Chương 4
Mô hình hồi qui bội

1. Mô hình :

Mô hình hồi qui tuyến tính k biến (PRF) :

$$E(Y/X_{2i},\ldots,X_{ki}) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + U_i$$

Trong đó :

Y - biến phụ thuộc

$X_2,\ldots,X_k$ - các biến độc lập

β_1 là hệ số tự do

β_j là các hệ số hồi qui riêng, cho biết khi X_j tăng 1 đvị thì trung bình của Y sẽ thay đổi β_j đvị trong trường hợp các yếu tố khác không đổi (j=2,…,k).

Khi k = 3 thì ta có mô hình hồi qui tuyến tính ba biến :

$$E(Y/X_2, X_3) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 \quad \text{(PRF)}$$

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

2. Các giả thiết của mô hình

- Giả thiết 1: Các biến độc lập phi ngẫu nhiên, giá trị được xác định trước.
- Giả thiết 2 : $E(U_i) = 0 \quad \forall i$
- Giả thiết 3 : $Var(U_i) = \sigma^2 \quad \forall i$
- Giả thiết 4 : $Cov(U_i, U_j) = 0 \quad i \neq j$
- Giả thiết 5 : $Cov(X_i, U_i) = 0 \quad \forall i$
- Giả thiết 6 : $U_i \sim N(0, \sigma^2) \quad \forall i$
- Giả thiết 7 : Không có hiện tượng cộng tuyến giữa các biến độc lập.

3. Ước lượng các tham số

a. Mô hình hồi qui ba biến :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i \qquad \text{(PRF)}$$

Hàm hồi qui mẫu :

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

Giả sử có một mẫu gồm n quan sát các giá trị (Y_i, X_{2i}, X_{3i}). Theo phương pháp OLS,

$\hat{\beta}_j$ (j= 1,2,3) phải thoả mãn :

$$\sum e_i^2 \rightarrow \min$$

Tức là :

$$\begin{cases} \dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 0 \\ \dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 0 \\ \dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_3} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-1) = 0 \\ \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0 \\ \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0 \end{cases}$$

Do $e_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$

Giải hệ ta có :

$$\hat{\beta}_2 = \frac{\sum x_{2i} y_i \sum x_{3i}^2 - \sum x_{2i} x_{3i} \sum x_{3i} y_i}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2}$$

$$\hat{\beta}_3 = \frac{\sum x_{3i} y_i \sum x_{2i}^2 - \sum x_{2i} x_{3i} \sum x_{2i} y_i}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2}$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_2 - \hat{\beta}_3 \overline{X}_3$$

* Phương sai của các hệ số ước lượng

$$Var(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\sum \left(\bar{X}_2 x_{3i} - \bar{X}_3 x_{2i} \right)^2}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i} \right)^2} \right] \times \sigma^2$$

$$Var(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i} \right)^2} \times \sigma^2$$

$$Var(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i} \right)^2} \times \sigma^2$$

Trong đó : $\sigma^2 = Var(U_i)$

σ^2 chưa biết nên dùng ước lượng của nó là :

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-3}$$

Với :

$$\sum e_i^2 = TSS - ESS = \sum y_i^2 - \hat{\beta}_2 \sum x_{2i} y_i - \hat{\beta}_3 \sum x_{3i} y_i$$

b. Mô hình hồi qui tuyến tính k biến

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + U_i \quad \text{(PRF)}$$

$$(i = 1, \ldots, n)$$

Hàm hồi qui mẫu :

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

Theo phương pháp OLS, $\hat{\beta}_j$ (j= 1,2,…,k) phải thoả mãn :

$$\sum e_i^2 \to \min$$

Tức là :

$$\begin{cases} \dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 0 \\ \vdots \\ \dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_k} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})(-1) = 0 \\ \vdots \\ \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})(-X_{ki}) = 0 \end{cases}$$

