

1. HỆ SỐ TƯƠNG QUAN MẪU

1.1. Định nghĩa

- Hệ số tương quan mẫu r là số đo mức độ phụ thuộc tuyến tính giữa hai mẫu ngẫu nhiên **cùng cỡ** X và Y .
- Giả sử ta có mẫu ngẫu nhiên cỡ n về vector ngẫu nhiên (X, Y) là $(x_i, y_i); i = 1; 2; \dots; n$. Khi đó, hệ số tương quan mẫu r được tính theo công thức:

$$r = \frac{\overline{xy} - \bar{x}.\bar{y}}{\hat{s}_x.\hat{s}_y}; \quad \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i.$$

1.2. Tính chất

1) $-1 \leq r \leq 1$.

2) Nếu $r = 0$ thì X, Y không có quan hệ tuyến tính;

Nếu $r = \pm 1$ thì X, Y có quan hệ tuyến tính tuyệt đối.

3) Nếu $r < 0$ thì quan hệ giữa X, Y là giảm biến.

4) Nếu $r > 0$ thì quan hệ giữa X, Y là đồng biến.

VD 1. Kết quả đo lường độ cholesterol (Y) có trong máu của 10 đối tượng nam ở độ tuổi (X) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tính hệ số tương quan mẫu giữa X và Y .

➤ Chương 8. Bài toán tương quan & Hồi quy

Giải. Từ số liệu ở bảng trên, ta tính được:

$$\overline{xy} = \frac{20 \times 1,9 + \dots + 49 \times 4,0}{10} = 167,26;$$

$$\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i = 43,9; \quad \hat{s}_x = 13,5385;$$

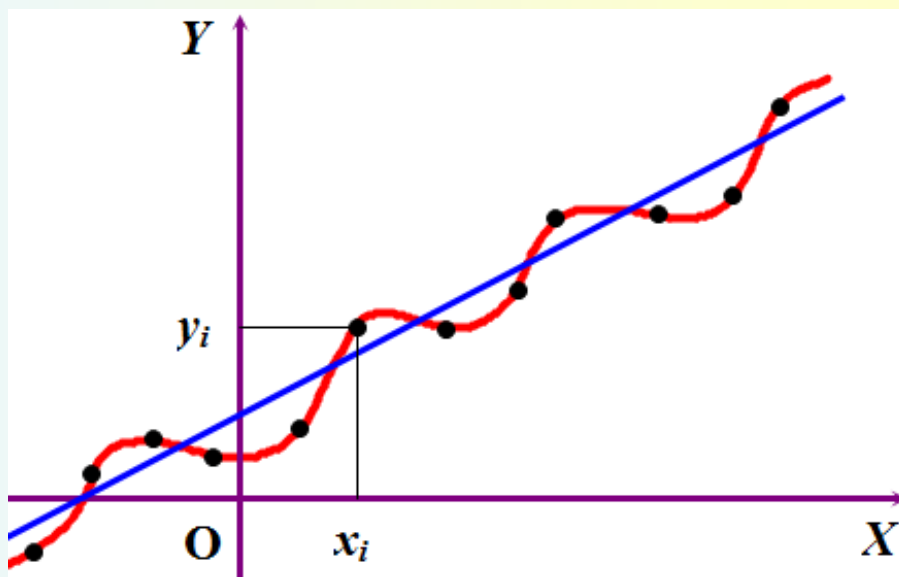
$$\overline{y} = \frac{1}{n} \sum_{i=1}^n y_i = 3,56; \quad \hat{s}_y = 0,8333.$$

$$\text{Vậy } r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\hat{s}_x \cdot \hat{s}_y} = 0,9729.$$

