

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2010

Các Phương Pháp Phân Tích Định Lượng

Gợi ý trả lời: Bài tập 11 (Phần 2)

BIẾN PHỤ THUỘC ĐỊNH TÍNH

Ngày Phát: Thứ Tư, 12/1/2011

Ngày Nộp: 8:20 sáng, Thứ Sáu, 14/01/2011

Bản in nộp tại Phòng Giáo Vụ

Bản điện tử gửi đến thầy Nguyễn Khánh Duy theo địa
chỉ duyнк@fetp.vnn.vn

Bài 1

Các bạn hãy xem số liệu trong bảng sau đây

Bảng 1

Số sinh viên	Điểm thi GRE	Được nhận vào chương trình	
	Định lượng (Q)	Ngôn ngữ (V) (1= được nhận; 0 nếu khác)	
1	760	550	1
2	600	350	0
3	720	320	0
4	710	630	1
5	530	430	0
6	650	570	0
7	800	500	1
8	650	680	1
9	520	660	0
10	800	250	0
11	670	480	0
12	670	520	1
13	780	710	1

Source: Donald F. Morrison, *Applied Linear Statistical Methods*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1983, p. 279 (adapted).

- a. Hãy sử dụng mô hình LPM để dự báo xác suất được nhận vào chương trình dựa trên điểm môn định lượng (Q) và môn ngôn ngữ (V) trong kỳ thi GRE?

Sau khi nhập liệu vào Eviews, bạn có thể có lệnh **LS Y C Q V** để ước lượng mô hình LPM.

Ta có phương trình $Y_i = \beta_1 + \beta_2 Q_i + \beta_3 V_i + u_i$ (1)

Gọi xác suất được nhận vào chương trình là p_i

$p_i = E(Y | Q_i, V_i) = P(Y=1 | Q_i, V_i) = \beta_1 + \beta_2 Q_i + \beta_3 V_i$

(1) có thể viết thành: $Y_i = E(Y | Q_i, V_i) + u_i = \beta_1 + \beta_2 Q_i + \beta_3 V_i + u_i = p_i + u_i$

Ước lượng của $E(Y | Q_i, V_i)$ là $\hat{Y}_i = -2.8673 + 0.0031 Q_i + 0.0023 V_i$

Hình 1.1 Mô hình 1

Dependent Variable: Y
 Method: Least Squares
 Date: 01/14/11 Time: 22:09
 Sample: 1 13
 Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.8673	0.8378	-3.4224	0.0065
Q	0.0031	0.0011	2.9758	0.0139
V	0.0023	0.0007	3.4405	0.0063

R-squared	0.6493	Mean dependent var	0.462
Adjusted R-squared	0.5792	S.D. dependent var	0.519
S.E. of regression	0.3366	Akaike info criterion	0.859
Sum squared resid	1.1329	Schwarz criterion	0.990
Log likelihood	-2.5853	Hannan-Quinn criter.	0.832
F-statistic	9.2584	Durbin-Watson stat	2.591
Prob(F-statistic)	0.0053		

Tại cửa sổ **Equation** của mô hình trên trong Eviews, bạn có thể tính \hat{Y}_i bằng cách bấm vào nút **Forecast** (Nếu có thêm một dòng dữ liệu thứ 14 lưu thông tin của Q, V thì bạn cũng sẽ dự đoán được giá trị của YF_LPM ở dòng thứ 14)

Hình 1.2

obs	Y	Q	V	YF_LPM
1	1	760	550	0.797
2	0	600	350	-0.172
3	0	720	320	0.133
4	1	710	630	0.828
5	0	530	430	-0.203
6	0	650	570	0.500
7	1	800	500	0.805
8	1	650	680	0.758
9	0	520	660	0.304
10	0	800	250	0.219
11	0	670	480	0.352
12	1	670	520	0.445
13	1	780	710	1.234

b. Bạn có thỏa mãn với mô hình này hay không? Nếu không bạn sẽ sử dụng mô hình nào?

