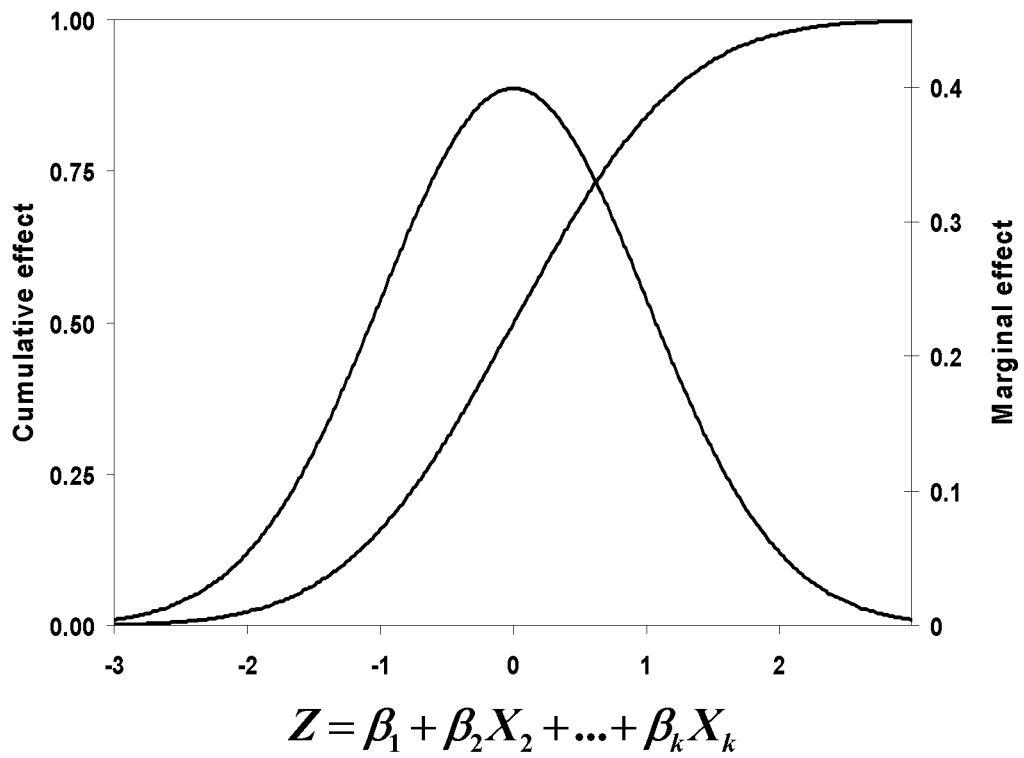


เอกสารประกอบการสอน E-Views

LOGIT AND PROBIT MODEL



นิติพงษ์ สงครีโรจน์

22 กุมภาพันธ์ 2550

การวิเคราะห์ความถดถอยโลจิสติกและโพรบิท

การตีความ Logistic Model กรณีใช้ข้อมูล Table 15.7 ของ Gujarati (2003) หน้า 604 ในตารางที่ 1 ซึ่งได้ผลการวิเคราะห์ตามตารางที่ 2

ตารางที่ 1 the Effect of "Personalized System of Instruction" (PSI) on Course Grades

OBS	GPA	TUCE	PSI	GRADE	LETTER	OBS	GPA	TUCE	PSI	GRADE	LETTER
1	2.66	20	0	0	C	17	2.75	25	0	0	C
2	2.89	22	0	0	B	18	2.83	19	0	0	C
3	3.28	24	0	0	B	19	3.12	23	1	0	B
4	2.92	12	0	0	B	20	3.16	25	1	1	A
5	4	21	0	1	A	21	2.06	22	1	0	C
6	2.86	17	0	0	B	22	3.62	28	1	1	A
7	2.76	17	0	0	B	23	2.89	14	1	0	C
8	2.87	21	0	0	B	24	3.51	26	1	0	B
9	3.03	25	0	0	C	25	3.54	24	1	1	A
10	3.92	29	0	1	A	26	2.83	27	1	1	A
11	2.63	20	0	0	C	27	3.39	17	1	1	A
12	3.32	23	0	0	B	28	2.67	24	1	0	B
13	3.57	23	0	0	B	29	3.65	21	1	1	A
14	3.26	25	0	1	A	30	4	23	1	1	A
15	3.53	26	0	0	B	31	3.1	21	1	0	C
16	2.74	19	0	0	B	32	2.39	19	1	1	A

