Taller 9: Modelos Logit y Probit
Econometría 06216

10-11-2010

Profesores: Carlos Giovanni González Espitia.
Monitores: Andrés Felipe Cuadros.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 AM del 18 de Octubre de 2010.
- Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.

Instrucciones:

- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo individual. Por tanto el taller debe reflejar únicamente el trabajo del estudiante.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos con todos ellos.
- Este taller debe ser escrito en computador.

Pregunta 1

El alcalde de Cali, preocupado por las altas tasas de natalidad en la ciudad, está interesado en estudiar la relación que existe entre la paternidad, la edad y los años de educación de los caleños. Para ello contrata a un investigador reconocido y pone a su disposición una base de datos de 50 hombres casados y 50 mujeres casadas con edades entre los 20 y 42 años, que se obtuvo a través de una gran encuesta realizada en 2009. El investigador está particularmente interesado en determinar cómo la probabilidad de la presencia de un niño en un hogar está relacionada con la edad y los niveles de educación de los encuestados. Con esa idea ha definido la variable *menor* que toma el valor de 1 cuando hay un menor de edad en el hogar y 0 de lo contrario, y las variables *edad* y *escolaridad* (medida en años de educación). Su idea es estimar dos modelos, uno para hombres jefes de hogar y otro para mujeres jefes de hogar.

a.) Estime un modelo de probabilidad lineal.
b.) Interprete los coeficientes asociados a este modelo.
c.) Explique por qué los mínimos cuadrados ordinarios (MCO) no deberían ser aplicados para estimar los parámetros poblacionales en este escenario. Sea breve y conciso.

Pregunta 2

Ahora considera un modelo probit en cada caso:
a.) Escriba los dos modelos que formalicen el planteamiento del investigador y haga los supuestos que considere necesarios.
b.) Estime cada modelo utilizando EasyReg y repórtelo en una misma tabla.

Pregunta 3

a) Explique cómo derivaría los efectos marginales de cada variable independiente sobre la probabilidad de la existencia de un niño en un hogar.
b) Además de ello, interprete a priori los efectos marginales y establezca cuáles son los signos esperados.
c) Argumente si es de esperar que el efecto marginal de la edad sea más alto para los hombres o para las mujeres, haga lo mismo con el nivel de escolaridad.

Pregunta 4

Calcule para ambos, hombres y mujeres, el efecto marginal promedio de cada variable explicatoria y repórtelos en la misma tabla de la pregunta 2. Muestre en la medida de lo posible, todo su procedimiento.

Pregunta 5

Usted cree que el alcalde se equivocó al pedir dos modelos distintos para hombres y mujeres, si lo que quería era determinar si el efecto marginal de la edad es significativamente diferente en los dos grupos, pero también si el efecto marginal de la escolaridad es significativamente diferente. Cómo probaría que el alcalde se equivocó. Sea claro. Recuerde que sus afirmaciones deben tener un nivel de confianza convencional.

Pregunta 6

Una variable estocástica X tiene una función de distribución acumulativa determinada por

$$f(x) = \lambda e^{-\lambda x} \; para \; x > 0 \tag{1}$$

Con un parámetro desconocido λ. Determine el estimador de máxima verosimilitud de λ, dada una muestra de n observaciones de X.

Taller 9: Modelos Logit y Probit
Econometría 06216

10-11-2010

Profesores: Julio César Alonso y Carlos Giovanni González Espitia.
Monitores: Carolina Restrepo, Sasha, Andrés Felipe Cuadros.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 AM del 18 de Octubre de 2010.
- Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.

Instrucciones:

- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo individual. Por tanto el taller debe reflejar únicamente el trabajo del estudiante.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos con todos ellos.
- Este taller debe ser escrito en computador.

