
Tarea 3

Econometría Aplicada

Nombre: Rafael Sandoval Fernández

Fecha: 7 de diciembre de 2020

Clave única: 143689

Tema: Logit Multinomial

Pregunta 1: Estadística descriptiva

A continuación, se presenta la tabla con estadísticas descriptivas de la base de datos *Stove*:

Tabla 1: Estadísticas descriptivas

variable	count	mean	sd	min	max
idcase	900	450.5	259.9519	1	900
stove	900	1.787778	1.240073	1	5
ic_gc	900	776.8266	115.563	431.83	1158.9
ic_gr	900	921.7702	138.0867	574.94	1344
ic_ec	900	824.5435	125.2739	469.61	1230.5
ic_er	900	983.928	147.1554	546.82	1496.3
ic_hp	900	1046.481	156.703	532.32	1679.8
oc_gc	900	172.1158	25.93978	84.016	248.43
oc_gr	900	154.4714	22.88574	77.863	227.92
oc_ec	900	476.803	73.15333	237.44	705.36
oc_er	900	429.7299	65.79263	179.95	664.43
oc_hp	900	219.2993	32.96955	120.97	318.58
income	900	4.641111	1.684543	2	7
age	900	42.94444	14.09475	20	65
num_people	900	4.424444	1.743847	2	7
region	900	2.494444	1.105639	1	4

Pregunta 2: Adquisición de estufas (Composición por región)

Tabla 2: Composición de estufas por región

stove	region				Total
	valley	scostl	mountn	ncostl	
gc	60.45	61.22	57.84	71.54	63.67
gr	15.25	16.07	16.67	10.38	14.33
ec	7.34	6.93	7.84	6.92	7.11
er	10.17	10.25	10.78	6.92	9.33
hp	6.78	5.54	6.86	4.23	5.56
Total	100.00	100.00	100.00	100.00	100.00

Pregunta 3: Estimación logit multinomial

En la tabla 3 se muestran los resultados de la regresión

$$Pr(stove_i = j | region) = P_j(region_i \cdot \beta)$$

En el excel se calcularon las probabilidades de forma manual, con ayuda de:

$$Pr(Y_i = j | X_i) = \frac{\exp(X_i' \beta_j)}{1 + \sum_{l=0}^J \exp(X_i' \beta_l)}$$
$$Pr(Y_i = 0 | X_i) = \frac{1}{1 + \sum_{l=0}^J \exp(X_i' \beta_l)}$$

Pregunta 4: Alternative-specific conditional logit

Cuando algunos de los regresores son específicos para alguna alternativa, el logit condicional específica que:

$$p_{ij} = \frac{\exp(x'_{ij}\beta + z'_i\gamma_j)}{\sum_{l=1}^m \exp(x'_{il}\beta + z'_i\gamma_l)}, \quad j = 1, \dots, m$$

Donde x_{ij} son regresores específicos para una alternativa y z_i son regresores específicos para un caso. En este ejemplo tenemos:

- depvar = *choice*
- indepvar = $\{oc, ic\} \leftarrow$ regresor específico para una alternativa ($x_{i,j}$)
- casevars = $\{age, num_people, income, i.region\} \leftarrow z_i$

Al final del documento, en la tabla 4 se presentan los resultados de la regresión, se omitieron los errores estándar para poder mostrar la tabla completa.

Tabla 3: Estimación logit multinomial

	stove
gr	
scostl	0.0393 (0.261)
mountn	0.133 (0.350)
ncostl	-0.553 (0.298)
Constant	-1.377*** (0.215)
ec	
scostl	-0.0714 (0.362)
mountn	0.110 (0.478)
ncostl	-0.227 (0.384)
Constant	-2.108*** (0.294)
er	
scostl	-0.00479 (0.311)
mountn	0.103 (0.416)
ncostl	-0.553 (0.355)
Constant	-1.782*** (0.255)
hp	
scostl	-0.215 (0.384)
mountn	0.0563 (0.502)
ncostl	-0.640 (0.435)
Constant	-2.188*** (0.304)
Observations	900

