

Chương 3

HỒI QUY TUYẾN TÍNH BỘI (tiếp theo)

III. HỒI QUY TUYẾN TÍNH K BIẾN

1. Hàm hồi quy tổng thể (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Trong đó

- Y là biến phụ thuộc
- $X_2, X_3, \dots, X_k$ là các biến độc lập
- U_i là các sai số ngẫu nhiên

• β_1 : Hệ số tự do

$\beta_2, \beta_3, \dots, \beta_k$ là các hệ số hồi quy riêng

III. HỒI QUY TUYẾN TÍNH K BIẾN

1. Hàm hồi quy tổng thể (PRF)

Quan sát thứ 1 :

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \dots + \beta_k X_{k1} + U_1$$

Quan sát thứ 2 :

$$Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \dots + \beta_k X_{k2} + U_2$$

.....

Quan sát thứ n :

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \dots + \beta_k X_{kn} + U_n$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

1. Hàm hồi quy tổng thể (PRF)

Ký hiệu

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{2n} & X_{3n} & \dots & X_{kn} \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

Ta có

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix}$$

$$\Rightarrow PRF : Y = X \cdot \beta + U$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

2. Các giả thiết của mô hình hồi quy k biến

Giả thiết 1 : Các biến độc lập $X_2, X_3, \dots, X_k$ biết trước và không ngẫu nhiên

Giả thiết 2 : Các sai số ngẫu nhiên U_i có giá trị trung bình bằng 0 và có phương sai không thay đổi

Giả thiết 3 : Không có sự tương quan giữa các sai số U_i

III. HỒI QUY TUYẾN TÍNH K BIẾN

2. Các giả thiết của mô hình hồi quy k biến

Giả thiết 4 : Không có hiện tượng cộng tuyến giữa các biến độc lập $X_2, X_3, \dots, X_k$

Giả thiết 5 : Không có tương quan giữa các biến độc lập $X_2, X_3, \dots, X_k$ với các sai số ngẫu nhiên U_i

III. HỒI QUY TUYẾN TÍNH K BIẾN

3. Ước lượng các tham số

Hàm hồi quy mẫu :

$$\text{SRF: } Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

$$\text{hoặc: } \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

Hay : (Viết dưới dạng ma trận)

$$Y = X \hat{\beta} + e$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

3. Ước lượng các tham số

Với

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$\text{SRF: } Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

$$\text{hoặc: } \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

Khi đó

$$e_i = (Y_i - \hat{Y}_i)$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

Theo nguyên lý của phương pháp OLS thì các tham số $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$ được chọn sao cho

$$\begin{aligned} \sum e_i^2 &= \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki})^2 \\ &\rightarrow \min \end{aligned}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

Người ta chứng minh được :

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i} X_{ki} \\ \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{2i} X_{ki} & \dots & \sum X_{ki}^2 \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum Y_i \\ \sum X_{2i} Y_i \\ \dots \\ \sum X_{ki} Y_i \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

3. Ví dụ minh họa

Bảng dưới đây cho các số liệu về lượng hàng bán được của một loại hàng hóa (Y), thu nhập của người tiêu dùng (X_2) và giá bán của loại hàng này (X_3)

Tìm hàm hồi quy tuyến tính

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

Y_i (tấn/tháng)	X_2 (triệu đồng/năm)	X_3 (ngàn đồng/kg)
20	8	2
18	7	3
19	8	4
18	8	4
17	6	5
17	6	5
16	5	6
15	5	7
13	4	8
12	3	8

III. HỒI QUY TUYẾN TÍNH K BIẾN

Giải Từ số liệu trên, ta tính được các tổng như sau :

$$\begin{aligned} \sum Y_i &= 165 & \sum X_{2i}^2 &= 388 \\ \sum X_{2i} &= 60 & \sum X_{2i} X_{3i} &= 282 \\ \sum X_{3i} &= 52 & \sum X_{3i}^2 &= 308 \\ \sum Y_i^2 &= 2781 & \bar{Y} &= 16,5 \\ \sum Y_i X_{3i} &= 813 & \bar{X}_2 &= 6 \\ \sum Y_i X_{2i} &= 1029 & \bar{X}_3 &= 5,2 \end{aligned}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i} X_{3i} \\ \sum X_{3i} & \sum X_{3i} X_{2i} & \sum X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 10 & 60 & 52 \\ 60 & 388 & 282 \\ 52 & 282 & 308 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 26.165 & -2.497 & -2.131 \\ -2.497 & 0.246 & 0.196 \\ -2.131 & 0.196 & 0.183 \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{2i} \\ \sum Y_i X_{3i} \end{pmatrix} = \begin{pmatrix} 165 \\ 1028 \\ 813 \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 14.992 \\ 0.762 \\ -0.589 \end{pmatrix} \quad \text{Vậy: } \begin{matrix} \hat{\beta}_1 = 14,992 \\ \hat{\beta}_2 = 0,762 \\ \hat{\beta}_3 = -0,589 \end{matrix}$$