Tuy nhiên mô hình 1 trên bị vi phạm các giả định của OLS như: u_i có phương sai thay đổi, u_i không có phân phối chuẩn, và ước lượng của $E(Y|Q_i, V_i)$ là \hat{Y}_i chưa chắc đã thỏa mãn điều kiện $0 \leq \hat{Y}_i \leq 1$. Cột Yf_LPM ở Hình 1.2 cũng cho thấy, giá trị \hat{Y}_i ở quan sát thứ 2, thứ 5 và thứ 13 nằm ngoài đoạn $[0,1]$.

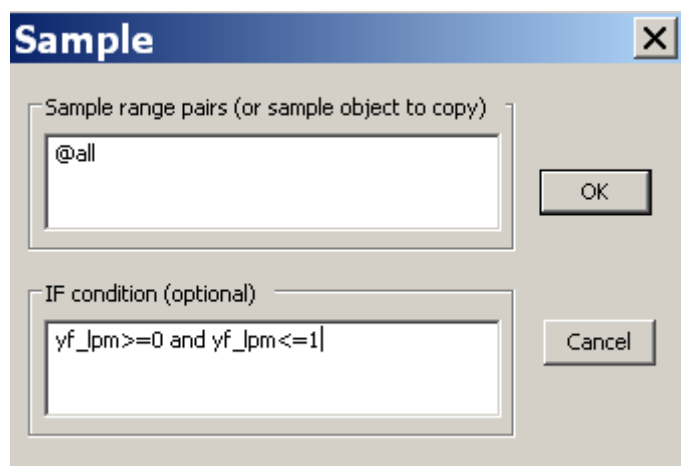
Người ta có thể sử dụng mô hình LPM có trọng số để khắc phục hiện tượng phương sai thay đổi, tuy nhiên khi đó mô hình cũng không tốt bằng mô hình Logit, hay Probit bởi lẽ LPM có trọng số cho rằng p_i có quan hệ tuyến tính theo các X_i (mà điều này không đúng về mặt kinh tế), và nó cũng không đảm bảo giá trị dự đoán của p_i nằm trong đoạn $[0,1]$ (Xem Hình 1.6 ở cột Yf) ... Vì thế, LPM, hay LPM có trọng số ít được sử dụng trong thực tiễn hiện nay.

a. Mô hình LPM có trọng số

Sau khi tính trọng số bằng lệnh **genr w=yf_lpm*(1-yf_lpm)**

tại **workfile**, bạn **định lại mẫu** phục vụ cho việc ước lượng mô hình bằng cách bấm nút **sample**, sau đó khai báo điều kiện if. Điều này giúp bạn loại trừ những \hat{Y}_i từ mô hình LPM nằm ngoài đoạn $[0,1]$ mà không phải xóa một cách thủ công khi dữ liệu lớn.

Hình 1.3



Gõ lệnh **LS y/sqr(w) 1/sqr(w) q/sqr(w) v/sqr(w)** vào cửa sổ lệnh

Hoặc chọn **Quick\Estimate Equation** và gõ dòng **y/sqr(w) 1/sqr(w) q/sqr(w) v/sqr(w)**

Hình 1.4 Mô hình LPM có trọng số

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1/SQR(W)	-5.352379	1.220316	-4.386059	0.0032
Q/SQR(W)	0.005631	0.001362	4.133006	0.0044
V/SQR(W)	0.003767	0.000748	5.037174	0.0015

R-squared	0.804435	Mean dependent var	1.200475
Adjusted R-squared	0.748559	S.D. dependent var	1.275902
S.E. of regression	0.639787	Akaike info criterion	2.187961
Sum squared resid	2.865290	Schwarz criterion	2.278737
Log likelihood	-7.939807	Hannan-Quinn criter.	2.088381
Durbin-Watson stat	1.824308		

Để dự đoán xác suất, tại cửa sổ Equation (Hình 2.1), bấm nút forecast bạn có thể dự đoán được xác suất mà không cần tính thủ công!

