

# Lý thuyết Xác suất và Thống kê Toán học

PGS.TS. Trần Lộc Hùng

Tp. Hồ Chí Minh, 10/ 2013

Ngày 12 tháng 10 năm 2013

TRƯỜNG ĐẠI HỌC TÀI CHÍNH-MARKETING  
KHOA CƠ BẢN, BỘ MÔN TOÁN-THỐNG KÊ

-----

PGS. TS. TRẦN LỘC HÙNG

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC

Tp. Hồ Chí Minh, 10/2013

# Lý thuyết Xác suất và Thống kê Toán học

PGS.TS. Trần Lộc Hùng

Tp. Hồ Chí Minh, 10/ 2013

Ngày 12 tháng 10 năm 2013

Chương 5. Mẫu ngẫu nhiên và các đặc trưng

# Từ khóa (Key Words)

- Tổng thể và mẫu

# Từ khóa (Key Words)

- Tổng thể và mẫu
- Hàm phân phối thực nghiệm

# Từ khóa (Key Words)

- Tổng thể và mẫu
- Hàm phân phối thực nghiệm
- Trung bình mẫu, phương sai mẫu, tần suất mẫu

# Từ khóa (Key Words)

- Tổng thể và mẫu
- Hàm phân phối thực nghiệm
- Trung bình mẫu, phương sai mẫu, tần suất mẫu
- Phân phối xác suất của các đặc trưng mẫu

# Lịch sử Thống kê

## 1 Thống kê thời cổ đại



# Lịch sử Thống kê

- 1 Thống kê thời cổ đại
- 2 Nghĩa Hán - Việt của từ "Thống kê"

# Lịch sử Thống kê

- 1 Thống kê thời cổ đại
- 2 Nghĩa Hán - Việt của từ "Thống kê"
- 3 Nghĩa tiếng Anh của từ "Statistics"

# Lịch sử Thống kê

- 1 Thống kê thời cổ đại
- 2 Nghĩa Hán - Việt của từ "Thống kê"
- 3 Nghĩa tiếng Anh của từ "Statistics"
- 4 Phân loại Thống kê (thống kê mô tả, thống kê cổ điển, thống kê lý thuyết, nguyên lý thống kê, Thống kê Bayes)

# Lịch sử Thống kê

- 1 Thống kê thời cổ đại
- 2 Nghĩa Hán - Việt của từ "Thống kê"
- 3 Nghĩa tiếng Anh của từ "Statistics"
- 4 Phân loại Thống kê (thống kê mô tả, thống kê cổ điển, thống kê lý thuyết, nguyên lý thống kê, Thống kê Bayes)
- 5 Thống kê tại Việt Nam

# Chương 5. Mẫu ngẫu nhiên và các đặc trưng

## ❶ Mẫu ngẫu nhiên

# Chương 5. Mẫu ngẫu nhiên và các đặc trưng

- ① Mẫu ngẫu nhiên
- ② Hàm phân phối thực nghiệm

# Chương 5. Mẫu ngẫu nhiên và các đặc trưng

- ① Mẫu ngẫu nhiên
- ② Hàm phân phối thực nghiệm
- ③ Trung bình mẫu

## Chương 5. Mẫu ngẫu nhiên và các đặc trưng

- 1 Mẫu ngẫu nhiên
- 2 Hàm phân phối thực nghiệm
- 3 Trung bình mẫu
- 4 Phương sai mẫu và phương sai điều chỉnh



## Chương 5. Mẫu ngẫu nhiên và các đặc trưng

- 1 Mẫu ngẫu nhiên
- 2 Hàm phân phối thực nghiệm
- 3 Trung bình mẫu
- 4 Phương sai mẫu và phương sai điều chỉnh
- 5 Tần suất mẫu

## Chương 5. Mẫu ngẫu nhiên và các đặc trưng

- 1 Mẫu ngẫu nhiên
- 2 Hàm phân phối thực nghiệm
- 3 Trung bình mẫu
- 4 Phương sai mẫu và phương sai điều chỉnh
- 5 Tần suất mẫu
- 6 Các phân phối xác suất của các đặc trưng mẫu

## Chương 5. Mẫu ngẫu nhiên và các đặc trưng

- 1 Mẫu ngẫu nhiên
- 2 Hàm phân phối thực nghiệm
- 3 Trung bình mẫu
- 4 Phương sai mẫu và phương sai điều chỉnh
- 5 Tần suất mẫu
- 6 Các phân phối xác suất của các đặc trưng mẫu
- 7 Bài tập

## 5.1 Mẫu ngẫu nhiên

tổng thể

$$\Omega = \{x_1, x_2, \dots, x_n, \dots\} -$$

tập tất cả các số liệu sinh ra từ một biến ngẫu nhiên  $X$

và

mẫu ngẫu nhiên

$$\omega_n = \{x_1, x_2, \dots, x_n\} \subseteq \Omega$$

là tập  $n$  các số liệu sinh ra từ  $n$  quan sát biến ngẫu nhiên  $X$ .

# Ví dụ

## Chiều cao sinh viên

$$\Omega = \{172, 168, \dots, 159, \dots\} -$$

tập tất cả các số liệu sinh ra từ một biến ngẫu nhiên  $X$ -chiều cao của sinh viên đại học

và

## mẫu số liệu của 150 sinh viên

$$\omega_{150} = \{158, 169, 171, \dots, 173\}$$

gồm 150 số liệu về chiều cao của 150 sinh viên.

# Giải thích

- 1 Biến ngẫu nhiên  $X$  đặc trưng cho một tính chất nào đó của tổng thể  $\Omega$ ) như chiều cao, trọng lượng, tuổi thọ, mức sống, ...