Pregunta 1

El alcalde de Cali, preocupado por las altas tasas de natalidad en la ciudad, está interesado en estudiar la relación que existe entre la paternidad, la edad y los años de educación de los caleños. Para ello contrata a un investigador reconocido y pone a su disposición una base de datos de 50 hombres casados y 50 mujeres casadas con edades entre los 20 y 42 años, que se obtuvo a través de una gran encuesta realizada en 2009. El investigador está particularmente interesado en determinar cómo la probabilidad de la presencia de un niño en un hogar está relacionada con la edad y los niveles de educación de los encuestados. Con esa idea ha definido la variable *menor* que toma el valor de 1 cuando hay un menor de edad en el hogar y 0 de lo contrario, y las variables *edad* y *escolaridad* (medida en años de educación). Su idea es estimar dos modelos, uno para hombres jefes de hogar y otro para mujeres jefes de hogar.

a.) Estime un modelo de probabilidad lineal.

Un modelo de probabilidad lineal que permitiera estimar el efecto marginal que la edad y la escolaridad sobre la probabilidad de que haya un menor en el hogar para cada género, podría ser el siguiente:

$$menor_i = \beta_0 + \beta_1 edad_i + \beta_2 escolaridad_i + \varepsilon_i$$

Las derivadas parciales de este modelo con respecto a cada variable explicativa corresponden a los cambios marginales de la probabilidad de que exista un menor en un hogar. Sin embargo, usted debió tener en cuenta que dicho modelo lineal para cada tipo de jefe de hogar presentaba problemas de heteroscedasticidad y más concretamente, que el error asociado al modelo tendría una varianza de la forma:

$$\text{var}(\varepsilon_i) = E[y_i/X_i]E\left[1 - E[y_i/X_i]\right]$$

Por lo tanto, para cada tipo de jefe de hogar se estima el modelo descrito en la ecuación 1 través del método de mínimos cuadrados ponderados, y los pesos en cada caso corresponderían a:

$$w_i = \sqrt{E[y_i/X_i]E\left[1 - E[y_i/X_i]\right]} = \sqrt{\hat{y}_i(1 - \hat{y}_i)}$$

Recuerde que para calcular el peso usted debió estimar una regresión auxiliar en la que pudiera determinar el valor de los coeficientes del modelo, y luego si efectuar la estimación formal por mínimos cuadrados ponderados. Para cada tipo de jefe de hogar se obtuvieron los siguientes resultados:

Table 1: Estimación de los modelos de probabilidad lineal

VARIABLE DEPENDIENTE: menor i		
	Estadísticos t entre paréntesis	
	Mujeres	Hombres
	MPL	MPL
	MCP	MCP
Constante	-0.417	1.665
	(-0,58)	(2.07)**
edad i	0.025	-0.0049
	(1,99)*	(-0.46)
escolaridad i	-0.021	-0.0069
	(-1,994)**	(-0.67)
R2	0.087	0.01216
R2 ajustado	0.047	-0.02987
F	2.209	0.2894
# de Obs.	49	50

(*) nivel de significancia: 10%
(**) nivel de significancia: 5%
(***) nivel de significancia: 1%
MCP: Mínimos cuadrados ponderados
MPL: Modelo de probabilidad lineal

b.) Interprete los coeficientes asociados a este modelo.
Para ambos casos, hombres y mujeres, los coeficientes de los modelos corresponden a los cambios marginales de la probabilidad de que exista un menor en un hogar. En el caso de las mujeres se tiene que:

- $$\beta_0 = -0.417$$

  Este intercepto no es significativo. Luego la probabilidad de que haya un niño en un hogar cuyo jefe sea una mujer, independientemente de la edad de la mujer o su grado de escolaridad, es cero.

- $$\beta_1 = 0.025$$

  Este coeficiente es significativo al 90%. Por cada aumento de un año en la edad de las mujeres jefes de hogar, la probabilidad de que haya un niño menor de edad se incrementa en 2.5 puntos porcentuales.