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

gc se tomó como categoría base

Tabla 4: Estimación logit condicional para alternativa específica

	choice
option	
ic	-0.00151*
oc	-0.00695***
gr	
age	0.00305
num_people	-0.0162
income	-0.110
scostl	0.0244
mountn	0.116
ncostl	-0.576
Constant	-0.840
ec	
age	0.00434
num_people	0.0522
income	0.00684
scostl	-0.0451
mountn	0.121
ncostl	-0.249
Constant	-0.445
er	
age	-0.0214*
num_people	0.0264
income	-0.0312
scostl	-0.0342
mountn	0.00836
ncostl	-0.599
Constant	1.212
hp	
age	-0.0151
num_people	0.00837
income	0.0663
scostl	-0.220
mountn	0.0459
ncostl	-0.641
Constant	-1.186
Observations	4500

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

gc se tomó como categoría base

Pregunta 5

Al correr el comando `estat mfx, varlist(num_people)` en Stata se obtiene que ante un aumento de 1 persona la probabilidad de elegir una estufa *er* aumenta en $0.003305 \times 100 = 0.3305$ puntos porcentuales. Entonces, ante un aumento de 2 personas en el hogar, la probabilidad de elegir una estufa *er* aumenta en $2 \times 0.3305 = 0.6610$ puntos porcentuales

. estat mfx, varlist(num_people)							
Pr(choice = gc 1 selected) = .61401941							
variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]
casevars							
num_people	-.002608	.009618	-0.27	0.786	-.021459	.016244	4.4244
Pr(choice = gr 1 selected) = .15321375							
variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]
casevars							
num_people	-.003134	.007206	-0.43	0.664	-.017257	.010989	4.4244
Pr(choice = ec 1 selected) = .06892472							
variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]
casevars							
num_people	.003305	.004836	0.68	0.494	-.006174	.012785	4.4244
Pr(choice = er 1 selected) = .09747322							
variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]
casevars							
num_people	<u>.002163</u>	.005891	0.37	0.713	-.009383	.01371	4.4244

Nota: Aquí se calculo el efecto parcial del hogar promedio (EPX), no logré averigüar cómo obtener el equivalente a `margins, dydx`, el cual nos daría el efecto parcial promedio (EPP)

Pregunta 6

En términos generales, la idea es que la R^2 puede ser vista como la variabilidad explicada o también puede ser vista como una mejoría del modelo sin regresores al modelo con regresores. Se puede usar la R^2 ajustada de McFadden, la cual toma en cuenta los dos elementos previos. Esta medida de ajuste se determina de acuerdo a la siguiente fórmula:

$$R_{MF}^2 = 1 - \frac{\ln \hat{L}(M_{full})}{\ln \hat{L}(M_{intercept})}$$

La log-verosimilitud del modelo sin regresores (intercept) es como la suma total de cuadrados y la log-verosimilitud del modelo con regresores (full) es como la suma de los errores al cuadrado. El

ratio de ambas verosimilitudes sugiere el nivel de mejora que ofrece el modelo full sobre el modelo con interceptos. Un menor ratio (y por lo tanto, una R_{MF}^2 mayor) indica que el modelo completo es mejor que el modelo sin regresores. Para el modelo obtenemos $R_{MF}^2 = 0.313$. Así mismo, obtenemos una $R_{adj;MF}^2 = 0.279$, la cual penaliza al modelo por incluir demasiados regresores, algo similar a la R_{adj}^2 de MCO.¹

¹Fuentes:

<https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/>

[https://www.statalist.org/forums/forum/general-stata-discussion/general/](https://www.statalist.org/forums/forum/general-stata-discussion/general/1425092-appropriate-method-for-analyzing-goodness-of-fit-after-asclogit-alternative-specific-condit)

[1425092-appropriate-method-for-analyzing-goodness-of-fit-after-asclogit-alternative-specific-condit](https://www.statalist.org/forums/forum/general-stata-discussion/general/1425092-appropriate-method-for-analyzing-goodness-of-fit-after-asclogit-alternative-specific-condit)