$$\hat{Y}_i = 14,992 + 0,762 X_{2i} - 0,589 X_{3i}$$

Kết quả hồi quy bằng Eviews như sau :

Equation: UNTITLED Workfile: HOIQUYBOI:Untitled1				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 04/01/09 Time: 10:15				
Sample: 1 10				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469
R-squared	0.960934	Mean dependent var	16.50000	
Adjusted R-squared	0.949773	S.D. dependent var	2.549510	
S.E. of regression	0.571382	Akaike info criterion	1.981807	
Sum squared resid	2.285340	Schwarz criterion	2.052582	
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226	
F-statistic	86.09278	Durbin-Watson stat	1.800073	
Prob(F-statistic)	0.000012			

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Hệ số xác định của mô hình

$$TSS = Y^T Y - n(\bar{Y})^2$$

$$ESS = \hat{\beta}^T X^T Y - n(\bar{Y})^2$$

$$RSS = TSS - ESS$$

$$\text{Hệ số xác định: } R^2 = \frac{ESS}{TSS}$$

$$\text{Hệ số xác định hiệu chỉnh: } \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

Kết quả hồi quy bằng Eviews như sau :

Equation: UNTITLED Workfile: HOIQUYBOI:Untitled1				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 04/01/09 Time: 10:15				
Sample: 1 10				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469
R-squared	0.960934	Mean dependent var	16.50000	
Adjusted R-squared	0.949773	S.D. dependent var	2.549510	
S.E. of regression	0.571382	Akaike info criterion	1.981807	
Sum squared resid	2.285340	Schwarz criterion	2.052582	
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226	
F-statistic	86.09278	Durbin-Watson stat	1.800073	
Prob(F-statistic)	0.000012			

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

Gọi c_{ij} là phần tử nằm ở dòng j cột i của ma trận $(X^T X)^{-1}$

$$\text{Khi đó: } \sigma_{\hat{\beta}_j}^2 = \sigma^2 \cdot c_{jj} \approx \hat{\sigma}^2 \cdot c_{jj}$$

$$se(\hat{\beta}_j) = \sqrt{\sigma_{\hat{\beta}_j}^2}$$

$$\text{Với } \hat{\sigma}^2 = \frac{RSS}{n-k} \quad (k \text{ là số tham số})$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

Khoảng tin cậy của β_j là

$$(\hat{\beta}_j - t_{\alpha/2} se(\hat{\beta}_j); \hat{\beta}_j + t_{\alpha/2} se(\hat{\beta}_j))$$

Hoặc tính giá trị tới hạn của β_j là

$$t = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)} \quad \text{Bậc tự do là } (n-k)$$

Kết quả hồi quy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469

R-squared	0.960934	Mean dependent var	16.50000
Adjusted R-squared	0.949773	S.D. dependent var	2.549510
S.E. of regression	0.571362	Akaike info criterion	1.961807
Sum squared resid	2.285340	Schwarz criterion	2.052582
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226
F-statistic	86.09278	Durbin-Watson stat	1.800073
Prob(F-statistic)	0.000012		

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

Kiểm định giả thiết về R^2 $H_0: R^2 = 0$

Với độ tin cậy 1- α $H_1: R^2 \neq 0$

Bước 1: tính $F = \frac{R^2(n-k)}{(k-1)(1-R^2)}$

Bước 2: Tra bảng tìm $F(k-1, n-k)$, mức ý nghĩa là α

Bước 3: Nếu $F > F(k-1, n-k)$, bác bỏ H_0 Nếu $F \leq F(k-1, n-k)$, chấp nhận H_0

Kết quả hồi quy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469

R-squared	0.960934	Mean dependent var	16.50000
Adjusted R-squared	0.949773	S.D. dependent var	2.549510
S.E. of regression	0.571362	Akaike info criterion	1.961807
Sum squared resid	2.285340	Schwarz criterion	2.052582
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226
F-statistic	86.09278	Durbin-Watson stat	1.800073
Prob(F-statistic)	0.000012		

