

II.3 ***Ma trận hiệp phương sai của các hệ số hồi quy***

Ma trận hiệp phương sai của hệ số hồi quy mẫu, kí hiệu $Cov(\hat{\beta})$ là ma trận được xác định như sau:

$$\begin{aligned} Cov(\hat{\beta}) &= E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)^T \right) \\ &= \begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & \dots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & Var(\hat{\beta}_k) \end{pmatrix} \end{aligned}$$

Ma trận hiệp phương sai của các hệ số hồi quy mẫu là ma trận vuông cấp k, đối xứng qua đường chéo chính và phần tử thứ j trên đường chéo chính là phương sai của $\hat{\beta}_j$



II.3

Ma trận hiệp phương sai của các hệ số hồi quy

Có thể chứng minh được rằng: $Cov((\hat{\beta})) = \sigma^2(X^T X)^{-1}$

với $V(U_i) = \sigma^2$

Khi σ^2 chưa biết, dùng ước lượng không chệch của nó:

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - k} \quad \text{với} \quad \sum_{i=1}^n e_i^2 = Y^T Y - \hat{\beta}^T X^T Y$$

Khai triển công thức này ta được:

$$\sum e_i^2 = \sum Y_i^2 - (\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum Y_i X_{2i} + \dots + \hat{\beta}_k \sum Y_i X_{ki})$$

Khi đó: $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim T(n-k) \quad j=1..k$ Phân phối Student
(n-k) bậc tự do

Xác định ma trận hiệp phương sai

- Bước 1: Tìm ma trận $(X^T X)^{-1}$ (xem lại cách xác định pt hồi quy)
- Bước 2: - σ^2 đã biết: $Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ (*)
 - σ^2 chưa biết: + Tìm $\hat{\beta}$ (xem lại cách xác định pt hồi quy)
 - + Tính $\sum Y_i^2, \sum Y_i, \sum Y_i X_{2i}, \dots, \sum Y_i X_{ki}$
 - + Tính $\sum e_i^2 = \sum Y_i^2 - (\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum Y_i X_{2i} + \dots + \hat{\beta}_k \sum Y_i X_{ki})$
 - + Tính ước lượng của σ^2 : $\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - k}$
 - + Thay σ^2 bởi $\hat{\sigma}^2$ trong các công thức (*)
- > $Var(\hat{\beta}_j)$ = phần tử thứ j nằm trên đường chéo chính

III. KHOẢNG TIN CẬY VÀ KIỂM ĐỊNH GIẢ THUYẾT

III.1 Khoảng tin cậy của β_j ($j = 1 \dots k$)

Cơ sở: $T_{\beta_j} = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim T(n-k) \quad j = 1 \dots k$

với $se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$, $V(\hat{\beta}_j)$ là phần tử thứ j trên đường chéo chính của ma trận $Cov(\hat{\beta})$

Nhắc lại: $Cov(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$ với $\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-k}$

$$\sum e_i^2 = \sum Y_i^2 - (\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum Y_i X_{2i} + \dots + \hat{\beta}_k \sum Y_i X_{ki})$$



III. KHOẢNG TIN CẬY VÀ KIỂM ĐỊNH GIẢ THUYẾT

III.1 *Khoảng tin cậy của β_j ($j = 1 \dots k$)*

Với hệ số tin cậy $1-\alpha$ ($\alpha \ll, <10\%$), ta xác định được $t_{\alpha/2}(n-k)$ (tra bảng phân phối Student) thỏa mãn:

$$P\left(\left|\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}\right| < t_{\alpha/2}(n-k)\right) = 1 - \alpha$$

Khoảng tin cậy $(1-\alpha)$ của β_j ($j = 1 \dots k$) cho bởi:

$$\left(\hat{\beta}_j - t_{\alpha/2}(n-k) * se(\hat{\beta}_j); \hat{\beta}_j + t_{\alpha/2}(n-k) * se(\hat{\beta}_j)\right)$$

