

CHƯƠNG V: THỐNG KÊ

§1 LÝ THUYẾT MẪU

I. Khái niệm:

Ví dụ:

Muốn nghiên cứu chiều cao trung bình của người Việt nam ở 1 độ tuổi nào đó, cách tốt nhất là đo chiều cao của tất cả công dân VN ở độ tuổi đó (đám đông hay tổng thể). Tuy nhiên cách này không thể thực hiện được vì:

Về kinh tế: tốn kém.

Việc xác định công dân ở độ tuổi đó khá khó khăn.

Thời gian: dài.

Thống kê đề nghị phương pháp:

- Chọn ra ngẫu nhiên n người (gọi là mẫu, kích thước mẫu: n), tính toán trên mẫu và từ đó suy rộng kết quả cho chiều cao trung bình của công dân VN ở độ tuổi đó.

- Tất nhiên sự suy rộng này có thể đúng cũng có thể sai. Để hạn chế sự sai lầm khi suy rộng, mẫu chọn phải khách quan.

1. Tổng thể (đám đông): là tập hợp tất cả các phần tử mà ta nghiên cứu.

Ví dụ: Tất cả công dân VN ở độ tuổi mà ta đang khảo sát.

2. *Mẫu*: 1 nhóm gồm n phần tử được chọn sao cho phản ánh trung thực đặc điểm của đám đông.

Mẫu ngẫu nhiên: 1 mẫu ngẫu nhiên cỡ n $\{x_1; x_2; \dots; x_n\}$ gồm n biến độc lập, có cùng phân phối X.

Đối với đám đông ta thường quan tâm đến 2 mặt: Lượng và chất.

Về lượng: thường đánh giá về trung bình (μ), Phương sai σ^2

Về chất: thường quan tâm đến tỷ lệ.

II. Các đặc trưng của mẫu ngẫu nhiên

1. Giả sử ta có mẫu ngẫu nhiên: $\{x_1; x_2; \dots; x_n\}$

a) *Trung bình mẫu*: $\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$

b) *Phương sai mẫu*: $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{X})^2$

Phương sai mẫu hiệu chỉnh: $\sigma_{n-1}^2 = \frac{n}{n-1} \sigma_n^2$

c) *Độ lệch mẫu*: $\hat{s} = \sqrt{\sigma_n^2}$, *độ lệch mẫu hiệu chỉnh*: $s = \sqrt{\sigma_{n-1}^2}$

d) Tỷ lệ mẫu: Cho mẫu định tính có kích thước n. Trong đó m phần tử có tính chất A, tỷ lệ mẫu được xác định như sau: $f = \frac{m}{n}$

2. Bảng mẫu thu gọn

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

a) Gía trị trung bình:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}, \quad n = n_1 + n_2 + \dots + n_k$$

$$\overline{x^2} = \frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k}{n}$$

b) Deviation $\hat{s}^2 = \overline{x^2} - (\bar{x})^2$

c) Độ lệch: $\hat{s} = \sqrt{\hat{s}^2}$

BÀI TẬP

Sử dụng máy tính bỏ túi tính các đặc trưng mẫu (trung bình, phương sai, độ lệch)

1) Nghiên cứu trọng lượng của 1 giống vịt mới người ta có kết quả sau:

Cân nặng(kg)	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Số con	2	6	24	35	39	24	14	6

Kết quả: $\bar{x} = 2.185$; $\hat{s}^2 = 0.142$; $\hat{s} = 0.377$

2) Cân nặng 45 con heo 3 tháng tuổi trong trại chăn nuôi ta được kết quả sau

x_i	35	37	39	41	43	45	47
n_i	2	6	10	11	8	5	3

a) Tính trung bình, độ lệch, phương sai.

b) Giả sử heo có trọng lượng $\geq 38kg$ là heo đạt tiêu chuẩn. Tính tỷ lệ heo đạt tiêu chuẩn của mẫu trên.

3) Nếu bảng phân bố cho dưới dạng ghép lớp, tính trung điểm của mỗi lớp

lớp	x_i	n_i
54.795-54.805	54.80	6
54.805-54.815	54.81	14
54.815-54.825	54.82	33
54.825-54.835	54.83	47
54.835-54.845	54.84	45
54.845-54.855	54.85	33
54.855-54.865	54.86	15
54.865-54.875	54.87	7

Trung điểm của mỗi lớp (là đại diện của lớp đó) $x_i = \frac{x_{i \min} + x_{i \max}}{2}$

§2 ĐẶC TRƯNG CỦA MẪU 2 CHIỀU

I. Các đặc trưng của mẫu 2 chiều

- Tổng thể 2 chiều(X,Y) là tập hợp các phần tử được quan sát đồng thời 2 dấu hiệu X,Y.
- Thực tế mẫu 2 chiều cho dưới dạng tương quan bảng:

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

a) $x_1 < x_2 < \dots < x_k$; $n_1 + n_2 + \dots + n_k = m_1 + m_2 + \dots + m_h = n$

b) $y_1 < y_2 < \dots < y_h$

$n_i; i = \overline{1, k}$ $m_j; j = \overline{1, h}$;

3. Các đặc trưng của mẫu 2 chiều:

a) Theo dấu hiệu X

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}, \quad n = n_1 + n_2 + \dots + n_k$$

$$\overline{x^2} = \frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k}{n}$$

b) Deviation $\hat{s}_x^2 = \overline{x^2} - (\bar{x})^2$

b) Theo dấu hiệu Y

$$\bar{y} = \frac{y_1 m_1 + y_2 m_2 + \dots + y_h m_h}{n}, \quad n = m_1 + m_2 + \dots + m_h$$

$$\overline{y^2} = \frac{y_1^2 m_1 + y_2^2 m_2 + \dots + y_h^2 m_h}{n}$$

b) Deviation $\hat{s}_y^2 = \overline{y^2} - (\bar{y})^2$

c) Trung bình của tích:

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

d) Hệ số tương quan mẫu

• Công thức: $r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x \cdot \hat{s}_y}$

• Ý NGHĨA: Hệ số tương quan mẫu dùng để đánh giá mức độ chặt chẽ của quan hệ X,Y trong mẫu.

$r_{xy} = 0 \Rightarrow X, Y$ không có quan hệ phụ thuộc.

$r_{xy} = \pm 1 \Rightarrow X, Y$ Có quan hệ phụ thuộc tuyến tính.

$r_{xy} \neq 0 \Rightarrow X, Y$ Có quan hệ phụ thuộc.

4. Đường hồi quy tuyến tính mẫu

a) Phương trình hồi quy tuyến tính mẫu Y theo X có dạng:

$$y = ax + b \text{ With } a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} \quad b = \bar{y} - a\bar{x}$$

b) Phương trình hồi quy tuyến tính mẫu X theo Y có dạng:

$$x = cy + d \text{ With } c = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_y^2} \quad d = \bar{x} - c\bar{y}$$

Ý NGHĨA:

Với phương trình: $y = ax + b$. Ta có thể dự báo được trung bình của Y khi X nhận giá trị x_0 (không có trong mẫu) nghĩa là ta có xấp xỉ $\bar{y}_{x_0} \approx ax_0 + b$

Tương tự với đường hồi quy tuyến tính X theo Y ta có xấp xỉ $\bar{x}_{y_0} \approx cy_0 + d$

Ví dụ:

Cho bảng mẫu 2 chiều như dưới đây

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X và hãy dự đoán giá trị của Y khi X=5

Y \ X	1	2	3	5	6	n_i
1	1	2	0	0	0	
2	0	2	2	0	0	
3	0	0	2	1	0	
4	0	0	0	0	1	
m_j						

GIẢI

X \ Y	1	2	3	5	6	n_i
1	1	2	0	0	0	3
2	0	2	2	0	0	4
3	0	0	2	1	0	3
4	0	0	0	0	1	1
m_j	1	4	4	1	1	n=11

a) Các đặc trưng mẫu theo X và Y

$$\bar{x} = 2.1818, \hat{s}_x^2 = 0.8761, \hat{s}_x = 0.936$$

$$\bar{y} = 2.909, \hat{s}_y^2 = 1.9, \hat{s}_y = 1.38$$

$$\overline{xy} = 7.455 \Rightarrow r_{xy} = \frac{7.455 - 2.1818 \times 2.909}{1.38 \times 0.936} = 0.86$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = \frac{7.455 - 2.1818 \times 2.909}{0.8761} = 1.26$$

$$b = \bar{y} - a\bar{x} = 2.909 - 1.26 \times 2.1818 = 0.16$$

$$\Rightarrow y = 1.26x + 0.16$$

Khi $x = 5 \Rightarrow y = 6.46$

BÀI TẬP

1) Để nghiên cứu sự phát triển của 1 loại cây trồng, người ta đo đường kính X(cm) và chiều cao Y(m) của 100 cây trồng. Kết quả cho ở bảng sau:

X \ Y	3	4	5	6	7	8
21	2	5	0	0	0	0
23	0	3	11	0	0	0
25	0	0	8	0	10	0
27	0	0	4	15	6	0
29	0	0	0	17	7	12

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X và X theo Y. Hãy dự đoán giá trị của Y khi X=30(cm) và Tính X khi Y=10(m).

