

MÔ HÌNH HỒI QUY HAI BIẾN

I

PHƯƠNG PHÁP
BÌNH PHƯƠNG
NHỎ NHẤT

II

CÁC GIẢ THIẾT
CƠ BẢN

III

KHOẢNG TIN CẬY
VÀ KIỂM ĐỊNH
GIẢ THUYẾT

IV

PHÂN TÍCH PHƯƠNG
SAI VÀ SỰ PHÙ HỢP
CỦA MÔ HÌNH

V

PHÂN TÍCH HỒI
QUY VÀ DỰ BÁO

MÔ HÌNH HỒI QUY 2 BIẾN $Y_i = \beta_1 + \beta_2 X_i + U_i$

Trong đó: Y_i : giá trị của biến phụ thuộc Y ($i=1..n$);

β_1 : hệ số chặn; β_2 : hệ số góc; U_i : sai số ngẫu nhiên

MÔ HÌNH HỒI QUY MẪU $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Xây dựng trên mẫu thử có n quan sát: $\{(X_i, Y_i), i = 1..n\}$

Trong đó:

$\hat{Y}_i$: ước lượng của Y_i hoặc $E(Y/X_i)$ ($i=1..n$)

$\hat{\beta}_1, \hat{\beta}_2$: ước lượng của hệ số hồi quy tổng thể

I. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT(OLS)

I.1 Nội dung phương pháp

Xác định $\widehat{\beta}_1, \widehat{\beta}_2$ sao cho:

Mô hình hồi quy tuyến tính đơn

$$S = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 \min \Leftrightarrow S = \sum_{i=1}^n (Y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 X_i))^2 \min$$

$$\Leftrightarrow \begin{cases} \frac{\partial S}{\partial \widehat{\beta}_1} = \sum_{i=1}^n 2(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i)(-1) = 0 \\ \frac{\partial S}{\partial \widehat{\beta}_2} = \sum_{i=1}^n 2(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i)(-X_i) = 0 \end{cases} \Leftrightarrow \begin{cases} \widehat{\beta}_1 = \bar{Y} - \widehat{\beta}_2 \bar{X} \\ \widehat{\beta}_2 = \frac{\overline{XY} - \bar{Y} \bar{X}}{\overline{X^2} - \bar{X}^2} \end{cases}$$
$$\left(\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \right)$$

Ý nghĩa các tham số

- β_1 : ảnh hưởng của các nguyên nhân khác không được kể đến trong mô hình tới kết quả Y
- β_2 : ảnh hưởng của nguyên nhân X tới kết quả Y.
 - $\beta_2 > 0$: X và Y có mối liên hệ thuận (cùng chiều)
 - $\beta_2 < 0$: X và Y có mối liên hệ nghịch (ngược chiều)

Khi X tăng lên 1 đơn vị thì Y tăng (giảm) trung bình β_2 đơn vị.



I.2 Tính chất

- i) $\widehat{\beta}_1, \widehat{\beta}_2$ xác định duy nhất (với 1 mẫu thử cụ thể). Chúng là các ước lượng điểm, các đại lượng ngẫu nhiên.
- ii) Hàm hồi quy tổng thể đi qua trung bình mẫu: $\bar{Y} = \widehat{\beta}_1 + \widehat{\beta}_2 \bar{X}$
- iii) Giá trị trung bình của $\widehat{Y}_i$ bằng trung bình các quan sát: $\bar{\widehat{Y}} = \bar{Y}$
- iv) Tổng các phần dư bằng 0: $\sum_{i=1}^n e_i = 0$
- v) Các phần dư không tương quan với $\widehat{Y}_i$: $\sum_{i=1}^n \widehat{Y}_i e_i = 0$
- vi) Các phần dư không tương quan với X_i : $\sum_{i=1}^n X_i e_i = 0$ 

Xác định phương trình hồi quy

- Bước 1: Tính cột X_i^2 -> trung bình X_i^2 : $\overline{X^2} = \sum_{i=1}^n X_i^2 / n$
- Bước 2: Tính cột giá trị $X_i Y_i$ -> trung bình $X_i Y_i$ ($\overline{XY}$)
- Bước 3: Tính trung bình của X_i , Y_i ($\bar{X}$, $\bar{Y}$)
- Bước 4: Tính
$$\hat{\beta}_2 = \frac{\overline{XY} - \bar{Y} \bar{X}}{\overline{X^2} - \bar{X}^2}$$
- Bước 5: Tính
$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$
- Bước 6: Phương trình hồi quy: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$



II. CÁC GIẢ THIẾT CƠ BẢN

II.1 *Các giả thiết cơ bản của phương pháp OLS*

Giả thiết 1. Biến giải thích X là phi ngẫu nhiên, giá trị của nó là xác định.