BẢNG PHÂN PHỐI STUDENT $t_{\alpha}(k)$

Bậc tự do	α				
	0,1	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947

Bậc tự do	α				
	0,1	0,05	0,025	0,01	0,005
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750

Cách xác định khoảng tin cậy của β_j ($j=1..k$)

- Bước 1: Tìm $\hat{\beta}_j, V(\hat{\beta}_j) \Rightarrow se(\hat{\beta}_j)$ (xem lại cách xác định phương trình hồi quy, ma trận phương sai)
- Bước 2: Với α cho trước, tra bảng phân phối Student tìm $t_{\alpha/2}(n-k)$
- Bước 3: Khoảng tin cậy $(1-\alpha)$ của β_j cho bởi:

$$\left(\hat{\beta}_j - t_{\alpha/2}(n-k) * se(\hat{\beta}_j); \hat{\beta}_j + t_{\alpha/2}(n-k) * se(\hat{\beta}_j) \right)$$

III.2 Kiểm định giả thuyết của β_j ($j = 1..k$)

- Giả sử với mức ý nghĩa α cho trước ta cần kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j \neq \beta_j^* \end{cases}$$

$$(H_1 : \beta_j < \beta_j^*, H_1 : \beta_j > \beta_j^*)$$
- Tiêu chuẩn kiểm định: $T = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)}$
 Nếu H_0 đúng thì $t \sim T(n-k)$
- Bảng kết quả: ($t_\alpha(k)$ được cho trong bảng phân phối Student)

Loại giả thiết	Giả thiết H_0	Giả thiết đối H_1	Miền bác bỏ H_0
Hai phía	$\beta_j = \beta_j^*$	$\beta_j \neq \beta_j^*$	$ T > t_{\frac{\alpha}{2}}(n - k)$
Phía trái	$\beta_j = \beta_j^*$	$\beta_j < \beta_j^*$	$T < -t_\alpha(n - k)$
Phía phải	$\beta_j = \beta_j^*$	$\beta_j > \beta_j^*$	$T > t_\alpha(n - k)$

VD



Các bước kiểm định giả thuyết của β_j ($j=1..k$) với β_j^*

- Bước 1: Biểu diễn các giả thuyết H_0 , H_1 dưới dạng toán học.
- Bước 2: Tính $\hat{\beta}_j, \hat{\sigma}^2 \Rightarrow se(\hat{\beta}_j)$ (xem lại cách xác định phương trình hồi quy, phương sai của các hệ số hồi quy)
- Bước 3: Tính $T = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)}$
- Bước 4: Với α cho trước, tra bảng phân phối Student tìm $t_{\frac{\alpha}{2}}(n-k)$ (kiểm định 2 phía) hoặc $t_{\alpha}(n-k)$ (kiểm định 1 phía)
- Bước 5: So sánh T và $t_{\frac{\alpha}{2}}(n-k)$ (hoặc $t_{\alpha}(n-k)$), dựa vào bảng kết quả đưa ra quyết định.

IV. PHÂN TÍCH PHƯƠNG SAI VÀ KIỂM ĐỊNH GIẢ THUYẾT ĐỒNG THỜI

Gọi: $TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n y_i^2$

$$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$$

$$ESS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum_{i=1}^n x_i^2$$

Mối liên hệ:

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

Tỉ lệ % toàn bộ sai lệch của Y so với $\bar{Y}$ gây ra bởi mô hình (hoặc các biến độc lập), biểu thị độ phù hợp của hàm hồi quy mẫu

Tỉ lệ % toàn bộ sai lệch của Y so với $\bar{Y}$ gây ra bởi sai số ngẫu nhiên và các yếu tố khác

IV.1

Hệ số xác định bội của hàm hồi quy

- Định nghĩa 1: Hệ số xác định bội R^2 được định nghĩa như sau:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \in [0;1]$$

Thước đo độ phù hợp của hàm hồi quy: R^2 lớn thì hàm hồi quy là phù hợp.