# Giải thích

- 1 Biến ngẫu nhiên  $X$  đặc trưng cho một tính chất nào đó của tổng thể  $\Omega$ ) như chiều cao, trọng lượng, tuổi thọ, mức sống, ...
- 2 Mẫu ngẫu nhiên cỡ  $n$ ,  $\omega_n$  sinh ra bởi  $n$  quan sát độc lập một biến ngẫu nhiên  $X$

# Giải thích

- 1 Biến ngẫu nhiên  $X$  đặc trưng cho một tính chất nào đó của tổng thể  $\Omega$ ) như chiều cao, trọng lượng, tuổi thọ, mức sống, ...
- 2 Mẫu ngẫu nhiên cỡ  $n$ ,  $\omega_n$  sinh ra bởi  $n$  quan sát độc lập một biến ngẫu nhiên  $X$
- 3 Mẫu ngẫu nhiên  $\omega_n = \{X_1, X_2, \dots, X_n\}$  gồm  $n$  biến ngẫu nhiên độc lập cùng phân phối



# Giải thích

- 1 Biến ngẫu nhiên  $X$  đặc trưng cho một tính chất nào đó của tổng thể  $\Omega$ ) như chiều cao, trọng lượng, tuổi thọ, mức sống, ...
- 2 Mẫu ngẫu nhiên cỡ  $n$ ,  $\omega_n$  sinh ra bởi  $n$  quan sát độc lập một biến ngẫu nhiên  $X$
- 3 Mẫu ngẫu nhiên  $\omega_n = \{X_1, X_2, \dots, X_n\}$  gồm  $n$  biến ngẫu nhiên độc lập cùng phân phối
- 4 Số  $n$  được gọi là kích thước mẫu hay cỡ mẫu

## 5.2 Hàm phân phối thực nghiệm (hàm phân phối mẫu)

### Định nghĩa

$$F_n(x) = \begin{cases} 0, & \text{nếu } x < x_{[1]} \\ \frac{n_j}{n}, & \text{nếu có } n_j \text{ giá trị của mẫu bé hơn } x \\ 1, & \text{nếu } x \geq x_{[n]}. \end{cases}$$

- ❶ Mẫu ngẫu nhiên  $\omega_n = \{X_{[1]}, X_{[2]}, \dots, X_{[n]}\}$  là mẫu có thứ tự

## 5.2 Hàm phân phối thực nghiệm (hàm phân phối mẫu)

### Định nghĩa

$$F_n(x) = \begin{cases} 0, & \text{nếu } x < x_{[1]} \\ \frac{n_j}{n}, & \text{nếu có } n_j \text{ giá trị của mẫu bé hơn } x \\ 1, & \text{nếu } x \geq x_{[n]}. \end{cases}$$

- 1 Mẫu ngẫu nhiên  $\omega_n = \{X_{[1]}, X_{[2]}, \dots, X_{[n]}\}$  là mẫu có thứ tự
- 2  $x_{[1]} = \min\{x_1, x_2, \dots, x_n\}$ - giá trị bé nhất của mẫu

## 5.2 Hàm phân phối thực nghiệm (hàm phân phối mẫu)

### Định nghĩa

$$F_n(x) = \begin{cases} 0, & \text{nếu } x < x_{[1]} \\ \frac{n_j}{n}, & \text{nếu có } n_j \text{ giá trị của mẫu bé hơn } x \\ 1, & \text{nếu } x \geq x_{[n]}. \end{cases}$$

- 1 Mẫu ngẫu nhiên  $\omega_n = \{X_{[1]}, X_{[2]}, \dots, X_{[n]}\}$  là mẫu có thứ tự
- 2  $x_{[1]} = \min\{x_1, x_2, \dots, x_n\}$ - giá trị bé nhất của mẫu
- 3  $x_{[n]} = \max\{x_1, x_2, \dots, x_n\}$ - giá trị lớn nhất của mẫu

# Định lý Glivenko-Cantelli

## Định lý

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathfrak{R}} |F_n(x) - F_X(x)| = 0\right) = 1$$

- 1 Hàm phân phối lý thuyết  $F_X(x) = P(X < x)$  còn được gọi là hàm phân phối chính xác

# Định lý Glivenko-Cantelli

## Định lý

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathfrak{R}} |F_n(x) - F_X(x)| = 0\right) = 1$$

- 1 Hàm phân phối lý thuyết  $F_X(x) = P(X < x)$  còn được gọi là hàm phân phối chính xác
- 2 Khi  $n$  lớn hàm phân phối thực nghiệm có giá trị gần đúng với hàm phân phối lý thuyết

$$F_n(x) \approx F_X(x)$$

# Định lý Glivenko-Cantelli

## Định lý

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathfrak{R}} |F_n(x) - F_X(x)| = 0\right) = 1$$

- 1 Hàm phân phối lý thuyết  $F_X(x) = P(X < x)$  còn được gọi là hàm phân phối chính xác
- 2 Khi  $n$  lớn hàm phân phối thực nghiệm có giá trị gần đúng với hàm phân phối lý thuyết

$$F_n(x) \approx F_X(x)$$

- 3 Làm thực nghiệm cần xử lý tập lớn số liệu để thông tin nhận được đủ tin cậy

## Lưu ý

Giả sử  $X$  là biến ngẫu nhiên đặc trưng cho tính chất nào đó (cần nghiên cứu) của tổng thể  $\Omega$  Khi đó,

- 1 Hàm phân phối lý thuyết  $F_X(x) = P(X \leq x)$  còn được gọi là hàm phân phối chính xác