Hình 1.5

Forecast equation: UNTITLED

Series to forecast: Y Y/SQR(W)

Forecast name: yf

S.E. (optional):

GARCH(optional):

Method: Static forecast (no dynamics in equation)

Structural (ignore ARMA)

Coef uncertainty in S.E. calc

Forecast sample: 1 13 if yf_lpm >= 0 and yf_lpm <= 1

Output: Forecast graph Forecast evaluation

Insert actuals for out-of-sample observations

OK Cancel

Hình 1.6

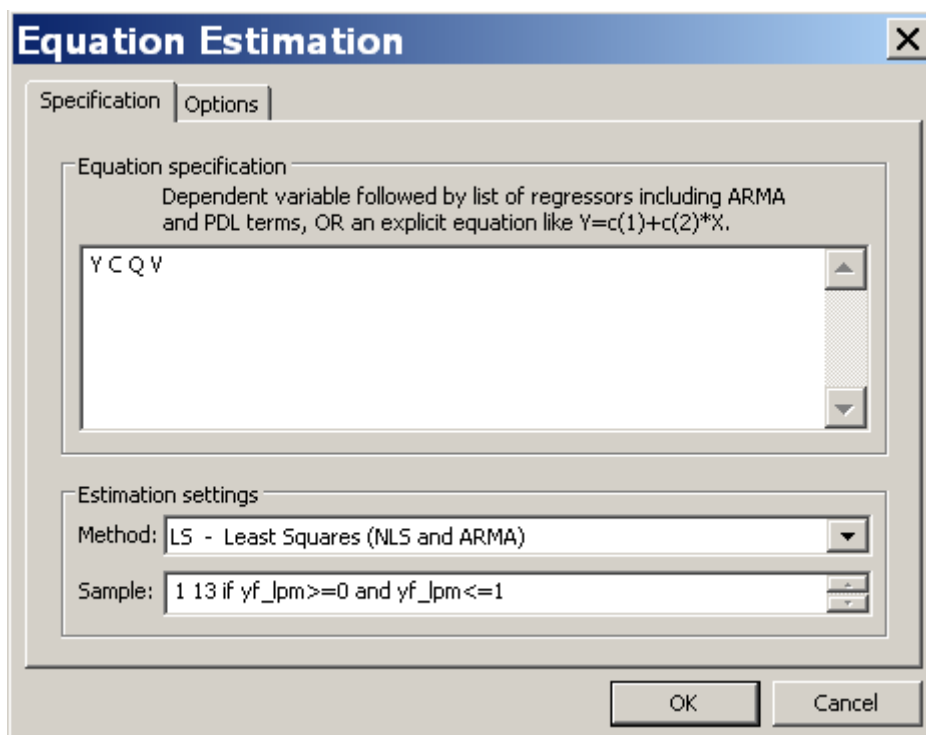
obs	Y	Q	V	YF_LPM	YF	
1	1	1	760	550	0.797	0.999
3	0	0	720	320	0.133	-0.093
4	1	1	710	630	0.828	1.019
6	0	0	650	570	0.500	0.455
7	1	1	800	500	0.805	1.036
8	1	1	650	680	0.758	0.870
9	0	0	520	660	0.304	0.062
10	0	0	800	250	0.219	0.094
11	0	0	670	480	0.352	0.229
12	1	1	670	520	0.445	0.379

Bạn cũng có thể ước lượng mô hình LMP có trọng số theo cách sau đây. Cách này sẽ thuận lợi hơn cho bạn về mặt thao tác để ước lượng và dự báo.

Bước 1. Định lại cỡ mẫu cho việc ước lượng

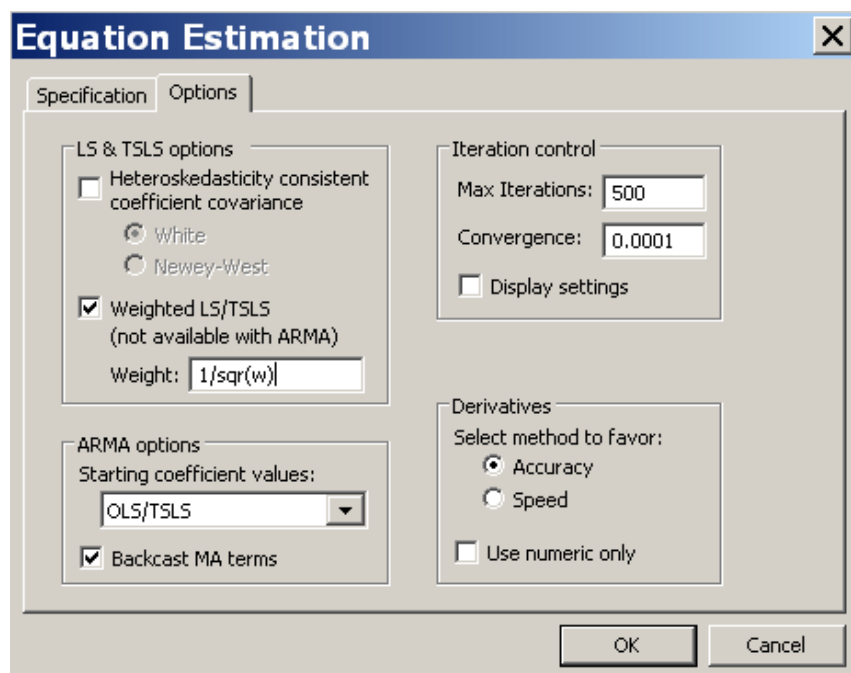
Bước 2: Quick\Estimate Equation → gõ biến vào Equation Specification