Viết hệ dưới dạng ma trận : $(X^T X)\hat{\beta} = X^T Y$

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

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \qquad X^T Y = \begin{bmatrix} \sum Y_i \\ \sum X_{2i} Y_i \\ \vdots \\ \sum X_{ki} Y_i \end{bmatrix}$$

$$X^T X = \begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i} X_{3i} & \dots & \sum X_{2i} X_{ki} \\ & \vdots & \vdots & & \\ \sum X_{ki} & \sum X_{ki} X_{2i} & \sum X_{ki} X_{3i} & \dots & \sum X_{ki}^2 \end{bmatrix}$$

4. Hệ số xác định

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

* <u>Chú ý</u> : Khi tăng số biến độc lập trong mô hình thì R^2 cũng tăng cho dù các biến độc lập tăng thêm có ảnh hưởng mô hình hay không . Do đó không thể dùng R^2 để quyết định có nên thêm biến vào mô hình hay không mà thay vào đó có thể sử dụng hệ số xác định được hiệu chỉnh :

$$\bar{R}^2 = 1 - \frac{\sum e_i^2/(n-k)}{\sum y_i^2/(n-1)}$$

Hay:

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

<u>Tính chất của $\bar{R}^2$:</u>

- Khi k > 1, $\bar{R}^2 \leq R^2 \leq 1$.

- $\bar{R}^2$ có thể âm, trong trường hợp âm, ta coi giá trị của nó bằng 0.

*** Cách sử dụng $\overline{R}^2$ để quyết định đưa thêm biến vào mô hình :**

Mô hình hai biến	Mô hình ba biến
$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i}$ **(1)**	$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$ **(2)**
$\downarrow$	$\downarrow$
R_1^2	R_2^2
$\downarrow$	$\downarrow$
$\overline{R}_1^2$	$\overline{R}_2^2$

- Nếu $\overline{R}_1^2 > \overline{R}_2^2$ thì chọn mô hình (1), tức là không cần đưa thêm biến X_3 vào mô hình. Ngược lại, ta chọn mô hình (2).

- <u>So sánh hai giá trị R^2</u> :

Nguyên tắc so sánh :

- Cùng cỡ mẫu n .
- Cùng các biến độc lập.
- Biến phụ thuộc phải ở dạng giống nhau. Biến độc lập có thể ở bất cứ dạng nào.

Ví dụ :

5. Ma trận tương quan

Xét mô hình : $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + ... + \hat{\beta}_k X_{ki}$

Gọi r_{tj} là hệ số tương quan tuyến tính giữa biến thứ t và thứ j. Trong đó Y được xem là biến thứ 1.

Ma trận tương quan tuyến tính có dạng :

$$\begin{bmatrix} 1 & r_{12} & ... & r_{1k} \\ r_{21} & 1 & ... & r_{2k} \\ ... & & ... & \\ r_{k1} & r_{k2} & ... & 1 \end{bmatrix}$$

6. Ma trận hiệp phương sai

$$\operatorname{cov}(\hat{\beta}) = \begin{bmatrix} \operatorname{var}(\hat{\beta}_1) & \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) & \ldots & \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_1) & \operatorname{var}(\hat{\beta}_2) & \ldots & \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \ldots & & \ldots & \\ \operatorname{cov}(\hat{\beta}_k, \hat{\beta}_1) & \operatorname{cov}(\hat{\beta}_k, \hat{\beta}_2) & \ldots & \operatorname{var}(\hat{\beta}_k) \end{bmatrix}$$

Để tính ma trận hiệp phương sai của các hệ số, áp dụng công thức :

$$\operatorname{cov}(\hat{\beta}) = (X^T X)^{-1} \sigma^2$$

7. Khoảng tin cậy của các hệ số hồi qui

Khoảng tin cậy của β_j (j =1,2, …, k) là :

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

Trong đó, **k** là số tham số trong mô hình.

