

Chương 8

TƯƠNG QUAN VÀ HỒI QUY MẪU

Với một mẫu đồng thời của vectơ ngẫu nhiên $Z = (X, Y)$ ta có các ước lượng của các đặc trưng của Z , có đường xấp xỉ của đường hồi quy của Y theo X (hoặc X theo Y). Sau đây ta xét một vài vấn đề cụ thể.

8.1 HỆ SỐ TƯƠNG QUAN MẪU

8.1.1 Bảng tương quan mẫu

Một mẫu kích thước n đồng thời về X và Y có dạng bảng số liệu sau đây, gọi là bảng tương quan mẫu.

Bảng 8.1

$\begin{matrix} Y \\ X \end{matrix}$	y_1	y_2	\dots	y_h	n_i
x_1	n_{11}	n_{12}	\dots	n_{1h}	n_1
x_2	n_{21}	n_{22}	\dots	n_{2h}	n_2
\vdots	\vdots	\vdots	\dots	\vdots	\vdots
x_k	n_{k1}	n_{k2}	\dots	n_{kh}	n_k
m_j	m_1	m_2	\dots	m_h	$\Sigma = n$

Trong bảng 8.1 ở trên:

x_i ($i = \overline{1, k}$) - các giá trị mà X nhận

y_j ($j = \overline{1, h}$) - các giá trị mà Y nhận

n_i ($i = \overline{1, k}$) - số lần X nhận x_i

m_j ($j = \overline{1, h}$) - số lần Y nhận y_j

$n_{i,j}$ ($i = \overline{1, k}, j = \overline{1, h}$) - số lần đồng thời X nhận x_i và Y nhận y_j

ta có:

$$\sum_{i=1}^k n_i = \sum_{j=1}^h m_j = \sum_{i=1}^k \sum_{j=1}^h n_{ij} = n$$

$$\sum_{j=1}^h n_{ij} = n_i; \sum_{i=1}^k n_{ij} = n_{ij}$$

từ đó ta có các đặc trưng sau đây:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_i; \quad \bar{x}^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i; \quad \hat{s}_X^2 = \bar{x}^2 - (\bar{x})^2$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^h y_j m_j; \quad \bar{y}^2 = \frac{1}{n} \sum_{j=1}^h y_j^2 n_j; \quad \hat{s}_Y^2 = \bar{y}^2 - (\bar{y})^2$$

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^h n_{ij} x_i y_j$$

8.1.2 Hệ số tương quan mẫu

Ta gọi:
$$r_{XY} = \frac{\overline{xy} - \bar{x} \bar{y}}{\hat{s}_X \hat{s}_Y}$$

là hệ số tương quan mẫu giữa X và Y .

Hệ số tương quan mẫu là một ước lượng của hệ số tương quan (xem mục 3.4.5 chương 3).

Ví dụ 8.1 Số vốn đầu tư X và lợi nhuận Y trong một đơn vị thời gian của 100 quan sát, được số liệu:

$\begin{matrix} \backslash \\ Y \end{matrix}$	0,3	0,7	1,0	n_i
$\begin{matrix} X \\ \backslash \end{matrix}$				
1	20	10		30
2		30	10	40
3		10	20	30
m_j	20	50	30	$n = 100$

Tìm hệ số tương quan giữa X và Y .

Giải

Ta có:
$$\bar{x} = \frac{1}{100} (1.30 + 2.40 + 3.30) = 2,00$$

$$\bar{x}^2 = \frac{1}{100} (1^2.30 + 2^2.40 + 3^2.30) = 4,60$$

$$\hat{s}_x^2 = 4,6 - 2,00^2 = 0,6$$

$$\bar{y} = \frac{1}{100} (0,3.20 + 0,7.50 + 1,0.30) = 0,71$$

$$\bar{y}^2 = \frac{1}{100} (0,3^2 \cdot 20 + 0,7^2 \cdot 50 + 1,0^2 \cdot 30) = 0,563$$