Hình 1.7



Bước 3. Chọn Options, sau đó khai báo biến trọng số ở khung Weight

Hình 1.8



Bước 4. Xem kết quả (Bạn sẽ thấy kết quả tương tự như Hình 1.4)

Dependent Variable: Y
 Method: Least Squares
 Date: 01/15/11 Time: 00:13
 Sample: 1 13 IF YF_LPM>=0 AND YF_LPM<=1
 Included observations: 10
 Weighting series: 1/SQR(W)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.352379	1.220316	-4.386059	0.0032
Q	0.005631	0.001362	4.133006	0.0044
V	0.003767	0.000748	5.037174	0.0015

Weighted Statistics

R-squared	0.797710	Mean dependent var	0.507910
Adjusted R-squared	0.739913	S.D. dependent var	0.539823
S.E. of regression	0.270688	Akaike info criterion	0.467627
Sum squared resid	0.512905	Schwarz criterion	0.558402
Log likelihood	0.661866	Hannan-Quinn criter.	0.368046
F-statistic	13.80190	Durbin-Watson stat	1.824308
Prob(F-statistic)	0.003723		

Unweighted Statistics

R-squared	0.726161	Mean dependent var	0.500000
Adjusted R-squared	0.647921	S.D. dependent var	0.527046
S.E. of regression	0.312729	Sum squared resid	0.684597
Durbin-Watson stat	1.813705		

b. Mô hình Logit

Hình 1.9 Mô hình Logit

Dependent Variable: Y

Method: ML - Binary Logit (Quadratic hill climbing)

Date: 01/14/11 Time: 23:13

Sample: 1 13

Included observations: 13

Convergence achieved after 8 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-80.900	111.728	-0.724	0.469
Q	0.083	0.123	0.678	0.498
V	0.047	0.057	0.818	0.414
McFadden R-squared	0.783	Mean dependent var	0.4615	
S.D. dependent var	0.519	S.E. of regression	0.2646	
Akaike info criterion	0.761	Sum squared resid	0.7000	
Schwarz criterion	0.891	Log likelihood	-1.9454	
Hannan-Quinn criter.	0.734	Restr. log likelihood	-8.9724	
LR statistic	14.054	Avg. log likelihood	-0.1496	
Prob(LR statistic)	0.001			
Obs with Dep=0	7.000	Total obs	13	
Obs with Dep=1	6.000			

c. Mô hình Probit

Hình 1.10 Mô hình Probit

Dependent Variable: Y
 Method: ML - Binary Probit (Quadratic hill climbing)
 Date: 01/14/11 Time: 23:15
 Sample: 1 13
 Included observations: 13
 Convergence achieved after 8 iterations
 Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-45.131	70.923	-0.636	0.525
Q	0.046	0.078	0.591	0.555
V	0.027	0.036	0.730	0.465
McFadden R-squared	0.788	Mean dependent var		0.462
S.D. dependent var	0.519	S.E. of regression		0.263
Akaike info criterion	0.754	Sum squared resid		0.690
Schwarz criterion	0.884	Log likelihood		-1.899
Hannan-Quinn criter.	0.727	Restr. log likelihood		-8.972
LR statistic	14.147	Avg. log likelihood		-0.146
Prob(LR statistic)	0.001			
Obs with Dep=0	7	Total obs		13
Obs with Dep=1	6			

Bài 2

Bảng 2

Giảm giá (X, cent)	Cỡ mẫu N_i	Số phiếu giảm giá được sử dụng (n_i)
5	500	100
7	500	122
9	500	147
11	500	176
13	500	211
15	500	244
17	500	277
19	500	310
21	500	343
23	500	372
25	500	391

Source: Douglas C. Montgomery and Elizabeth A. Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, 1982, p. 243 (notation changed).