- $$\beta_2 = -0.021$$

  Este coeficiente es significativo al 95%. Por lo tanto se concluye con por cada aumento en un año de escolaridad de los jefes de familia que son mujeres, la probabilidad de que haya un menor en el hogar se incrementa en 2.1 puntos porcentuales.

En el caso de los hombres se tiene que:

- $$\alpha_0 = 1.665$$

  Este coeficiente es significativo al 95%. Por lo tanto, la probabilidad de que haya un niño en un hogar cuyo jefe sea un hombre, independientemente de la edad del hombre o su grado de escolaridad, es 166,5%. Note que este resultado no tiene mucho sentido.

- $$\alpha_1 = -0.0049$$

  Este coeficiente es estadísticamente igual a cero. Luego el cambio marginal derivado del aumento en un año en la edad de los hombres jefes de hogar, sobre la probabilidad de que en su hogar haya un menor de edad es cero.

- $$\alpha_2 = -0.0069$$

  Este coeficiente es estadísticamente igual a cero. Luego el cambio marginal derivado del aumento en un año de escolaridad de los hombres jefes de hogar, sobre la probabilidad de que en su hogar haya un menor de edad es cero.

c.) Explique por qué los mínimos cuadrados ordinarios (MCO) no deberían ser aplicados para estimar los parámetros poblacionales en este escenario. Sea breve y conciso.

Una estimación por MCO supone de entrada que los cambios marginales sobre la probabilidad derivados de los cambios en las variables explicativas son constantes para todas las observaciones. Esta suposición es bastante fuerte y a menudo errónea. En el ejemplo que se analiza, claramente el efecto sobre la probabilidad de tener un hijo debería ser mayor cuando por ejemplo, una mujer de 25 años pasa a ser un año mayor, comparado con lo que aumentaría la probabilidad de que otra mujer de 19 pase a ser un año mayor.

Por otra parte, los modelos de probabilidad lineal que se estiman a través de MCO SIEMPRE presentan problemas de heteroscedasticidad. Y además de ello, al suponer una distribución de probabilidad lineal no es posible restringir los valores de predicción entre cero y uno, lo cual implica la posibilidad de obtener probabilidades sin sentido. Pero además también existe la posibilidad de obtener varianzas negativas que impliquen deshacerse de algunos datos, es decir, reducir el tamaño de la muestra.

Pregunta 2

Ahora considera un modelo probit en cada caso:
a.) Escriba los dos modelos que formalicen el planteamiento del investigador y haga los supuestos que considere necesarios.

$P(menor = 1|x_i) = \Phi(\beta^T x_i^T) donde:$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-v/2} dv$$

$x_i = \begin{bmatrix} 1 & edad_i & escolaridad_i \end{bmatrix}$

$\beta_i^T x_i^T = \beta_0 + \beta_2 edad_i + \beta_3 escolaridad_i$

b.) Estime cada modelo utilizando EasyReg y repórtelo en una misma tabla.

Table 2: Estimación de los modelos Probit para Hombres y Mujeres

	VARIABLE DEPENDIENTE: menor i			
	Estadísticos t entre paréntesis			
	Mujeres Probit EMV		Hombres Probit EMV	
	Coeficientes	Efecto Marginal	Coeficientes	Efecto Marginal
Constante	-2.349428	–	0.6847	–
	(1.98) *		(0.70)	
edad i	0.0738	0.0230	-0.0101	-0.0039
	(-2.23) **		(-0.39)	
escolaridad i	-0.0373	-0.0116	-0.0136	-0.0053
	(-0.93)		(-0.52)	
LRI	0.10441		0.00590	
Wald	14.74 **		1.58	
ln(L)	-27.35426		-33.81406	
# de Obs.	50		50	

(*) nivel de significancia: 10%
(**) nivel de significancia: 5%
(***) nivel de significancia: 1%
Wald: corresponde al estadístico de Wald que comprueba la significancia conjunta de todas las pendientes.
EMV: Estimadores de Máxima Verosimilitud.

