

Taller 10: Modelos con variable dependiente cualitativa
Econometría 06216
11-04-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 18 de abril de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta**(no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

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

La Asociación de Médicos (AMF) presentará durante el Congreso Mundial de la Salud una ponencia que describe un trabajo de investigación que relaciona linealmente la probabilidad que tiene un individuo de sufrir un paro cardiorrespiratorio (PC_i), con algunas características propias del individuo y otras de su entorno. Para ello, el grupo encargado del análisis econométrico cuenta con el concepto médico suministrado por algunos representantes de la Asociación, quienes sugieren que el riesgo de sufrir este tipo de anomalía está asociado con las siguientes variables: la edad ($Edad_i$), si la persona tiene trabajo (T_i); si es fumador (F_i), si realiza algún deporte (D_i), si tiene antecedentes familiares con paros cardiorrespiratorios (A_i), si sufre de hipertensión (H_i), si sufre de altos niveles de colesterol (C_i), si sufre de sobrepeso (S_i), si sufre de diabetes (B_i) y finalmente si es de género masculino (M_i). Para realizar el estudio se cuenta con datos de una muestra representativa que se encuentran adjuntos en el archivo T10-01-11.xls.

Pregunta 1

- a. De acuerdo al enunciado escriba correctamente el modelo econométrico a estimar
- b. Interprete a priori los coeficientes del modelo.

Taller 10: Modelos con variable dependiente cualitativa
Econometría 06216
11-04-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 18 de abril de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta**(no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

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

La Asociación de Médicos (AMF) presentará durante el Congreso Mundial de la Salud una ponencia que describe un trabajo de investigación que relaciona linealmente la probabilidad que tiene un individuo de sufrir un paro cardiorrespiratorio (PC_i), con algunas características propias del individuo y otras de su entorno. Para ello, el grupo encargado del análisis econométrico cuenta con el concepto médico suministrado por algunos representantes de la Asociación, quienes sugieren que el riesgo de sufrir este tipo de anomalía está asociado con las siguientes variables: la edad ($Edad_i$), si la persona tiene trabajo (T_i); si es fumador (F_i), si realiza algún deporte (D_i), si tiene antecedentes familiares con paros cardiorrespiratorios (A_i), si sufre de hipertensión (H_i), si sufre de altos niveles de colesterol (C_i), si sufre de sobrepeso (S_i), si sufre de diabetes (B_i) y finalmente si es de género masculino (M_i). Para realizar el estudio se cuenta con datos de una muestra representativa que se encuentran adjuntos en el archivo T10-01-11.xls.

Pregunta 1

- a. De acuerdo al enunciado escriba correctamente el modelo econométrico a estimar

El modelo de probabilidad lineal que vamos a estimar es:

$$PC_i = \alpha_0 + \alpha_1 Edad_i + \alpha_2 T_i + \alpha_3 F_i + \alpha_4 D_i + \alpha_5 A_i + \alpha_6 H_i + \alpha_7 C_i + \alpha_8 S_i + \alpha_9 B_i + \alpha_{10} M_i + \varepsilon_i$$

$$i = 1, 2, 3, \dots, 1220$$

Donde la variable dependiente PC_i es una variable dicotómica definida como:

$$PC_i = \begin{cases} 1 & \text{Si el individuo } i \text{ sufre de un paro cardiorrespiratorio} \\ 0 & \text{o.w} \end{cases}$$

b. Interprete a priori los coeficientes del modelo.

α_0 = Es la probabilidad que tiene el individuo i de sufrir un paro cardiorrespiratorio que no depende de las otras variables.

α_1 = Cuando la edad del individuo i aumenta un año, en promedio, se espera que la probabilidad que tiene de sufrir un paro cardiorrespiratorio varíe en $\alpha_1 * 100$ puntos porcentuales.