Một vài kết quả hồi quy khác bằng Eviews

Dependent Variable: Y_CNHN	
Method: Least Squares	
Date: 04/01/09 Time: 10:23	
Sample: 1 120	
Included observations: 120	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.477215	2.207579	-1.575126	0.1180
X2_CNMS	0.080708	0.444005	0.181772	0.8561
X3_COM	0.543628	0.382117	1.422675	0.1576
X4_SUA	0.005008	0.001267	3.951200	0.0001
X5_CD	0.086614	0.009635	8.989518	0.0000
X6_RAU	0.011712	0.015035	0.779036	0.4376
X7_CB	0.143078	0.019623	7.291392	0.0000

R-squared	0.702570	Mean dependent var	17.42500
Adjusted R-squared	0.686778	S.D. dependent var	3.364902
S.E. of regression	1.883210	Akaike info criterion	4.160395
Sum squared resid	400.7523	Schwarz criterion	4.322999
Log likelihood	-242.6237	Hannan-Quinn criter.	4.226430
F-statistic	44.48686	Durbin-Watson stat	1.881293
Prob(F-statistic)	0.000000		

Theo kết quả bài tập của nhóm 13 lớp KK1_05 trường Đại học Hồng Bàng

Các yếu tố ảnh hưởng đến giá bán 1 căn nhà

Dependent Variable: Y	
Method: Least Squares	
Date: 04/01/09 Time: 10:32	
Sample: 1 225	
Included observations: 225	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.485774	0.648155	3.835155	0.0002
X2	0.003751	0.004468	0.839525	0.4021
D1	0.402730	0.294895	1.365721	0.1734
D2	0.480869	0.241677	1.989719	0.0479
D3	-1.285296	0.563680	-2.280187	0.0236
D4	1.122022	0.532487	2.107134	0.0362
D5	2.560528	0.366850	6.979768	0.0000

R-squared	0.669341	Mean dependent var	3.515511
Adjusted R-squared	0.660240	S.D. dependent var	2.844026
S.E. of regression	1.657752	Akaike info criterion	3.879419
Sum squared resid	599.0949	Schwarz criterion	3.985698
Log likelihood	-429.4346	Hannan-Quinn criter.	3.922313
F-statistic	73.54620	Durbin-Watson stat	1.764107
Prob(F-statistic)	0.000000		

X2 : diện tích
D1 : môi trường
D2 : khu vực
kinh doanh
D3 : nhu cầu bán
D4 : an ninh khu
vực
D5 : vị trí nhà
D6 : thị trường
đồng bằng

Theo kết quả bài tập của nhóm 4 lớp KK2_05 trường Đại học Hồng Bàng

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Vấn đề dự báo

$$\text{Cho } X_o = \begin{pmatrix} 1 \\ X_2^0 \\ \dots \\ X_k^0 \end{pmatrix}$$

Yêu cầu dự báo giá trị Y_0 của Y

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Vấn đề dự báo

Dự báo điểm :

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_2 X_2^0 + \hat{\beta}_3 X_3^0 + \dots + \hat{\beta}_k X_k^0$$

Dự báo khoảng :

$$(\hat{Y}_0 - t_{\frac{\alpha}{2}} se(\hat{Y}_0); \hat{Y}_0 + t_{\frac{\alpha}{2}} se(\hat{Y}_0))$$

Bậc tự do là $(n-k)$

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Vấn đề dự báo

$$\sigma_{\hat{Y}_0}^2 = \hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0$$

$$se(\hat{Y}_0) = \sqrt{\sigma_{\hat{Y}_0}^2}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

Ví dụ : Tính khoảng tin cậy của β_2 theo số liệu của ví dụ trước với độ tin cậy 95%

III. HỒI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

yêu cầu kiểm định các giả thiết

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

Với độ tin cậy 95%

III. HỒI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

Yêu cầu kiểm định các giả thiết

$$H_0: R^2 = 0$$

$$H_1: R^2 \neq 0$$

Với độ tin cậy 95%

III. HỎI QUY TUYẾN TÍNH K BIẾN



5. Ví dụ (số liệu trước)

Yêu cầu dự báo giá trị của Y khi $X_2=9$
và $X_3=9$ với độ tin cậy 95%

Hết