Giả thiết 2. Trung bình ảnh hưởng các yếu tố ngẫu nhiên đến biến phụ thuộc bằng 0:

$$E(U_i) = E(U|X_i) = 0 \rightarrow E(Y_i) = E(Y|X_i) = \beta_1 + \beta_2 X_i$$

Giả thiết 3. Các yếu tố ngẫu nhiên có phương sai bằng nhau (phương sai thuần nhất):

$$Var(U_i) = Var(U|X_i) = \sigma^2 \forall i \rightarrow Var(Y_i) = Var(Y|X_i) = \sigma^2 \forall i$$

II. CÁC GIẢ THIẾT CƠ BẢN

II.1 Các giả thiết cơ bản của phương pháp OLS

Giả thiết 4. Không có sự tương quan giữa các U_i :

$$\text{Cov}(U_i, U_j) = 0 \quad \forall i \neq j$$

Giả thiết 5. Không có sự tương quan giữa U_i và X_i :

$$\text{Cov}(U_i, X_i) = 0 \quad \forall i$$

Định lý Gauss-Markov:

Với các giả thiết từ 1 đến 5, các ước lượng bình phương nhỏ nhất là các ước lượng tuyến tính, không chệch và có phương sai nhỏ nhất trong lớp các ước lượng tuyến tính không chệch.

Cụ thể:

- Ước lượng: $\widehat{\beta}_1 = \bar{Y} - \widehat{\beta}_2 \bar{X}$; $\widehat{\beta}_2 = \frac{\overline{XY} - \bar{Y} \bar{X}}{\overline{X^2} - \bar{X}^2}$
- Đặt $x_i = X_i - \bar{X}$; $y_i = Y_i - \bar{Y} \Rightarrow \widehat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$
- Phương sai, độ lệch chuẩn:

$$V(\widehat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}; V(\widehat{\beta}_1) = \frac{\overline{X^2}}{\sum_{i=1}^n x_i^2} \sigma^2; (V(U_i) = \sigma^2)$$

$$se(\widehat{\beta}_2) = \frac{\sigma}{\sqrt{\sum_{i=1}^n x_i^2}}; se(\widehat{\beta}_1) = \sigma \sqrt{\frac{\overline{X^2}}{\sum_{i=1}^n x_i^2}};$$

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

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - 2} \quad \hat{\sigma}: \text{sai số tiêu chuẩn của đường hồi quy (Standard Error)}$$



Xác định phương sai của các hệ số hồi quy

- Bước 1: Tìm $\bar{X}$, $\overline{X^2}$ ($\bar{X} = \sum_{i=1}^n X_i / n$)
- Bước 2: Tính cột giá trị x_i^2 ($x_i = X_i - \bar{X}$) $\rightarrow \sum_{i=1}^n x_i^2$
- Bước 3: - σ^2 đã biết: $V(\widehat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$; $V(\widehat{\beta}_1) = \frac{\overline{X^2}}{\sum_{i=1}^n x_i^2} \sigma^2 (*)$
- σ^2 chưa biết:
 - + Tìm $\widehat{\beta}_1, \widehat{\beta}_2$ (xem lại cách xác định pt hồi quy)
 - + Tính cột giá trị $\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i \rightarrow e_i^2$ ($e_i = Y_i - \widehat{Y}_i$) $\rightarrow \sum_{i=1}^n e_i^2$
 - + Tính ước lượng của σ^2 : $\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - 2}$
 - + Thay σ^2 bởi $\hat{\sigma}^2$ trong các công thức (*)

