

## Chương 4: Lý thuyết mẫu và ước lượng

### I. Tổng thể và mẫu

- Thống kê là môn khoa học bao gồm các phương pháp thu thập, trình bày, tóm tắt dữ liệu, phân tích dự đoán, đưa ra quyết định.
- Tổng thể là tập hợp tất cả các phần tử mà ta quan tâm trong một nghiên cứu cụ thể.
- Số phần tử của tổng thể:  $N$ .
- Mẫu là một tập con của tổng thể mà ta chọn ra để nghiên cứu.
- Số phần tử của mẫu:  $n$ .

- Vì sao phải lấy mẫu?

#### Ví dụ 114

Tổng điều tra dân số và nhà ở năm 2019 là một cuộc điều tra toàn bộ diễn ra 10 năm 1 lần.  
⇒ tốn tiền, thời gian, nhưng phải làm.

#### Ví dụ 115

Kiểm tra tỉ lệ sữa không đạt tiêu chuẩn trong lô sữa sản xuất ngày 8/3/2018.  
⇒ không thể kiểm tra từng hộp.  
⇒ cứ 100 hộp, lấy 1 hộp để kiểm tra (mẫu).

- Một số phương pháp lấy mẫu

- Lấy mẫu theo xác suất.
  - Lấy mẫu ngẫu nhiên.
  - Lấy mẫu hệ thống.
  - Lấy mẫu phân tầng.

- ..
- Lấy mẫu phi xác suất
  - Lấy mẫu thuận tiện.
  - Lấy mẫu phán đoán.

...

### II. Các đặc trưng tổng thể

- Trung bình tổng thể:  $\mu = E(X)$ .
- Phương sai tổng thể:  $\sigma^2 = Var(X)$ .
- Tỉ lệ phần tử có tính chất A trong tổng thể:

$$p = \frac{m}{N},$$

trong đó

$m$ : số phần tử có tính chất A.

### III. Các đặc trưng mẫu

Giả sử  $x_1, x_2, \dots, x_m$  là các kết quả quan sát.

#### 1. Bảng số liệu

- Dạng liệt kê:  $x_1, x_2, \dots, x_m$  trong đó có các số liệu  $x_i$  được lặp lại.
- Bảng tần số

X	$x_1$	$x_2$	...	$x_k$
Tần số $n_i$	$n_1$	$n_2$	...	$n_k$

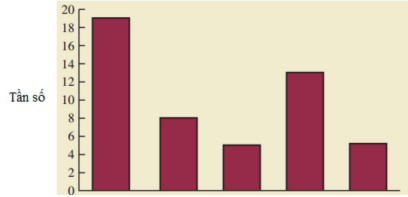
- Bảng dạng khoảng

X	$a_1 - b_1$	$a_2 - b_2$	...	$a_k - b_k$
Tần số $n_i$	$n_1$	$n_2$	...	$n_k$

Nếu gặp dạng bảng liệt kê và dạng khoảng, ta chuyển về dạng tần số để tính toán.

- Dạng liệt kê  $\Rightarrow$  đếm  $\Rightarrow$  xếp lại.
- Dạng khoảng, lấy điểm giữa  $x_i = \frac{a_i + b_i}{2}$ .

## 2. Biểu đồ tần số - tần suất



## 3. Các đặc trưng mẫu

Cho bảng tần số

X	$x_1$	$x_2$	...	$x_k$
Tần số $n_i$	$n_1$	$n_2$	...	$n_k$

Gọi  $n = n_1 + n_2 + \dots + n_k$ . Ta có

▷ **Trung bình mẫu**

$$\bar{x} = \frac{\sum_{i=1}^n n_i \cdot x_i}{n}$$

▷ **Phương sai mẫu**

$$s^2 = \frac{\sum (x_i)^2 - n \cdot (\bar{x})^2}{n - 1} = \frac{n}{n - 1} (\overline{x^2} - (\bar{x})^2)$$

trong đó  $\overline{x^2} = \frac{1}{n} \sum_{i=1}^k n_i x_i^2$ .

▷ **Độ lệch chuẩn mẫu ( $s$ ):**  $s = \sqrt{s^2}$ .

▷ **Tỉ lệ mẫu:**  $f = \frac{m}{n}$  (m: số phần tử có tính chất A)

### Ví dụ 116

Cho dữ liệu về năng suất lúa (tạ/ha) của 30 hộ dân ở một địa phương được ghi nhận như sau

30 35 37 40 41 42 43 46 47 49  
32 35 38 41 41 42 44 46 48 49  
33 35 38 41 42 43 44 46 48 51

- Lập bảng tần số.
- Tính trung bình, phương sai, độ lệch chuẩn mẫu.