- Trong thực hành, người ta dùng công thức:

$$\begin{aligned} R^2 &= \frac{\hat{\beta}^T X^T Y - n\bar{Y}^2}{Y^T Y - n\bar{Y}^2} = 1 - \frac{\sum e_i^2}{\sum Y_i^2 - n\bar{Y}^2} \\ &= \frac{\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum Y_i X_{2i} + \dots + \hat{\beta}_k \sum Y_i X_{ki} - n\bar{Y}^2}{\sum Y_i^2 - n\bar{Y}^2} \end{aligned}$$

➤ Một số tính chất của hệ số xác định bội R^2 :

1. $0 \leq R^2 \leq 1$

2. R^2 là hàm không giảm của số biến giải thích có trong mô hình. Nếu tăng số biến giải thích trong mô hình thì R^2 tăng.

→ Không thể dùng R^2 làm tiêu chuẩn để xét việc đưa thêm hay không đưa thêm biến độc lập mới vào mô hình mà phải dùng hệ số xác định bội đã điều chỉnh.



IV.2

Hệ số xác định bội đã điều chỉnh

Định nghĩa 2: Hệ số xác định bội đã điều chỉnh, ký hiệu $\overline{R^2}$ được định nghĩa như sau:

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

Tính chất:

1. Nếu $k > 1$ thì $\overline{R^2} \leq R^2 \leq 1$ và $\overline{R^2}$ cũng là hàm không giảm đối với số biến giải thích có trong mô hình tuy nhiên tốc độ tăng của $\overline{R^2}$ chậm hơn so với R^2
2. $\overline{R^2}$ có thể nhận giá trị âm dù R^2 luôn dương

Khi nào cần đưa thêm biến mới vào mô hình?

Khi đáp ứng đồng thời 2 yêu cầu:

- Khi đưa thêm biến mới vào, $\overline{R^2}$ tăng lên.

- Hệ số hồi quy của biến mới X_j là $\beta_j \neq 0$

-> Kiểm định giả thuyết:
$$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$$

Nếu H_0 bị bác bỏ, mô hình có biến mới là phù hợp hơn, ta cần thêm biến mới vào mô hình.

Cách tìm hệ số xác định bội của hàm hồi quy

- Bước 1: Tìm $\bar{Y}, \sum Y_i^2, \sum e_i^2$ (xem lại cách xác định ma trận hiệp phương sai)
- Bước 2: Tính $R^2 = 1 - \frac{\sum e_i^2}{\sum Y_i^2 - n\bar{Y}^2}$

cuu duong than cong. com

Cách tìm hệ số xác định bội đã điều chỉnh

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

cuu duong than cong. com

IV.2 Kiểm định sự phù hợp của mô hình

- Kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_2 = \dots = \beta_k = 0 \\ H_1 : \exists j \in \{2, \dots, k\} \text{ sao cho } \beta_j \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0 : R^2 = 0 \rightarrow \text{Không phù hợp} \\ H_1 : R^2 > 0 \rightarrow \text{Phù hợp} \end{cases}$$

- Tiêu chuẩn kiểm định: $F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$

Nếu H_0 đúng thì $F \sim F(k-1, n-k)$

- Với α cho trước, xác định được $F_\alpha(k-1, n-k)$ thỏa:

$$P(F < F_\alpha(k-1, n-k)) = 1 - \alpha$$

- Điều kiện bác bỏ $H_0 : F > F_\alpha(k-1, n-k)$: hàm hồi quy là phù hợp.



Các bước kiểm định sự phù hợp của mô hình

- Bước 1: Tìm hệ số xác định bội R^2 (xem lại cách tìm hệ số xác định bội của hàm hồi quy)
- Bước 2: Tính $F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$
- Bước 3: Với α cho trước, tra bảng phân phối Fisher tìm $F_\alpha(k-1, n-k)$.
- Bước 4: Nếu $F > F_\alpha(k-1, n-k)$, hàm hồi quy phù hợp.
Nếu $F \leq F_\alpha(k-1, n-k)$, hàm hồi quy không phù hợp.