α_2 = La diferencia en la probabilidad de que un individuo trabajo sufra un paro cardiorrespiratorio frente a uno que no trabaje es $\alpha_2 * 100$ puntos porcentuales

α_3 = La diferencia en la probabilidad de que un individuo fumador sufra un paro cardiorrespiratorio frente a uno no fumador es $\alpha_3 * 100$ puntos porcentuales

α_4 = La diferencia en la probabilidad de que un individuo que realiza deporte sufra un paro cardiorrespiratorio frente a uno que no es $\alpha_4 * 100$ puntos porcentuales

α_5 = La diferencia en la probabilidad de que un individuo que tiene antecedentes familiares con paros cardiacos sufra un paro cardiorrespiratorio frente a uno que no tenga estos antecedentes es $\alpha_5 * 100$ puntos porcentuales

α_6 = La diferencia en la probabilidad de que un individuo que sufre de hipertensión sufra un paro cardiorrespiratorio frente a uno que no sufra esta enfermedad es $\alpha_6 * 100$ puntos porcentuales

α_7 = La diferencia en la probabilidad de que un individuo que sufre de altos niveles de colesterol sufra un paro cardiorrespiratorio frente a uno que no es $\alpha_7 * 100$ puntos porcentuales

α_8 = La diferencia en la probabilidad de que un individuo que sufre de sobrepeso sufra un paro cardiorrespiratorio frente a uno que no es $\alpha_8 * 100$ puntos porcentuales

α_9 = La diferencia en la probabilidad de que un individuo que sufre de diabetes sufra un paro cardiorrespiratorio frente a uno que no es $\alpha_9 * 100$ puntos porcentuales

α_{10} = La diferencia en la probabilidad de que un hombre sufra un paro cardiorrespiratorio frente a una mujer es $\alpha_{10} * 100$ puntos porcentuales

Pregunta 2

- a. Estime el modelo presentado en el apartado anterior y reporte sus resultados en una tabla.

Tabla 1. Estimación modelo MLP

Variable Dependiente: Pc_i

MCO		
Variables	Parámetro	Significancia
Constante	0,6582173 (11,46)	***
Edad_i	0,0036207 (3,809)	***
T_i	-0,0929973 (-2,057)	**
F_i	-0,1278667 (-2.814)	***
D_i	-0,3157115 (-5,090)	***
A_i	-0,4010508 (-6,074)	***
H_i	-0,4684018 (-3,862)	***
C_i	-0,0318258 (-1,012)	
S_i	-0,0368347 (-0,630)	
B_i	-0,107360 (-1,876)	*
M_i	-0,3780526 (-2,695)	***
R²	0,0956	
R²-ajustado	0,0881	
F	12,78	***
n	1220	

(*) Nivel de significancia: 10%

(**) Nivel de significancia: 5%

(***) Nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

La estimación anterior fue por el método MCO, sin embargo, la estimación del modelo correcto es la que se presenta a continuación en la tabla 2, pues esta corresponde a la corrección de MCP al dividir por la variable $w_i = \sqrt{\hat{y}_i(1 - \hat{y}_i)}$ donde $\hat{y}_i = (\hat{\beta}_{OLS})^T X_i^T$, ya que como sabemos por la construcción del modelo de probabilidad lineal, se obtiene un error heteroscedástico, su varianza es de la forma: $\sigma_i^2 = E[y_i|x_i](1 - E[y_i|x_i])$

VARIABLE DEPENDIENTE: pc/w	
MCP	
Constante	1,38631 (-11,71) ***
edad/w	0,00614 (7,921) ***
T/w	-0,15928 (-3,66) ***
F/w	-0,22159 (-4,92) ***
D/w	-0,46566 (-7,110) ***
A/w	-0,4575 (-7,097) ***
H/w	-0,52268 (-5,298) ***
C/w	-0,04683 (-1,424)
S/w	-0,02615 (-0,432)
B/w	-0,12146 (-2,010) **
M/w	0,03652 (0,463)
R2	0,1528
R2 ajustado	0,1458
No. Obs.	1217

(*) Nivel de significancia: 10%
 (**) Nivel de significancia: 5%
 (***) Nivel de significancia: 1%
 MCP: Mínimos cuadrados ponderados

- b. ¿Los resultados obtenidos son confiables? Explique su respuesta.

