez. **(DO + Feuille)**
- Utilisez la commande « margin » de STATA pour déterminer la moyenne des effets marginaux sur l'échantillon (AME). Interprétez. **(DO + Feuille)**

Exercice 2 : modèle logit ordonné et « chaleur » de la relation mère-enfant

Considérons l'exemple consistant à expliquer la « chaleur » de la relation mère-enfant se trouvant dans le cours magistral.

- warm** est la variable qui capte cette relation mère-enfant en quatre (04) modalités codifiées de 1 (« froid ») à 4 (« chaud »).
- Les variables explicatives sont :
 - **yr89** : variable binaire indiquant si l'enquête est réalisée en 1989¹
 - **male** : variable binaire indiquant si l'enfant est de sexe masculin
 - **white** : variable binaire indiquant si la mère est de race blanche
 - **age** : variable indiquant l'âge de la mère
 - **ed** : niveau d'éducation de la mère
 - **prst** : prestige occupationnel de la mère (échelle 12-82)

Nous considérons que la variable explicative (**warm**) et les variables explicatives (**yr89**, **male**, **white**, **age**, **ed**, **prst**) sont liées comme suit :

$$\mathbf{warm} = \begin{cases} 1 & \text{si} & \mathbf{w}^* < c_1 \\ 2 & \text{si} & c_1 \leq \mathbf{w}^* < c_2 \\ 3 & \text{si} & c_2 \leq \mathbf{w}^* < c_3 \\ 4 & \text{si} & c_3 \leq \mathbf{w}^* \end{cases}$$

avec $\mathbf{w}^* = a_1 * \mathbf{yr89} + a_2 * \mathbf{male} + a_3 * \mathbf{white} + a_4 * \mathbf{age} + a_5 * \mathbf{ed} + a_6 * \mathbf{prst} + u$

et u suit une loi logistique.

Les résultats d'un modèle **logit ordonné** sont obtenus à l'aide de STATA et sont présentés dans le tableau 1.

Q1: Comment appelle-t-on la variable w^* ?

Q2: A partir du tableau 1, donnez les valeurs de a_2 et a_6 . Interprétez.

¹ Notez que la base de données provient de deux enquêtes, la première étant réalisée en 1979 et la seconde en 1989.

Examen : Econométrie des variables qualitatives

Année Universitaire : 2020-2021

Durée : 1h45min

Q3: A partir du tableau 1, donnez la valeur de c_1 , c_2 et c_3 .

Q4: Expliquez pourquoi il n'y a pas de constante dans l'expression de w^* .

Tableau 1 : résultats logit ordonné

```
. ologit warm yr89 male white age ed prst
```

```
Iteration 0:  log likelihood = -2995.7704
Iteration 1:  log likelihood = -2846.4532
Iteration 2:  log likelihood = -2844.9142
Iteration 3:  log likelihood = -2844.9123
Iteration 4:  log likelihood = -2844.9123
```

Ordered logistic regression

Number of obs = 2,293

LR chi2(6) = 301.72

Prob > chi2 = 0.0000

Pseudo R2 = 0.0504

Log likelihood = -2844.9123

warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
yr89	.5239025	.0798989	6.56	0.000	.3673036	.6805014
male	-.7332997	.0784827	-9.34	0.000	-.887123	-.5794765
white	-.3911595	.1183808	-3.30	0.001	-.6231816	-.1591373
age	-.0216655	.0024683	-8.78	0.000	-.0265032	-.0168278
ed	.0671728	.015975	4.20	0.000	.0358624	.0984831
prst	.0060727	.0032929	1.84	0.065	-.0003813	.0125267
/cut1	-2.465362	.2389128			-2.933622	-1.997102
/cut2	-.630904	.2333156			-1.088194	-.1736138
/cut3	1.261854	.234018			.8031871	1.720521

Fin de l'épreuve.