# Chương 4
# THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ

BỘ MÔN TOÁN ỨNG DỤNG[1]

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC
ĐẠI HỌC BÁCH KHOA HÀ NỘI

SAMI.HUST – 2023

[1] Phòng BIS.201-D3.5

# GIỚI THIỆU CHƯƠNG 4

Chương này trình bày về bài toán ước lượng tham số của tổng thể. Nội dung bao gồm:

- Mẫu ngẫu nhiên, thống kê và phân phối mẫu.
- Ước lượng điểm cho tham số, một số tiêu chuẩn lựa chọn hàm ước lượng.
- Khoảng tin cậy cho kỳ vọng, phương sai, tỷ lệ; xác định kích thước mẫu.

# 4.1. MẪU NGẪU NHIÊN VÀ PHÂN PHỐI MẪU

# Tổng thể

**Khái niệm 1**

Khi nghiên cứu các vấn đề tự nhiên, kinh tế, xã hội và nhiều lĩnh vực khác thường dẫn đến khảo sát một hay nhiều dấu hiệu (định tính hoặc định lượng) trên nhiều phần tử. Tập hợp tất cả các phần tử này gọi là tổng thể hay đám đông.

✍

- Số phần tử trong tổng thể có thể là hữu hạn hoặc vô hạn.
- Cần chú ý rằng ta không nghiên cứu trực tiếp các phần tử của tổng thể mà chỉ nghiên cứu dấu hiệu nào đó của nó.
- Ký hiệu $N$ là số phần tử của tổng thể; $\mathcal{X}$ là dấu hiệu cần nghiên cứu.

# Tổng thể

**Ví dụ 1**

**(a)** Muốn điều tra thu nhập bình quân của các hộ gia đình ở Hà Nội thì

- Tổng thể cần nghiên cứu là toàn bộ các hộ gia đình ở Hà Nội;
- Dấu hiệu nghiên cứu là thu nhập của từng hộ gia đình (dấu hiệu định lượng).

**(b)** Một doanh nghiệp muốn nghiên cứu các khách hàng của mình. Tổng thể là toàn bộ các khách hàng của doanh nghiệp.

- Dấu hiệu định tính có thể là mức độ hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp;
- Dấu hiệu định lượng là số lượng sản phẩm mà khách hàng mua của doanh nghiệp.

# Tổng thể

**Ví dụ 2**

Một nhà máy sản xuất 5.000.000 sản phẩm. Ta muốn đánh giá tỷ lệ phế phẩm trong các sản phẩm của nhà máy.

- Tổng thể cần nghiên cứu là 5.000.000 sản phẩm của nhà máy.
- Dấu hiệu nghiên cứu là một sản phẩm có phải là phế phẩm hay không.

# Tổng thể

✍ Một số lý do không thể khảo sát toàn bộ tổng thể

- Do quy mô của tổng thể cần nghiên cứu quá lớn nên việc nghiên cứu toàn bộ sẽ đòi hỏi nhiều kinh phí và thời gian.
- Trong nhiều trường hợp không thể biết được toàn bộ các phần tử của tổng thể cần nghiên cứu.
- Có thể trong quá trình điều tra sẽ phá hủy đối tượng nghiên cứu...

# Tập mẫu

- Thay vì khảo sát tổng thể, ta chỉ cần chọn ra một tập nhỏ để khảo sát và đưa ra quyết định. Việc chọn ra từ tổng thể một tập hợp con nào đó được gọi là phép lấy mẫu.
- Tập hợp con được chọn được gọi là tập mẫu.
- Số phần tử trong tập mẫu được gọi là kích thước mẫu hoặc cỡ mẫu, ký hiệu là $n$.

# Tập mẫu

### Ví dụ 3

Với số liệu trong Ví dụ 2, ta không có đủ thời gian và tiền bạc để xem xét toàn bộ 5.000.000 sản phẩm.

