Chương 4

MÔ HÌNH HỒI QUY BỘI

Bảng 3.1

Doanh số bán Y_i (tr.đ)	Chi phí chào hàng X_{2i} (tr.đ)	Chi phí quảng cáo X_{3i} (tr.đ)
1270	100	180
1490	106	248
1060	60	190
1626	160	240
1020	70	150
1800	170	260
1610	140	250
1280	120	160
1390	116	170
1440	120	230
1590	140	220
1380	150	150

Dependent Variable: Y
Method: Least Squares
Date: 09/04/04 Time: 10:46
Sample: 1 12
Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

R-squared	0.967693	Mean dependent var	1413.000
Adjusted R-squared	0.960514	S.D. dependent var	231.7420
S.E. of regression	46.04989	Akaike info criterion	10.70965
Sum squared resid	19085.33	Schwarz criterion	10.83087
Log likelihood	-61.25787	F-statistic	134.7884
Durbin-Watson stat	2.493309	Prob(F-statistic)	0.000000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

HSHQ

Sai số chuẩn của các hshq – $se(\hat{\beta}_i)$

TK t

P-value

$$\hat{Y}_i = 328{,}1383 + 4{,}649511X_{2i} + 2{,}560152X_{3i}$$

- Khi chi phí chào hàng và chi phí quảng cáo đều bằng 0 thì doanh số bán trung bình của một khu vực bán hàng là 328,1383 triệu đồng/năm.

- Nếu giữ chi phí quảng cáo không đổi, khi chi phí chào hàng tăng lên 1 triệu đồng/năm sẽ làm cho doanh thu trung bình của một khu vực bán hàng tăng lên 4,6495 triệu đồng/năm.

* Nếu giữ chi phí chào hàng không đổi, khi chi phí quảng cáo tăng lên 1 triệu đồng/năm sẽ làm cho doanh thu trung bình của một khu vực bán hàng tăng 2,56 triệu đồng/năm.

MÔ HÌNH HỒI QUY TUYẾN TÍNH K BIẾN

Hàm hồi quy tổng thể (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots$$

$$+ \beta_k X_{ki} + U_i$$

β_1 – Hệ số tự do

β_1 cho biết giá trị TB của biến phụ thuộc (Y) bằng bao nhiêu khi tất cả các biến độc lập X_j đều bằng 0.

β_j (j = 2, 3, . . . k) - Hệ số hồi quy riêng của biến X_j

β_j (j = 2, 3, . . . k) cho biết TB của Y sẽ tăng (giảm) bao nhiêu đơn vị khi X_j tăng (hay giảm) 1 đơn vị,

1- Hệ số xác định:

$$R^2 = \frac{ESS}{TSS}$$

So sánh giá trị R^2

Xét 2 mô hình sau:

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + U_i$$

Ta không thể so sánh R^2 của hai mô hình trên được.

2- Hệ số xác định có hiệu chỉnh

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Thí dụ 4.1:

① Mô hình hồi qui 2 biến (Y và X_2) ta tính được:

$$R^2 = 0,80425$$

$$\Rightarrow \overline{R}^2 = 1 - (1 - 0,80425)\frac{12-1}{12-2}$$

$$= 0,78467$$

Đối với hàm 3 biến :

$\Longrightarrow$ Thêm biến giải thích mới (X_3) vào MH.

$R^2 = 0,9677$

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

$$= 1 - (1 - 0,9677)\frac{12-1}{12-3} = 0,9605$$

$\overline{R}^2 = 0,9605 > 0,78467$

Tức $\overline{R}^2$ có tăng lên

② Ta tiến hành kiểm định giả thiết: $H_0: \beta_3 = 0$; $H_1: \beta_3 \neq 0$

$$t = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{2,560152}{0,37941} = 6,748$$

Vì $t > t_{0,025}(9) = 2,262$ nên ta bác bỏ giả thiết H_0.

Vậy việc thêm biến chi phí quảng cáo (X_3) vào mô hình là cần thiết.

3- Khoảng tin cậy của các hệ số hồi qui

$$\hat{\beta}_j \pm t_{\alpha/2}(n-k).se(\hat{\beta}_j)$$

Thí dụ 4.1:

Với số liệu cho ở thí dụ 4.1, tìm khoảng tin cậy của β_2 và β_3 với hệ số tin cậy 95%?