## Lưu ý

Giả sử  $X$  là biến ngẫu nhiên đặc trưng cho tính chất nào đó (cần nghiên cứu) của tổng thể  $\Omega$  Khi đó,

- ① Hàm phân phối lý thuyết  $F_X(x) = P(X \leq x)$  còn được gọi là hàm phân phối chính xác
- ② Kỳ vọng toán  $E(X) = \mu = \sum_{j=1}^n x_j p_j$  còn được gọi là trung bình tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)

## Lưu ý

Giả sử  $X$  là biến ngẫu nhiên đặc trưng cho tính chất nào đó (cần nghiên cứu) của tổng thể  $\Omega$  Khi đó,

- ① Hàm phân phối lý thuyết  $F_X(x) = P(X \leq x)$  còn được gọi là hàm phân phối chính xác
- ② Kỳ vọng toán  $E(X) = \mu = \sum_{j=1}^n x_j p_j$  còn được gọi là trung bình tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)
- ③ Phương sai  $D(X) = \sigma^2 = \sum_{j=1}^n |x_j - \mu|^2 p_j$  còn được gọi là phương sai tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)

## Lưu ý

Giả sử  $X$  là biến ngẫu nhiên đặc trưng cho tính chất nào đó (cần nghiên cứu) của tổng thể  $\Omega$  Khi đó,

- 1 Hàm phân phối lý thuyết  $F_X(x) = P(X \leq x)$  còn được gọi là hàm phân phối chính xác
- 2 Kỳ vọng toán  $E(X) = \mu = \sum_{j=1}^n x_j p_j$  còn được gọi là trung bình tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)
- 3 Phương sai  $D(X) = \sigma^2 = \sum_{j=1}^n |x_j - \mu|^2 p_j$  còn được gọi là phương sai tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)
- 4 Xác suất  $p_j = P(X = x_j) = p, j = 1, 2, \dots, n$  còn được gọi là tần suất tổng thể

## Lưu ý

Giả sử  $X$  là biến ngẫu nhiên đặc trưng cho tính chất nào đó (cần nghiên cứu) của tổng thể  $\Omega$  Khi đó,

- 1 Hàm phân phối lý thuyết  $F_X(x) = P(X < x)$  còn được gọi là hàm phân phối chính xác
- 2 Kỳ vọng toán  $E(X) = \mu = \sum_j^n x_j p_j$  còn được gọi là trung bình tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)
- 3 Phương sai  $D(X) = \sigma^2 = \sum_j^n |x_j - \mu|^2 p_j$  còn được gọi là phương sai tổng thể (xét trường hợp  $X$  là biến ngẫu nhiên rời rạc)
- 4 Xác suất  $p_j = P(X = x_j) = p, j = 1, 2, \dots, n$  còn được gọi là tần suất tổng thể
- 5 Các đặc trưng  $\mu, \sigma^2, p$  được gọi là các tham số tổng thể (tham số lý thuyết)

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Trung bình mẫu

$$E_n(X) = \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

- 1 Trung bình mẫu  $\bar{X}$  là một biến ngẫu nhiên.

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Trung bình mẫu

$$E_n(X) = \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

- 1 Trung bình mẫu  $\bar{X}$  là một biến ngẫu nhiên.
- 2 Không phụ thuộc vào cỡ mẫu, kỳ vọng của trung bình mẫu (thực nghiệm) bằng trung bình lý thuyết:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n} \sum_{j=1}^n E(X_j) = \frac{1}{n} \sum_{j=1}^n \mu = \mu$$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Trung bình mẫu

$$E_n(X) = \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

- 1 Trung bình mẫu  $\bar{X}$  là một biến ngẫu nhiên.
- 2 Không phụ thuộc vào cỡ mẫu, kỳ vọng của trung bình mẫu (thực nghiệm) bằng trung bình lý thuyết:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n} \sum_{j=1}^n E(X_j) = \frac{1}{n} \sum_{j=1}^n \mu = \mu$$

- 3 Phương sai của trung bình mẫu  $D(\bar{X}) = \frac{\sigma^2}{n} \downarrow 0$  khi  $n \uparrow \infty$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Phương sai mẫu

$$D_n(X) = S_n^2 = \frac{1}{n} \sum_{j=1}^n |X_j - \bar{X}|^2$$

- 1 Phương sai mẫu  $S_n^2$  là một biến ngẫu nhiên.



## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Phương sai mẫu

$$D_n(X) = S_n^2 = \frac{1}{n} \sum_{j=1}^n |X_j - \bar{X}|^2$$

- ❶ Phương sai mẫu  $S_n^2$  là một biến ngẫu nhiên.
- ❷ Kỳ vọng của phương sai mẫu không bằng phương sai lý thuyết:

$$E(S_n^2) = E\left(\frac{1}{n} \sum_{j=1}^n |X_j - \bar{X}|^2\right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Phương sai mẫu

$$D_n(X) = S_n^2 = \frac{1}{n} \sum_{j=1}^n |X_j - \bar{X}|^2$$

- 1 Phương sai mẫu  $S_n^2$  là một biến ngẫu nhiên.
- 2 Kỳ vọng của phương sai mẫu không bằng phương sai lý thuyết:

$$E(S_n^2) = E\left(\frac{1}{n} \sum_{j=1}^n |X_j - \bar{X}|^2\right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

- 3 Phải sử dụng phương sai mẫu điều chỉnh  $\hat{S}_n^2$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Phương sai mẫu điều chỉnh

$$D_n(\hat{X}) = \hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n |X_j - \bar{X}|^2$$

- 1 Phương sai mẫu điều chỉnh  $\hat{S}_n^2$  là một biến ngẫu nhiên.