- Ta chọn ra một mẫu gồm 500 sản phẩm để kiểm tra và phát hiện có 20 sản phẩm mắc lỗi.
- Tỷ lệ phế phẩm trong mẫu kiểm tra này là $20/500 = 0,04 = 4\%$. Từ đó, ta nhận định tỷ lệ phế phẩm của nhà máy này khoảng $4\%$.

### Ví dụ 4

Ta muốn đánh giá số giờ trong một ngày mà một kỹ sư các ngành kỹ thuật sử dụng điện thoại. Vì số kỹ sư các ngành kỹ thuật rất lớn, nên ta không thể điều tra trên tất cả các kỹ sư được.

- Ta chọn ngẫu nhiên một mẫu gồm $n = 50$ kỹ sư để khảo sát và tìm được số giờ trung bình dùng điện thoại của 50 kỹ sư này, chẳng hạn, là 2,7 giờ.
- Con số 2,7 giờ cho ta một thông tin về việc sử dụng điện thoại của các kỹ sư các ngành kỹ thuật.

# Một số kỹ thuật chọn mẫu cơ bản

**Ví dụ 5**

Để điều tra mức thu nhập trung bình của sinh viên tốt nghiệp đại học mới ra trường, nếu mẫu được chọn trong số các sinh viên tốt nghiệp ngành Công nghệ thông tin thì rõ ràng mức lương trung bình trong mẫu không phản ánh trung thực mức lương trung bình của sinh viên mới ra trường nói chung.

✍ Vấn đề chọn mẫu

- Các kết luận suy diễn từ mẫu có đáng tin cậy chỉ đạt được nếu mẫu được chọn phản ánh trung thực, thực sự đại diện cho tổng thể. Do đó, vấn đề chọn mẫu là một vấn đề rất quan trọng của thống kê.
- Các kỹ thuật chọn mẫu đúng đắn sẽ giúp ta đảm bảo được tính đại diện trung thực cho tổng thể.
- Để trả lời cho câu hỏi "làm sao chọn được tập mẫu có tính chất tương tự như tổng thể để các kết luận của tập mẫu có thể dùng cho tổng thể" ta sử dụng một trong những cách chọn mẫu sau.

# Một số kỹ thuật chọn mẫu cơ bản

- Lấy mẫu ngẫu nhiên: mỗi cá thể của tổng thể được chọn một cách độc lập với xác suất như nhau.
- Lấy mẫu theo khối: Tổng thể được chia làm $M$ khối. Chọn ngẫu nhiên ra $m$ khối trong $M$ khối đó. Tập hợp tất cả các cá thể của $m$ khối được chọn sẽ được lập thành một mẫu để khảo sát.
- Lấy mẫu phân tầng: Chia tổng thể ra một số tầng, sao cho các phần tử trong mỗi tầng khác nhau càng ít càng tốt. Mỗi tầng được coi là một tổng thể con. Trong mỗi tầng ta sẽ thực hiện việc lấy mẫu ngẫu nhiên.

# Một số kỹ thuật chọn mẫu cơ bản

## Ví dụ 6

Một doanh nghiệp có 20.000 kỹ sư được tuyển chọn từ các hệ đào tạo khác nhau, trong đó có 10.000 học đại học chính quy, 2.000 học hệ liên thông, 2.000 học văn bằng hai, 5.000 học tại chức và 1.000 học sau đại học. Để tiến hành khảo sát về mức độ hài lòng của doanh nghiệp đối với chất lượng công việc của các kỹ sư, người ta chọn ngẫu nhiên 1.000 kỹ sư tham gia khảo sát. Bảng dưới đây trình bày một ví dụ về việc chọn mẫu theo tầng, ở đó mỗi hệ đào tạo được xem là một tầng.

| Hệ đào tạo | Số kỹ sư | Tỷ lệ % | Số kỹ sư được chọn |
|---|---|---|---|
| Đại học chính quy | 10.000 | 50 | 500 |
| Đại học liên thông | 2.000 | 10 | 200 |
| Đại học văn bằng hai | 2.000 | 10 | 200 |
| Đại học tại chức | 2.000 | 10 | 200 |
| Sau đại học | 1.000 | 5 | 50 |
| **Tổng** | **20.000** | **100** | **1.000** |

# Biểu diễn dữ liệu

✍ Tại sao cần biểu diễn dữ liệu?

