

Chương 8: Tương quan và hồi quy mẫu

§1. Hệ số tương quan mẫu.

Định nghĩa 1.1: Hệ số tương quan mẫu giữa X và Y là:

$$r_{XY} = \frac{xy - x \cdot y}{S_X \cdot S_Y}$$

Hệ số tương quan mẫu là một ước lượng của hệ số tương quan giữa X và Y ở chương 3, §6.

\$2.Đường hồi quy

1.Đường hồi quy mẫu.

Định nghĩa 2.1: Ký hiệu

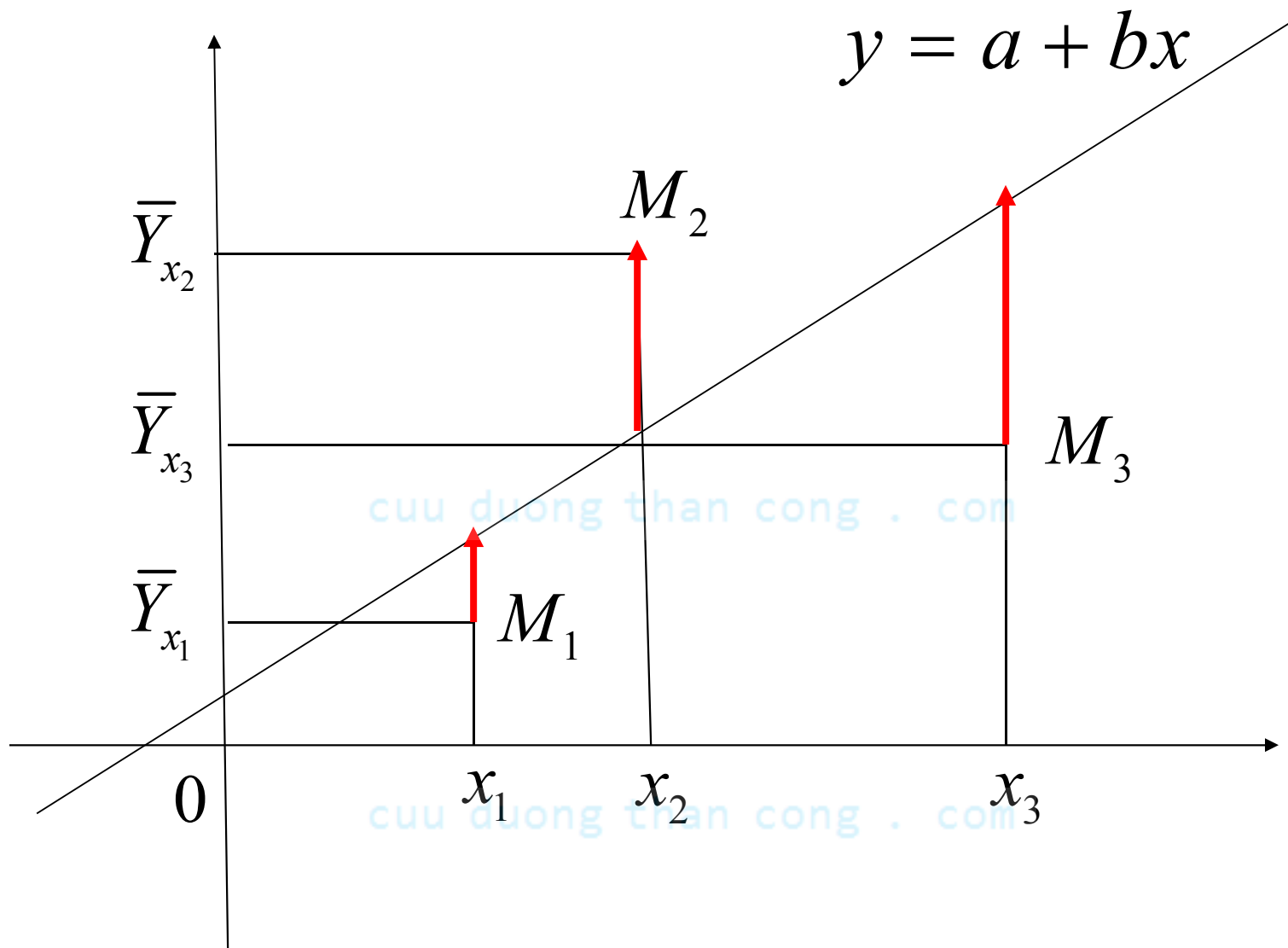
$$\bar{Y}_{x_i} = (\bar{Y} / X = x_i) = \frac{1}{n_i} \sum_{j=1}^h y_j \cdot n_{ij}, i = \overline{1, k} \Rightarrow M_i(x_i, \bar{Y}_{x_i}), \overline{1, k}$$

Đường gấp khúc $M_1 M_2 \dots M_k$ được gọi là đường hồi quy mẫu của Y theo X.

2. Đường hồi quy tuyến tính mẫu.