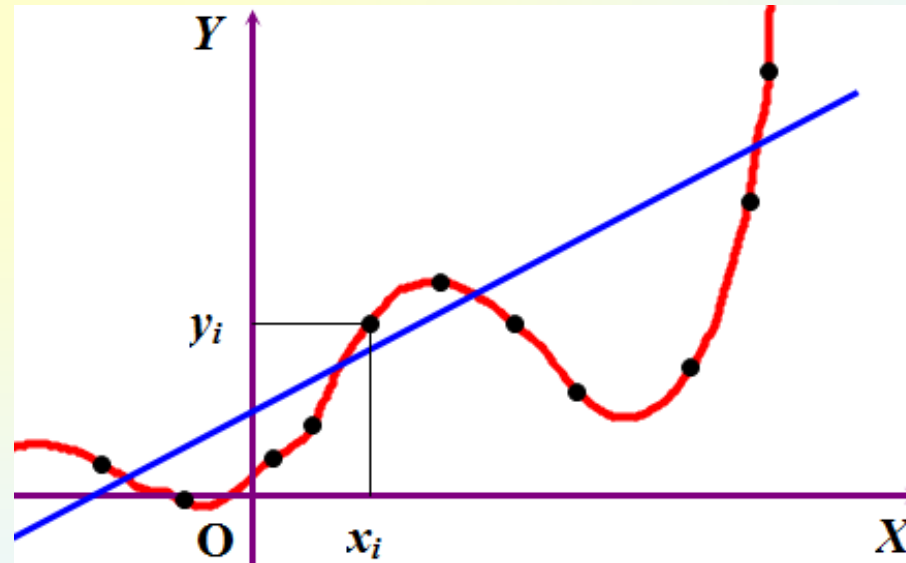
➤ Chương 8. Bài toán tương quan & Hồi quy

2. Đường hồi quy trung bình tuyến tính thực nghiệm

- Từ mẫu thực nghiệm về vector ngẫu nhiên (X, Y) , ta biểu diễn các cặp điểm (x_i, y_i) lên mpOxy. Khi đó, đường cong nối các điểm là đường cong phụ thuộc của Y theo X mà ta cần tìm (xem hình a), b)).



Hình a



Hình b

➤ Chương 8. Bài toán tương quan & Hồi quy

- Đường thẳng là *đường hồi quy* thực nghiệm xấp xỉ tốt nhất các điểm mẫu đã cho, cũng là xấp xỉ đường cong cần tìm. Trong hình *a*) ta thấy xấp xỉ tốt (phụ thuộc tuyến tính chặt), hình *b*) xấp xỉ không tốt.

2.1. Phương pháp bình phương bé nhất

- Khi có sự phụ thuộc tuyến tính tương đối chặt giữa hai biến ngẫu nhiên X và Y ta cần tìm biểu thức $a + bX$ xấp xỉ Y tốt nhất theo nghĩa cực tiểu sai số bình phương trung bình $E(Y - a - bX)^2$, phương pháp này được gọi là *bình phương bé nhất*.

- Với mỗi cặp điểm (x_i, y_i) thì sai số xấp xỉ là:

$$\varepsilon_i = y_i - (a + bx_i) \text{ (xem hình c)).}$$

➤ Chương 8. Bài toán tương quan & Hồi quy

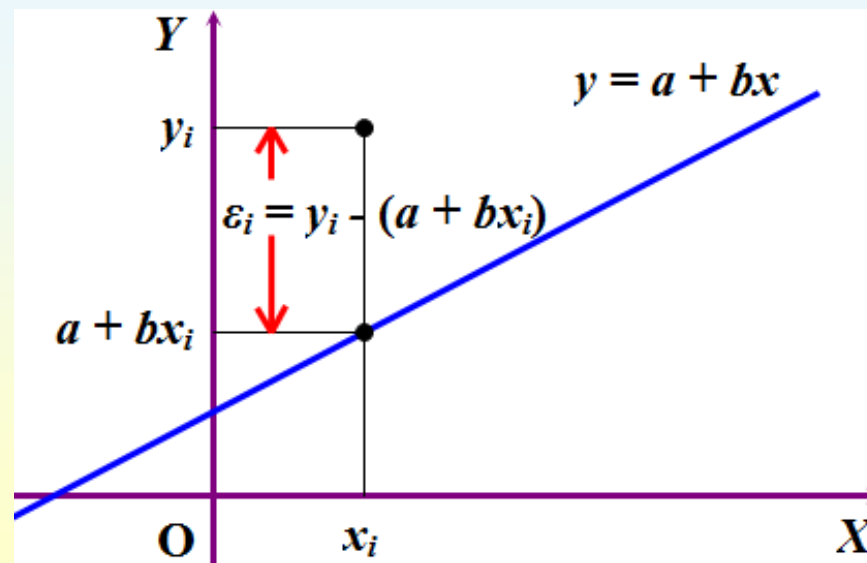
Ta đi tìm các ước lượng a, b

sao cho $\sum_{i=1}^n \varepsilon_i^2$ đạt cực tiểu.

$$\text{Đặt } Q = \sum_{i=1}^n \varepsilon_i^2$$

$$= \sum_{i=1}^n [y_i - (a + bx_i)]^2, \text{ ta có:}$$

$$\begin{cases} Q'_a = 0 \\ Q'_b = 0 \end{cases} \Leftrightarrow \begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i & (1) \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i & (2) \end{cases}$$