Giải:

Với hệ số tin cậy $1 - \alpha = 95\%$

$\Rightarrow \alpha = 5\% \Rightarrow \alpha/2 = 0{,}025$

Tra bảng phân phối Student ta được $t_{\alpha/2}(n-3) = t_{0,025}(9) = 2,262$

Khoảng tin cậy của β_2:

$$4,64951 \pm 2,262 \times 0,469148$$

Hay:

$(3,588 < \beta_2 < 5,711)$

Khoảng tin cậy của β_3:

$$2{,}560152 \pm 2{,}262 \times 0{,}379407$$

Hay:

$$(1{,}702 < \beta_3 < 3{,}418)$$

4- Kiểm định g.thiết về hệ số hồi qui

Để kiểm định giả thiết:

$H_0: \beta_j = B_0$; $H_1: \beta_j \neq B_0$ (j = 1, 2, . . . k)

Với mức ý nghĩa α

Có thể sử dụng một trong 3 phương pháp sau:

1. **Ph.pháp khoảng tin cậy**
2. **Ph.pháp kiểm định mức ý nghĩa**
3. **Ph.pháp kiểm định bằng p-value**

Thí dụ:

Với số liệu cho ở thí dụ 4.1

Dùng Eviews ta có kết quả:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

5- K.định g.thiết đồng thời

Xét giả thiết:

$$H_0: \beta_2 = \beta_3 = \ldots = \beta_k = 0$$

Giả thiết trên tương đương với giả thiết:

$$H_0: R^2 = 0$$

Qui tắc kiểm định:

* Tính:

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

* Nếu $F > F_{\alpha}(k-1, n-k)$ thì bác bỏ H_0.

* Nếu $F \le F_{\alpha}(k-1, n-k)$ thì chấp nhận H_0.

- Với số liệu cho ở thí dụ 4.1

F-statistic	134.7884
Prob(F-statistic)	0.000000

7- Ma trận tương quan

- Với số liệu cho ở thí dụ 4.1

	X2	X3	Y
X2	1.00000	0.478766	0.89680
X3	0.478766	1.00000	0.784293
Y	0.89680	0.784293	1.00000

Hết chương 4

Dạng ma trận:

$$\boxed{Y = X\beta + U}$$

Trong đó:

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

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

2- Các giả thiết của mô hình

❶ $E(U_i) = 0 \quad (\forall i)$

❷ $$E(U_i.U_j) = \begin{cases} 0 & (i \neq j) \\ \sigma^2 & (i = j) \end{cases}$$

hay $E(UU^T) = \sigma^2 I$

❸ **X_2, X_3, . . . , X_k đã được xác định hay ma trận X đã xác định.**

❹ **Không xảy ra hiện tượng <u>*cộng tuyến*</u> giữa các biến giải thích hay hạng của ma trận X bằng k.**

❺ **$U_i \sim N(0, \sigma^2)$**

3- Ước lượng các tham số

Hàm hồi quy mẫu có dạng:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots\ldots + \hat{\beta}_k X_{ki}$$

Dạng ma trận:

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

trong đó:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \ldots \\ \hat{\beta}_k \end{pmatrix}$$

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix} = Y - X\hat{\beta}$$

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

Trong đó ma trận $(X^T X)$ có dạng như sau:

$$X^TX = \begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{ki} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & & \sum X_{ki}^2 \end{bmatrix}$$

Thí dụ:

Số liệu quan sát của một mẫu cho ở bảng dưới đây:

y_i	x_{2i}	x_{3i}
20	8	2
18	7	3
19	8	4
18	8	4
17	6	5

y_i	x_{2i}	x_{3i}
17	6	5
16	5	6
15	5	7
13	4	8
12	3	8

Trong đó:

Y là lượng hàng bán được của một loại hàng (tấn/tháng)

X_2 là thu nhập của người tiêu dùng (triệu đ/năm)

X_3 là giá bán (ngàn đ/kg)

Tìm hàm hồi quy tuyến tính mẫu của Y theo X_2 và X_3.