II. CÁC GIẢ THIẾT CƠ BẢN

II.2 Giả thiết về phân phối xác suất của U_i

Giả thiết 6. Sai số ngẫu nhiên U_i có phân phối chuẩn, tức là:

$$U_i \sim N(0, \sigma^2) \Leftrightarrow Z = \frac{U_i}{\sigma} \sim N(0, 1)$$

Với các giả thiết 1-6, mô hình hồi quy 2 biến có các tính chất sau:

$$1. \hat{\beta}_1 \sim N(\beta_1, \text{Var}(\hat{\beta}_1)) \quad 3. \chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$$

$$2. \hat{\beta}_2 \sim N(\beta_2, \text{Var}(\hat{\beta}_2)) \quad 4. Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$$

5. Các ước lượng OLS $\widehat{\beta}_1, \widehat{\beta}_2$ của β_1, β_2 là các ước lượng hiệu quả.

III. KHOẢNG TIN CẬY VÀ KIỂM ĐỊNH GIẢ THUYẾT

III.1 Khoảng tin cậy của β_j ($j=1$ hoặc 2)

Cơ sở: $T_{\beta_j} = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim T(n-2) \quad j=1..2$ Phân phối Student
(n-2) bậc tự do

$$\text{Với } se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n x_i^2}}; \quad se(\hat{\beta}_1) = \hat{\sigma} \sqrt{\frac{\overline{X^2}}{\sum_{i=1}^n x_i^2}} \quad \left(\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} \right)$$

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

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

Khoảng tin cậy
($1-\alpha$) của β_j
($j = 1..2$) cho bởi:

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

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 ($j=1;2$)

- Bước 1: 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 2: Với α cho trước, tra bảng phân phối Student tìm $t_{\frac{\alpha}{2}}(n-2)$
- Bước 3: Khoảng tin cậy $(1-\alpha)$ của β_j cho bởi:

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

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

- 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 $t \sim T(n-2)$
- 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-2)$
Phía trái	$\beta_j = \beta_j^*$	$\beta_j < \beta_j^*$	$t < -t_\alpha(n-2)$
Phía phải	$\beta_j = \beta_j^*$	$\beta_j > \beta_j^*$	$t > t_\alpha(n-2)$

VD

Các bước kiểm định giả thuyết của β_j ($j=1;2$) 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 $t_{\frac{\alpha}{2}}(n-2)$ (kiểm định 2 phía) hoặc $t_{\alpha}(n-2)$ (kiểm định 1 phía)
- Bước 5: So sánh t và $t_{\frac{\alpha}{2}}(n-2)$ (hoặc $t_{\alpha}(n-2)$), dựa vào bảng kết quả đưa ra quyết định.

III.3 Khoảng tin cậy của σ^2

- Cơ sở: $\chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2) \rightarrow$ Phân phối khi bình phương (n-2) bậc tự do

- Khoảng tin cậy $(1-\alpha)$ của σ^2 được xác định từ:

$$P\left(\chi^2_{1-\frac{\alpha}{2}}(n-2) < \frac{(n-2)\hat{\sigma}^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(n-2)\right) = 1-\alpha$$

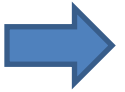
$\downarrow$
 Giá trị được cho trong [bảng phân phối](#) khi bình phương

$\downarrow$

- Khoảng tin cậy $(1-\alpha)$ của σ^2 là: $\left(\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}(n-2)}; \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}(n-2)}\right)$



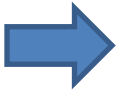
BẢNG PHÂN PHỐI $\chi^2_\alpha(k)$



Bậc tự do	α				
	0,1	0,05	0,025	0,01	0,005
1	2,706	3,841	5,024	6,635	7,879
2	4,605	5,991	7,378	9,210	10,597
3	6,251	7,815	9,348	11,345	12,838
4	7,779	9,488	11,143	13,277	14,860
5	9,236	11,070	12,833	15,086	16,750
6	10,645	12,592	14,449	16,812	18,548
7	12,017	14,067	16,013	18,475	20,278
8	13,362	15,507	17,535	20,090	21,955
9	14,684	16,919	19,023	21,666	23,589
10	15,987	18,307	20,483	23,209	25,188
11	17,275	19,675	21,920	24,725	26,757
12	18,549	21,026	23,337	26,217	28,300
13	19,812	22,362	24,736	27,688	29,819
14	21,064	23,685	26,119	29,141	31,319
15	22,307	24,996	27,488	30,578	32,801