Hình c

➤ Chương 8. Bài toán tương quan & Hồi quy

$$(1) \Leftrightarrow a = \frac{1}{n} \sum_{i=1}^n y_i - b \cdot \frac{1}{n} \sum_{i=1}^n x_i = \bar{y} - b \cdot \bar{x}.$$

Thay a vào (2), ta được:

$$(\bar{y} - b \cdot \bar{x}) \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

$$\Leftrightarrow b \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x} \cdot \frac{1}{n} \sum_{i=1}^n x_i \right) = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{y} \cdot \frac{1}{n} \sum_{i=1}^n x_i \right)$$

$$\Leftrightarrow b \left(\overline{x^2} - \bar{x}^2 \right) = \overline{xy} - \bar{x} \cdot \bar{y} \Leftrightarrow b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2}.$$

➤ Chương 8. Bài toán tương quan & Hồi quy

- Vậy $b = \frac{\overline{xy} - \bar{x}.\bar{y}}{\hat{s}_x^2}$, $a = \bar{y} - b.\bar{x}$.

Đường hồi quy tuyến tính của Y theo X là:

$$y = a + bx.$$

- Tương tự: $b = \frac{\overline{xy} - \bar{x}.\bar{y}}{\hat{s}_y^2}$, $a = \bar{x} - b.\bar{y}$.

Đường hồi quy tuyến tính của X theo Y là:

$$x = a + by.$$

➤ Chương 8. Bài toán tương quan & Hồi quy

VD 2. Đo chiều cao (X : m) và khối lượng (Y : kg) của 5 học sinh nam, ta có kết quả:

X	1,45	1,60	1,50	1,65	1,55
Y	50	55	45	60	55

- 1) Tìm hệ số tương quan r .
- 2) Lập phương trình hồi quy tuyến tính của Y theo X .
- 3) Dự đoán nếu một học sinh cao 1,62m thì nặng khoảng bao nhiêu kg?

Giải. 1) $\bar{x} = 1,55$; $\hat{s}_x = 0,0707$; $\bar{y} = 53$; $\hat{s}_y = 5,099$;

$$\overline{xy} = 82,45 \Rightarrow r = \frac{82,45 - 1,55 \times 53}{0,0707 \times 5,099} = 0,8322.$$

➤ Chương 8. Bài toán tương quan & Hồi quy

$$2) b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = \frac{82,45 - 1,55 \times 53}{(0,0707)^2} = 60,0181;$$

$$a = \bar{y} - b\bar{x} = 53 - 60,0181 \times 1,55 = -40,0281.$$

$$\text{Vậy } y = -40,0281 + 60,0181x.$$

3) Học sinh cao 1,62m thì nặng khoảng:

$$y = -40,0281 + 60,0181 \times 1,62 = 57,2012 \text{ kg.}$$

➤ Chương 8. Bài toán tương quan & Hồi quy

VD 3. Số vốn đầu tư (X : triệu đồng) và lợi nhuận thu được (Y : triệu đồng) trong một đơn vị thời gian của 100 quan sát là:

$X \backslash Y$	0,3	0,7	1,0
1	20	10	
2		30	10
3		10	20

- 1) Lập phương trình hồi tuyến tính của X theo Y .
- 2) Dự đoán nếu muốn lợi nhuận thu được là 0,5 triệu đồng thì cần đầu tư bao nhiêu?

➤ Chương 8. Bài toán tương quan & Hồi quy

Giải. 1) Ta có $\bar{x} = 2$; $\hat{s}_x = 0,7746$; $\bar{y} = 0,71$;

$$\hat{s}_y = 0,2427; \overline{xy} = 1,56.$$

$$\Rightarrow b = \frac{\overline{xy} - \bar{x}\bar{y}}{\hat{s}_y^2} = \frac{1,56 - 0,71 \times 2}{(0,2427)^2} = 2,3768;$$

$$a = \bar{x} - b\bar{y} = 2 - 2,3768 \times 0,71 = 0,3125.$$

$$\text{Vậy } x = 0,3125 + 2,3768y.$$

2) Nếu muốn lợi nhuận thu được là 0,5 triệu thì cần đầu tư khoảng:

$$x = 0,3125 + 2,3768 \times 0,5 = 1,5009 \text{ triệu đồng.}$$

➤ Chương 8. Bài toán tương quan & Hồi quy

VD 4. Số thùng bia (Y : thùng) được bán ra phụ thuộc vào giá bán (X : triệu đồng/ thùng). Điều tra 100 đại lý về 1 loại bia trong một đơn vị thời gian có bảng số liệu:

$X \backslash Y$	100	110	120
0,150	5	15	30
0,160	10	25	
0,165	15		

- 1) Tính hệ số tương quan r .
- 2) Lập phương trình hồi tuyến tính của X theo Y .
- 3) Dự đoán nếu muốn bán được 115 thùng bia thì giá bán mỗi thùng cỡ bao nhiêu?