### Ví dụ 117

Đường kính của 100 chi tiết do 1 máy SX có số liệu

Đường kính (mm)	Số chi tiết
19,8 - 19,85	3
19,85 - 19,90	5
19,90 - 19,95	16
19,95 - 20,00	28
20,00 - 20,05	23
20,05 - 20,10	14
20,10 - 20,15	7
20,15 - 20,20	4

Hãy tính các đặc trưng mẫu.

Giải

Đường kính (mm)	Số chi tiết	$x_i$	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
19,8 - 19,85	3			
19,85 - 19,90	5			
19,90 - 19,95	16			
19,95 - 20,00	28			
20,00 - 20,05	23			
20,05 - 20,10	14			
20,10 - 20,15	7			
20,15 - 20,20	4			

## Dùng máy tính tính các đặc trưng mẫu (Fx 570VN Plus, Vinacal 570ES Plus II)

### Bước chuẩn bị:

Xóa bộ nhớ:  $\boxed{\text{Shift}} \Rightarrow \boxed{9} \Rightarrow \boxed{2} \Rightarrow \boxed{=}$

Khai báo cột tần số:  $\boxed{\text{Shift}} \Rightarrow \boxed{\text{Mode}} \Rightarrow \boxed{\downarrow} \Rightarrow$

$\boxed{4}$  (chọn STAT)  $\Rightarrow \boxed{1}$  (chọn ON)

### Nhập dữ liệu

$\boxed{\text{Mode}} \Rightarrow \boxed{3}$  (chọn STAT)  $\Rightarrow \boxed{1}$  (chọn 1-VAR)

nhập dữ liệu xong nhấn  $\boxed{\text{AC}}$ .

### Xuất kết quả

$\boxed{\text{Shift}} \Rightarrow \boxed{1} \Rightarrow \boxed{4}$  (chọn var)  $\Rightarrow$  chọn kết quả cần xuất ( $n, \bar{x}, \sigma_x, S_x$ ).

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

## V. Ước lượng điểm

Kết quả cần ước lượng được cho bởi 1 số.

$$\mu \approx \bar{x}, \quad \sigma^2 \approx s^2, \quad p \approx f.$$

### Ví dụ 118

Ta lấy mẫu và ước lượng chiều cao trung bình của người Việt Nam. Kết luận chiều cao trung bình là 160cm là một ước lượng điểm.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

Gọi  $\theta$  là đại lượng cần ước lượng  $\theta \in \{\mu, \sigma^2, p\}$ .

Ta cần tính

$$P(\theta \in (a, b)) = \gamma$$

trong đó

- $(a, b)$  khoảng tin cậy.
- $P(\theta \in (a, b))$  xác suất để giá trị  $\theta$  nằm trong đoạn  $(a, b)$ .
- $\gamma$  độ tin cậy.
- $\alpha = 1 - \gamma$  mức ý nghĩa.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

## IV. Lý thuyết ước lượng

Ước lượng là phỏng đoán một giá trị chưa biết của tổng thể bằng cách dựa vào các quan sát trên mẫu lấy ra từ tổng thể.

Thông thường, ta cần ước lượng:

- Trung bình:  $\mu$ .
- Phương sai:  $\sigma^2$ .
- Tỷ lệ:  $p$ .

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

## VI. Ước lượng khoảng

Kết quả cần được ước lượng được cho bởi 1 khoảng  $(a, b)$ .

### Ví dụ 119

Ta lấy mẫu và ước lượng chiều cao trung bình của người Việt Nam. Kết luận chiều cao trung bình trong khoảng  $(158\text{cm}, 162\text{cm})$  là một ước lượng khoảng.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

## VII. Ước lượng khoảng cho trung bình $\mu$

Giả thuyết:

Cho cỡ mẫu  $n$ , tính được  $\bar{x}$  và  $s$ .

Cho độ tin cậy  $\gamma$  (hoặc mức ý nghĩa  $\alpha$ ).

Mục tiêu:

Tìm sai số ước lượng  $\varepsilon$  sao cho

$\mu \in (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$ : khoảng tin cậy đối xứng

$\mu \in (-\infty, \bar{x} + \varepsilon)$ : khoảng tin cậy tối đa.