Để nghiên cứu hiệu quả của chương trình giảm giá bằng các phiếu giảm giá dành cho một lít 6 chai nước ngọt 2 lít, Montgomery và Peck đã thu thập được số liệu cho trong Bảng 2. Một mẫu gồm 5500 người tiêu dùng được chia một cách ngẫu nhiên vào 11 chính sách giảm giá khác nhau, 500 người cho một chính sách. Biến phụ thuộc là người tiêu dùng có sử dụng phiếu giảm giá này hay không trong vòng một tháng (mỗi người chỉ có một phiếu).

- a. Các bạn hãy xem xét mô hình logit (với tỷ lệ phiếu giảm giá được sử dụng là biến phụ thuộc và mức giảm giá là biến độc lập) có phù hợp với dữ liệu không?

Bạn hãy gõ những dòng lệnh sau vào cửa sổ lệnh của Eviews, sẽ ước lượng được mô hình Logit cho trường hợp này

```

genr phat=n01/n
genr w=n*phat*(1-phat)
genr sqrtw=sqr(w)
ls sqrtw*log(phat/(1-phat)) sqrtw sqrtw*x
    
```

Hình 2.1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SQRTW	-2.084570	0.014517	-143.5967	0.0000
SQRTW*X	0.135713	0.000895	151.6210	0.0000

R-squared	0.999608	Mean dependent var	-0.424914
Adjusted R-squared	0.999565	S.D. dependent var	8.668768
S.E. of regression	0.180830	Akaike info criterion	-0.419551
Sum squared resid	0.294296	Schwarz criterion	-0.347206
Log likelihood	4.307530	Hannan-Quinn criter.	-0.465154
Durbin-Watson stat	1.785213		

Mô hình trên cho thấy

$$\sqrt{w_i} \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = -2.085 + 0.136 \sqrt{w_i} X_i$$

Hàm trên có một biến độc lập, có Adjusted $R^2 = 0.99$ là rất cao, đồng thời P-value ($\sqrt{w_i} X_i$) có ý nghĩa thống kê, nên có thể nói rằng mô hình trên phù hợp với dữ liệu.

b. Hãy kiểm tra xem mô hình probit cũng phù hợp với số liệu như mô hình logit hay không?

Bạn có thể gõ tiếp các câu lệnh để ước lượng mô hình Probit với dữ liệu theo nhóm.

genr i=@qnorm(phat)

Khi đó biến I ở từng quan sát sẽ như sau

Hình 2.2

obs	PHAT	I	X	
1	0.20	-0.84	5	
2	0.24	-0.69	7	
3	0.29	-0.54	9	
4	0.35	-0.38	11	
5	0.42	-0.20	13	
6	0.49	-0.03	15	
7	0.55	0.14	17	
8	0.62	0.31	19	
9	0.69	0.48	21	
10	0.74	0.66	23	
11	0.78	0.78	25	

Theo Gujarati (2003, trang 610), I_i âm khi P_i nhỏ hơn 0.5, vì vậy, khi ước lượng mô hình Probit với dữ liệu nhóm, về mặt thực hành, người ta thường cộng thêm số 5 vào I. (bạn không cộng vào không sai). Vì vậy bạn có thể gõ những lệnh sau để ước lượng mô hình

genr i5=i+5

ls i5 c x

ls i c x

Hình 2.3 Mô hình Probit khi biến phụ thuộc là I5

Dependent Variable: I5

Method: Least Squares

Date: 01/15/11 Time: 01:32

Sample: 1 11

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.724	0.012	318.707	0.000
X	0.083	0.001	115.769	0.000