Giải:

Từ bảng số liệu đã cho ta tính được các tổng:

$$\sum_{i=1}^{10} Y_i = 165 \quad ; \quad \sum_{i=1}^{10} X_{2i} = 60 \quad ; \quad \sum_{i=1}^{10} X_{3i} = 52$$

$$\sum_{i=1}^{10} Y_i^2 = 2781 \quad ; \quad \sum_{i=1}^{10} X_{2i}^2 = 388 \quad ; \quad \sum_{i=1}^{10} X_{3i}^2 = 308$$

$$\sum_{i=1}^{10} X_{2i} X_{3i} = 282 \quad ; \quad \sum_{i=1}^{10} Y_i X_{2i} = 1029$$

$$\sum_{i=1}^{10} Y_i X_{3i} = 813$$

$$\left(X^T X\right)^{-1} = \begin{pmatrix} 10 & 60 & 52 \\ 60 & 388 & 282 \\ 52 & 282 & 308 \end{pmatrix}^{-1}$$

$$= \frac{1}{1528} \begin{pmatrix} 39980 & -3816 & -3256 \\ -3816 & 376 & 300 \\ -3256 & 300 & 280 \end{pmatrix}$$

$$\hat{\beta} = \frac{1}{1528}\begin{pmatrix} 39980 & -3816 & -3256 \\ -3816 & 356 & 300 \\ -3256 & 300 & 280 \end{pmatrix} . \begin{pmatrix} 165 \\ 1029 \\ 813 \end{pmatrix}$$

$$= \begin{pmatrix} 22908/1528 \\ 1164/1528 \\ -900/1528 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} 14,99215 \\ 0,76178 \\ -0,58901 \end{pmatrix}$$

Hàm hồi quy tuyến tính mẫu của Y theo X_2 và X_3 là:

$$\hat{Y}_i = 14,99215 + 0,76178\,X_{2i} - 0,58901\,X_{3i}$$

❶ Ph.pháp k.đ bằng k.tin cậy

Cần kiểm định giả thiết:

$$H_0: \beta_j = B_0; H_1: \beta_j \neq B_0$$

với mức ý nghĩa α

* **Trước hết ta tìm khoảng tin cậy với độ tin cậy (1- α) cho β_j . Chẳng hạn khoảng này là (α_1, α_2)**

- Nếu $B_0 \in (\alpha_1, \alpha_2)$ thì chấp nhận gt H_0.
- Nếu $B_0 \notin (\alpha_1, \alpha_2)$ thì bác bỏ gt H_0.

❷ Ph.pháp k.đ mức ý nghĩa:

Để KĐ giả thiết:

$$H_0: \beta_j = B_0; H_1: \beta_j \neq B_0$$

với mức ý nghĩa α

*** Tính**

$$t = (\hat{\beta}_j - B_0)/se(\hat{\beta}_j)$$

* Với mức ý nghĩa α, tra bảng (hoặc dùng hàm TINV trong Excel) để tìm $t_{\alpha/2}(n-k)$

* Nếu $|t| > t_{\alpha/2}(n-k)$ thì bác bỏ giả thiết H_0.

* Nếu $|t| \leq t_{\alpha/2}(n-k)$ thì chấp nhận giả thiết H_0.

③ Ph.pháp k.đ bằng p-value

$$\text{p-value} = P(|T| > |t|)$$

Các phần mềm K.tế lượng đều tính sẵn p-value

Kiểm định giả thiết:

$$H_0: \beta_j = 0;\ H_1: \beta_j \neq 0$$

với mức ý nghĩa α

11- Dự báo

Dự báo giá trị trung bình:

Cho $X_2 = X^0{}_2$; $X_3 = X^0{}_3$; . . . , $X_k = X^0{}_k$,

cần dự báo giá trị TB của biến phụ thuộc, tức dự báo:

$$E(Y/X^0) = \beta_1 + \beta_2 X^0{}_2 + \ldots + \beta_k X^0{}_k$$

Dự báo điểm (ước lượng điểm) của E(Y/X^0) là $\hat{Y}_0$.

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_2^0 + \ldots + \hat{\beta}_k X_k^0$$

Dự báo khoảng của E(Y/X^0) tính theo công thức:

$$\boxed{\hat{Y}_0 \pm t_{\alpha/2}(n-k)se(\hat{Y}_0)}$$

Trong đó:

$$se(\hat{Y}_0) = \sqrt{var(\hat{Y}_0)}$$

$$var(\hat{Y}_0) = \hat{\sigma}^2 X^{0T}(X^T X)^{-1} X^0$$

Dự báo giá trị cá biệt:

Ta tìm dự báo khoảng cho giá trị của biến phụ thuộc Y khi $X = X^0$ với độ tin cậy $1 - \alpha$, tức tìm khoảng tin cậy cho Y_0

Dự báo khoảng của Y_0 với độ tin cậy (1– α) là:

$$\hat{Y}_0 \pm t_{\alpha/2}(n-k).se(Y_0 - \hat{Y}_0)$$

Trong đó:

$$se(Y_0 - \hat{Y}_0) = \sqrt{var(Y_0 - \hat{Y}_0)}$$

$$\mathrm{var}(Y_0 - \hat{Y}_0) = \mathrm{var}\left(\hat{Y}_0\right) + \hat{\sigma}^2$$

Thí dụ:

Với số liệu cho ở thí dụ 4.1, hãy dự báo d.số bán tr.b của một khu vực bán hàng khi chi phí chào hàng là 165 triệu đ/năm và chi phí quảng cáo là 200 triệu đ/năm với độ tin cậy 95%.