$$\hat{s}_y^2 = 0,563 - 0,71^2 = 0,0589$$

$$\overline{xy} = \frac{1}{100} (0,3 \cdot 20 + 0,7 \cdot 10 + 1,4 \cdot 30 + 2 \cdot 10 + 2,1 \cdot 10 + 3 \cdot 20) = 1,56$$

từ đó hệ số tương quan mẫu là:

$$r_{XY} = \frac{1,56 - 2,00 \cdot 0,71}{\sqrt{0,60 \cdot 0,0589}} = 0,7447$$

8.2 ĐƯỜNG HỒI QUY

8.2.1 Đường hồi quy mẫu

Tiếp tục xét bảng 8.1. Với mỗi $i = \overline{1, k}$, đặt:

$$\bar{Y}_{x_i} = (\bar{Y} | X = x_i) = \frac{1}{n_i} \sum_{j=1}^h n_{ij} y_j$$

là trung bình mẫu của Y khi $X = x_i$.

Biểu diễn các điểm (x_i, \bar{Y}_{x_i}) lên mặt phẳng tọa độ và nối các điểm (x_i, \bar{Y}_{x_i}) và $(x_{i+1}, \bar{Y}_{x_{i+1}})$ bằng một đoạn thẳng ($i = \overline{1, k-1}$), ta được một đường gấp khúc, gọi là đường hồi quy mẫu Y theo X .

Đường hồi quy mẫu là một xấp xỉ của đường hồi quy, tức là đồ thị của hàm $f(x) = E(Y | X = x)$ (xem mục 3.4 chương 3).

Tương tự ta cũng có đường hồi quy mẫu X theo Y .

Ví dụ 8.2 Điểm kiểm tra chất lượng môn toán (X) và môn văn (Y) của 100 học sinh, có số liệu:

$\begin{matrix} Y \\ X \end{matrix}$	3	4	5	6	7	8	9	n_i
3	1	2						3
4		3	5	4				12
5			11	13	4			28
6				7	23	7		37
7					4	8	3	15
8						3	2	5
n_j	1	5	16	24	31	18	5	$n = 100$

Lập đường hồi quy mẫu Y theo X .

Giải Ta có:

$$\bar{Y}_3 = \frac{1}{3}(3.1 + 4.2) = 3,67$$

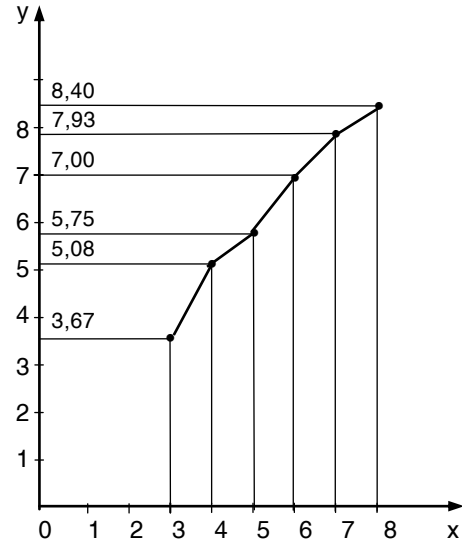
$$\bar{Y}_4 = \frac{1}{12}(4.3 + 5.5 + 6.4) = 5,08$$

$$\bar{Y}_5 = \frac{1}{28}(5.11 + 6.13 + 7.4) = 5,75$$

$$\bar{Y}_6 = \frac{1}{37}(6.7 + 7.23 + 8.7) = 7,00$$

$$\bar{Y}_7 = \frac{1}{15}(7.4 + 8.8 + 9.3) = 7,93$$

$$\bar{Y} = \frac{1}{5}(8.3 + 9.2) = 8,40$$



Ta có đường hồi quy mẫu Y theo X như trong hình vẽ.