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Phương sai mẫu điều chỉnh

$$D_n(\hat{X}) = \hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n |X_j - \bar{X}|^2$$

- 1 Phương sai mẫu điều chỉnh  $\hat{S}_n^2$  là một biến ngẫu nhiên.
- 2 Kỳ vọng của phương sai mẫu bằng phương sai lý thuyết:

$$E(\hat{S}_n^2) = E\left(\frac{1}{n-1} \sum_{j=1}^n |X_j - \bar{X}|^2\right) = \sigma^2$$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Phương sai mẫu điều chỉnh

$$D_n(\hat{X}) = \hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n |X_j - \bar{X}|^2$$

- 1 Phương sai mẫu điều chỉnh  $\hat{S}_n^2$  là một biến ngẫu nhiên.
- 2 Kỳ vọng của phương sai mẫu bằng phương sai lý thuyết:

$$E(\hat{S}_n^2) = E\left(\frac{1}{n-1} \sum_{j=1}^n |X_j - \bar{X}|^2\right) = \sigma^2$$

- 3 Trong thực tế thường sử dụng phương sai mẫu điều chỉnh  $\hat{S}_n^2$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

Phương sai mẫu điều chỉnh và phương sai mẫu

$$\hat{S}_n^2 = \frac{n}{n-1} S_n^2$$

- ① Khi  $n$  lớn, hai giá trị đó xấp xỉ nhau, vì

$$\lim_{n \rightarrow \infty} \frac{n}{n-1} = 1$$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

Phương sai mẫu điều chỉnh và phương sai mẫu

$$\hat{S}_n^2 = \frac{n}{n-1} S_n^2$$

- ① Khi  $n$  lớn, hai giá trị đó xấp xỉ nhau, vì

$$\lim_{n \rightarrow \infty} \frac{n}{n-1} = 1$$

- ② Trong nhiều phần mềm thống kê,  $\hat{S}_n^2$  được thay cho  $S_n^2$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

Dùng trong tính toán

$$S_n^2 = \frac{1}{n} \sum_{j=1}^n X_j^2 - (\bar{X})^2$$



## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Tần suất mẫu

$$f_n = \frac{k}{n}$$

- ① n là cỡ mẫu, k là số các thành phần mẫu thỏa mãn điều kiện của người chọn mẫu

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Tần suất mẫu

$$f_n = \frac{k}{n}$$

- 1 n là cỡ mẫu, k là số các thành phần mẫu thỏa mãn điều kiện của người chọn mẫu
- 2 Khi n lớn, tần suất thực nghiệm (mẫu) xấp xỉ tần suất lý thuyết p

$$f_n \approx p$$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Tần suất mẫu

$$f_n = \frac{k}{n}$$

- 1 n là cỡ mẫu, k là số các thành phần mẫu thỏa mãn điều kiện của người chọn mẫu
- 2 Khi n lớn, tần suất thực nghiệm (mẫu) xấp xỉ tần suất lý thuyết p

$$f_n \approx p$$

- 3 Đây là bản chất của xác suất theo quan điểm thống kê và là Luật yếu các số lớn Bernoulli

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

Mô men mẫu

$$\nu_k = \frac{1}{n} \sum_{j=1}^n X_j^k$$

Mô men trung tâm mẫu

$$\mu_k = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^k$$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Trung vị mẫu (Median mẫu)

Trong một mẫu thứ tự  $\{x_{[1]}, x_{[2]}, \dots, x_{[n]}\}$  trung vị mẫu là số đứng ở giữa nếu  $n$  lẻ, là trung bình cộng hai số đứng giữa nếu  $n$  chẵn

- ① Mẫu  $A = \{3, 4, 4, 5, 5, 5, 6, 7, 8, 9, 10\}$  có trung vị mẫu  $\text{med}=5$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Trung vị mẫu (Median mẫu)

Trong một mẫu thứ tự  $\{x_{[1]}, x_{[2]}, \dots, x_{[n]}\}$  trung vị mẫu là số đứng ở giữa nếu  $n$  lẻ, là trung bình cộng hai số đứng giữa nếu  $n$  chẵn

- ① Mẫu  $A = \{3, 4, 4, 5, 5, 5, 6, 7, 8, 9, 10\}$  có trung vị mẫu  $\text{med}=5$
- ② Mẫu  $B = \{3, 4, 4, 5, 5, 5, 6, 7, 8, 9, 10, 11\}$  có trung vị mẫu  $\text{med}=(5+6)/2=5.5$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Số trội mẫu (Mod mẫu)

Trong một mẫu  $\{x_{[1]}, x_{[2]}, \dots, x_{[n]}\}$  số trội mẫu là các số xuất hiện nhiều lần nhất

- 1 Mẫu  $A = \{3, 4, 4, 5, 5, 5, 6, 7, 8, 9, 10\}$  có số trội mẫu  $\text{mod}=5$

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Số trội mẫu (Mod mẫu)

Trong một mẫu  $\{x_{[1]}, x_{[2]}, \dots, x_{[n]}\}$  số trội mẫu là các số xuất hiện nhiều lần nhất

- 1 Mẫu  $A = \{3, 4, 4, 5, 5, 5, 6, 7, 8, 9, 10\}$  có số trội mẫu  $\text{mod}=5$
- 2 Mẫu  $B = \{3, 4, 4, 4, 5, 5, 6, 7, 8, 9, 10, 11\}$  có số trội mẫu  $\text{mod}=4$



## 5.3 Các đặc trưng của mẫu ngẫu nhiên

Hệ số bất đối xứng mẫu

$$\gamma_1 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}}$$

Hệ số nhọn mẫu

$$\gamma_2 = \frac{\mu_4}{(\mu_2)^2} - 3$$

Giả sử  $\omega_n$  là một mẫu ngẫu nhiên có các đặc trưng mẫu. Khi đó,

- 1 Đặc trưng về vị trí: trung bình mẫu, trung vị mẫu, số trội mẫu.