the Effect of "Personalized System of Instruction" (PSI) on Course Grades

OBS = Observation

GPA = Entering Grade Point Average

TUCE = Score on an Examination Given at the Beginning of the Term to

Test Entering Knowledge of Macroeconomics

PSI = Personal System of Instruction

where: PSI = 1 if the new method is used

= 0 otherwise

GRADE = 1 if the Letter Grade is A

0 Otherwise

LETTER = Letter Grade

ตารางที่ 2 ผลการวิเคราะห์ด้วย Logistic Model

Dependent Variable: GRADE				
Method: ML - Binary Logit (Quadratic hill climbing)				
Date: 02/22/07 Time: 21:27				
Sample: 1 32				
Included observations: 32				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-13.02135	4.931317	-2.640541	0.0083
GPA	2.826113	1.262940	2.237725	0.0252
TUCE	0.095158	0.141554	0.672235	0.5014
PSI	2.378688	1.064563	2.234426	0.0255
Mean dependent var	0.343750	S.D. dependent var		0.482559
S.E. of regression	0.384716	Akaike info criterion		1.055602
Sum squared resid	4.144171	Schwarz criterion		1.238819
Log likelihood	-12.88963	Hannan-Quinn criter.		1.116333
Restr. log likelihood	-20.59173	Avg. log likelihood		-0.402801
LR statistic (3 df)	15.40419	McFadden R-squared		0.374038
Probability(LR stat)	0.001502			
Obs with Dep=0	21	Total obs		32
Obs with Dep=1	11			

ตัวแบบโลจิสติก มีสิ่งที่สังเกตอยู่ 4 ประการ คือ

- (1) เนื่องจากวิธี MLE ใช้กับตัวอย่างที่มีจำนวนมาก ดังนั้นในการทดสอบนัยสำคัญของสัมประสิทธิ์จึงใช้ Z Statistic เนื่องจากถ้ามีตัวอย่างขนาดใหญ่จะทำให้ การแจกแจงแบบ t เปลี่ยนเป็นการแจกแจงแบบปกติ
- (2) ในกรณีตัวแบบโลจิสติก McFadden R-squared จะมีความสำคัญเป็นอันดับสอง โดยที่ไม่ต้องไปให้ความสำคัญในบทบาทมาก แต่สิ่งที่สำคัญที่สุดก็คือการพิจารณาที่เครื่องหมายและนัยสำคัญทางสถิติของค่าสัมประสิทธิ์
- (3) กรณีที่มีการพยากรณ์ หากค่าความน่าจะเป็นที่ทำนาย มากกว่า 0.5 นั้นเราจะกำหนดให้เป็น 1 หากมีค่าน้อยกว่า 0.5 เราจะกำหนดให้เป็น 0
- (4) การทดสอบว่าค่าสัมประสิทธิ์ทั้งหมดเป็นศูนย์หรือไม่ เราจะใช้ LR (Likelihood Ratio) Statistic ซึ่งมีการแจกแจงเป็นแบบไคสแควร์

