

BÀI 5

KHÁI NIỆM PHƯƠNG PHÁP MẪU

TS. Nguyễn Mạnh Thế

TÌNH HUỐNG KHỞI ĐỘNG

Tình huống

Điều tra mức thu nhập cá nhân trong một tháng (triệu đồng) ở huyện Đông Anh, ta có bảng số liệu mẫu sau:

Thu nhập	1-2	2-3	3-4	4-5	5-6	6-7
Số người	10	8	5	7	3	2

Cần phải tính thu nhập bình quân đầu người và độ chênh lệch thu nhập để xác định mức sống của người dân và mức độ đồng đều về thu nhập trong vùng.

Câu hỏi gợi mở

Câu 1: Thu nhập bình quân đầu người là bao nhiêu?

Câu 2: Độ chênh lệch thu nhập là bao nhiêu?

Câu 3: Độ chênh lệch bình quân hiệu chỉnh?

TÌNH HUỐNG KHỞI ĐỘNG

Kết luận

1. Thống kê:

$$\bar{X} = G(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

được gọi là trung bình mẫu.

2. Thống kê:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

được gọi là độ lệch chuẩn mẫu.

3. Thống kê:

$$S' = \sqrt{S'^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

được gọi là độ lệch chuẩn mẫu hiệu chỉnh.

MỤC TIÊU

- Cơ sở lý thuyết mẫu;
- Tổng thể nghiên cứu;
- Mẫu ngẫu nhiên;
- Thống kê;
- Mẫu ngẫu nhiên hai chiều;
- Quy luật phân phối xác suất của một số thống kê.



1. CỞ SỞ LÝ THUYẾT MẪU

Khái niệm phương pháp mẫu:

Bài toán: Cần nghiên cứu tính chất định tính hoặc định lượng của các phần tử trong một tập hợp nào đó.

Ta có hai phương pháp thực hiện nghiên cứu:

Nghiên cứu toàn bộ

- Chi phí lớn về kinh tế, có thể phá hủy toàn bộ tập hợp cần nghiên cứu.
- Không thể nghiên cứu được toàn bộ.

Vậy ta thấy nghiên cứu toàn bộ tập hợp là không khả thi.

Nghiên cứu bộ phận

- Ta lấy ra một tập con và nghiên cứu toàn bộ các phần tử trong tập con đó.
- Đưa ra kết luận cho các phần tử trong tập hợp nghiên cứu.

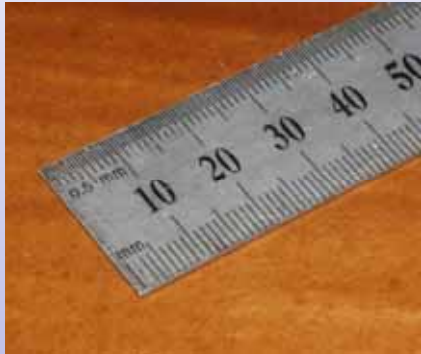
Đây là phương pháp nghiên cứu mẫu (Sampling).

2. TỔNG THỂ NGHIÊN CỨU

Định nghĩa:

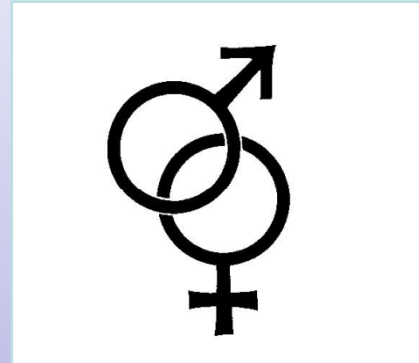
Tổng thể là tập hợp các phần tử cần nghiên cứu tính chất định tính hoặc định lượng, số phần tử trong tổng thể gọi là cỡ của tổng thể, ký hiệu là N .

a) Biến định lượng



Mã hóa: Lấy giá trị của biến định lượng làm mã của biến.

b) Biến định tính



Mã hoá: Gán tính chất định tính của biến ứng với các số nguyên.