Pregunta 3

a) Explique cómo derivaría los efectos marginales de cada variable independiente sobre la probabilidad de la existencia de un niño en un hogar.
En el caso de los modelos de elección discreta, en este caso el modelo probit, los efectos marginales se derivan de los coeficientes asociados al modelo que corresponden a la solución del siguiente problema de maximización:

$$\begin{aligned} \underset{\beta}{Max} \left(\ln \left[\prod_{\forall i\, tal\, que\, y_i=1} \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-}^{-Z_i} e^{-\frac{t^2}{2}} dt \right) \prod_{\forall i\, tal\, que\, y_i=0} \left(\frac{1}{\sqrt{2\pi}} \int_{-}^{-Z_i} e^{-\frac{t^2}{2}} dt \right) \right] \right) \\ = \underset{\beta}{Max} \left(\ln \left[\prod_{\forall i\, tal\, que\, y_i=1} (1 - \Phi(-Z_i)) \prod_{\forall i\, tal\, que\, y_i=0} (\Phi(-Z_i)) \right] \right) \end{aligned} \tag{4}$$

De dicho problema se obtiene el vector de coeficientes que maximizan la probabilidad de que la distribución de la muestra provenga de una función de densidad normal. En este caso, se tiene que:

$$E\left[y_i\left|\mathbf{x}_i\right.\right] = 1 \bullet \left(1 - \Phi\left(-\mathbf{Z_i}\right)\right) + 0 \bullet \Phi\left(-\mathbf{Z_i}\right) = 1 - \Phi\left(-\mathbf{Z_i}\right)$$

y por lo tanto, los efectos marginales estarían dados por:

$$\frac{\partial E\left[y_{\mathbf{i}}\left|\mathbf{x}_{i_{\mathbf{i}}}\right.\right]}{\partial \mathbf{X_j}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathbf{z_i^2}}{2}} \beta_j$$

No obstante, recuerde que los efectos marginales dependen de cada observación, motivo por el cual en la práctica se promedian.

b) Además de ello, interprete a priori los efectos marginales y establezca cuáles son los signos esperados.

- **Edad:** Tanto para hombres como para mujeres, el efecto marginal de la edad debería ser positivo y se interpreta como que en promedio un año más de edad implica un incremento de $\frac{\partial E[y_{\mathbf{i}}|\mathbf{Edad}_{i_{\mathbf{i}}}]}{\partial \mathbf{Edad_j}} * 100$ puntos porcentuales en la probabilidad de que haya un menor de edad en el hogar.
- **Escolaridad:** Tanto para hombres como para mujeres jefes de hogar, el efecto marginal de la escolaridad también debería ser positivo. La interpretación de este efecto corresponde a que en promedio, un año más de escolaridad implicaría un incremento de $\frac{\partial E[y_{\mathbf{i}}|\mathbf{Escolaridad}_{i_{\mathbf{i}}}]}{\partial \mathbf{Escolaridad_j}} * 100$ puntos porcentuales en la probabilidad de que haya un menor de edad en el hogar.

c) Argumente si es de esperar que el efecto marginal de la edad sea más alto para los hombres o para las mujeres, haga lo mismo con el nivel de escolaridad.

En teoría es de esperar que el efecto marginal de la edad sobre la probabilidad de que haya un niño en un hogar donde el jefe de familia sea una mujer fuera un valor muy similar al efecto que tendría esta misma variable si el jefe de hogar fuera un hombre. El argumento apela básicamente a la condición biológica del ser humano: entre más años tenga la persona, más apta es su condición para tener un primer hijo.

Ahora bien, en el caso de la escolaridad, más años de escolaridad representan una mayor posibilidad de que la persona genere mayores ingresos. Tanto en el caso de las mujeres como de los hombres, mayores ingresos incrementarían la probabilidad de que haya un menor en el hogar. Sin embargo, algunos podrían argumentar que el efecto marginal de la escolaridad en las mujeres sobre la probabilidad de que sean madres podría ser menor con respecto al mismo análisis pero con los hombres.

