



Phần 2. Thống
kê.

Bài 1. Lý
thuyết mẫu



PHẦN II. LÝ THUYẾT THỐNG KÊ (Statistical theory)

Chương VI. MẪU THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ

§1. Lý thuyết mẫu

§2. Ước lượng điểm

§3. Ước lượng khoảng

§1. LÝ THUYẾT MẪU

1.1. Mẫu và tổng thể

- Tập hợp tất cả phần tử là các đối tượng mà ta nghiên cứu được gọi là **tổng thể**. Số phần tử của tổng thể được gọi là **kích thước** của tổng thể (thường rất lớn).

➤ Chương 6. Mẫu thống kê & Ước lượng tham số

- Từ tổng thể ta chọn ra n phần tử thì n phần tử đó được gọi là một ***mẫu có kích thước n (cỡ mẫu)***.
- Mẫu được chọn ngẫu nhiên một cách khách quan được gọi là ***mẫu ngẫu nhiên***.
- Có hai cách lấy mẫu:
 - ***Mẫu có hoàn lại***: phần tử vừa quan sát xong được trả lại cho tổng thể trước khi quan sát lần sau.
 - ***Mẫu không hoàn lại***: Phần tử vừa quan sát xong không được trả lại cho tổng thể.

Khi mẫu có kích thước lớn thì ta không phân biệt mẫu có hoàn lại hay không hoàn lại.

- **Mẫu định tính** là mẫu mà ta chỉ quan tâm đến các phần tử của nó có tính chất A nào đó hay không.
- **Mẫu định lượng** là mẫu mà ta quan tâm đến các yếu tố về lượng (như chiều dài, cân nặng,...) của các phần tử có trong mẫu.
- Gọi X_1, X_2, \dots, X_n là những kết quả quan sát. Ta xem như đã quan sát n lần, mỗi lần ta được một biến ngẫu nhiên X_i ($i = 1, \dots, n$).

Do ta thường lấy mẫu trong tổng thể có rất nhiều phần tử nên X_1, X_2, \dots, X_n được xem là độc lập và có cùng phân phối xác suất.

1.2. Sắp xếp mẫu dựa vào số liệu thực nghiệm

a) Sắp xếp theo dạng bảng

VD 1. Kiểm tra ngẫu nhiên 50 sinh viên. Ta sắp xếp điểm số X thu được theo thứ tự tăng dần và số sinh viên n có điểm tương ứng vào bảng như sau:

X (điểm)	2	4	5	6	7	8	9	10
n (số SV)	4	6	20	10	5	2	2	1

b) Sắp xếp theo dạng khoảng

VD 2. Đo chiều cao X (cm) của $n = 100$ thanh niên. Vì chiều cao khác nhau nên để tiện việc sắp xếp, người ta chia chiều cao thành nhiều khoảng.

Các thanh niên có chiều cao trong cùng 1 khoảng được xem là cao như nhau. Khi đó, ta có bảng số liệu ở dạng khoảng như sau:

X	148-152	152-156	156-160	160-164	164-168
n	5	20	35	25	15

Khi cần tính toán, người ta chọn *số trung bình* của mỗi khoảng để đưa số liệu trên về dạng bảng:

X	150	154	158	162	166
n	5	20	35	25	15

Chú ý

Đối với trường hợp số liệu được cho dưới dạng liệt kê thì ta sắp xếp lại ở dạng bảng.

VD 3. Theo dõi mức nguyên liệu hao phí để sản xuất ra một đơn vị sản phẩm ở một nhà máy, ta thu được các số liệu sau (đơn vị: gam):

20; 22; 21; 20; 22; 22; 20; 19; 20; 22; 21;
19; 19; 20; 18; 19; 20; 20; 18; 19; 20; 20;
21; 20; 18; 19; 19; 21; 22; 21; 21; 20; 19.

Hãy sắp xếp số liệu trên dưới dạng bảng ?

1.3. Các đặc trưng mẫu

Xét một mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) , ta có các đặc trưng mẫu như sau.

a) Trung bình mẫu

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Để đơn giản, ta dùng ký hiệu $\bar{X} = \bar{X}_n$.

b) Phương sai mẫu

- Phương sai mẫu:

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

- Phương sai mẫu hiệu chỉnh:

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

- Trong tính toán cụ thể, ta sử dụng công thức:

$$S^2 = \frac{n}{n-1} \left[\overline{X^2} - (\overline{X})^2 \right] = \frac{n}{n-1} \hat{S}^2.$$

$$\text{Với } \overline{X^2} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

c) Tỷ lệ mẫu

Xét mẫu định tính với các biến X_i ($i = 1, \dots, n$) có phân phối Bernoulli $B(1; p)$:

$$X_i = \begin{cases} 0, & \text{nếu phaà tồû khoâng coù tính chaấA} \\ 1, & \text{nếu phaà tồû coù tính chaấA.} \end{cases}$$

Nếu mẫu có m phần tử có tính chất A thì tỷ lệ mẫu là:

$$F = F_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{m}{n}.$$

d) Liên hệ giữa đặc trưng của mẫu và tổng thể

Các đặc trưng mẫu \bar{X} , S^2 , F là các thống kê dùng để nghiên cứu các đặc trưng μ , σ^2 , p tương ứng của tổng thể. Từ luật số lớn ta có:

$$F \rightarrow p, \bar{X} \rightarrow \mu, S^2 \rightarrow \sigma^2 \text{ (theo xác suất).}$$

Ví dụ 1. Để khảo sát trọng lượng X của một loại vật nuôi trong nông trại, người ta quan sát một mẫu và thu được kết quả sau

$X(\text{kg})$	36	42	48	54	60	66	72
Số con	15	12	25	18	10	10	10

- Tính cỡ mẫu, trung bình mẫu, phương sai chưa hiệu chỉnh, phương sai hiệu chỉnh của mẫu.
- Vật nuôi được xếp loại A nếu trọng lượng lớn hơn 49 (kg). Tính tỉ lệ vật nuôi loại A có trong mẫu.

Ví dụ 2. Cân thử 100 trái quýt trong vườn, ta thu được kết quả sau

X(g)	40	50	60	70	80	90	100	110
Số trái	3	10	12	15	28	16	11	5

- Tính cỡ mẫu, trung bình mẫu, phương sai chưa hiệu chỉnh, phương sai hiệu chỉnh của mẫu.
- Quýt được xếp loại A nếu trọng lượng lớn hơn 75 (g). Tính tỉ lệ quýt loại A có trong mẫu.

Ví dụ 3. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát trong kho và thu được kết quả như sau

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	25	18	18	14	18

- Tính cỡ mẫu, trung bình mẫu, phương sai chưa hiệu chỉnh, phương sai hiệu chỉnh của mẫu.
- Sản phẩm được xếp loại A nếu kích thước lớn hơn 25 (cm). Tính tỉ lệ sản phẩm loại A có trong mẫu.

Ví dụ 4. Thời gian hoàn thành một sản phẩm của công nhân là X (phút) được cho trong bản sau

$X(\text{phút})$	12-14	14-16	16-18	18-20	20-22
Số công nhân	1	5	14	20	10

- Tính cỡ mẫu, trung bình mẫu, phương sai chưa hiệu chỉnh, phương sai hiệu chỉnh của mẫu.
- Công nhân được cho là thạo nghề nếu thời gian hoàn thành một sản phẩm dưới 18 phút. Tính tỉ lệ công nhân thạo nghề