Vậy khi nghiên cứu tổng thể ta luôn có thể giả sử là các các phần tử có dấu hiệu định lượng.

2. TỔNG THỂ NGHIÊN CỨU (tiếp theo)

Mô tả tổng thể

Cho tổng thể với các phần tử $\{x_1, x_2, \dots, x_n\}$. Với N_i là số lần giá trị x_i xuất hiện trong tổng thể, ta có: $N_1 + N_2 + \dots + N_k = N$

Đặt $f_i = \frac{N_i}{N}$ ($i = 1 \dots k$), f_i được gọi là tần suất của x_i trong tổng thể:

x_i	x_1	x_2	...	x_k
f_i	f_1	f_2	...	f_k

Dễ thấy: $f_1 + f_2 + \dots + f_k = 1$

Trung bình tổng thể:

$$m = \frac{1}{N} \sum_{i=1}^k N_i x_i = \sum_{i=1}^k f_i x_i$$

Phương sai tổng thể:

$$s = \frac{1}{N} \sum_{i=1}^k (x_i - m)^2 = \sum_{i=1}^k f_i x_i^2 - (m)^2$$



MẪU NGẪU NHIÊN

- Các phương pháp lấy mẫu;
- Định nghĩa mẫu ngẫu nhiên;
- Mô tả mẫu ngẫu nhiên:
 - Theo biểu đồ tần suất;
 - Theo tổ chức đồ.



3.1. CÁC PHƯƠNG PHÁP LẤY MẪU

Mục đích

- Ta không thể nghiên cứu cận kề từng phần tử của tổng thể, do đó phải nghiên cứu hạn chế trên một nhóm nhỏ rút ra từ tổng thể gọi là mẫu, từ đó rút ra kết luận cho tổng thể.
- Một mẫu được gọi là đại diện tốt nhất cho tổng thể nếu nó là mẫu ngẫu nhiên (random sample).

Các phương pháp lấy mẫu

- Cách chọn mẫu ngẫu nhiên đơn giản:
 - Chọn mẫu ngẫu nhiên có hoàn lại;
 - Chọn mẫu ngẫu nhiên không hoàn lại.
- Chọn mẫu phân cấp: Chia tổng thể ra thành k tổng thể bộ phận và ta thực hiện cách lấy mẫu ngẫu nhiên đơn giản trên mỗi tổng thể thành phần.

3.2. ĐỊNH NGHĨA MẪU NGẪU NHIÊN

Một mẫu ngẫu nhiên cỡ n của biến ngẫu nhiên X là một tập các biến ngẫu nhiên X_1, X_2, \dots, X_n độc lập và có cùng phân phối với biến ngẫu nhiên X .

Trong đó mỗi X_k là một quan sát về biến ngẫu nhiên X .

Quan sát X_k nhận giá trị x_k

Khi đó bộ giá trị x_1, x_2, \dots, x_n gọi là giá trị cụ thể của mẫu ngẫu nhiên X_1, X_2, \dots, X_n ;

Ví dụ:

Khi gieo con xúc xắc 5 lần ta được một mẫu ngẫu nhiên $(X_1, X_2, X_3, X_4, X_5)$ với các giá trị của mỗi lần gieo là $(3, 5, 2, 3, 1)$.



3.3. MÔ TẢ MẪU NGẪU NHIÊN

Biểu đồ tần suất:

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

Với n_i là số lần giá trị x_i xuất hiện trong mẫu.

Ta có $n_1 + n_2 + \dots + n_k = n$

Đặt $f_i = \frac{n_i}{n}$, f_i gọi là tần suất của x_i trong mẫu.