Pregunta 4

Calcule para ambos, hombres y mujeres, el efecto marginal promedio de cada variable explicatoria y repórtelos en la misma tabla de la pregunta 2. Muestre en la medida de lo posible, todo su procedimiento.
Los resultados obtenidos para cada jefe de hogar se reportaron en la tabla 1.

Pregunta 5

Usted cree que el alcalde se equivocó al pedir dos modelos distintos para hombres y mujeres, si lo que quería era determinar si el efecto marginal de la edad es significativamente diferente en los dos grupos, pero también si el efecto marginal de la escolaridad es significativamente diferente. Cómo probaría que el alcalde se equivocó. Sea claro. Recuerde que sus afirmaciones deben tener un nivel de confianza convencional.

Note que para establecer si los efectos marginales de la edad y/o la escolaridad sobre la variable dependiente son estadísticamente diferentes para hombres y mujeres, NO se pueden utilizar los dos modelos que se venían trabajando por separado. La respuesta es que las estimaciones de uno y otro modelo no son comparables.

Por otra parte, la introducción de una variable dummy (D) que tome el valor de 1 cuando el individuo de la muestra es mujer, y cero de lo contrario, permitiría establecer si existen diferencias entre los géneros o no. El modelo que se propone es entonces:

$$P(menor_i/X_i) = \phi(\beta_0 + \beta_1 edad_i + \beta_2 escolaridad_i + \beta_3 D_i + \beta_4 D_i * edad_i + \beta_5 D_i * escolaridad_i) + \varepsilon_i$$

De encontrarse que β_3, β_4,β_5 o alguno de ellos es estadísticamente diferente de cero, habremos encontrado que si existen diferencias en la probabilidad de la existencia de un menor derivadas del hecho de que el jefe de familia sea un hombre o una mujer.
La estimación del modelo probit correspondiente se reporta en la siguiente tabla:

Table 3: Estimación Probit incluyendo la variable D i

VARIABLE DEPENDIENTE: menor i
Estadísticos t entre paréntesis
Probit
EMV

	Coeficientes		Efecto Marginal
Constante	0.68467		–
	(0.70)		
edad i	-0.0101		-0.0035
	(-0.39)		
escolaridad i	-0.0136		-0.0048
	(-0.52)		
D i	-3.0341		-1.0613
	(-1.87)	*	
Diescolaridad i	0.0840		0.0294
	(-0.49)		
Diescolaridad i	-0.0237		-0.0083
	(1.99)	**	
LRI	0.11110		
Wald	16.32	**	
ln(L)	-61.16832		
# de Obs.	100		

(*) nivel de significancia: 10%
(**) nivel de significancia: 5%
(***) nivel de significancia: 1%
EMV: Estimadores de Máxima Verosimilitud

Noten que en efecto existen diferencias entre los géneros. Tanto la variable D_I como el término de interacción $D_I * escolaridad_i$ son estadísticamente significativas para el modelo. Por lo tanto, puede decirse con un nivel de confianza del 95% que en efecto existen diferencias en la probabilidad de que haya un menor de edad en un hogar derivada de que el jefe de hogar fuera un hombre o una mujer. El alcalde se había equivocado al pedir que se estimaran dos modelos diferentes para constatar esta hipótesis.

Pregunta 6

Una variable estocástica X tiene una función de distribución acumulativa determinada por

$$f(x) = \lambda e^{-\lambda x} \; para \; x > 0 \tag{7}$$

Con un parámetro desconocido λ . Determine el estimador de máxima verosimilitud de λ, dada una muestra de n observaciones de X.