Bậc tự do	α				
	0,1	0,05	0,025	0,01	0,005
16	23,542	26,296	28,845	32,000	34,267
17	24,769	27,587	30,191	33,409	35,718
18	25,989	28,869	31,526	34,805	37,156
19	27,204	30,144	32,852	36,191	38,582
20	28,412	31,410	34,170	37,566	39,997
21	29,615	32,671	35,479	38,932	41,401
22	30,813	33,924	36,781	40,289	42,796
23	32,007	35,172	38,076	41,638	44,181
24	33,196	36,415	39,364	42,980	45,559
25	34,382	37,652	40,646	44,314	46,928
26	35,563	38,885	41,923	45,642	48,290
27	36,741	40,113	43,195	46,963	49,645
28	37,916	41,337	44,461	48,278	50,993
29	39,087	42,557	45,722	49,588	52,336
30	40,256	43,773	46,979	50,892	53,672

BẢNG PHÂN PHỐI $\chi^2_\alpha(k)$



Bậc tự do	α				
	0,9	0,95	0,975	0,99	0,995
1	0,016	0,004	0,001	0,000	0,000
2	0,211	0,103	0,051	0,020	0,010
3	0,584	0,352	0,216	0,115	0,072
4	1,064	0,711	0,484	0,297	0,207
5	1,610	1,145	0,831	0,554	0,412
6	2,204	1,635	1,237	0,872	0,676
7	2,833	2,167	1,690	1,239	0,989
8	3,490	2,733	2,180	1,646	1,344
9	4,168	3,325	2,700	2,088	1,735
10	4,865	3,940	3,247	2,558	2,156
11	5,578	4,575	3,816	3,053	2,603
12	6,304	5,226	4,404	3,571	3,074
13	7,042	5,892	5,009	4,107	3,565
14	7,790	6,571	5,629	4,660	4,075
15	8,547	7,261	6,262	5,229	4,601

Bậc tự do	α				
	0,9	0,95	0,975	0,99	0,995
16	9,312	7,962	6,908	5,812	5,142
17	10,085	8,672	7,564	6,408	5,697
18	10,865	9,390	8,231	7,015	6,265
19	11,651	10,117	8,907	7,633	6,844
20	12,443	10,851	9,591	8,260	7,434
21	13,240	11,591	10,283	8,897	8,034
22	14,041	12,338	10,982	9,542	8,643
23	14,848	13,091	11,689	10,196	9,260
24	15,659	13,848	12,401	10,856	9,886
25	16,473	14,611	13,120	11,524	10,520
26	17,292	15,379	13,844	12,198	11,160
27	18,114	16,151	14,573	12,879	11,808
28	18,939	16,928	15,308	13,565	12,461
29	19,768	17,708	16,047	14,256	13,121
30	20,599	18,493	16,791	14,953	13,787

Cách xác định khoảng tin cậy của σ^2

- Bước 1: Tính $\hat{\sigma}^2$ (xem lại cách xác định phương sai của các hệ số hồi quy)
- Bước 2: Với α cho trước, tra bảng phân phối χ^2 tìm $\chi^2_{1-\frac{\alpha}{2}}(n-2)$, $\chi^2_{\frac{\alpha}{2}}(n-2)$
- Bước 3: Khoảng tin cậy $(1-\alpha)$ của σ^2 cho bởi:

$$\left(\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}(n-2)} ; \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}(n-2)} \right)$$

III.4 Kiểm định giả thuyết của σ^2

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

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

$$(H_1 : \sigma^2 > \sigma_0^2, H_1 : \sigma^2 < \sigma_0^2)$$
- Tiêu chuẩn kiểm định: $\chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma_0^2}$
 Nếu H_0 đúng thì $\chi^2 \sim \chi^2(n-2)$
- Bảng kết quả: ($\chi_\alpha^2(k)$ được cho trong bảng phân phối χ^2)