V. PHÂN TÍCH HỒI QUY VÀ DỰ BÁO

Xét hàm hồi quy tổng thể và hàm hồi quy mẫu có dạng:

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

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

Viết dưới dạng ma trận: $Y = X\beta; \hat{Y} = X\hat{\beta}$

Biết $X_2 = X_{20}, X_3 = X_{30}, \dots, X_k = X_{k0}$ hay kí hiệu: $X_0 = \begin{pmatrix} 1 \\ X_{20} \\ \dots \\ X_{k0} \end{pmatrix}$

Có 2 loại dự báo cơ bản:

- Dự báo trung bình có điều kiện của Y khi $X=X_0 : E(Y | X_0)$
- Dự báo giá trị cá biệt của Y khi $X=X_0 : Y_0$

Ước lượng điểm của Y_0 hay của $E(Y | X_0)$ cho bởi:

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$$

V.1

Dự báo giá trị trung bình

- Thống kê $T = \frac{\hat{Y}_0 - E(Y / X_0)}{se(\hat{Y}_0)} \sim T(n-k)$ với

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$$

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

- Với α cho trước ($\alpha \ll 1$, $< 10\%$), xác định được $t_{\alpha/2}(n-k)$ dựa vào bảng phân phối Student thỏa mãn:

$$P(|T| < t_{\alpha/2}(n-k)) = 1 - \alpha$$

- Khoảng tin cậy $(1-\alpha)$ của $E(Y/X_0)$ được xác định:

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



Các bước dự báo giá trị trung bình có điều kiện

- Bước 1: Tìm $(X^T X)^{-1}$, $\hat{\sigma}^2$, $\hat{\beta}$ (xem lại cách xác định pt hồi quy, ma trận hiệp phương sai).
- Bước 2: Tính $(X^T X)^{-1} X_0$ rồi tính $X_0^T (X^T X)^{-1} X_0$
- Bước 3: Tính $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$
$$se(\hat{Y}_0) = \sqrt{\hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0}$$
- Bước 4: Với α , tra bảng phân phối Student tìm $t_{\alpha/2}(n-k)$
- Bước 5: Khoảng tin cậy $(1-\alpha)$ của $E(Y/X_0)$ cho bởi:

$$\left(\hat{Y}_0 - t_{\alpha/2}(n-k) * se(\hat{Y}_0); \hat{Y}_0 + t_{\alpha/2}(n-k) * se(\hat{Y}_0) \right)$$

V.2

Dự báo giá trị cá biệt

- Thống kê $T = \frac{Y_0 - \hat{Y}_0}{se(Y_0 - \hat{Y}_0)} \sim T(n-k)$ với

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$$

$$se(Y_0 - \hat{Y}_0) = \sqrt{\hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0 + \hat{\sigma}^2}$$

- Với α cho trước, xác định được $t_{\alpha/2}(n-k)$ dựa vào bảng phân phối Student thỏa mãn:

$$P(|T| < t_{\alpha/2}(n-k)) = 1 - \alpha$$

- Khoảng tin cậy $(1-\alpha)$ của Y_0 được xác định:

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



Các bước dự báo giá trị cá biệt

- Bước 1: Tìm $(X^T X)^{-1}$, $\hat{\sigma}^2$, $\hat{\beta}$ (xem lại cách xác định pt hồi quy, ma trận hiệp phương sai).
- Bước 2: Tính $(X^T X)^{-1} X_0$ rồi tính $X_0^T (X^T X)^{-1} X_0$
- Bước 3: Tính $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$
$$se(Y_0 - \hat{Y}_0) = \sqrt{\hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0 + \hat{\sigma}^2}$$
- Bước 4: Với α , tra bảng phân phối Student tìm $t_{\alpha/2}(n-k)$
- Bước 5: Khoảng tin cậy $(1-\alpha)$ của $E(Y/X_0)$ cho bởi:

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