Giả sử  $\omega_n$  là một mẫu ngẫu nhiên có các đặc trưng mẫu. Khi đó,

- ① Đặc trưng về vị trí: trung bình mẫu, trung vị mẫu, số trội mẫu.
- ② Đặc trưng về sự phân tán: phương sai mẫu, phương sai mẫu điều chỉnh, độ rộng mẫu ( $R_n = x_{[n]} - x_{[1]}$ )

Giả sử  $\omega_n$  là một mẫu ngẫu nhiên có các đặc trưng mẫu. Khi đó,

- ① Đặc trưng về vị trí: trung bình mẫu, trung vị mẫu, số trội mẫu.
- ② Đặc trưng về sự phân tán: phương sai mẫu, phương sai mẫu điều chỉnh, độ rộng mẫu ( $R_n = x_{[n]} - x_{[1]}$ )
- ③ Đặc trưng về hình dáng của phân phối (của hàm mật độ): hệ số bất đối xứng, hệ số nhọn

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Ví dụ

Trọng lượng tính bằng kg của 45 sinh viên

43, 50, 58, 65, 55, 60, 53, 65, 71, 63, 66, 76, 54, 61, 66

49, 55, 58, 65, 54, 61, 54, 65, 70, 64, 61, 75, 56, 69, 67

54, 55, 51, 61, 59, 61, 56, 65, 71, 64, 61, 70, 59, 61, 65

- 1 Tính các đặc trưng của mẫu trên

## 5.3 Các đặc trưng của mẫu ngẫu nhiên

### Ví dụ

Trọng lượng tính bằng kg của 45 sinh viên

43, 50, 58, 65, 55, 60, 53, 65, 71, 63, 66, 76, 54, 61, 66

49, 55, 58, 65, 54, 61, 54, 65, 70, 64, 61, 75, 56, 69, 67

54, 55, 51, 61, 59, 61, 56, 65, 71, 64, 61, 70, 59, 61, 65

- 1 Tính các đặc trưng của mẫu trên
- 2 Vẽ biểu đồ mô tả mẫu trên

## Giải:

Sử dụng ngôn ngữ R, với lệnh tạo mẫu  $x < -c()$

① lệnh `summary(x)` cho biết

$$\text{Min} = 43. \quad \text{1stQu.} = 55 \quad \text{Median} = 61.$$

$$\text{Mean} = 60.93 \quad \text{3rdQu.} = 65 \quad \text{Max.} = 76$$

## Giải:

Sử dụng ngôn ngữ R, với lệnh tạo mẫu  $x < -c()$

- 1 lệnh `summary(x)` cho biết

$$\text{Min} = 43. \quad \text{1stQu.} = 55 \quad \text{Median} = 61.$$

$$\text{Mean} = 60.93 \quad \text{3rdQu.} = 65 \quad \text{Max.} = 76$$

- 2 Lệnh `mean(x)` cho biết  $\bar{X} = 60.93333$



## Giải:

Sử dụng ngôn ngữ R, với lệnh tạo mẫu  $x < -c()$

- 1 lệnh `summary(x)` cho biết

$$\text{Min} = 43. \quad \text{1stQu.} = 55 \quad \text{Median} = 61.$$

$$\text{Mean} = 60.93 \quad \text{3rdQu.} = 65 \quad \text{Max.} = 76$$

- 2 Lệnh `mean(x)` cho biết  $\bar{X} = 60.93333$
- 3 Lệnh `var(x)` cho biết  $S_{45}^2 = 49.65455$

## Giải:

Sử dụng ngôn ngữ R, với lệnh tạo mẫu  $x < -c()$

- 1 lệnh `summary(x)` cho biết

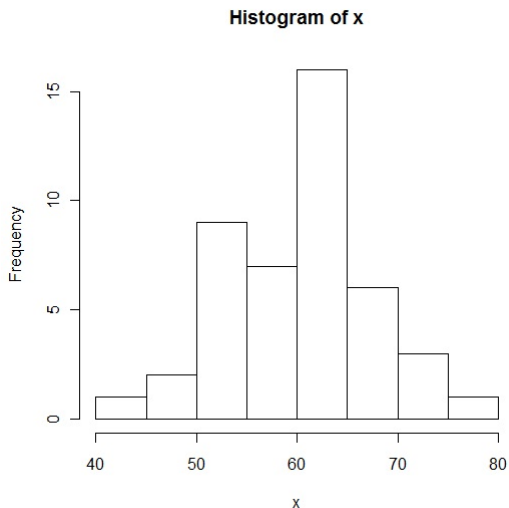
$$\text{Min} = 43. \quad \text{1stQu.} = 55 \quad \text{Median} = 61.$$

$$\text{Mean} = 60.93 \quad \text{3rdQu.} = 65 \quad \text{Max.} = 76$$

- 2 Lệnh `mean(x)` cho biết  $\bar{X} = 60.93333$
- 3 Lệnh `var(x)` cho biết  $S_{45}^2 = 49.65455$
- 4 Lệnh `range(x)` cho biết  $R_{45} = 76 - 43 = 33$

