

MỘT SỐ HƯỚNG DẪN THÊM VỀ CHƯƠNG BIẾN PHỤ THUỘC ĐỊNH TÍNH

1. Ước lượng mô hình xác suất tuyến tính (sử dụng Table 15.1 về thu nhập và sở hữu nhà)

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (1)$$

Vì phương sai của u_i thay đổi nên chúng ta ước lượng phương trình hồi qui chuyển đổi sau

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_1}{\sqrt{w_i}} + \frac{\beta_2 X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \quad (2)$$

Các bước tiến hành:

- ✚ Hồi qui OLS phương trình (1), sau đó tính yhat (\hat{Y} hay \hat{Y} ước lượng)
- ✚ Loại bỏ các số liệu nhỏ hơn 0 và lớn hơn 1 trong yhat
- ✚ Tính $w = \text{yhat} * (1 - \text{yhat})$
- ✚ Tính sqrtw (căn bậc 2 của w)
- ✚ Tính $\text{ystar} = y / \text{sqrtw}$
- ✚ Tính $\text{xstar} = x / \text{sqrtw}$
- ✚ Tính $\text{betastar} = 1 / \text{sqrtw}$

Kết quả hồi qui

$$\frac{\hat{Y}_i}{\sqrt{w_i}} = -1,2456 \frac{1}{\sqrt{w_i}} + 0,1196 \frac{X_i}{\sqrt{w_i}}$$

t = (-10,332) (17,454) $R^2 = 0.98$

Diễn giải kết quả hồi qui:

- ✚ Ý nghĩa của hệ số ước lượng 0,1196 là nếu các yếu tố khác không đổi khi thu nhập có trọng số thay đổi 1 đơn vị (\$1000) thì xác suất sở hữu nhà có trọng số sẽ tăng lên 11,96%.
- ✚ **Tính xác suất sở hữu nhà với một mức thu nhập cho trước.** Hãy tính xác suất sở hữu nhà của các hộ gia đình có mức thu nhập là \$18.000?

Căn cứ vào bộ dữ liệu ta thấy ở mức thu nhập \$18.000 thì sqrtw (căn bậc 2 của w) bằng 0,31
 $\hat{Y}(X = 18,000) = \hat{P} = -1,2456 * 0,31 + 0,1196 * 0,31 * 18 = 0,28 = 28\%$

2. Ước lượng mô hình Logit (Logistic)

Sử dụng Table 15.7

Chúng ta không thể sử dụng phương pháp OLS một cách trực tiếp vì khi tính $L = \ln(P^*/(1-P^*))$ thì chúng ta sẽ có $\ln(0)$ và $\ln(\infty) \rightarrow$ không có ý nghĩa.

\Rightarrow Chúng ta cần sử dụng phương pháp maximum likelihood.

Mô hình hồi qui

$$GRADE_i = \beta_1 + \beta_2 GPA_i + \beta_3 TUCE_i + \beta_4 PSI_i + u_i$$

trong đó

$$\begin{aligned} GRADE &= 1 \text{ nếu điểm môn kinh tế vi mô là A} \\ &= 0 \text{ nếu khác} \end{aligned}$$

GPA = điểm trung bình các môn học

TUCE = điểm kiểm tra đầu kỳ với kiến thức môn kinh tế vĩ mô

PSI = phương pháp dạy học mới

Các bước thực hiện hồi qui mô hình Logit sử dụng Maximum Likelihood

Mở file Table 15.7

Vào **Quick**, chọn **Estimation Equation**

Trong hộp **Estimation Equation** nhập lệnh hồi qui **GRADE C GPA TUCE PSI**

Trong hộp **Estimation settings** chọn **BINARY**

Dưới hộp Estimation Equation chọn **LOGIT**

Kết quả hồi qui như sau:

Dependent Variable: GRADE
 Method: ML - **Binary Logit** (Quadratic hill climbing)
 Date: 01/13/11 Time: 17:57
 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	4.931324	-2.640537	0.0083
GPA	2.826113	1.262941	2.237723	0.0252
TUCE	0.095158	0.141554	0.672235	0.5014
PSI	2.378688	1.064564	2.234424	0.0255
McFadden R-squared	0.374038	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	
LR statistic	15.40419	Avg. log likelihood	-0.402801	
Prob(LR statistic)	0.001502			
Obs with Dep=0	21	Total obs	32	
Obs with Dep=1	11			

Diễn giải kết quả hồi qui:

- ✚ Hệ số hồi qui ước lượng của PSI = 2,3786, do đó $e^{2,3786} = 10,7897 \rightarrow$ Nếu các yếu tố khác không đổi thì một SV được giảng dạy bằng phương pháp mới sẽ có xác suất đạt được điểm A cao gấp 10 lần so với các SV được dạy bằng phương pháp cũ.
- ✚ Làm tương tự cho các hệ số hồi qui ước lượng khác

Kiểm định mô hình hồi qui trong phương pháp Maximum Likelihood

- Để đánh giá độ thích hợp của mô hình chúng ta dùng

$$\text{Pseudo } R^2 = \text{Mc Fadden } R^2 = 1 - (\text{LLF}_{UR} - \text{LLF}_R) \quad (=0,374 \text{ trong kết quả trên})$$
- Kiểm tra ý nghĩa thống kê các hệ số hồi qui: **sử dụng thống kê z thay vì t-student** (vì thường thì chúng ta có cỡ mẫu lớn)
- Kiểm định ý nghĩa chung của toàn bộ mô hình (giả thuyết $H_0: \beta_0 = \beta_1 = \dots = \beta_K = 0$)

Chúng ta sử dụng **thống kê chi-square**

$$\text{LR (Likelihood ratio)} = 2(\text{LLF}_{UR} - \text{LLF}_R) \quad (=15,54 \text{ trong kết quả hồi qui trên})$$

Thường thì chúng ta sử dụng p-value (bác bỏ H_0 nếu p-value < 0,05)

3. Mô hình Probit

Mô hình Probit sử dụng hàm **cdf chuẩn** thay vì hàm cdf thường như trong mô hình Logit.

Sử dụng dữ liệu trong Table 15.7

Các bước thực hiện hồi qui mô hình với dữ liệu cá nhân sử dụng **Maximum Likelihood**

Mở file Table 15.7

Vào **Quick**, chọn **Estimation Equation**

Trong hộp **Estimation Equation** nhập lệnh hồi qui **GRADE C GPA TUCE PSI**

Trong hộp **Estimation settings** chọn **BINARY**

Dưới hộp Estimation Equation chọn **PROBIT**

Kết quả hồi qui

Dependent Variable: GRADE
Method: ML - **Binary Probit** (Quadratic hill climbing)
Date: 01/13/11 Time: 18:22
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	-2.931132	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
McFadden R-squared	0.377478	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.	1.111906	Restr. log likelihood	-20.59173	
LR statistic	15.54585	Avg. log likelihood	-0.400588	
Prob(LR statistic)	0.001405			
Obs with Dep=0	21	Total obs	32	
Obs with Dep=1	11			

Diễn giải kết quả hồi qui: tương tự như trong mô hình Logit.

Kiểm định kết quả hồi qui: tương tự như trong mô hình Logit.

So sánh giữa mô hình logit và probit: thực chất chúng rất giống nhau. Đem các hệ số hồi qui ước lượng của mô hình probit nhân với 1,81 chúng ta sẽ được hệ số hồi qui ước lượng trong mô hình logit.