Định nghĩa 2.2: Đường hồi quy tuyến tính mẫu của Y theo X là đường thẳng $y=a+bx$ sao cho:

$$Q(a, b) = \sum_{i=1}^k \left[\bar{Y}_{x_i} - (a + bx_i) \right]^2 \cdot n_i \rightarrow \min$$



Định lý:

$$b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\hat{S}_X^2}, a = \overline{y} - b \overline{x}$$

Ý nghĩa: Đường hồi quy tuyến tính mẫu là đường thẳng xấp xỉ nội suy từ bảng số liệu của X và Y theo phương pháp bình phương tối thiểu. Nếu X và Y có tương quan xấp xỉ tuyến tính thì đường hồi quy tuyến tính mẫu cho ta một dự báo đơn giản:

$$X = x_0 \Rightarrow Y \approx y_0 = a + bx_0$$

3.Cách dùng máy tính bỏ túi:.

a)Loại ES: • Mở tần số(1 lần): Shift Mode Stat On(Off)

MODE	STAT	a+bx	x_i	y_j	n_{ij}

SHIFT START VAR

$$\overline{x} \rightarrow \overline{x}, \quad \overline{y} \rightarrow \overline{y}$$

SHIFT START REG

$$r \rightarrow r_{xy}$$

SHIFT START REG

$$a \rightarrow a$$

SHIFT START REG

$$b \rightarrow b$$

SHIFT START SUM

$$\sum x y \rightarrow n \cdot \overline{xy}$$

b)Loại MS :: MODE REG LIN

Cách xóa dữ liệu cũ : SHIFT CLR SCL =

Cách nhập dữ liệu : $x_i, y_j; n_{ij} \quad M +$

.SHIFT START S-VAR $\overline{x} = \rightarrow \overline{x} \triangleright \overline{y} = \rightarrow \overline{y}$

.SHIFT START S-VAR $\triangleright \triangleright r = \rightarrow r_{xy}$

.SHIFT START S-VAR $\triangleright \triangleright a = \rightarrow a$

.SHIFT START S-VAR $\triangleright \triangleright b = \rightarrow b$

.SHIFT START S-SUM $\triangleright \sum xy = \rightarrow n . \overline{xy}$

...

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

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

Bảng số liệu đề bài tương đương với bảng sau:

x_i	y_j	n_{ij}
1	0,3	20
1	0,7	10
2	0,7	30
2	1	10
3	0,7	10
3	1	20

Nhập vào ta có:(Chú ý:Nếu muốn tìm đường hồi qui tuyến tính của X theo Y thì phải đổi chỗ 2 cột dữ liệu X,Y.Lúc đó phương trình là $x=a+by$)

$$n \Rightarrow n = 100; \bar{x} \Rightarrow \bar{x} = 2; \bar{y} \Rightarrow \bar{y} = 0,71$$

$$r \Rightarrow r_{xy} = 0,7447$$

$$a \Rightarrow a = 0,2433$$

$$b \Rightarrow b = 0,2333$$

$$\Rightarrow y = a + bx = 0,2433 + 0,2333x$$

$$\sum xy \Rightarrow n.\overline{xy} = 156 \Rightarrow \overline{xy} = 1,56$$

$$S_X = x\sigma n - 1 = 0,7785$$

$$S_Y = y\sigma n - 1 = 0,2439$$

1. Với độ tin cậy 0.95, hãy ước lượng số vốn đầu tư và lợi nhuận trung bình

- Ta có bài toán ước lượng trung bình, TH 2

$$\varepsilon_X = \frac{S_X}{\sqrt{n}} Z_\alpha = 0,15259$$

$$\bar{x} - \varepsilon_X < a_X < \bar{x} + \varepsilon_X$$

$$\varepsilon_Y = \frac{S_Y}{\sqrt{n}} Z_\alpha = 0,04781$$

$$\bar{y} - \varepsilon_Y < a_Y < \bar{y} + \varepsilon_Y$$

2. Trước đây lợi nhuận trung bình là 0,6. Với mức ý nghĩa 0.05, hãy kiểm tra ý kiến cho rằng lợi nhuận trung bình đã tăng lên .

- Ta có bài toán kiểm định trung bình, TH 2

$$H : a = a_0 = 0,6$$

$$u_{qs} = \frac{(\bar{y} - 0,6).10}{S_Y} = 4,51 > Z_{0,05} = 1,96$$

$$\Rightarrow a > a_0 = 0,6$$

- Vậy lợi nhuận trung bình đã tăng lên.

3. Lợi nhuận lớn hơn 0,7 là lợi nhuận cao. Với mức ý nghĩa 0.01, hãy kiểm tra ý kiến cho rằng tỷ lệ lợi nhuận cao là 0,32.

$$H : p = p_0 = 0,32$$

$$f = \frac{30}{100} = 0,3$$

$$u_{qs} = \frac{(0,3 - 0,32) \cdot 10}{\sqrt{0,32 \cdot 0,68}} = -0,43$$

$$|u_{qs}| < Z_{0,01} = 2,575 \Rightarrow p = p_0 = 0,32$$

- Vậy ý kiến đã cho là đúng.

4. Lợi nhuận lớn hơn 0,7 là lợi nhuận cao. Với độ tin cậy 0.9, hãy ước lượng số vốn trung bình cho lợi nhuận cao .

Ta có bảng phân phối của số vốn cho lợi nhuận cao là:

X	2	3
n_i	10	20

$$n = 30, \bar{x} = 2,66667, S_X = 0,47946$$

$$\epsilon_X = \frac{S_X}{\sqrt{n}} Z_{0,1} = 0,144$$

$$\bar{x} - \epsilon_X < a_X < \bar{x} + \epsilon_X$$