Loại giả thiết	Giả thiết H_0	Giả thiết đối H_1	Miên bác bỏ H_0
Hai phía	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\begin{cases} \chi^2 > \chi_{\frac{\alpha}{2}}^2(n-2) \\ \chi^2 < \chi_{1-\frac{\alpha}{2}}^2(n-2) \end{cases}$
Phía phải	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi_\alpha^2(n-2)$
Phía trái	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi_{1-\alpha}^2(n-2)$

Các bước kiểm định giả thuyết của σ^2 đối với σ_0^2

- Bước 1: Biểu diễn các giả thuyết H_0 , H_1 dưới dạng toán học.
- Bước 2: Tính $\hat{\sigma}^2$ (xem lại cách xác định phương sai của các hệ số hồi quy)
- Bước 3: Tính $\chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma_0^2}$
- Bước 4: Với α cho trước, tra bảng phân phối χ^2 tìm

$$\chi_{\frac{\alpha}{2}}^2(n-2) \text{ và } \chi_{1-\frac{\alpha}{2}}^2(n-2)$$

- Bước 5: So sánh χ^2 với $\chi_{\frac{\alpha}{2}}^2(n-2)$ và $\chi_{1-\frac{\alpha}{2}}^2(n-2)$, dựa vào bảng kết quả đưa ra quyết định.

IV. PHÂN TÍCH PHƯƠNG SAI VÀ ĐỘ PHÙ HỢP CỦA HÀM HỒI QUY MẪU

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 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 của hàm hồi quy

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

$$r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = \widehat{\beta}_2^2 \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n y_i^2} \in [0;1] \quad \longrightarrow \quad \text{Thước đo độ phù hợp của hàm hồi quy}$$

- Định nghĩa 2: Hệ số tương quan r được xác định:

$$\left\{ \begin{array}{l} \text{- Độ lớn:} \\ |r| = \sqrt{r^2} = \sqrt{\frac{ESS}{TSS}} = \sqrt{1 - \frac{RSS}{TSS}} \\ \text{- Dấu: cùng dấu với } \beta_2 \end{array} \right. \quad \longrightarrow \quad \text{Thước đo sự **phụ thuộc tuyến tính** giữa các biến X và Y, không có ý nghĩa trong các quan hệ phi tuyến.}$$

➤ Một số tính chất của hệ số tương quan r :

1. $|r| \leq 1$

- $r > 0$ nếu X, Y quan hệ cùng chiều $\leftrightarrow \hat{\beta}_2 > 0$

- $r < 0$ nếu X, Y quan hệ ngược chiều $\leftrightarrow \hat{\beta}_2 < 0$

2. Tính đối xứng: $r(X, Y) = r(Y, X)$

3. Cho a, b, c, d với $ab > 0$. Đặt $X^* = aX + c$; $Y^* = bY + d$.

Thế thì: $r(X^*, Y^*) = r(X, Y)$

4. Nếu X, Y độc lập với nhau thì $r(X, Y) = 0$.

Ngược lại không đúng.



Cách tìm hệ số xác định của hàm hồi quy

- Bước 1: Tìm $\hat{\beta}_2, \bar{X}, \bar{Y}$ (xem lại cách xác định phương trình hồi quy)
- Bước 2: Tính các cột $x_i = X_i - \bar{X}; y_i = Y_i - \bar{Y}$ suy ra $\sum_{i=1}^n x_i^2$ và $\sum_{i=1}^n y_i^2$
- Bước 3: Tính $r^2 = \widehat{\beta}_2^2 \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n y_i^2}$

IV.2

Kiểm định sự phù hợp của mô hình

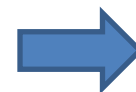
- Thực chất là kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0 : r^2 = 0 \longrightarrow \text{Mô hình không phù hợp} \\ H_1 : r^2 > 0 \longrightarrow \text{Mô hình phù hợp} \end{cases}$$

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

Nếu H_0 đúng thì $F \sim F(1, n-2)$: phân phối Fisher

- Với α cho trước, xác định được $F_\alpha(1, n-2)$ (tra bảng phân phối Fisher) thỏa: $P(F < F_\alpha(1, n-2)) = 1 - \alpha$
- Điều kiện bác bỏ $H_0 : F > F_\alpha(1, n-2)$: hàm hồi quy là phù hợp.