8.2.2 Đường hồi quy tuyến tính mẫu

1- Phương trình đường hồi quy tuyến tính mẫu

Đường hồi quy tuyến tính mẫu Y theo X là đường thẳng có phương trình:

$$y = Bx + A$$

“gần” với đường hồi quy mẫu Y theo X nhất, theo nghĩa (B, A) là điểm cực tiểu của hàm:

$$Q(B, A) = \sum_{i=1}^k n_i [\bar{Y}_{x_i} - (Bx_i + A)]^2$$

Định lý 8.1 Phương trình hồi quy mẫu Y theo X là:

$$y = Bx + A$$

trong đó: $B = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_X^2}$; $A = \bar{y} - B\bar{x}$

Chứng minh $Q(B, A)$ là hàm hai biến có các đạo hàm riêng liên tục, bị chặn dưới bởi 0 nên Q có giá trị nhỏ nhất. Ta sẽ chỉ ra Q có một điểm dừng duy nhất, khi đó tọa độ của điểm dừng chính là (B, A) muốn tìm.

Ta có:

$$\begin{aligned} \frac{\partial Q}{\partial B} &= -2 \sum_{i=1}^k n_i (\bar{Y}_{x_i} - Bx_i - A) \cdot x_i = 2 \left(B \sum_{i=1}^k x_i^2 n_i + A \sum_{i=1}^k n_i x_i - \sum_{i=1}^k n_i x_i \bar{Y}_{x_i} \right) \\ &= 2 \left(nB\bar{x}^2 + nA\bar{x} - \sum_{i=1}^k \sum_{j=1}^k n_{ij} x_i x_j \right) = 2n \left(B\bar{x}^2 + A\bar{x} - \overline{xy} \right) \end{aligned}$$

$$\begin{aligned}\frac{\partial Q}{\partial A} &= -2 \sum_{i=1}^k n_i (\bar{Y}_{x_i} - Bx_i - A) \cdot x_i = 2 \left(B \sum_{i=1}^k n_i x_i + A \sum_{i=1}^k n_i - \sum_{i=1}^k n_i \bar{Y}_{x_i} \right) \\ &= 2 \left(nB\bar{x} + nA - \sum_{j=1}^k n_{ij} y_j \right) = 2n(B\bar{x} + A - \bar{y})\end{aligned}$$

từ đó ta có điểm dừng là nghiệm của hệ:

$$\begin{cases} \frac{\partial Q}{\partial B} = 0 \\ \frac{\partial Q}{\partial A} = 0 \end{cases} \Leftrightarrow \begin{cases} B\bar{x}^2 + A\bar{x} = \overline{xy} \\ B\bar{x} + A = \bar{y} \end{cases} \Leftrightarrow \begin{cases} B = \frac{\overline{xy} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} \\ A = \bar{y} - B\bar{x} \end{cases}$$

hay: $B = \frac{\overline{xy} - \bar{x}\bar{y}}{\bar{s}_X^2}; \quad A = \bar{y} - B\bar{x}$

Hoàn toàn tương tự ta có đường hồi quy tuyến tính mẫu X theo Y .

Ví dụ 8.3 Với số liệu như trong ví dụ 8.2, hãy tìm đường hồi quy tuyến tính Y theo X .

Giải Ta có:

$$\bar{x} = \frac{1}{100} (3.3 + 4.12 + 5.28 + 6.37 + 7.15 + 8.5) = 5,64$$

$$\bar{x}^2 = \frac{1}{100} (3^2.3 + 4^2.12 + 5^2.28 + 6^2.37 + 7^2.15 + 8^2.5) = 33,06$$

$$\bar{s}_X^2 = \bar{x}^2 - \bar{x}^2 = 33,6 - 5,64^2 = 1,25$$

$$\bar{y}^2 = \frac{1}{100} (3.1 + 4.5 + 5.16 + 6.24 + 7.31 + 8.18 + 9.5) = 6,53$$

$$\begin{aligned}\overline{xy} &= \frac{1}{100} (3.3.1 + 3.4.2 + 4.4.3 + 4.5.5 + 4.6.4 + 5.5.11 + 5.6.13 + 5.7.4 + \\ &\quad + 6.6.7 + 6.7.23 + 6.8.7 + 7.7.4 + 7.8.8 + 7.9.3 + 8.8.3 + 8.9.2) \\ &= 38,05\end{aligned}$$

từ đó: $B = \frac{38,05 - 5,64.6,53}{1,25} = 0,98; \quad A = 6,53 - 0,98.5,64 = 1,00$