8. Kiểm định giả thiết

a. Kiểm định $H_0 : \beta_j = a$ (=const)

(j = 1, 2, …, k)

Phần này hoàn toàn tương tự như ở mô hình hồi qui hai biến, khác duy nhất ở chỗ ***bậc tự do của thống kê t là (n-k).***

b. Kiểm định giả thiết đồng thời :

$\mathbf{H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0} \Leftrightarrow H_0 : R^2 = 0$

$\mathbf{H_1 : \exists\, \beta_j \neq 0\ (2 \leq j \leq k)} \Leftrightarrow H_1 : R^2 \neq 0$

Cách kiểm định :

-Tính

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

$$\left[\begin{array}{l} \text{Nếu } p(F^* > F) \leq \alpha \\ \text{Nếu } F > F_\alpha(k-1, n-k) \end{array}\right. \Rightarrow \text{bác bỏ } H_0,$$

Tức là các hệ số hồi qui không đồng thời bằng 0 hay hàm hồi qui phù hợp.

c. Kiểm định Wald

Xét mô hình (U) sau đây :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + U_i$$

(U) được xem là mô hình không hạn chế.

Ví dụ 1 : Với mô hình (U), cần kiểm định

$$H_0 : \beta_2 = \beta_5 = 0$$

Áp đặt giả thiết H_0 lên mô hình (U), ta có mô hình hạn chế (R) như sau :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_4 X_{4i} + U_i \qquad (R)$$

Để kiểm định H_0, ta dùng kiểm định Wald.

Các bước kiểm định Wald :

- Hồi qui mô hình (U) → thu được RSS_U.
- Hồi qui mô hình (R) → thu được RSS_R.
- Tính

$$F = \frac{(RSS_R - RSS_u)/(df_R - df_U)}{RSS_U / df_U}$$

df_U : bậc tự do của (U)

df_R : bậc tự do của (R)

- $\left\{\begin{array}{l}\text{Nếu } p\,(F^* > F) \le \alpha \\ \text{Nếu } F > F_\alpha(df_R - df_U,\ df_U)\end{array}\right. \Rightarrow \text{bác bỏ } H_0,$

Ví dụ 2 : Với mô hình (U), kiểm định

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

Áp đặt H_0 lên (U), ta có mô hình (R):

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_2 X_{3i} + \beta_2 X_{4i} + \beta_5 X_{5i} + U_i$$

hay

$$Y_i = \beta_1 + \beta_2 (X_{2i} + X_{3i} + X_{4i}) + \beta_5 X_{5i} + U_i$$

Đến đây, áp dụng các bước kiểm định Wald cho giả thiết H_0.

Ví dụ 3 : Với mô hình (U), kiểm định

$$H_0 : \beta_2 + \beta_3 = 1$$

Thực hiện tương tự như các ví dụ trên, bằng các áp đặt H_0 lên (U), ta có mô hình hạn chế (R) :

$$Y_i = \beta_1 + \beta_2 X_{2i} + (1 - \beta_2) X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + U_i$$

$$(Y_i - X_{3i}) = \beta_1 + \beta_2 (X_{2i} - X_{3i}) + \beta_4 X_{4i} + \beta_5 X_{5i} + U_i$$

* Chú ý : Trong Eviews, thủ tục kiểm định Wald được viết sẵn, bạn chỉ cần gõ vào giả thiết bạn muốn kiểm định rồi đọc kết quả.

9. Dự báo :

a. Dự báo giá trị trung bình

Cho X_2^0, X_3^0, …, X_k^0. Dự báo E(Y).

- Dự báo điểm của E(Y) là :

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

- Dự báo khoảng của E(Y) :

$$[\hat{Y}_0 - se(\hat{Y}_0)t_{\alpha/2}(n-k)\ ;\ \hat{Y}_0 + se(\hat{Y}_0)t_{\alpha/2}(n-k)]$$

Trong đó :

$$Var(\hat{Y}_0) = X^{0T}(X^TX)^{-1}X^0\sigma^2 \qquad X^0 = \begin{bmatrix} 1 \\ X_2^0 \\ \vdots \\ X_k^0 \end{bmatrix}$$

b. Dự báo giá trị cá biệt của Y khi X=X⁰.

$$[\hat{Y}_0 - se(Y_0 - \hat{Y}_0)t_{\alpha/2}(n-k) \;;\; \hat{Y}_0 + se(Y_0 - \hat{Y}_0)t_{\alpha/2}(n-k)]$$

Trong đó : $Var(Y_0 - \hat{Y}_0) = Var(\hat{Y}_0) + \sigma^2$