# Biểu đồ 1

- Sử dụng ngôn ngữ R, với lệnh tạo `hist(x)`



## 5.4 Các phân phối của các đại lượng thống kê

- ❶ Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu ngẫu nhiên

Định nghĩa

$$f(\omega_n) = f(X_1, X_2, \dots, X_n)$$

## 5.4 Các phân phối của các đại lượng thống kê

- 1 Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu ngẫu nhiên
- 2 Mọi hàm của mẫu được gọi là một thống kê

Định nghĩa

$$f(\omega_n) = f(X_1, X_2, \dots, X_n)$$

# Các ví dụ

- ① Trung bình mẫu  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  là một thống kê

## Các ví dụ

- 1 Trung bình mẫu  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  là một thống kê
- 2 Phương sai mẫu  $S_n^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$  là một thống kê

## Các ví dụ

- ① Trung bình mẫu  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  là một thống kê
- ② Phương sai mẫu  $S_n^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$  là một thống kê
- ③ Phương sai mẫu điều chỉnh  $\hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$  là một thống kê



## Các ví dụ

- ① Trung bình mẫu  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  là một thống kê
- ② Phương sai mẫu  $S_n^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$  là một thống kê
- ③ Phương sai mẫu điều chỉnh  $\hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$  là một thống kê
- ④ Tần suất mẫu  $f_n = \frac{k}{n}$  là một thống kê

# Phân phối của trung bình mẫu $\bar{X}$

## Định lý

Nếu  $(X_1, X_2, \dots, X_n)$  là một mẫu sinh từ biến ngẫu nhiên có phân phối chuẩn chính tắc  $\mathcal{N}(0, 1)$ , thì

- 1 Trung bình mẫu  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  có phân phối chuẩn  $\mathcal{N}(\mu, \sigma^2)$ .

# Phân phối của trung bình mẫu $\bar{X}$

## Định lý

Nếu  $(X_1, X_2, \dots, X_n)$  là một mẫu sinh từ biến ngẫu nhiên có phân phối chuẩn chính tắc  $\mathcal{N}(0, 1)$ , thì

- 1 Trung bình mẫu  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  có phân phối chuẩn  $\mathcal{N}(\mu, \sigma^2)$ .
- 2 Thống kê  $Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$  có phân phối chuẩn  $\mathcal{N}(0, 1)$ .

# Phân phối của phương sai mẫu điều chỉnh $\hat{S}_n^2$

## Định lý 1

Nếu  $(X_1, X_2, \dots, X_n)$  là một mẫu sinh từ biến ngẫu nhiên có phân phối chuẩn chính tắc  $\mathcal{N}(0, 1)$ , thì

- 1 Tổng các bình phương  $U = X_1^2 + \dots + X_n^2$  có phân phối  $\chi_n^2$  (khi bình phương) với  $n$  bậc tự do.

# Phân phối của phương sai mẫu điều chỉnh $\hat{S}_n^2$

## Định lý 1

Nếu  $(X_1, X_2, \dots, X_n)$  là một mẫu sinh từ biến ngẫu nhiên có phân phối chuẩn chính tắc  $\mathcal{N}(0, 1)$ , thì

- 1 Tổng các bình phương  $U = X_1^2 + \dots + X_n^2$  có phân phối  $\chi_n^2$  (khi bình phương) với  $n$  bậc tự do.
- 2 Hàm mật độ của phân phối  $\chi_n^2$  có dạng

$$f_U(x, n) = \begin{cases} 0, & \text{nếu } x \leq 0 \\ \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{nếu } x > 0, n > 0. \end{cases}$$

# Phân phối của phương sai mẫu điều chỉnh $\hat{S}_n^2$

## Định lý 1

Nếu  $(X_1, X_2, \dots, X_n)$  là một mẫu sinh từ biến ngẫu nhiên có phân phối chuẩn chính tắc  $\mathcal{N}(0, 1)$ , thì

- 1 Tổng các bình phương  $U = X_1^2 + \dots + X_n^2$  có phân phối  $\chi_n^2$  (khi bình phương) với  $n$  bậc tự do.
- 2 Hàm mật độ của phân phối  $\chi_n^2$  có dạng

$$f_U(x, n) = \begin{cases} 0, & \text{nếu } x \leq 0 \\ \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{nếu } x > 0, n > 0. \end{cases}$$

- 3 Trong đó hàm Gamma  $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, (a > 1)$

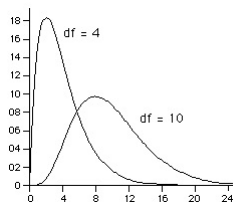
# Phân phối của phương sai mẫu điều chỉnh $\hat{S}_n^2$

## Định lý 2

Nếu  $(X_1, X_2, \dots, X_n)$  là một mẫu sinh từ biến ngẫu nhiên có phân phối chuẩn  $\mathcal{N}(\mu, \sigma^2)$ , thì  $\frac{n-1}{\sigma^2} \hat{S}_n^2$  có phân phối  $\chi_{n-1}^2$  với  $(n-1)$  bậc tự do

# Hàm mật độ phân phối $\chi_n^2$

- Hàm mật độ phân phối  $\chi_n^2$  với  $n=4$  và  $n=10$





# Phân phối Student (t-phân phối)

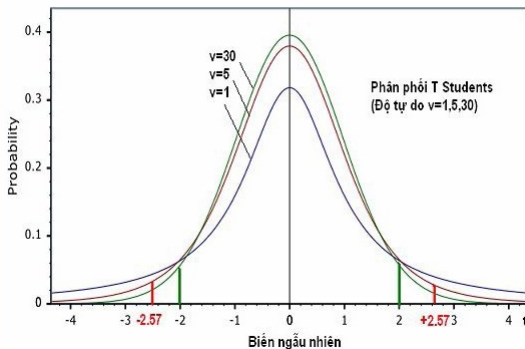
## Định nghĩa

Một biến ngẫu nhiên  $T$  được gọi là có phân phối Student với  $\nu$  bậc tự do, nếu hàm mật độ có dạng

$$f(t, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \frac{1}{(1 + \frac{t^2}{\nu})^{\frac{\nu+1}{2}}}, \quad \nu > 0, t \in \mathfrak{R}$$

# Hàm mật độ phân phối Student

- Khi  $\nu$  tăng, thì hàm mật độ  $f(t, \nu)$  tiệm cận chuẩn.