➤ Chương 8. Bài toán tương quan & Hồi quy

Giải. 1) $\bar{x} = 0,1558$; $\hat{s}_x = 0,006$; $\bar{y} = 110$; $\hat{s}_y = 7,746$;

$$\overline{xy} = 17,1 \Rightarrow r = \frac{17,1 - 0,1558 \times 110}{0,006 \times 7,746} = -0,8176.$$

$$2) b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_y^2} = \frac{17,1 - 0,1558 \times 110}{(7,746)^2} = -0,0006;$$

$$a = \bar{x} - b\bar{y} = 0,1558 + 0,0006 \times 110 = 0,2218.$$

$$\text{Vậy } x = 0,2218 - 0,0006y.$$

3) Nếu muốn bán được 115 thùng bia thì giá bán mỗi thùng khoảng:

$$x = 0,2218 - 0,0006 \times 115 = 0,1528 \text{ triệu đồng.}$$

1. Số liệu không có tần số

a) Máy tính f_x500MS, f_x570MS

VD 1. Bài toán cho ở dạng cặp (x_i, y_i) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tìm hệ số r , đường hồi quy Y theo X : $y = a + bx$.

Nhập số liệu:

MODE → **REG** → **LIN**

$X, Y \rightarrow M^+$

20, 1.9 → M^+

52, 4.0 → M^+

... ..

49, 4.0 → M^+

Xuất kết quả:

SHIFT → **2** → (dịch chuyển mũi tên phải 2 lần)
→ **1** (A chính là a trong phương trình)
→ **2** (B chính là b trong phương trình)
→ **3** (r chính là r).

Đáp số: $r = 0,9729$; $y = 0,9311 + 0,0599x$.

b) Máy tính f_x500ES, f_x570ES

Xét lại VD 1 ở trên.

Nhập số liệu:

SHIFT → **MODE** → dịch chuyển mũi tên tìm chọn
mục **Stat** → **2** (chế độ không tần số)

MODE → **3** (stat) → **2** ($A+Bx$) → (nhập các giá trị
của X, Y vào 2 cột)

➤ Sử dụng máy tính bỏ túi tìm đường hồi quy

X	Y
20	1.9
52	4.0
...	...
49	4.0

Xuất kết quả:

SHIFT → **1** → **7** → **1** (A chính là a trong phương trình)

SHIFT → **1** → **7** → **2** (B chính là b trong phương trình)

SHIFT → **1** → **7** → **3** (r chính là r trong phương trình).

2. Số liệu có tần số

a) Máy tính f_x500MS, f_x570MS

VD 2. Tìm hệ số r , đường hồi quy thực nghiệm Y theo

$X: y = a + bx$ với bài toán cho ở dạng bảng như sau:

➤ Sử dụng máy tính bỏ túi tính đường hồi quy

$Y \backslash X$	21	23	25
3	2		
4	5	3	
5		11	8

Nhập số liệu:

MODE → **REG** → **LIN**

$X, Y; n \rightarrow M^+$

21, 3; 2 → M^+

21, 4; 5 → M^+

...

...

25, 5; 8 → M^+

Xuất kết quả: làm như 1a).

Đáp số: $r = 0,7326; y = -2,6694 + 0,3145x.$

➤ Sử dụng máy tính bỏ túi tìm đường hồi quy

b) Máy tính f_x500ES , f_x570ES

Xét lại VD 2 ở trên

Nhập số liệu:

SHIFT → **MODE** → dịch chuyển mũi tên tìm chọn

Mục **Stat** → **1** (chế độ có tần số)

MODE → **3** (stat) → **2** ($A+Bx$) → (nhập các giá trị của X , Y , tần số vào 3 cột)

X	Y	FREQ
21	3	2
21	4	5
...
25	5	8

Xuất kết quả: làm như 1b).

.....**Hết**.....