- Trong thống kê, sau khi thu thập dữ liệu, bước tiếp theo là biểu diễn, mô tả và tổng kết dữ liệu.
- Việc biểu diễn dữ liệu sẽ giúp trực quan hóa; dễ dàng hơn trong việc trích xuất thông tin quan trọng so với việc sử dụng tập dữ liệu thô ban đầu.

# Bảng dữ liệu

Từ tổng thể, xét một mẫu có kích thước $n$. Ta có thể biểu diễn mẫu theo một số cách sau.

1. Dạng liệt kê: các số liệu thu được được ghi lại thành dãy $x_1, x_2, \ldots, x_n$.
2. Dạng rút gọn: khi trong mẫu có nhiều giá trị trùng nhau, ta có thể sử dụng bảng tần số hoặc bảng tần suất:

- Dạng tần số $(n_1 + n_2 + \ldots + n_k = n)$

| Giá trị | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| Tần số | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

(1)

- Dạng tần suất $(f_k = n_k/n)$

| Giá trị | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| Tần suất | $f_1$ | $f_2$ | $\ldots$ | $f_k$ |

(2)

# Bảng dữ liệu

**3.** Dạng khoảng: dữ liệu thu được nhận các giá trị khác nhau nhưng lại khá gần nhau trong khoảng $(a, b)$. Ta chia đều khoảng $(a, b)$ thành $k$ miền con bởi các điểm chia $a_0 = a < a_1 < a_2 < \cdots < a_{k-1} < a_k = b$.

- Dạng tần số $(n_1 + n_2 + \ldots + n_k = n)$

| Giá trị | $(a_0, a_1]$ | $(a_1, a_2]$ | $\ldots$ | $(a_{k-1}, a_k]$ |
|---|---|---|---|---|
| Tần số | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

(3)

- Dạng tần suất $(f_k = n_k/n)$

| Giá trị | $(a_0, a_1]$ | $(a_1, a_2]$ | $\ldots$ | $(a_{k-1}, a_k]$ |
|---|---|---|---|---|
| Tần suất | $f_1$ | $f_2$ | $\ldots$ | $f_k$ |

(4)

Trong tính toán, ta thường lấy giá trị chính giữa của mỗi khoảng làm giá trị đại diện, $x_i = \frac{1}{2}(a_{i-1} + a_i)$. Khi đó, bảng (3) và (4) có thể đưa về dạng (1) và (2).

# Biểu đồ

Một câu ngạn ngữ Trung Hoa "Một hình ảnh có tác dụng bằng một nghìn lời nói". Để có được một hình ảnh rõ ràng và dễ nhớ về mẫu các giá trị của biến ngẫu nhiên $X$, ta dùng các đồ thị và các biểu đồ để thể hiện chúng.

1. Biểu đồ hình cột: là biểu đồ nhằm biểu diễn cho dữ liệu được phân nhóm (thường dùng cho dữ liệu định tính) như các tháng trong năm, các nhóm tuổi... Các nhóm được biểu diễn xuất hiện theo trục hoành, trục tung là chiều cao của các hình chữ nhật tỷ lệ với giá trị được biểu diễn. Mục tiêu của việc dùng biểu đồ hình cột là đưa ra so sánh giữa các nhóm.
2. Biểu đồ hình quạt: cũng được dùng để biểu diễn dữ liệu được phân nhóm, nhưng các nhóm được biểu diễn bằng các hình quạt trong hình tròn. Số lượng hoặc tỷ lệ của mỗi hạng mục (mỗi nhóm) tỷ lệ với diện tích hình quạt biểu diễn nó. Biểu đồ này thường dùng để phân tích hoặc so sánh ở mức độ tổng thể.