# Phân phối Student

## Định lý 1

Biến ngẫu nhiên  $T$  có dạng

$$T = \frac{Z}{\sqrt{U}} \sqrt{n}$$

có phân phối Student; trong đó  $Z$  độc lập với  $U$ ,  $Z \in \mathfrak{N}(0, 1)$ ;  $U \in \chi_n^2$

## Định lý 2

Nếu  $X \in \mathfrak{N}(\mu, \sigma^2)$ , thì biến ngẫu nhiên  $T = \frac{\bar{X} - \mu}{\hat{S}_n^2} \sqrt{n}$  có phân phối Student với  $(n-1)$  bậc tự do

# Phân phối F (Fisher R.A - Snedecor G.W)

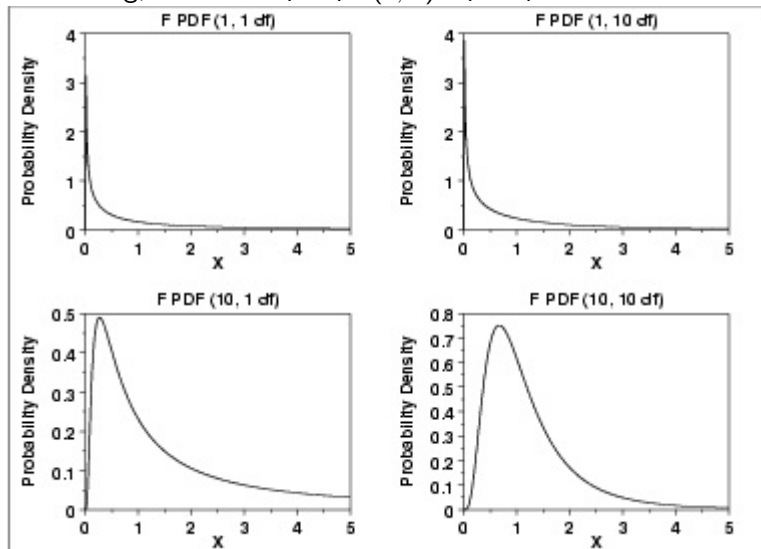
## Định nghĩa

Một biến ngẫu nhiên  $X$  được gọi là có phân phối F với  $\nu_1$  và  $\nu_2$  bậc tự do, nếu với  $x > 0, \nu_1 > 0, \nu_2 > 0$ , hàm mật độ xác suất có dạng

$$f(x, \nu_1, \nu_2) = \frac{\Gamma(\frac{\nu_1 + \nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2} - 1} \left(1 + \frac{\nu_1}{\nu_2} x\right)^{-\frac{\nu_1 + \nu_2}{2}}$$

# Hàm mật độ phân phối F

- Khi  $\nu$  tăng, thì hàm mật độ  $f(t, \nu)$  tiệm cận chuẩn.



# Phân phối F

## Định lý 1

Biến ngẫu nhiên F có dạng

$$F = \frac{U_1}{U_2} \times \frac{\nu_2}{\nu_1}$$

có phân phối F với  $\nu_1$  và  $\nu_2$  bậc tự do; trong đó  $U_1$  và  $U_2$  là các biến ngẫu nhiên độc lập có phân phối  $\chi^2$  với  $\nu_1$  và  $\nu_2$  bậc tự do

## Định lý 2

Nếu  $X, Y \in \mathfrak{N}(\mu, \sigma^2)$ , với  $X, Y$  độc lập, thì biến ngẫu nhiên  $F = \frac{\hat{S}_1^2}{\hat{S}_2^2}$  có phân phối F với  $(\nu_1 - 1)$  và  $(\nu_2 - 1)$  bậc tự do

# Phân phối tiệm cận chuẩn

## Định lý 1

Giả sử  $(X_1, X_2, \dots, X_n)$  là mẫu sinh ra từ  $n$  quan sát độc lập của một biến ngẫu nhiên có  $E(X) = \mu$  và  $D(X) = \sigma^2$ , thì biến ngẫu nhiên  $Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$  có phân phối tiệm cận chuẩn khi  $n \rightarrow \infty$ , tức là

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X} - \mu}{\sigma} \sqrt{n} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

# Phân phối tiệm cận chuẩn

## Định lý 2

Giả sử tiến hành  $n$  quan sát độc lập, với xác suất thành công của mỗi quan sát là  $p$ ,  $p \in (0, 1)$ . Gọi  $k$  là số quan sát thành công trong  $n$  quan sát,  $0 \leq k \leq n$ . Khi đó, biến ngẫu nhiên  $Z = \frac{\frac{k}{n} - p}{\sqrt{p(1-p)}} \sqrt{n}$  có phân phối tiệm cận chuẩn khi  $n \rightarrow \infty$ , tức là

$$\lim_{n \rightarrow \infty} P\left(\frac{\frac{k}{n} - p}{\sqrt{p(1-p)}} \sqrt{n} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$