Các bước kiểm định sự phù hợp của mô hình

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

IV. PHÂN TÍCH HỒI QUY VÀ DỰ BÁO

V.1

Dự báo giá trị trung bình

- Thống kê $T = \frac{\hat{Y}_0 - E(Y / X_0)}{se(\hat{Y}_0)} \sim T(n-2)$ với $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$

$$se(\hat{Y}_0) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n x_i^2} \right)}$$

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

$$P\left(|T| < t_{\frac{\alpha}{2}}(n-2)\right) = 1 - \alpha$$

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

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



Các bước dự báo giá trị trung bình có điều kiện

- Bước 1: Tính $\widehat{\beta}_1, \widehat{\beta}_2, \hat{\sigma}^2, \sum_{i=1}^n x_i^2, \bar{X}$ (xem lại cách xác định phương trình hồi quy, phương sai của các hệ số hồi quy).
- Bước 2: Tính $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0, se(\hat{Y}_0) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n x_i^2} \right)}$
- Bước 3: Với α cho trước, tra bảng phân phối Student tìm $t_{\frac{\alpha}{2}}(n-2)$
- Bước 4: Khoảng tin cậy $(1-\alpha)$ của $E(Y/X_0)$ cho bởi:

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

V.1

Dự báo giá trị cá biệt

- Thống kê $T = \frac{Y_0 - \hat{Y}_0}{se(Y_0 - \hat{Y}_0)} \sim T(n-2)$ với $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$
$$se(Y_0 - \hat{Y}_0) = \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n x_i^2} \right)}$$
- Với α cho trước, xác định được $t_{\frac{\alpha}{2}}(n-2)$ dựa vào bảng phân phối Student thỏa mãn:

$$P\left(|T| < t_{\frac{\alpha}{2}}(n-2)\right) = 1 - \alpha$$

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

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



Các bước dự báo giá trị cá biệt

- Bước 1: Tính $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}^2, \sum_{i=1}^n x_i^2, \bar{X}$ (xem lại cách xác định phương trình hồi quy, phương sai của các hệ số hồi quy).

- Bước 2: Tính $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0, se(Y_0 - \hat{Y}_0) = \sqrt{\widehat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n x_i^2} \right)}$
- Bước 3: Với α cho trước, tra bảng

phân phối Student tìm $t_{\frac{\alpha}{2}}(n-2)$

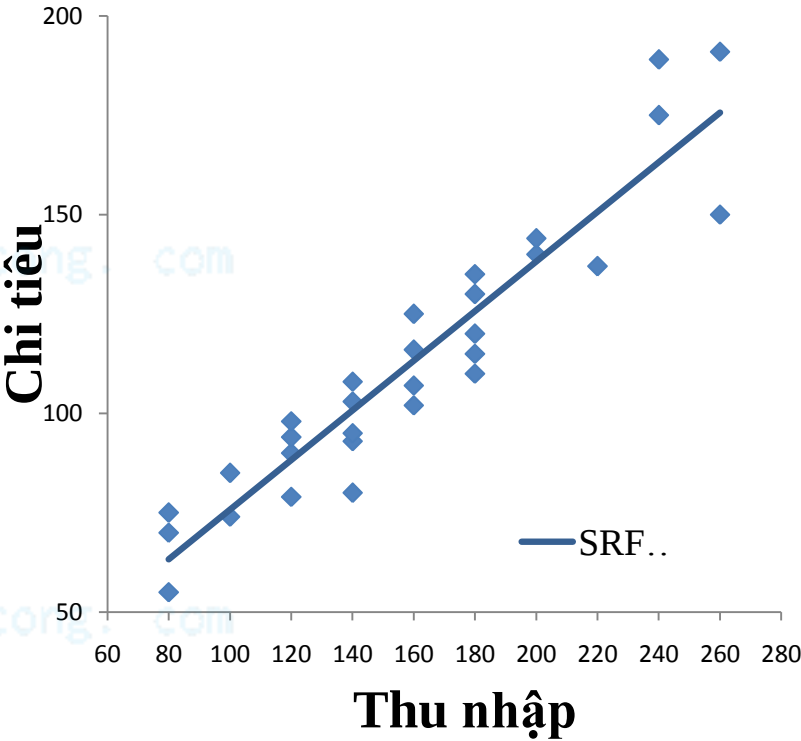
- Bước 4: Khoảng tin cậy $(1-\alpha)$ của $E(Y/X_0)$ cho bởi:

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

STT	Thu nhập X_i	Chi tiêu Y_i	STT	Thu nhập X_i	Chi tiêu Y_i
1	80	55	16	160	107
2	80	70	17	160	125
3	80	75	18	160	116
4	100	74	19	180	110
5	100	85	20	180	115
6	120	79	21	180	120
7	120	90	22	180	130
8	120	94	23	180	135
9	120	98	24	200	140
10	140	93	25	200	144
11	140	80	26	220	137
12	140	95	27	240	175
13	140	103	28	240	189
14	140	108	29	260	150
15	160	102	30	260	191

Cho thu nhập và chi tiêu trong 1 tuần (đơn vị tính: \$) của 30 hộ gia đình trong bảng bên.

Đồ thị:



$$\text{SRF: } \widehat{CT} = \widehat{\beta}_1 + \widehat{\beta}_2 * TN$$

Ước lượng hệ số của mô hình hồi quy

$$\widehat{\beta}_2 = \frac{\overline{XY} - \bar{Y} \bar{X}}{\overline{X^2} - \bar{X}^2}$$

$$\widehat{\beta}_1 = \bar{Y} - \widehat{\beta}_2 \bar{X}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{4780}{30} = 159,333 \quad \bar{Y} = \frac{3385}{30} = 112,833$$

$$\overline{X^2} = \frac{838000}{30} = 27933,333 \quad \overline{XY} = \frac{587040}{30} = 19568$$

n	30
$\sum X_i$	4780
$\sum Y_i$	3385
$\sum X_i^2$	838000
$\sum X_i Y_i$	587040

$$\widehat{\beta}_2 = \frac{19568 - 159,333 * 112,833}{27933,333 - 159,333^2} = 0,624;$$

$$\widehat{\beta}_1 = 112,833 - 0,624 * 159,333 = 13,409$$

$$\Rightarrow \text{SRF: } \widehat{CT} = 13,409 + 0,624 * TN$$

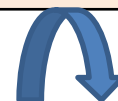
$\widehat{\beta}_1 = 13,409$: ảnh hưởng của các nhân tố khác ngoài thu nhập tới chi tiêu.
Người không có thu nhập vẫn chi tiêu ở mức tối thiểu 13,409\$ mỗi tuần.

$\widehat{\beta}_2 = 0,624$: ảnh hưởng của thu nhập tới chi tiêu: khi thu nhập tăng thêm 1\$ thì chi tiêu tăng trung bình 0,624\$



STT	X_i	Y_i	$\hat{Y}_i = 13,409 + 0,624 * X_i$
1	80	55	63,329
2	80	70	63,329
3	80	75	63,329
4	100	74	75,809
5	100	85	75,809
6	120	79	88,289
7	120	90	88,289
8	120	94	88,289
9	120	98	88,289
10	140	93	100,769
11	140	80	100,769
12	140	95	100,769
13	140	103	100,769
14	140	108	100,769
15	160	102	113,249

STT	X_i	Y_i	$\hat{Y}_i = 13,409 + 0,624 * X_i$
16	160	107	113,249
17	160	125	113,249
18	160	116	113,249
19	180	110	125,729
20	180	115	125,729
21	180	120	125,729
22	180	130	125,729
23	180	135	125,729
24	200	140	138,209
25	200	144	138,209
26	220	137	150,689
27	240	175	163,169
28	240	189	163,169
29	260	150	175,649
30	260	191	175,649



Phương sai, độ lệch chuẩn của các hệ số hồi quy

	X_i	X_i^2	e_i^2	x_i^2
Tổng	4780	838000	3871,861	76386,667
TB	159,333	27933,333	129,062	2546,222