R-squared	0.999	Mean dependent var	4.971
Adjusted R-squared	0.999	S.D. dependent var	0.551
S.E. of regression	0.015	Akaike info criterion	-5.391
Sum squared resid	0.002	Schwarz criterion	-5.319
Log likelihood	31.650	Hannan-Quinn criter.	-5.437
F-statistic	13402.500	Durbin-Watson stat	1.417
Prob(F-statistic)	0.000		

Hình 2.4 Mô hình probit khi biến phụ thuộc là I

Dependent Variable: I
 Method: Least Squares
 Date: 01/15/11 Time: 01:35
 Sample: 1 11
 Included observations: 11

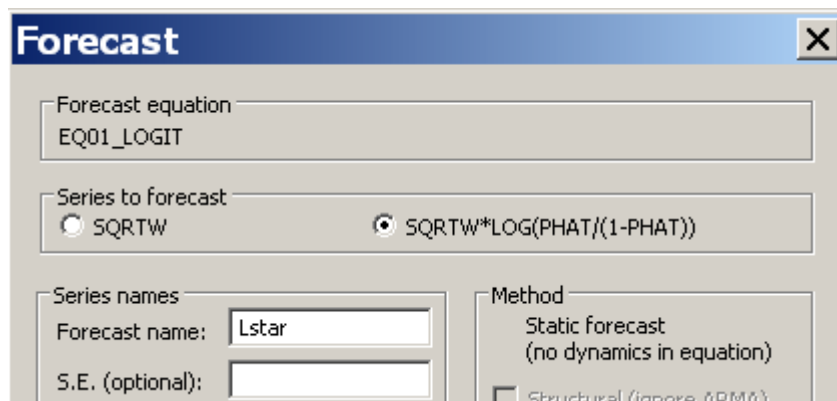
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.276	0.0117	-109.1887	0.0000
X	0.083	0.0007	115.7692	0.0000
R-squared	0.999	Mean dependent var		-0.029
Adjusted R-squared	0.999	S.D. dependent var		0.551
S.E. of regression	0.015	Akaike info criterion		-5.391
Sum squared resid	0.002	Schwarz criterion		-5.319
Log likelihood	31.650	Hannan-Quinn criter.		-5.437
F-statistic	13402.500	Durbin-Watson stat		1.417
Prob(F-statistic)	0.000			

Với Mô hình Probit ở Hình 2.3, hoặc 2.4, ta thấy biến X có ý nghĩa thống kê ở độ tin cậy 95%; Adjusted R^2 rất lớn (0.99), và Kiểm định F có ý nghĩa thống kê nên mô hình Probit phù hợp với dữ liệu.

c. Nếu mức giảm giá là 17 cent thì tỷ lệ phiếu giảm giá được sử dụng dự báo là bao nhiêu?

Giả sử bạn sử dụng Mô hình Logit để dự đoán, trong các giá trị của X có giá trị 17. Vì vậy, để dự đoán xác suất từ mô hình Logit, bạn chỉ cần bấm nút Forecast

Hình 2.5



Hình 2.6

X	PHAT	SQRTW	LSTAR
5	0.2	8.944	-12.576
7	0.244	9.604	-10.896
9	0.294	10.187	-8.793
11	0.352	10.679	-6.319
13	0.422	11.043	-3.537
15	0.488	11.177	-0.546
17	0.554	11.115	2.474
19	0.62	10.854	5.362
21	0.686	10.378	7.943
23	0.744	9.759	10.118
25	0.782	9.232	12.078

Tính toán Pi

$$\ln(P_i/(1-P_i))=2.474/11.115= 0.223$$

$$P_i/(1-P_i)=\exp(0.223)= 1.249$$

$$P_i=1.249/(1+1.249)= 0.555$$

Như vậy, với X=17 cent, tỷ lệ phiếu giảm giá được sử dụng là 55.5%

d. Hãy tính mức giảm giá mà ở đó có 70% số phiếu giảm giá sẽ được sử dụng?

Từ Mô hình Logit

$$\sqrt{W_i} \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = -2.085 + 0.136 \sqrt{W_i} X_i$$

Đơn giản trọng số ở các vế, ta có thể tính Xi

$$\ln\left(\frac{0.7}{1-0.7}\right) = -2.085 + 0.136X$$

$$0.847 = -2.085 + 0.136X \rightarrow X = 21.56 \text{ cent}$$