

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:

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

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:

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

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

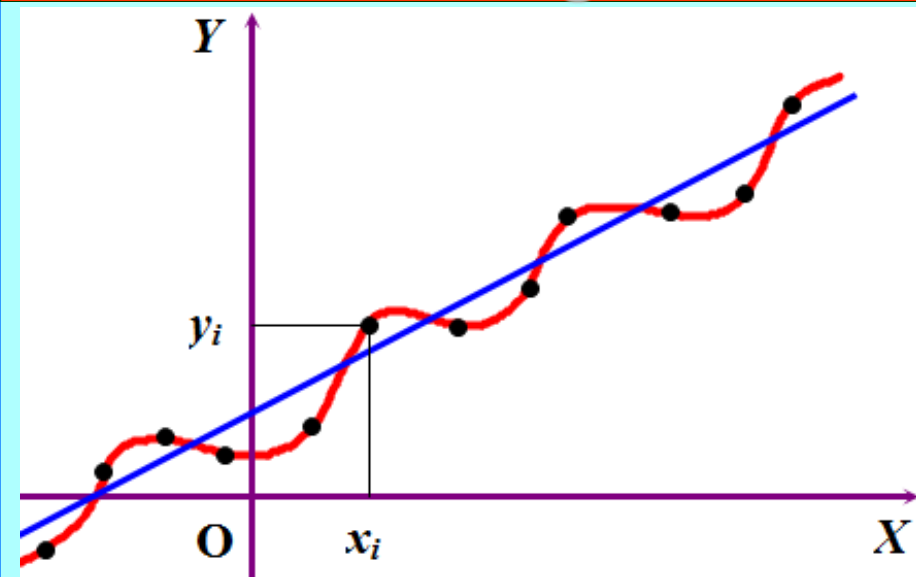
$$\overline{y} = \frac{1}{n} \sum_{i=1}^n y_i = 3,56; \quad \hat{s}_y = 0,8333.$$

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

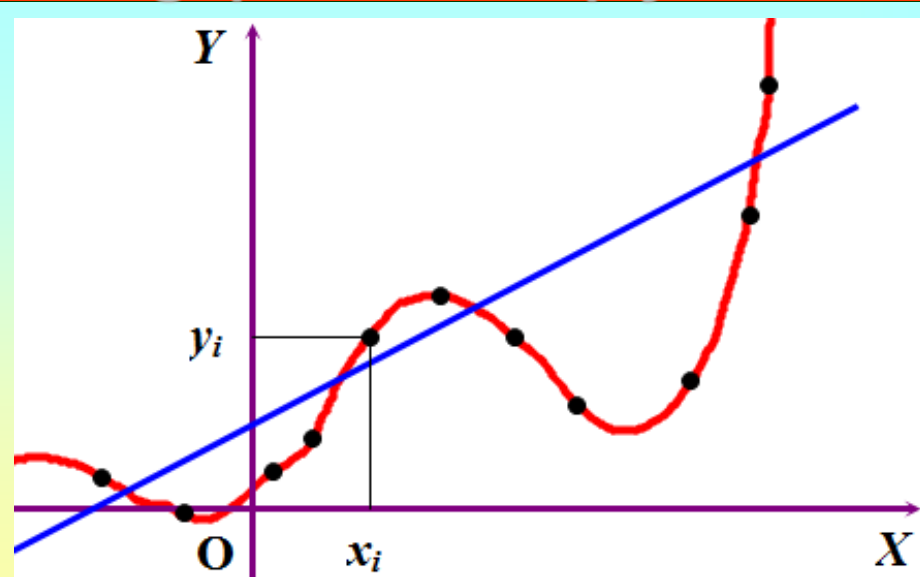
2. Đường hồi quy trung bình tuyến tính thực nghiệm

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

➤ Chương 8. Bài toán tương quan & Hồi quy



Hình a



Hình b

- 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

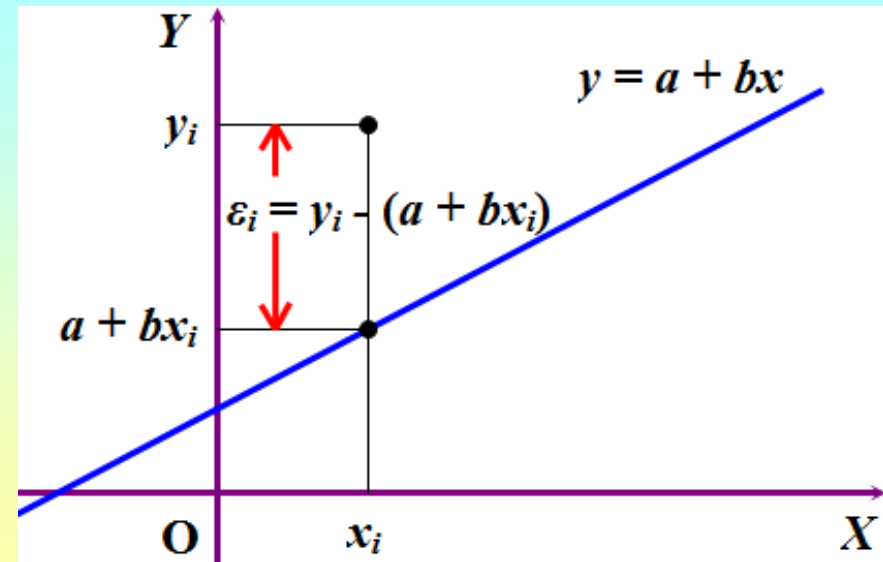
Ta đi tìm các ước lượng a, b

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

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

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

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



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

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