Ta có bảng biểu diễn tần suất mẫu:

x_i	x_1	x_2	...	x_k
f_i	f_1	f_2	...	f_k

trong đó: $f_1 + f_2 + \dots + f_k = \frac{n_1 + n_2 + \dots + n_k}{n} = 1$

3.3. MÔ TẢ MẪU NGẪU NHIÊN (tiếp theo)

Tổ chức đồ:

khoảng	$a_0 - a_1$	$a_1 - a_2$...	$a_{k-1} - a_k$
n_i	n_1	n_2	...	n_k

n_i là số giá trị mẫu rơi vào khoảng $(a_{i-1}, a_i]$

$$n_1 + n_2 + \dots + n_k = n$$

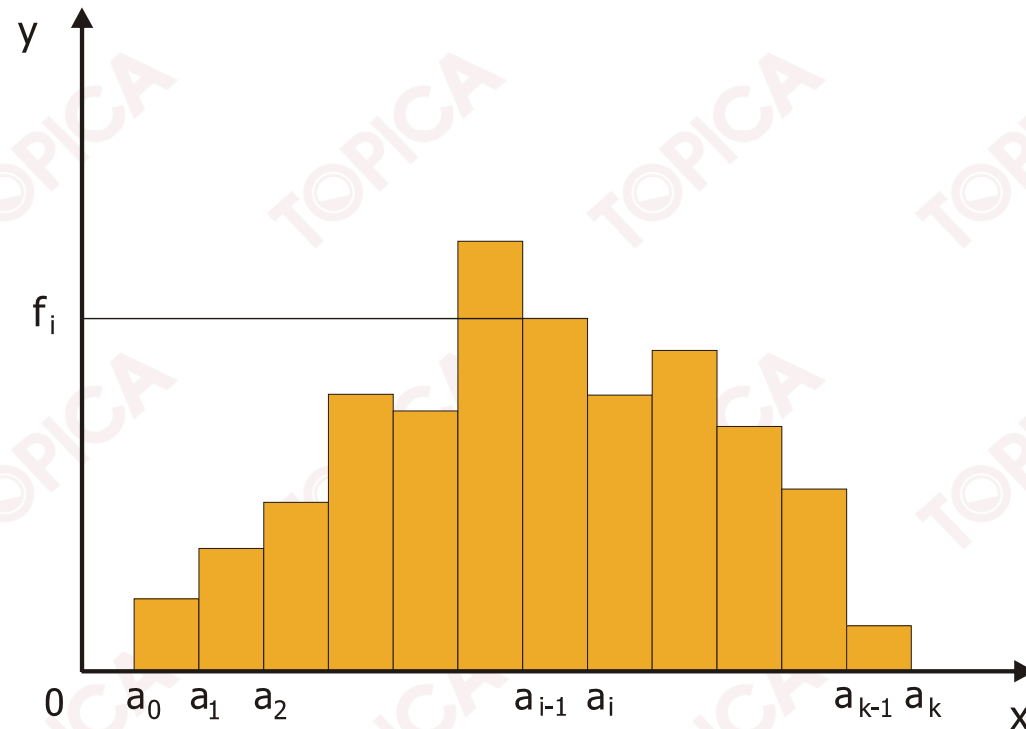
Mỗi khối chữ nhật có:

- Chiều cao:

$$y_i = \frac{n_i}{n \cdot h_i}$$

- Độ dài:

$$h_i = a_i - a_{i-1}$$



4. THỐNG KÊ



Phương sai mẫu hiệu chỉnh

Độ lệch chuẩn mẫu và độ lệch chuẩn mẫu hiệu chỉnh

Phần giới thiệu

Cho biến ngẫu nhiên X với mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) và giá trị mẫu (x_1, x_2, \dots, x_n) .

Định nghĩa

Thống kê là một hàm của các quan sát trong mẫu ngẫu nhiên, ký hiệu là $G(X_1, X_2, \dots, X_n)$. Khi mẫu nhận giá trị cụ thể (x_1, x_2, \dots, x_n) thì thống kê G nhận giá trị g được xác định bởi $g = G(x_1, x_2, \dots, x_n)$.