Los resultados obtenidos no son confiables pues no es apropiado aplicar el método de mínimos cuadrados (MCO) en el caso anterior, ya que presenta problemas como la no normalidad de la distribución de los errores y la heteroscedasticidad. Por lo cual, se emplea el método de MCP para eliminar este último problema. No obstante, en un caso de variable dependiente dicotómica como este, no es la mejor opción emplear este método de estimación, pues las probabilidades que predice el Modelo Lineal de Probabilidad no están acotadas entre 0 y 1, lo que genera en ocasiones que se obtengan probabilidades sin sentido al no poder restringir los valores de predicción entre 0 y 1.

Pregunta 3

- a. Un experto le sugiere que existe una mejor forma de hacer la estimación y le propone que la realice suponiendo una distribución normal acumulativa.

Tabla 3 Estimación por EMV

Variable dependiente: PC _i		
EMV – Modelo Probit		
Variable	Parámetro	Efectos mg
Constante	0,435174 (2,57) **	
Edad _i	0,010566 (3,85) ***	0,00362
T _i	-0,293814 (-2,11) **	-0,01007
F _i	-0,397434 (-2,84) ***	-0,13621
D _i	-0,889446 (-4,93) ***	-0,30484
A _i	-1,116602 (-5,65) ***	-0,38270
H _i	-1,363492 (-3,6) ***	-0,46731
C _i	-0,091270 (-1,02)	-0,03128
S _i	-0,091004 (-0,54)	-0,03119
B _i	-0,285903 (-1,74) *	-0,09799
M _i	-1,403401 (-2,08) **	-0,48099
LRI	0,9907	
Wald	97,70 ***	
n	1220	

***Nivel de significancia 1%

**Nivel de significancia 5%

*Nivel de significancia 10%

Fuente: Calculos propios en EasyReg

- b. Interprete los coeficientes estimados teniendo en cuenta su significancia.

Como se vio en clase, estamos trabajando un modelo Logit, y por tanto no podemos interpretar los coeficientes estimados, pero si podemos comentar su significancia individual:

- Los coeficientes asociados a las variables $edad_i$, F_i , D_i , A_i y H_i son significativos al 99% de confianza.
- Los coeficientes asociados a la constante y a las variables T_i y M_i son significativos al 95% de confianza.
- El coeficiente asociado a la variable B_i es significativo al 90% de confianza.
- Los coeficientes asociados a las variables C_i y S_i no son significativos a ningún nivel de confianza.

Pregunta 4

- a. Explique cómo obtendría los efectos marginales de cada variable independiente sobre la probabilidad de sufrir un paro cardiorrespiratorio sin realizar cálculo alguno.

Para obtener el efecto marginal promedio de cada variable explicatoria sobre la probabilidad de sufrir un paro cardiorrespiratorio, es necesario evaluar cada una de las observaciones en la siguiente formula

$$\frac{\delta E(y_i | \beta^T x_i^T)}{\delta X_{j,i}} = \varphi(\beta^T x_i^T) \beta_j = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\beta^T x_j^T)^2}{2}} \beta_j$$

En principio, es necesario obtener $(\beta^T x_i^T)^2$. Para ello, multiplicamos el vector traspuesto de coeficientes estimados β^T con cada una de las observaciones X_i . Luego de ingresar este resultado en la formula donde el último termino de ella es β_j , que es el coeficiente estimado asociado a la variable explicativa, obtenemos un resultado para cada observación y para cada variable explicativa.

Finalmente, se calcula el promedio ponderado de los resultados para cada variable que será el efecto marginal de esa variable sobre la variable dependiente.

- b. Interprete a priori los efectos marginales y explique cuáles son los signos esperados.