# Biểu đồ

3. Biểu đồ tần suất: thường được dùng để biểu thị tần số hay tần suất xuất hiện của các giá trị trong mỗi khoảng.

- Nếu độ rộng các khoảng bằng nhau, thì chiều cao của hình chữ nhật dựng trên mỗi khoảng chính là tần số hay tần suất tương ứng của khoảng.
- Nếu độ rộng các khoảng không bằng nhau, chiều cao của hình chữ nhật dựng trên mỗi khoảng được tính toán sao cho diện tích mỗi hình chữ nhật tỷ lệ với tần số hoặc tần suất của khoảng đó.

4. Đa giác tần số, tần suất: dùng khi dữ liệu là liên tục và khoảng dữ liệu rất rộng. Tại mỗi giá trị của dữ liệu $x_i$ và tần số $n_i$ ta chấm một điểm có tọa độ $(x_i, n_i)$. Nối các điểm này với nhau ta được đa giác tần số. Nếu muốn có đa giác tần suất ta thay $n_i$ bằng $f_i = n_i/n$.

## Ví dụ 7

Khảo sát 400 khách hàng đánh giá, xếp hạng một loại sản phẩm trên thị trường, ta nhận được bảng dữ liệu sau:

| Xếp hạng | A | B | C | D |
|---|---|---|---|---|
| Số khách hàng | 35 | 260 | 93 | 12 |

- Tổng số khách hàng được khảo sát $n = 400$.

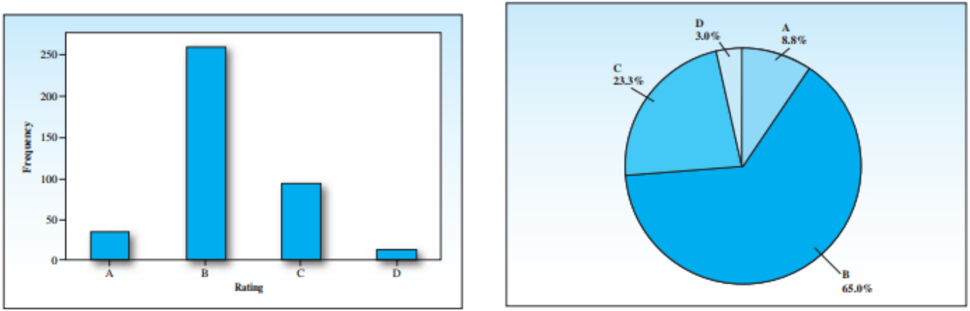
  35 người xếp hạng A chiếm 8,75%;

  260 người xếp hạng B chiếm 65%;

  93 người xếp hạng C chiếm 23,25%;

  12 người xếp loại D chiếm 3%.
- Biểu đồ hình cột cho tập dữ liệu này biểu diễn ở Hình 1 (bên trái).
- Biểu đồ hình quạt cho tập dữ liệu này biểu diễn ở Hình 1 (bên phải).

# Biểu đồ




**Hình 1:** Biểu đồ hình cột và hình quạt cho dữ liệu trong Ví dụ 7

# Biểu đồ

## Ví dụ 8

Khảo sát số lần đến một cửa hàng của 25 khách hàng trong một tuần ta được bảng số liệu:

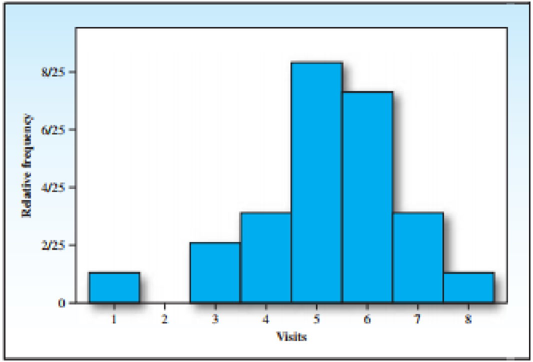
$$
\begin{array}{ccccccccccccc}
6 & 7 & 1 & 5 & 6 & 4 & 6 & 4 & 6 & 8 & 6 & 5 & \\
6 & 3 & 4 & 5 & 5 & 5 & 7 & 6 & 3 & 5 & 7 & 5 & 5
\end{array}
$$