PROPERTIES

Allow user to leave interaction:

Show 'Next Slide' Button:

Completion Button Label:

Anytime

Don't show

Next Slide



Properties...



Edit in Engage

**Articulate Quizmaker Quiz
Placeholder - XSTK_chuong
5_Quiz_Slide 12_17_10_08**

4. THỐNG KÊ

Cách tính các giá trị thống kê đặc trưng mẫu cho mẫu thu gọn

Nếu mẫu cho dưới dạng khoảng ta chọn mỗi khoảng điểm đại diện

$$x_i = \frac{a_{i-1} + a_i}{2}, i = 1, 2, \dots, k \text{ lúc đó ta có mẫu thu gọn.}$$

Lập một bảng tính như sau:

Khoảng giá trị mẫu	x_i	n_i	$n_i \cdot x_i$	$n_i \cdot x_i^2$
$a_0 - a_1$	x_1	n_1	$n_1 x_1$	$n_1 x_1^2$
$a_1 - a_2$	x_2	n_2	$n_2 x_2$	$n_2 x_2^2$
\vdots	\vdots	\vdots	\vdots	\vdots
$a_{i-1} - a_i$	x_i	n_i	$n_i x_i$	$n_i x_i^2$
\vdots	\vdots	\vdots	\vdots	\vdots
$a_{k-1} - a_k$	x_k	n_k	$n_k x_k$	$n_k x_k^2$
Σ		n	A	B

$$\bar{x} = \frac{A}{n}$$

$$s^2 = \frac{B}{n} - (\bar{x})^2$$

$$s'^2 = \frac{n}{n-1} s^2$$

4. THỐNG KÊ (tiếp theo)

Ví dụ:

Điều tra mức thu nhập cá nhân trong một tháng (triệu đồng), ta có bảng số liệu mẫu sau:

Thu nhập	1-2	2-3	3-4	4-5	5-6	6-7
Số người	10	8	5	7	3	2

Tính các giá trị đặc trưng mẫu.

$$\bar{x}, s^2, s'^2, s, s'$$

4. THỐNG KÊ (tiếp theo)

Bài giải:

Ta lập bảng tính sau

Khoảng thu nhập	x_i	n_i	$n_i \cdot x_i$	$n_i \cdot x_i^2$
1-2	1,5	10	15	22,5
2-3	2,5	8	20	50
3-4	3,5	5	17,5	61,25
4-5	4,5	7	31,5	141,75
5-6	5,5	3	16,5	90,75
6-7	6,5	2	13	84,5
Σ		$n = 35$	113,5	450,75

$$\bar{x} = 113,5 / 35 = 3,243$$

$$s^2 = 450,75 / 35 - (3,243)^2 = 2,363$$

$$s = \sqrt{2,363} = 1,573$$

$$s' = \sqrt{2,43} = 1,559$$

$$s'^2 = \frac{n}{n-1} s^2 = \frac{35}{34} 2,363 = 2,43$$

5. MẪU NGẪU NHIÊN HAI CHIỀU

Khái niệm:

Một mẫu ngẫu nhiên cỡ n của véc tơ ngẫu nhiên (X, Y) là một tập các véc tơ ngẫu nhiên $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ độc lập và có cùng phân phối với biến ngẫu nhiên (X, Y) .

Trong đó mỗi véc tơ (X_i, Y_i) là một quan sát thứ i về véc tơ ngẫu nhiên (X, Y) .

Ký hiệu (x_i, y_i) là giá trị của mẫu (X_i, Y_i) ($i = 1, 2, \dots, n$)

Bộ giá trị $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ gọi là giá trị cụ thể của $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$.