$\mu \in (\bar{x} - \varepsilon, +\infty)$ : khoảng tin cậy tối thiểu.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

### Phương pháp

**Bước 1.** xác định  $\sigma^2$  và  $n$ .

**Bước 2.** tìm  $C$

	$\begin{cases} \sigma^2 \text{ biết} \\ \sigma^2 \text{ chưa biết} \\ n \geq 30. \end{cases}$	$\begin{cases} \sigma^2 \text{ chưa biết} \\ n < 30. \end{cases}$
ULTB DX	$\varphi(C) = \frac{\gamma}{2} \Rightarrow C$	$C = t(n-1, \frac{\alpha}{2})$
ULTB TĐ	$\varphi(C) = 0,5 - \alpha \Rightarrow C$	$C = t(n-1, \alpha)$
ULTB TT	$\varphi(C) = 0,5 - \alpha \Rightarrow C$	$C = t(n-1, \alpha)$

**Bước 3.** Tìm độ chính xác

$$\varepsilon = \frac{C \cdot \square}{\sqrt{n}}$$

trong đó

- $\square = \sigma$  nếu  $\sigma$  biết.
- $\square = s$  nếu  $\sigma$  chưa biết.

**Bước 4.** Kết luận

$$\mu \in ( \quad , \quad )$$

### Một số bài toán thường gặp

$$\varepsilon = \frac{C \cdot \square}{\sqrt{n}}$$

**Bài toán 1.** Biết  $\gamma = 1 - \alpha$  và  $n$ . Tìm  $\varepsilon$ .

**Bài toán 2.** Bài toán ngược. Biết  $\varepsilon, n$  tìm

$$\gamma = 1 - \alpha.$$

**Bài toán 3.** Biết  $1 - \alpha$  và  $\varepsilon$ . Tìm kích thước mẫu cần điều tra thêm.

$$n = \left\lceil \frac{C^2 \cdot s^2}{\varepsilon^2} \right\rceil + 1$$

trong đó  $\lceil x \rceil$ : phần nguyên trên của  $x$ .  $n$  là kích thước mẫu cần.

### Ví dụ 120

Chủ một kho cung cấp sơn muốn ước lượng lượng sơn chứa trong một thùng được sản xuất từ một dây chuyền công nghệ quốc gia. Biết rằng theo tiêu chuẩn của dây chuyền công nghệ đó, độ lệch tiêu chuẩn của lượng sơn là 0,08 thùng. Điều tra một mẫu 50 thùng được lượng sơn trung bình là 0,97 thùng. Với độ tin cậy 99% hãy ước lượng.

1. Lượng sơn trung bình chứa trong 1 thùng.
2. Lượng sơn trung bình tối thiểu chứa trong 1 thùng.
3. Nếu chủ kho muốn ước lượng sơn trung bình trong thùng đảm bảo độ tin cậy 99% và độ chính xác 2% thì cần điều tra thêm bao nhiêu thùng nữa.

### Giải

$$1. \bar{x} = \quad \sigma = \quad n =$$

Gọi  $\mu$  (thùng) là lượng sơn trung bình chứa trong 1 thùng.

$$\gamma = \quad \Rightarrow \alpha =$$
$$\varphi(C) = \quad \Rightarrow C =$$

$$\Rightarrow \varepsilon =$$

$$\Rightarrow \mu \in$$

$$2. \varphi(C) = 0,5 - \alpha =$$

$$\Rightarrow C =$$

$$\Rightarrow \varepsilon =$$

$$\Rightarrow \mu \in$$

$$3. \varphi(C) = \frac{\gamma}{2} =$$

$$\Rightarrow C =$$

$$\Rightarrow n =$$

### Ví dụ 121

Giá bán của một loại thiết bị (USD) trên thị trường là biến ngẫu nhiên có phân phối chuẩn. Một người định mua loại thiết bị này, khảo sát ngẫu nhiên tại 8 cửa hàng được bán giá trung bình 137,75 USD với độ lệch chuẩn 7,98 USD. Với độ tin cậy 90%, hãy ước lượng giá bán trung bình của thiết bị.