Đường hồi quy tuyến tính mẫu Y theo X là:

$$y = 0,98x + 1,00$$

2- Ý nghĩa của đường hồi quy tuyến tính mẫu

Nếu X và Y có tương quan xấp xỉ tuyến tính thì đường hồi quy cho ta khả năng dự báo một cách đơn giản:

$$X = x_o \Rightarrow Y \approx y_o = Bx_o + A$$

Ví dụ 8.4 Với số liệu như trong ví dụ 8.2, ta có:

$$X = 9 \Rightarrow Y \approx 0,98.9 + 1,00 = 9,82$$

Điều này có ý nghĩa là, một học sinh có điểm toán 9 thì có khả năng điểm môn văn là 10.

BÀI TẬP

8.1 Điều tra độ mòn lưỡi dao của một xưởng kim khí, ta đo độ dày lưỡi dao và thu được bảng số liệu sau:

Y (chiều dày lưỡi dao: mm)	X (thời gian sử dụng: ngày)		
	2	3	5
2,6		1	2
2,8	1	2	
3,0	3	2	

- a) Tại sao nói X và Y là phụ thuộc tương quan?
b) Viết phương trình đường hồi quy tuyến tính của Y đối với X .
- 8.2** Tính hệ số tương quan và tìm hàm hồi quy ước lượng dạng tuyến tính của Y đối với X dựa vào số liệu cho trong bảng tương quan sau đây:

x_i	1	2	2	3	3	4
y_i	3	4	5	5	6	7
n_i	3	2	1	2	1	1

8.3 Tìm hàm ước lượng của hàm hồi quy tuyến tính của Y đối với X dựa vào các bảng số liệu sau:

- a)
- | Y | X | | | |
|----|-----|------|-----|------|
| | 0,1 | 0,15 | 0,2 | 0,25 |
| 10 | 1 | | | |
| 12 | 2 | 3 | 2 | 1 |
| 14 | 1 | 4 | 2 | 4 |
- b)
- | Y | X | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 5 | | | | 2 |
| 4 | | | 1 | 2 |
| 3 | | 2 | 2 | |

2		1	2	
1	2	1		

8.4 Theo dõi trọng lượng Y và tháng tuổi X của một loại con giống thu được bảng số liệu sau:

X	Y			
	5	6	7	9
1	8	2		
2	1	6	4	4
3			8	7
4			5	5

a) Hãy vẽ đường hồi quy thực nghiệm của Y đối với X

b) Tìm hàm hồi quy ước lượng dạng tuyến tính của Y đối với X .

8.5 Hãy tìm hàm hồi quy ước lượng dạng tuyến tính biểu diễn sự phụ thuộc giữa mức suy giảm hàm lượng đường Y (%) và thời gian chế biến X trên cơ sở bảng số liệu sau:

X	Y				
	30	35	40	45	50
2	1				
4	1	3	1	2	
6		1	2	3	
8			2	1	1
10					2

8.6 Tìm hàm hồi quy ước lượng dạng tuyến tính của Y đối với X dựa vào các bảng số liệu sau đây:

a)

X	Y			
	10	20	30	40
0,4	3		6	7
0,6		2	6	4
0,8	3	19		

b)

X	Y		
	15	20	25
1			1
2		1	1

3	1	2	
4	4		

8.7 Bảng dưới đây chỉ kết quả thu hoạch Y theo lượng phân bón X của một loại hoa màu trên 100 thửa ruộng:

Y	X				
	1	2	3	4	5
14	10	8			
15		12	7		
16			28	6	
17				8	9
18					12

- a) Hãy vẽ đường hồi quy thực nghiệm của Y đối với X
- b) Tìm hệ số tương quan mẫu
- c) Tìm phương trình hồi quy tuyến tính của Y đối với X .
- 8.8 Nghiên cứu mối liên hệ giữa X là số tiền đầu tư cho việc phòng bệnh tính theo đầu người và Y là tỉ lệ người mắc bệnh ở 50 địa phương ta thu được bảng tương quan thực nghiệm sau đây:

X (đồng)	Y (%)				
	2	2,5	3	3,5	4
100				2	3
200			3	6	2
300		4	6	3	
400	1	6	4	1	
500	6	3			

- a) Tìm phương trình hồi quy tuyến tính của Y đối với X qua mẫu trên
- b) Tìm hệ số tương quan tuyến tính
- c) Nếu năm sau đầu tư cho phòng bệnh là 600 đ/người thì tỷ lệ mắc bệnh là bao nhiêu phần trăm?
- 8.9 Tìm hàm ước lượng của hàm hồi quy tuyến tính của Y đối với X và của X đối với Y dựa vào các bảng tương quan sau đây:

a)

X	Y		
	10	20	30
1			1
2		1	1

3	1	2	
4	4		

b)

X	Y			
	1	2	3	4
10	8	3		
12	1	7	1	
14	1		9	10

8.10 Cho bảng tương quan:

Y	X				
	1	2	3	4	5
1	3				3
2	2	2		2	2
3		1	2	1	
4			2		

a) Vẽ đường hồi quy thực nghiệm của Y đối với X b) Tính hệ số tương quan mẫu r c) Từ các kết quả trên hãy nhận xét về dạng hàm hồi quy của Y đối với X .8.11 Cho bảng tương quan các số liệu về mức độ tăng năng suất lao động $X(\%)$ và tổng sản lượng của các xí nghiệp công nghiệp Y (% so với năm trước).

X	Y					
	85	95	105	115	125	135
85	1	1				
95	1	2	2			
105	1	3	6	5		
115		1	4	4	2	2
125			3	2	2	1
135				1	2	1
145					1	2

a) Vẽ đường hồi quy thực nghiệm của Y đối với X b) Tìm hàm ước lượng của hàm hồi quy tuyến tính của Y đối với X .

8.12 Các số liệu của việc phân tích 100 mẫu quặng sắt được cho ở bảng dưới. Hàm

lượng oxyt sắt X (%), hàm lượng tạp chất Y (%).

X	Y					
	3	9	15	21	27	33
35				1		1
45			1	5	4	5
55			2	18	10	2
65		6	14	2	2	
75		6	3			
85	4	8				
95	6					

a) Vẽ đường hồi quy thực nghiệm của Y đối với X

b) Ước lượng hàm hồi quy tuyến tính của Y đối với X .

8.13 Kết quả của việc theo dõi mối quan hệ giữa chiều cao và trọng lượng Y của học sinh, ta có bảng số liệu sau đây:

X (cm)	Y (cm)				
	24	27	30	33	36
120	1	3			
125		2	6	1	
130		1	5	5	
135		1	6	7	2
140			1	4	2
145				1	1
150					1

a) Tìm hệ số tương quan mẫu

b) Tìm ước lượng của hàm hồi quy tuyến tính của Y đối với X .

8.14 Nghiên cứu về sự ảnh hưởng của thu nhập X đối với mức độ tiêu dùng Y về một loại thực phẩm, người ta điều tra ở 200 gia đình và thu được bảng số liệu sau đây:

Y	X					
	10	20	30	40	50	60
15	5	7				
25		20	23			
35			30	47	2	
45			10	11	20	6
55				9	7	3

Giả thiết Y phụ thuộc tuyến tính vào X .

a) Hãy ước lượng hàm hồi quy tuyến tính của Y đối với X và đánh giá mức độ chặt chẽ của sự phụ thuộc đó

b) Hãy dự đoán mức tiêu dùng nếu thu nhập là 80 nghìn/đầu người.

8.15 Cho bảng tương quan:

X	Y				
	1,25	1,5	1,75	2	2,25
8			1	2	3
13			1	4	3
18		4	7	1	
23	2	7	5		
28	6	4			

a) Tìm hệ số tương quan mẫu

b) Tìm phương trình hồi quy tuyến tính mẫu.

TÀI LIỆU SƯU TẬP
BỞI HCMUT-CNCP