

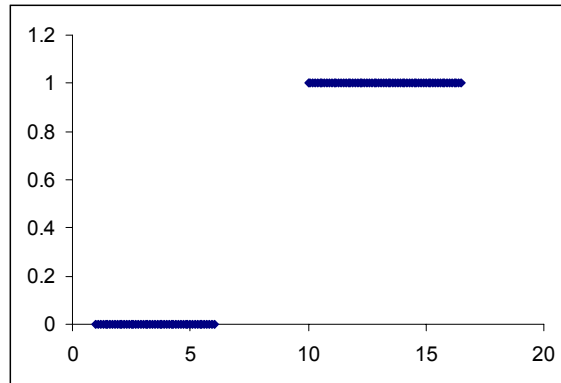
Maestría en Marketing
Métodos y técnicas de análisis cuantitativo y cualitativo

Estimación para Variables Binarias

Otoño 2004

Problemas a tratar

- Si la variable explicada es binaria:
 - Sirve el modelo lineal?
 - Hay alternativas mejores?
 - Cómo se interpretan los resultados?



Esperanza de una variable binaria

- Una variable es binaria si

$$Y_i = \begin{cases} 1 & \text{si fuma} \\ 0 & \text{si no fuma} \end{cases}$$

- Entonces, su esperanza es:

$$E(Y) = 1P(Y = 1) + 0P(Y = 0) = P(Y = 1)$$

- Luego, hacer un modelo para el valor esperado de una variable binaria es equivalente a describir un modelo de la probabilidad de ocurrencia de la característica que tiene un 1

Modelo lineal de probabilidad

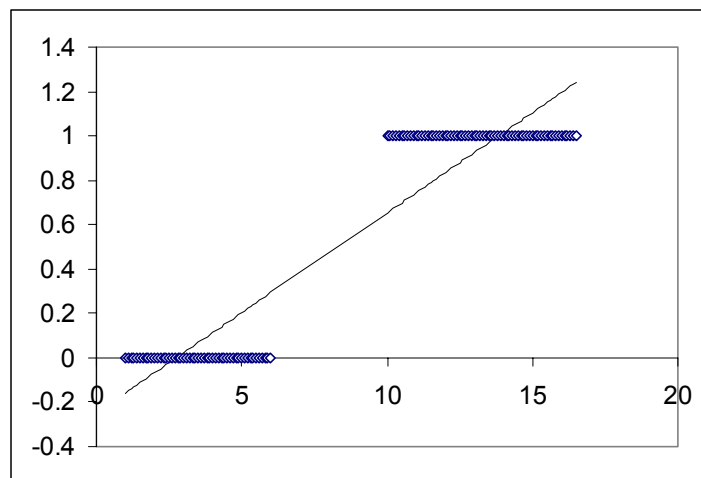
- Tal como lo venimos haciendo, una alternativa es pensar en un modelo lineal para estimar la probabilidad

$$P(Y = 1) = \alpha + \beta X$$

- Esta alternativa presenta varios problemas:
 - Las probabilidades toman valores entre 0 y 1, mientras que con la alternativa lineal no estamos haciendo restricciones para que esto pase
 - Valores intermedios

Métodos y técnicas de análisis cuantitativo y cualitativo

4



Métodos y técnicas de análisis cuantitativo y cualitativo

5

Alternativa: Modelo probit

- Una alternativa es pasarnos a un modelo no lineal en parámetros

$$P(Y = 1) = F(\alpha + \beta X)$$

- Donde F sea una función acotada en el intervalo $[0,1]$. Una función que cumple estas características es la función de distribución acumulada normal estandar.
- Cuando utilizamos esta alternativa, el modelo se denomina *probit*

Interpretación de los coeficientes

- Lo que sucede en esta alternativa es que el efecto del cambio en X sobre la variable explicada (la probabilidad) ya no es constante como en el caso lineal.
- En este caso dependerá del punto en el que lo evaluemos.
- Si embargo, lo que si puede mostrarse es que el signo del efecto esta determinado exclusivamente por el signo de la variable explicada.
- De este modo, una interpretación de los coeficientes es cual es el signo del impacto de un aumento de la variable explicativa sobre la probabilidad

Ejemplo: Tabaco

Dependent Variable: FUMA
 Method: ML - Binary Probit (Quadratic hill climbing)
 Date: 06/19/04 Time: 00:18
 Sample: 1 2724
 Included observations: 2724
 Convergence achieved after 5 iterations
 Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
LNX	0.2759	0.0633	-4.3609	0.0000
NADULTS	0.1458	0.0351	4.1572	0.0000
NKIDS2	-0.2827	0.1206	-2.3433	0.0191
AGE	-0.1397	0.0233	-5.9955	0.0000
TOMA	0.2357	0.0689	3.4233	0.0006
BLUECOL	0.1985	0.0830	2.3930	0.0167
WHITECOL	-0.0237	0.0688	-0.3443	0.7306
C	3.3181	0.8231	4.0312	0.0001
Mean dependent var	0.380323	S.D. dependent var		0.485555
S.E. of regression	0.477463	Akaike info criterion		1.297887
Sum squared resid	619.1701	Schwarz criterion		1.315244
Log likelihood	-1759.722	Hannan-Quinn criter.		1.304161
Restr. log likelihood	-1809.341	Avg. log likelihood		-0.646007
LR statistic (7 df)	99.23728	McFadden R-squared		0.027424
Probability(LR stat)	0			
Obs with Dep=0	1688	Total obs		2724
Obs with Dep=1	1036			

Métodos y técnicas de análisis cuantitativo y cualitativo

8