การตีความผลกรณี Logistic Model

เพื่อให้การตีความหมายได้ดีขึ้น เราต้องทำค่าสัมประสิทธิ์ในตารางที่ 2 อยู่ในรูปของ Odds ซึ่งทำได้โดยการ antilog ของค่าสัมประสิทธิ์ตัวแปรนั้น เช่น ค่าสัมประสิทธิ์ PSI เท่ากับ 2.3786 จะได้ $e^{2.3786}$ ประมาณเท่ากับ 10.7879 ซึ่งหมายความว่า นักเรียนที่ได้รับการใช้วิธีการสอนแบบใหม่นั้นมีความเป็นไปได้ที่จะได้เกรด A ถึง 10 เท่าของนักเรียนที่ได้รับการสอนวิธีแบบเดิม โดยที่ให้ปัจจัยอย่างอื่นคงที่ หากจะตีความตรงจากสัมประสิทธิ์โดยตรงเช่น 2.3786 หมายความว่า หากใช้วิธีการสอนเปลี่ยนจากแบบเดิมไปแบบใหม่นั้นจะทำให้ log odds ของนักเรียนเพิ่มขึ้น 2.3786 หน่วย ซึ่งผู้อ่านโดยทั่วไปจะเข้าใจความหมายได้ไม่ดีนัก กรณีสัมประสิทธิ์ของ TUCE เท่ากับ 0.095 จะได้ $e^{0.095}$ เท่ากับ 1.099 หมายความว่านักเรียนที่มีคะแนนการสอบเพิ่มขึ้น 1 หน่วย จะมีความน่าจะเป็นที่นักเรียนคนนั้นจะได้เกรด A เพิ่มขึ้น $1.099-1=0.099$ หรือร้อยละ 9.9

การตีความผลกรณี Probit Model

กรณีเป็นแบบ Probit นั้นหมายความว่าเราได้สมมติให้การแจกแจงเป็นแบบปกติ เมื่อเราพิจารณาค่าสัมประสิทธิ์ ก็คือการหา Marginal effect ของตัวแปรนั้นซึ่งก็คือการ Derivative ซึ่งจะทำให้ได้ค่า $f(\beta_1 + \beta_2 X_i)$ ซึ่งเทอมนี้เรียกว่า Standard normal probability density function และค่าสัมประสิทธิ์จากการประมาณโพรบิทโมเดลนั้นเป็นผลมาจาก Cumulative normal function ของความน่าจะเป็นที่ $Y=1$ เราจึงต้องตีความในรูปค่า Z-score โดยผลของการวิเคราะห์โพรบิทแสดงในตารางที่ 3

ตารางที่ 3 ผลการวิเคราะห์ด้วย Probit Model

Dependent Variable: GRADE				
Method: ML - Binary Probit (Quadratic hill climbing)				
Date: 02/22/07 Time: 22:07				
Sample: 1 32				
Included observations: 32				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-7.452320	2.542472	-2.931131	0.0034
GPA	1.625810	0.693882	2.343063	0.0191
TUCE	0.051729	0.083890	0.616626	0.5375
PSI	1.426332	0.595038	2.397045	0.0165
Mean dependent var	0.343750	S.D. dependent var		0.482559
S.E. of regression	0.386128	Akaike info criterion		1.051175
Sum squared resid	4.174660	Schwarz criterion		1.234392
Log likelihood	-12.81880	Hannan-Quinn criter.		1.111907
Restr. log likelihood	-20.59173	Avg. log likelihood		-0.400588
LR statistic (3 df)	15.54585	McFadden R-squared		0.377478
Probability(LR stat)	0.001405			
Obs with Dep=0	21	Total obs		32
Obs with Dep=1	11			

จากตารางที่ 3 กรณีค่าสัมประสิทธิ์ตัวแปร GPA เท่ากับ 1.62581 หมายความว่า เกรดเฉลี่ยของนักเรียนเพิ่มขึ้น 1 หน่วยจะทำให้ค่า Z-score เพิ่มขึ้น 1.62581 หน่วย ซึ่งต้องแปลงค่า Z-score ดังกล่าวให้เป็นความน่าจะเป็นโดยการใช้ตาราง Standard Normal Distribution หรือใช้สูตร $=@dnorm(x)$ ใน E-Views นั่นคือ $=@dnorm(1.62581)$ เท่ากับ 0.1064 หรือร้อยละ 10.64 หมายความว่า เกรดเฉลี่ยของนักเรียนเพิ่มขึ้น 1 หน่วยจะทำให้ค่าความน่าจะเป็นที่จะได้เกรด A เพิ่มขึ้นร้อยละ 10.64 กรณี PSI นักเรียนที่ได้รับการสอนแบบใหม่จะมีความน่าจะเป็นที่จะได้เกรด A มากกว่าวิธีการสอนแบบเดิม ร้อยละ 14.43