GIẢI

a) Các đặc trưng mẫu theo X và Y

$$\bar{x} = 26.38, \hat{s}_x^2 = 6.5356, \hat{s}_x = 2.556$$

$$\bar{y} = 6.02, \hat{s}_y^2 = 1.44, \hat{s}_y = 1.2$$

$$\overline{xy} = 161.04 \Rightarrow r_{xy} = \frac{161.04 - 26.38 \times 6.02}{2.556 \times 1.2} = 0.728$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = \frac{161.04 - 26.38 \times 6.02}{6.5356} = 1.7648$$

$$b = \bar{y} - a\bar{x} = 6.02 - 1.7648 \times 26.38 = -40.535$$

$$\Rightarrow y = 1.7648x - 40.535$$

$$\text{Khi } x = 30(\text{cm}) \Rightarrow y = 12.4(\text{m})$$

2) Tính các đặc trưng của mẫu sau:

a)

x_i	50	60	70	80	90	100	110
n_i	6	9	22	33	25	4	1

b)

x_i	3	4	5	6	7	8	10	11	12
n_i	7	10	30	35	25	16	10	8	3

c)

x_i	n_i
100-110	8
110-120	12
120-130	15
130-140	25
140-150	22
150-160	18
160-170	7

d)

x_i	n_i
100-200	5
200-250	12
250-300	56
300-350	107
350-400	75
400-450	70
450-500	35
500-550	30
550-700	10

3) Xét mẫu tương quan cặp cho như sau:

X	5	10	10	10	15	15	15	20	20	20
Y	20	20	30	30	30	40	50	50	60	60

a) Tìm r_{xy}

b) Tìm Lập phương trình hồi quy tuyến tính Y theo X và X theo Y.

4) Nghiên cứu mối liên hệ giữa số tiền đầu tư cho việc phòng bệnh theo đầu người X (đồng/ người) và tỷ lệ người mắc bệnh Y(%) ở 50 địa phương, người ta thu được bảng số liệu sau:

X \ 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) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X .Hãy dự đoán giá trị của Y khi X=600 đồng/người.

GIẢI

$$\bar{x} = 318, \hat{s}_x^2 = 15475.36, \hat{s}_x = 124.4$$

$$\bar{y} = 2.95, \hat{s}_y^2 = 0.362404, \hat{s}_y = 0.602$$

$$\overline{xy} = 876 \Rightarrow r_{xy} = -0.83$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = -0.004$$

$$b = \bar{y} - a\bar{x} = 4.2$$

$$\Rightarrow y = -0.004x + 4.2$$

$$\text{Khi } x = 600(\text{đ} / \lambda) \Rightarrow y = 1.8\%$$

5) Nghiên cứu mối liên hệ giữa thu nhập X với mức tiêu dùng Y về 1 loại thực phẩm ta có bảng sau:

X \ Y	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	

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X. Hãy dự đoán giá trị của Y khi X=80.

ĐÁP SỐ:

$$\bar{x} = 35.75, \hat{s}_x^2 = 122.44, \hat{s}_x = 11.06$$

$$\bar{y} = 35.9, \hat{s}_y^2 = 106.19, \hat{s}_y = 10.3$$

$$\overline{xy} = 1371.75 \Rightarrow r_{xy} = 0.78$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 0.72$$

$$b = \bar{y} - a\bar{x} = 10.2$$

$$\Rightarrow y = 0.72x + 10.2$$

$$\text{Khi } x = 80 \Rightarrow y = 67.76$$

6) Nghiên cứu mối liên hệ của doanh số bán Y(Tỷ đồng/ năm) theo chi phí chào hàng X(Triệu đồng /năm) của 1 công ty thương mại tại 1 số khu vực bán hàng ta có bảng sau:

Y \ X	28	29	30	32	35	36
50	5	3	2			
55	2	7	9	2		
60		2	8	7	3	
65		3	5	5	4	1

- a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu
b) Lập phương trình hồi quy tuyến tính Y theo X .
c) Nếu chi phí chào hàng là 100 triệu đồng/ năm thì doanh số bán được là bao nhiêu(tỷ đồng/ năm).

ĐÁP SỐ:

$$\bar{x} = 58.38, \hat{s}_x^2 = 26.1, \hat{s}_x = 5.1$$

$$\bar{y} = 30.59, \hat{s}_y^2 = 4.2, \hat{s}_y = 2.05$$

$$\overline{xy} = 1791.6 \Rightarrow r_{xy} = 0.56$$

b)PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 0.22$$

$$b = \bar{y} - a\bar{x} = 17.75$$

$$\Rightarrow y = 0.22x + 17.75$$

$$x = 100(\text{million VND}) \Rightarrow y = 39.75 \text{Bilion VND}$$

7) Để nghiên cứu sự ảnh hưởng của phân bón X (Kg)đến năng suất lúa Y(Kg). Người ta tiến hành thí nghiệm trên 100 mảnh ruộng và thu được kết quả như sau