Giải:

Dùng Eviews, ta tính được:

$$\hat{Y}_0 = 1607{,}388\ ;$$

$$se(\hat{Y}_0) = 25{,}2017;$$

$$t_{0.025}(9) = TINV(0.05,9)$$

$$= 2{,}26216.$$

Vậy dự báo khoảng cho doanh số bán trung bình của một khu vực bán hàng với độ tin cậy 95% là:

$$1607,338 \pm 2,26216 \times 25,2017$$

hay

$$(1550,328 < E(Y/X^0) < 1664,3483)$$

III. MỘT SỐ DẠNG HÀM

1- Hàm sản xuất Cobb – Douglas:

$$Yi = \beta_1 X_{2i}^{\beta 2} X_{3i}^{\beta 3} e^{Ui}$$

trong đó:

Y- sản lượng; X2 - lượng lao động; X3 - lượng vốn;   Ui - sai số ng.nhiên

Lôgarit cả hai vế, ta được:

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$

$$= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$

(hàm hồi qui tuyến tính lôgarit)

Đặc điểm của hàm Cobb- Douglas

❶ β_2 là hệ số co giãn của sản lượng đối với lao động, nó cho biết sản lượng tăng (hay giảm) bao nhiêu % khi lượng lao động tăng (giảm) 1%, với đ.kiện lượng vốn không đổi.

Lôgarit cả hai vế, ta được:

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$

$$= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$

(hàm hồi qui tuyến tính lôgarit)

❷ β_3 là hệ số co giãn của sản lượng đối với vốn, nó cho biết sản lượng tăng (hay giảm) bao nhiêu % khi lượng vốn tăng (giảm) 1%, với điều kiện lượng lao động không đổi.

❸ Tổng ($\beta_2+\beta_3$) cho biết thông tin để đánh giá việc tăng qui mô s/x.

- Nếu ($\beta_2+\beta_3$) = 1 thì tăng qui mô không hiệu quả
- Nếu ($\beta_2+\beta_3$)< 1 thì tăng qui mô kém hiệu quả
- Nếu ($\beta_2+\beta_3$)> 1 thì tăng qui mô có hiệu quả

<u>Thí dụ :</u>

Các dữ liệu của khu vực nông nghiệp Đài Loan giai đoạn 1958-1972 (Bảng 4.44)

Y- Tổng sản lượng (triệu NT$- Đô la Đài loan)

X_2- Ngày lao động (triệu ngày);

X_3- Lượng vốn (triệu NT$)

Hồi qui lnY theo $\ln X_2$ và $\ln X_3$:

$$\ln Y_i = -3{,}38194 + 1{,}4887 \ln X_{2i} + 0{,}49997 \ln X_{3i} + e_i$$

Kết quả trên cho biết, trong g/đ 1958 – 1972, trong khu vực nông nghiệp của Đài loan, khi tăng 1% lượng lao động sẽ làm tăng TB khoảng 1,5% sản lượng, nếu giữ lượng vốn không đổi.

Nếu giữ lượng lao động không đổi, khi lượng vốn tăng 1% thì sản lượng tăng TB khoảng 0,5%.

Tổng (1,4887+0,49997) = 1,98867 cho biết trong g/đ này việc tăng qui mô là có hiệu quả.

2.MÔ HÌNH ĐA THỨC

$$Y_i=\beta_0+ \beta_1X_i + \beta_2X_i^2 +....+\beta_kX_i^k +U_i$$

- **Gọi là hồi qui đa thức bậc k.**
- **Mô hình này có các số hạng X_2, X_3.. đều là hàm không t. tính của X.**
- **Vì vậy chúng không vi phạm về giả thuyết đa cộng tuyến**

Nếu giữ lượng lao động không đổi, khi lượng vốn tăng 1% thì sản lượng tăng TB khoảng 0,5%.

Tổng (1,4887+0,49997) = 1,98867 cho biết trong g/đ này việc tăng qui mô là có hiệu quả.