La función de densidad conjunta de la función objetivo es

$$\lambda^n e^{-\lambda \sum x_i}$$

Tomando esta función como la función de máxima verosimilitud, note que es más fácil maximizar su logaritmo. En este caso la expresión que se obtiene es

$$n \log \lambda - \lambda \sum x_i$$

La condición de primer orden de esta expresión con respecto a λ es:

$$n/\lambda - \sum x_i = 0$$

Por lo tanto, despejando λ se obtiene que el estimador de máxima verosimilitud es:

$$\lambda = \sum x_i/n$$

menor	edad	escolaridad			
1	31,78658	22,6138726			
1	38,92094	1,8450902			
1	26,00397	13,7684528			
0	29,00571	8,9763032			
0	27,01691	13,4291148			
0	30,02236	1,8734941			
1	29,77524	21,6115356			
1	26,06926	7,5509925			
0	25,78425	19,6114736			
0	42,9231	0,4871371			
1	40,33319	17,9733047			
1	43,82815	6,0216974			
0	40,99101	1,0457203	-0,01012632	-0,01363921	0,68467043
0	42,10202	13,8219629			
0	39,05412	12,4116958	0,68467043		
1	34,24374	15,7766595	-0,01012632		
0	36,10147	14,5432854	-0,01363921		
1	25,74152	6,6302168			
1	31,72801	10,3314991			
0	35,42707	19,8848234			
1	28,95113	5,0127746			
0	28,516	14,7329432			
0	26,02933	2,8840698			
0	25,70049	14,2713999			
1	44,87969	6,323439			
0	44,27181	18,644415			
1	22,93512	18,2550156			
1	43,48727	16,2613464			
0	40,74837	6,5770389			
0	30,24375	21,1873171			
1	39,02111	16,7512462			
1	22,34627	13,1788225			
1	30,99443	0,7138584			
0	23,26906	18,840786			
1	35,22544	3,3983822			
1	24,1638	0,2831391			
1	23,97691	1,8759833			
1	28,40743	1,814666			
0	23,54939	12,8236767			
1	37,08241	5,7494029			
1	20,56809	22,5916692			
1	34,87757	1,3259041			
1	42,92365	13,5179992			
0	31,30884	11,1215597			
1	30,80041	17,7889792			
0	40,5949	11,5877464			
1	28,85615	0,6946111			
1	38,35307	15,3050215			
1	38,18957	18,0819472			
0	35,66659	1,6915697			

Menor	edad	escolaridad
1	43,88904	6,037487
0	39,02642	10,709586
1	33,84022	16,025769
0	21,38403	21,00354
0	38,88186	10,435232
0	37,7305	19,008752
1	22,73745	17,732313
1	43,98692	5,880274
0	36,00828	5,149875
0	24,74316	20,414413
0	31,60528	22,723701
1	42,32941	7,111182
0	35,04544	11,039246
0	26,43035	7,471414
0	34,93085	7,946363
1	29,86047	8,233897
0	36,47296	19,811471
1	41,92036	15,121524
0	33,73005	11,36498
0	32,86976	20,736299
0	30,3463	15,222044
0	28,35951	22,556688
0	26,76245	8,538063
0	27,31797	12,321968
0	27,20096	12,145287
0	28,81743	18,664016
0	32,62928	20,627268
1	36,66833	16,742284
0	26,05601	14,502491
0	37,06382	17,573664
0	38,26657	22,41771
0	25,04216	7,085359
1	30,0025	18,249211
0	22,32431	19,75053
1	41,31735	5,454419
1	33,4898	20,289774
0	29,69786	10,041061
0	30,41388	10,977004
0	29,16799	18,507814
0	22,62303	19,925733
1	37,61406	7,682372
0	27,56057	15,124544
1	40,77295	19,507172
0	22,44475	6,61543
1	26,40747	7,844131
0	24,41014	17,399342
1	33,82239	5,319708
1	24,76926	16,524298
0	24,03409	12,05679
0	37,98729	12,185993