Y \ X	10	12	14	16
100	22			
150	8	10	3	
200	14	15	12	
250				16

- a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X .Hãy dự đoán giá trị của Y khi X=20(Kg).

$$\bar{x} = 12.06, \hat{s}_x^2 = 4.9, \hat{s}_x = 2.2$$

$$\bar{y} = 175.5, \hat{s}_y^2 = 2524.75, \hat{s}_y = 50.25$$

$$\overline{xy} = 2199 \Rightarrow r_{xy} = 0.75$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 16.83$$

$$b = \bar{y} - a\bar{x} = -27.48$$

$$\Rightarrow y = 16.83x - 27.48$$

$$x = 20(\text{kg}) \Rightarrow y = 309.12(\text{kg})$$

8) X(Cm) , Y(Kg) là 2 chỉ tiêu chất lượng của 1 loại sản phẩm. Điều tra ở 1 số sản phẩm ta có kết quả sau:

X \ Y	21	23	25	27	29
3	13				
5	1	28			
8		2	31	3	
10			4	17	1

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X .Hãy dự đoán giá trị của Y khi X=31(cm).

ĐÁP SỐ:

$$\bar{x} = 24.28, \hat{s}_x^2 = 3.88, \hat{s}_x = 1.97$$

$$\bar{y} = 6.92, \hat{s}_y^2 = 5.57, \hat{s}_y = 2.36$$

$$\overline{xy} = 172.4 \Rightarrow r_{xy} = 0.94$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 1.13$$

$$b = \bar{y} - a\bar{x} = -20.49$$

$$\Rightarrow y = 1.13x - 20.49$$

$$\text{Khi } x = 31(\text{cm}) \Rightarrow y = 14.54(\text{kg})$$

9) Điều tra 1 số sản phẩm của xí nghiệp về chiều dài X(cm) và hàm lượng Y(%) ta có kết quả sau:

X \ Y	8	10	12	14	16
100	5	5			
110	4	6	7		
120		5	9	8	
130			4	6	9
140				5	7

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X .Hãy dự đoán giá trị của Y khi X=160(cm).

ĐÁP SỐ:

$$\bar{x} = 120.75, \hat{s}_x^2 = 154.44, \hat{s}_x = 12.43$$

$$\bar{y} = 12.425, \hat{s}_y^2 = 6.569, \hat{s}_y = 2.56$$

$$\overline{xy} = 1526.5 \Rightarrow r_{xy} = 0.823$$

b)PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 0.17$$

$$b = \bar{y} - a\bar{x} = -8.1$$

$$\Rightarrow y = 0.17x - 8.1$$

$$\text{Khi } x = 160(\text{cm}) \Rightarrow y = 19.1(\%)$$

10) Chiều cao X(cm) và cân nặng Y(Kg) của 100 học sinh ta có kết quả ở bảng sau: ta có kết quả sau:

X \ Y	147	152	157	162	167
37	3				
42	5	10			
47		14	20	6	
52			15	12	5
57				6	4

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X. Hãy dự đoán giá trị của Y khi X=62(kg).

ĐÁP SỐ:

$$\bar{x} = 157.1, \hat{s}_x^2 = 28.99, \hat{s}_x = 5.38$$

$$\bar{y} = 48.55, \hat{s}_y^2 = 22.35, \hat{s}_y = 4.73$$

$$\overline{xy} = 7647.55 \Rightarrow r_{xy} = 0.8$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 0.7$$

$$b = \bar{y} - a\bar{x} = -61.7$$

$$\Rightarrow y = 0.7x - 61.7$$

$$\text{Khi } x = 62(\text{kg}) \Rightarrow y = 176.7(\text{cm})$$

11) Tiến hành quan sát về độ chảy X và độ bền Y của 1 kim loại ta có kết quả sau:

X \ Y	35	45	55	65	75
80	7	4			
100	6	13	20		
120		12	15	10	
140		8	8	5	3
160			1	2	2

a) Tính các đặc trưng mẫu theo X, Y và Tính hệ số tương quan mẫu

b) Lập phương trình hồi quy tuyến tính Y theo X. Hãy dự đoán giá trị của Y khi X=80.

ĐÁP SỐ:

$$\bar{x} = 51.9, \hat{s}_x^2 = 98.99, \hat{s}_x = 9.95$$

$$\bar{y} = 115.34, \hat{s}_y^2 = 416.26, \hat{s}_y = 20.4$$

$$\overline{xy} = 6102.59 \Rightarrow r_{xy} = 0.57$$

b) PTHQTT

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{s}_x^2} = 1.18$$

$$b = \bar{y} - a\bar{x} = 54.1$$

$$\Rightarrow y = 1.18x + 54.1$$

$$\text{Khi } x = 100 \Rightarrow y = 172.1$$