Ví dụ:

Lấy mẫu điều tra thu nhập và tiêu dùng (triệu đồng/tháng) của 10 hộ gia đình ta thu được giá trị mẫu:

$(2; 1.4), (2, 1.5), (3; 1.8), (4; 1.8),$
 $(2; 1.5), (4; 3.5), (7; 5.5), (3; 1.4),$
 $(4; 3.5), (5; 3.7).$



5.1. PHƯƠNG PHÁP MÔ TẢ MẪU

Dạng 1: Lập một bảng hai dòng như sau:

x_i	x_1	x_2	...	x_n
y_i	y_1	y_2	...	y_n

Dạng 2: Ta thu gọn mẫu và biểu diễn dưới dạng bảng chữ nhật như sau:

y_i x_i	y_1	y_2	... y_j ...	y_h	a_i
x_1	n_{11}	n_{12}	... n_{1j} ...	n_{1h}	a_1
x_2	n_{21}	n_{22}	... n_{2j} ...	n_{2h}	a_2
...
x_i	n_{i1}	n_{i2}	n_{ij}	n_{ih}	a_i
...
x_k	n_{k1}	n_{k2}	... n_{kj} ...	n_{kh}	a_k
b_j	b_1	b_2	b_j	b_h	$\sum \sum = n$

Ta có:

$$a_i = \sum_{j=1}^h n_{ij} \quad b_j = \sum_{i=1}^k n_{ij} \quad \text{và} \quad \sum_{i=1}^k \sum_{j=1}^h n_{ij} = n$$

5.2. THỐNG KÊ ĐẶC TRƯNG MẪU HAI CHIỀU

Trung bình mẫu:

- Véc tơ ngẫu nhiên hai chiều (\bar{X}, \bar{Y}) gọi là trung bình mẫu của véc tơ ngẫu nhiên (X, Y) , trong đó \bar{X} và \bar{Y} là các trung bình mẫu của biến ngẫu nhiên thành phần X và Y .
- Giá trị thống kê mẫu của mẫu ngẫu nhiên hai chiều là (\bar{x}, \bar{y}) .



5.2. THỐNG KÊ ĐẶC TRƯNG MẪU HAI CHIỀU (tiếp theo)

Hệ số tương quan mẫu

Định nghĩa. Hệ số tương quan mẫu của mẫu ngẫu nhiên hai chiều ký hiệu là R được xác định bởi:

$$R = \frac{\overline{XY} - (\bar{X})(\bar{Y})}{S_X S_Y} \quad \text{trong đó thống kê } \overline{XY} = \frac{1}{n} \sum_{k=1}^n X_k Y_k$$

Giá trị của hệ số tương quan mẫu đối với mẫu cụ thể:

$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ là r

$$r = \frac{\overline{xy} - (\bar{x})(\bar{y})}{s_X s_Y}$$

$$\bar{x} = \frac{1}{n} \sum_{l=1}^n x_l$$

$$\bar{y} = \frac{1}{n} \sum_{l=1}^n y_l$$

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

$$s_X = \sqrt{\frac{1}{n} \sum_{l=1}^n x_l^2 - (\bar{x})^2}$$

$$s_Y = \sqrt{\frac{1}{n} \sum_{l=1}^n y_l^2 - (\bar{y})^2}$$

**Articulate Quizmaker Quiz
Placeholder -
XSTK_chuong5_Quiz_Slide
19_17_10_08**

6. QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ

Định nghĩa

Phân phối xác suất của một thống kê được gọi là phân phối mẫu.

Ví dụ:

Phân phối xác suất của \bar{X} được gọi là phân phối mẫu của thống kê trung bình mẫu.



6.1. BIẾN NGẪU NHIÊN GỐC CÓ PHÂN PHỐI 0-1

Cho $X \sim A(p)$

Xét mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) rút ra từ X .

Định lý 1: Thống kê $U = \frac{\bar{X} - p}{\sqrt{p(p-1)}} \sqrt{n}$ có quy luật phân phối $N(0,1)$ khi n đủ lớn.