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - 2} = \frac{3871,861}{30 - 2} \approx 138,281$$

$$V(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2} = \frac{138,281}{76386,667} \approx 0,002$$

$$V(\hat{\beta}_1) = \frac{\bar{X}^2}{\sum_{i=1}^n x_i^2} \hat{\sigma}^2 = 27933,333 * \frac{138,281}{76386,667} \approx 50,567$$

$$se(\hat{\beta}_2) = \sqrt{V(\hat{\beta}_2)} \approx 0,045 \quad se(\hat{\beta}_1) = \sqrt{V(\hat{\beta}_1)} \approx 7,111$$

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

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$x_i = X_i - \bar{X}$$



Khoảng tin cậy của các hệ số hồi quy

- Lấy $\alpha = 5\%$
- Khoảng tin cậy 95% của β_1 là:

$$\left(\hat{\beta}_1 - t_{\frac{\alpha}{2}}(n-2) * se(\hat{\beta}_1); \hat{\beta}_1 + t_{\frac{\alpha}{2}}(n-2) * se(\hat{\beta}_1) \right)$$

$$(13,409 - 2,048 * 7,111; 13,409 + 2,048 * 7,111)$$

$$(-1,154; 27,972)$$

với $\hat{\beta}_1 = 13,409; se(\hat{\beta}_1) = 7,111; \hat{\beta}_2 = 0,624; se(\hat{\beta}_2) = 0,045$

$$t_{\frac{\alpha}{2}}(n-2) = t_{0,025}(28) = 2,048$$

- Khoảng tin cậy 95% của β_2 là: (0,532; 0,716)

$$(0,624 - 2,048 * 0,045; 0,624 + 2,048 * 0,045)$$



Kiểm định giả thuyết của các hệ số hồi quy

- Giả sử có ý kiến cho rằng, thu nhập không ảnh hưởng (hoặc ảnh hưởng không đáng kể) đến chi tiêu. Để xem nhận định này đúng hay không, ta phải kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_2 = 0 \longrightarrow \text{thu nhập không có ảnh hưởng đến chi tiêu} \\ H_1 : \beta_2 \neq 0 \longrightarrow \text{thu nhập có ảnh hưởng đến chi tiêu} \end{cases}$$

- Tiêu chuẩn kiểm định: $t = \frac{\hat{\beta}_2 - \beta_2^*}{se(\hat{\beta}_2)} = \frac{0,624 - 0}{0,045} = 13,867$

với $\hat{\beta}_2 = 0,624; se(\hat{\beta}_2) = 0,045; \beta_2^* = 0$

- Với $\alpha=5\%$, kiểm định 2 phía, $t_{\frac{\alpha}{2}}(n-2) = t_{0,025}(28) = 2,048$

$$\rightarrow |t| > t_{\frac{\alpha}{2}}(n-2) \longrightarrow \text{bác bỏ } H_0: \text{thu nhập ảnh hưởng đến chi tiêu}$$

- Giả sử có ý kiến cho rằng, thu nhập ảnh hưởng thuận đến chi tiêu (thu nhập tăng kéo theo chi tiêu tăng). Để xem nhận định này phù hợp hay không, ta phải kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_2 = 0 \longrightarrow \text{thu nhập không có ảnh hưởng đến chi tiêu} \\ H_1 : \beta_2 > 0 \longrightarrow \text{thu nhập có ảnh hưởng thuận đến chi tiêu} \end{cases}$$

- Tiêu chuẩn kiểm định: $t = \frac{\hat{\beta}_2 - \beta_2^*}{se(\hat{\beta}_2)} = \frac{0,624 - 0}{0,045} = 13,867$

với $\hat{\beta}_2 = 0,624$; $se(\hat{\beta}_2) = 0,045$; $\beta_2^* = 0$

- Với $\alpha = 5\%$, kiểm định phía phải, $t_\alpha(n-2) = t_{0,05}(28) = 1,701$

$\rightarrow t > t_\alpha(n-2) \longrightarrow$ bác bỏ H_0 : thu nhập có ảnh hưởng thuận đến chi tiêu