*Giải*

$$\bar{x} = \quad s = \quad n =$$

$$\gamma = \quad \Rightarrow \alpha =$$

$$C = t \left( n - 1, \frac{\alpha}{2} \right) =$$

$$\Rightarrow \varepsilon =$$

$$\Rightarrow \mu \in ( \quad , \quad )$$

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

*Giải*

### Ví dụ 122

Cho mẫu 25 cá thể, đem cân ta có các kết quả

0.63	0.65	0.73	0.60	0.65	0.77	0.82	0.64	0.66
0.72	0.79	0.74	0.45	0.66	0.52	0.62	0.76	0.66
0.84	0.87	0.76	0.54	0.64	0.75	0.77		

1. Tính kỳ vọng, phương sai mẫu.
2. Tìm khoảng tin cậy của cân nặng trung bình với độ tin cậy 95%.
3. Để có độ chính xác 0.01 cần có cỡ mẫu bao nhiêu.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

## VIII. Ước lượng tỉ lệ $p$

*Giả thuyết:* Cho

- Cỡ mẫu  $n$ ,
- Tỉ lệ mẫu  $f = \frac{m}{n}$ ,  $m$  số phần tử có tính chất A,
- Độ tin cậy  $\gamma$ .

*Mục tiêu:*

Tìm sai số ước lượng  $\varepsilon$  sao cho

$p \in (f - \varepsilon, f + \varepsilon)$ : khoảng tin cậy đối xứng

$p \in (-\infty, f + \varepsilon)$ : khoảng tin cậy tối đa.

$p \in (f - \varepsilon, +\infty)$ : khoảng tin cậy tối thiểu.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

### Phương pháp

**Bước 1.** xác định  $n$  và  $f$ .

**Bước 2.** tìm  $C$

ƯLTL ĐX	$\varphi(C) = \frac{\gamma}{2} \Rightarrow C$
ƯLTL TĐ	$\varphi(C) = 0,5 - \alpha \Rightarrow C$
ƯLTL TT	$\varphi(C) = 0,5 - \alpha \Rightarrow C$

**Bước 3.** Tìm độ chính xác

$$\varepsilon = C \cdot \sqrt{\frac{f(1-f)}{n}}$$

**Bước 4.** Kết luận

$$p \in ( \quad , \quad )$$

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

### Một số bài toán thường gặp

$$\varepsilon = C \cdot \sqrt{\frac{f(1-f)}{n}}$$

**Bài toán 1.** Biết  $\gamma = 1 - \alpha$  và  $n$ . Tìm  $\varepsilon$ .

**Bài toán 2.** Bài toán ngược. Biết  $\varepsilon, n$  tìm

$$\gamma = 1 - \alpha.$$

**Bài toán 3.** Biết  $1 - \alpha$  và  $\varepsilon$ . Tìm kích thước mẫu cần điều tra thêm.

$$n = \left\lceil f(1-f) \cdot \frac{C^2}{\varepsilon^2} \right\rceil + 1$$

trong đó  $\lceil x \rceil$ : phần nguyên trên của  $x$ .  $n$  kích thước mẫu cần điều tra.

GV: Hoàng Đức Thắng hdtthang@sgu.edu.vn

XÁC SUẤT THỐNG KÊ A (864001, 45 Tiết) (Lý thuyết 30)

### Ví dụ 123

Quan sát ngẫu nhiên 200 lọ thuốc trong một lô hàng lớn, ta thấy có 17 lọ không đạt tiêu chuẩn. Hãy ước lượng tỉ lệ thuốc không đạt chuẩn ở độ tin cậy 95%.

*Giải*

$$- f = \frac{17}{200} = 0,085$$

$$- \varphi(C) = \frac{\gamma}{2} = 0,475 \Rightarrow C = 1,96$$

$$- \varepsilon = 1,96 \cdot \sqrt{\frac{0,085(1 - 0,085)}{200}} = 0,039.$$

$$- p \in (0,085 - 0,039; 0,085 + 0,039) = (0,046; 0,124)$$

### Ví dụ 124

Trong số liệu từ cuộc bầu cử tổng thống năm 2004, một bang quan trọng là bang Ohio đã cho kết quả sau đây: đã có 2020 người trả lời trong các cuộc thăm dò xuất cảnh và 768 là sinh viên tốt nghiệp đại học. Trong số các sinh viên tốt nghiệp đại học có 412 bầu cho George Bush. Xây dựng khoảng tin cậy 95% cho tỉ lệ sinh viên tốt nghiệp đại học bầu cho George Bush.

### Ví dụ 125

Lấy ngẫu nhiên 200 sản phẩm trong một kho hàng thấy có 25 phế phẩm.

- Nếu muốn độ chính xác của ước lượng là  $\varepsilon = 0.035$  thì độ tin cậy của ước lượng là bao nhiêu?
- Nếu muốn độ chính xác là 0.001, độ tin cậy 95% thì cần kiểm tra thêm bao nhiêu sản phẩm.

*Giải*