Chú ý: Thống kê U cũng có thể viết lại dưới dạng $U = \frac{f - p}{\sqrt{p(p-1)}} \sqrt{n}$

trong đó $f = \frac{k}{n}$, với k là số lần mẫu nhận giá trị 1.

6.2. HAI BIẾN NGẪU NHIÊN GỐC CÓ PHÂN PHỐI 0-1

Cho $X \sim A(p_1)$ và $Y \sim A(p_2)$

Xét hai mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) và (Y_1, Y_2, \dots, Y_m) rút ra từ X và Y .

Định lý 2: Thống kê $U = \frac{\bar{X} - \bar{Y} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}}$ có quy luật phân phối $N(0,1)$

Chú ý: \bar{X} và \bar{Y} có thể được thay bằng f_1 và f_2

trong đó $f_1 = \frac{k_1}{n}$; $f_2 = \frac{k_2}{m}$,

với k_1 là số lần mẫu ngẫu nhiên của X nhận giá trị 1,
 k_2 là số lần mẫu ngẫu nhiên của Y nhận giá trị 1.

6.3. BIẾN NGẪU NHIÊN GỐC CÓ PHÂN PHỐI CHUẨN

Cho mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) được rút ra từ biến ngẫu nhiên X có quy luật phân phối chuẩn $N(\mu, \sigma^2)$.

Định lý 3: Thống kê trung bình mẫu \bar{X} có phân phối chuẩn $N(\mu_{\bar{X}}, \sigma_{\bar{X}}^2)$

$$\mu_{\bar{X}} = \frac{\mu + \mu + \dots + \mu}{n} = \mu$$
$$\sigma_{\bar{X}}^2 = \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n^2} = \sigma^2 / n$$
$$\Rightarrow U = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

Định lý 4:

- Thống kê $T = \frac{\bar{X} - \mu}{S'} \sqrt{n}$ có quy luật phân bố student với $n-1$ bậc tự do.
- Thống kê $\chi^2 = \frac{(n-1)S'^2}{\sigma^2}$ có quy luật phân phối khi bình phương với $n-1$ bậc tự do.

6.4. HAI BIẾN NGẪU NHIÊN GỐC CÓ PHÂN PHỐI CHUẨN

Cho hai biến ngẫu nhiên độc lập X và Y

$$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$$

Xét mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) rút ra từ X
và mẫu ngẫu nhiên (Y_1, Y_2, \dots, Y_m) rút ra từ Y .

Định lý 5: Thống kê
$$U = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

có quy luật phân phối chuẩn $N(0,1)$.

Định lý 6: Thống kê
$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{nS_X^2 + mS_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$$

có quy luật phân phối student với $n + m - 2$ bậc tự do.

Các thống kê U và T sẽ được sử dụng trong phần ước lượng và kiểm định tiếp theo.

TÓM TẮT CUỐI BÀI

Nội dung chính

1. Khái niệm mẫu ngẫu nhiên và tổng thể nghiên cứu.
2. Các phương pháp lấy mẫu.
3. Thống kê và các thống kê đặc trưng của mẫu.
4. Quy luật phân phối xác suất của một số thống kê.



A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Select a term:

Độ lệch chuẩn mẫu

Độ lệch chuẩn mẫu hiệu chỉnh

Giá trị cụ thể của mẫu ngẫu n...

Hệ số tương quan mẫu

Mẫu ngẫu nhiên

Phương sai mẫu

Phương sai mẫu hiệu chỉnh

Thống kê (Statistics)

Tổng thể (population)

Trung bình mẫu

Độ lệch chuẩn mẫu

Độ lệch chuẩn mẫu

Thống kê $S = \sqrt{S^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$
 được gọi là độ lệch chuẩn mẫu.

**PROPERTIES**

Allow user to leave interaction:

Anytime

Show 'Next Slide' Button:

Don't show

Completion Button Label:

Next Slide



Properties...



Edit in Engage