En promedio, la probabilidad de que un individuo sufra un paro cardiorrespiratorio cambia en X puntos porcentuales, ante un cambio de 1 año en la edad del individuo. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo que trabaja sufra un paro cardiorrespiratorio con respecto a uno que no trabaja cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo que fuma sufra un paro cardiorrespiratorio con respecto a uno que no fuma cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo que realiza algún deporte sufra un paro cardiorrespiratorio con respecto a uno que no hace deporte cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo que tiene antecedentes familiares sufra un paro cardiorrespiratorio con respecto a uno que no tiene estos antecedentes cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo que tenga hipertensión sufra un paro cardiorrespiratorio con respecto a uno que no tiene hipertensión cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo que tiene altos niveles de colesterol sufra un paro cardiorrespiratorio con respecto a uno que no tiene colesterol alto cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo con sobrepeso sufra un paro cardiorrespiratorio con respecto a uno sin sobrepeso cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un individuo diabético sufra un paro cardiorrespiratorio con respecto a uno que no tiene diabetes cambia en X puntos porcentuales. Se espera signo positivo.

En promedio, la diferencia en la probabilidad de que un hombre sufra un paro cardiorrespiratorio con respecto a una mujer cambia en X puntos porcentuales. Se espera signo positivo.

Pregunta 5

- a. Calcule el efecto marginal promedio correspondiente a cada variable independiente y reporte sus resultados en la misma tabla del punto 3

Los resultados aparecen en la tabla 3

Pregunta 6

- a. Otro experto le sugiere una estimación que supone una distribución logística. Realícela y repórtela en una tabla.

Tabla 4 Estimación por EMV

Variable dependiente: PC _i	
EMV - Modelo Logit	
Variable	Parámetro
Constante	0,713608 (2,46) **
Edad _i	0,017503 (3,78) ***
T _i	-0,498591 (-2,08) **
F _i	-0,666546 (-2,79) ***
D _i	-1,461303 (-4,82) ***
A _i	-1,825943 (-5,59) ***
H _i	-2,241295 (0,639455) ***
C _i	-0,153514 (-1,04)
S _i	-0,146985 (-0,54)
B _i	-0,462799 (-1,77) *
M _i	-2,315835 (-2,16) **
LRI	99,07
Wald	97,75 ***
n	1220

***Nivel de significancia 1%

**Nivel de significancia 5%

*Nivel de significancia 10%

Fuente: Calculos propios en EasyReg

- b.** Finalmente usted tendrá resultados de tres modelos estimados. Explique con cuál de los tres modelos usted se quedaría para concluir el análisis

Como se explicó en clase, los resultados de los modelos Logit y Probit no son comparables entre sí. Si bien se obtiene el LRI que nos dice qué tan bueno es el modelo, no podemos comparar el LRI de un modelo Logit contra el de un modelo Probit (solo es comparable entre modelos Probit o entre modelos Logit). Los resultados obtenidos a través de la estimación por MCO no son confiables por los argumentos expuestos en la pregunta 2. Por tanto, la recomendación para concluir el análisis es utilizar los resultados de la estimación del modelo Logit, o bien los del modelo Probit, sin realizar ningún tipo de diferenciación.

Pregunta 2

- a. Estime el modelo presentado en el apartado anterior y reporte sus resultados en una tabla.
- b. ¿Los resultados obtenidos son confiables? Explique su respuesta.

Pregunta 3

- a. Un experto le sugiere que existe una mejor forma de hacer la estimación y le propone que la realice suponiendo una distribución normal acumulativa.
- b. Interprete los coeficientes estimados teniendo en cuenta su significancia.

Pregunta 4

- a. Explique cómo obtendría los efectos marginales de cada variable independiente sobre la probabilidad de sufrir un paro cardiorrespiratorio sin realizar cálculo alguno.
- b. Interprete a priori los efectos marginales y explique cuáles son los signos esperados.

Pregunta 5

- a. Calcule el efecto marginal promedio correspondiente a cada variable independiente y reporte sus resultados en la misma tabla del punto 3

Pregunta 6

- a. Otro experto le sugiere una estimación que supone una distribución logística. Realícela y repórtela en una tabla.
- b. Finalmente usted tendrá resultados de tres modelos estimados. Explique con cuál de los tres modelos usted se quedaría para concluir el análisis