# Phân phối tiệm cận chuẩn

## Định lý 3

Giả sử  $X$  có phân phối  $\chi_n^2$  với  $n$  bậc tự do. Khi đó, các biến ngẫu nhiên  $Z = \frac{X-n}{\sqrt{2n}}$  và  $W = (\sqrt{2X} - \sqrt{2n-1})$  có phân phối tiệm cận chuẩn khi  $n \rightarrow \infty$ , tức là

$$\lim_{n \rightarrow \infty} P\left(\frac{X-n}{\sqrt{2n}} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

và

$$\lim_{n \rightarrow \infty} P\left(\sqrt{2X} - \sqrt{2n-1} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

# Phân phối tiệm cận chuẩn

## Định lý 4

Giả sử  $T$  có phân phối Student với  $n$  bậc tự do. Khi đó, phân phối của  $T$  là tiệm cận chuẩn khi  $n \rightarrow \infty$ , tức là

$$\lim_{n \rightarrow \infty} P(T < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

# Chú ý

- Nếu mẫu cỡ  $n$  chia thành  $k$  lớp với  $n_1 + n_2 + \dots + n_k = n$ , thì trung bình mẫu

$$\bar{X} = \frac{1}{n} \sum_{j=1}^k n_j X_j$$

và phương sai mẫu

$$S_n^2 = \frac{1}{n} \sum_{j=1}^k n_j X_j^2 - \bar{X}_n^2$$

## Chú ý

- Nếu mẫu chia thành  $k$  khoảng có độ dài bằng nhau

$$[x_0; h + x_0], [x_1; h + x_1], \dots [x_{k-1}; h + x_{k-1}]$$

thì trung bình mẫu có dạng

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^k n_i X'_i$$

và phương sai mẫu

$$S_n^2 = \frac{1}{n} \sum_{i=1}^k n_i X_i'^2 - \overline{X_n}'^2$$

với  $X'_i = \frac{1}{2}[x_{i-1} + x_i]$ , các điểm giữa các khoảng

# Chú ý

- Nếu mẫu chia thành  $k$  khoảng có độ dài bằng nhau và các giá trị khoảng lớn, có thể sử dụng phép biến đổi tuyến tính

$$Y_i = \frac{(X_i - X_0)}{h}$$

và

$$\overline{X_n} = h\overline{Y} + X_0; \quad S_X^2 = h^2 S_Y^2,$$

trong đó  $X_0 \in \{X_1, X_2, \dots, X_n\}$  và  $h$  là độ rộng các khoảng

# Bài tập chương 5

## Bài tập 1

Kiểm tra thời điểm nhân viên không làm việc tại một đơn vị, người ta có thống kê sau, với thời gian quy định kết thúc giờ làm việc là 5:00 pm.

5 : 10   4 : 55   5 : 05   4 : 37   4 : 58   4 : 55   4 : 59

5 : 11   4 : 50   5 : 00   4 : 35   4 : 59   4 : 54   4 : 58

- 1 Hãy tính các đặc trưng của mẫu (trung bình, phương sai mẫu, phương sai mẫu điều chỉnh, trung vị, mod, cỡ mẫu, độ rộng mẫu)

# Bài tập chương 5

## Bài tập 1

Kiểm tra thời điểm nhân viên không làm việc tại một đơn vị, người ta có thống kê sau, với thời gian quy định kết thúc giờ làm việc là 5:00 pm.

5 : 10   4 : 55   5 : 05   4 : 37   4 : 58   4 : 55   4 : 59

5 : 11   4 : 50   5 : 00   4 : 35   4 : 59   4 : 54   4 : 58

- 1 Hãy tính các đặc trưng của mẫu (trung bình, phương sai mẫu, phương sai mẫu điều chỉnh, trung vị, mod, cỡ mẫu, độ rộng mẫu)
- 2 Cho ý kiến nhận xét

# Bài tập chương 5

## Bài tập 2

Thống kê số phút đi làm muộn của một cơ quan, có kết quả sau X số phút muộn: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9  
 $n_j$ —tần số: 183, 15, 16, 14, 13, 2, 1, 1, 1, 1

- 1 Hãy tính các đặc trưng của mẫu (trung bình, phương sai mẫu, phương sai mẫu điều chỉnh, trung vị, mod, cỡ mẫu, độ rộng mẫu)



# Bài tập chương 5

## Bài tập 2

Thống kê số phút đi làm muộn của một cơ quan, có kết quả sau X số phút muộn: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9  
 $n_j$ —tần số: 183, 15, 16, 14, 13, 2, 1, 1, 1, 1

- 1 Hãy tính các đặc trưng của mẫu (trung bình, phương sai mẫu, phương sai mẫu điều chỉnh, trung vị, mod, cỡ mẫu, độ rộng mẫu)
- 2 Tính trung vị và trung bình các độ lệch so với trung vị

$$\frac{\sum_{j=1}^n n_j |X_j - Med|}{n}$$

# Bài tập chương 5

## Bài tập 2

Thống kê số phút đi làm muộn của một cơ quan, có kết quả sau X số phút muộn: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9  
 $n_j$ —tần số: 183, 15, 16, 14, 13, 2, 1, 1, 1, 1

- 1 Hãy tính các đặc trưng của mẫu (trung bình, phương sai mẫu, phương sai mẫu điều chỉnh, trung vị, mod, cỡ mẫu, độ rộng mẫu)
- 2 Tính trung vị và trung bình các độ lệch so với trung vị

$$\frac{\sum_{j=1}^n n_j |X_j - Med|}{n}$$

- 3 So sánh kết quả ở (1) và (2)