Biến đo lường "số lần đến cửa hàng trong một tuần" là biến ngẫu nhiên rời rạc chỉ nhận giá trị nguyên. Bảng dưới đây cho thấy các lớp, tần số và tần suất tương ứng:

| Số lượt đến chuỗi cửa hàng | Tần số | Tần suất |
|---|---|---|
| 1 | 1 | 0,04 |
| 2 | 0 | 0,00 |
| 3 | 2 | 0,08 |
| 4 | 3 | 0,12 |
| 5 | 8 | 0,32 |
| 6 | 7 | 0,28 |
| 7 | 3 | 0,12 |
| 8 | 1 | 0,04 |

Biểu đồ tần suất tương đối được thể hiện trong Hình 2.

# Biểu đồ




**Hình 2:** Biểu đồ tần suất cho dữ liệu trong Ví dụ 8

# Trung bình mẫu

## Định nghĩa 1

Trung bình mẫu là trung bình cộng các giá trị mà ta quan sát được. Nếu ta có mẫu kích thước $n$ về biến ngẫu nhiên $X$, thì trung bình mẫu, ký hiệu là $\overline{x}$, được xác định như sau.

- Nếu mẫu cho dưới dạng liệt kê $x_1, x_2, \ldots, x_n$ kích thước $n$, thì

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{5}$$

- Nếu mẫu cho ở dạng rút gọn (1), thì

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i. \tag{6}$$

# Trung vị mẫu

✍ Giá trị trung bình mẫu thường được dùng để đo vị trí trung tâm của mẫu. Tuy nhiên, trong một số trường hợp, việc dùng đặc trưng này sẽ không còn chính xác. Trong bộ dữ liệu xuất hiện các giá trị bất thường, người ta thường dùng giá trị trung vị thay cho giá trị trung bình.

**Định nghĩa 2**

Trung vị (median) mẫu, ký hiệu là $\widetilde{x}$, là một số thỏa mãn số các giá trị của mẫu bé hơn hay bằng $\widetilde{x}$ bằng số các giá trị của mẫu lớn hơn hay bằng $\widetilde{x}$.

✍ Với một mẫu kích thước $n$ cho ở dạng liệt kê, ta sắp xếp các giá trị theo thứ tự tăng dần $x_1 \leq x_2 \leq \cdots \leq x_n$. Khi đó,

$$\widetilde{x} = \begin{cases} x_{(n+1)/2}, & \text{nếu } n \text{ lẻ}, \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{nếu } n \text{ chẵn}. \end{cases} \tag{7}$$

# Mốt

**Định nghĩa 3**

Mốt là giá trị trong mẫu xuất hiện với tần số lớn nhất.

✍

- Đa phần các tập dữ liệu có một mốt. Một số tập dữ liệu có hai mốt.
- Mốt có thể dùng làm số đo vị trí trung tâm của nhiều loại dữ liệu khác nhau.

# Mốt

## Ví dụ 9

Bảng sau đây cho biết hoạt động được ưu tiên của 9 sinh viên được chọn ngẫu nhiên

| Sinh viên | Hoạt động |
|---|---|
| 1 | Tình nguyện |
| 2 | Học lập trình |
| 3 | Học tiếng Anh |
| 4 | Tình nguyện |
| 5 | Tình nguyện |
| 6 | Học tiếng Pháp |
| 7 | Học võ |
| 8 | Học vẽ |
| 9 | Tình nguyện |

Mốt của các hoạt động của sinh viên là Tình nguyện với tần số là 4.

# Phương sai và độ lệch chuẩn mẫu

✍ Trong thống kê, số đặc trưng thông dụng để đo mức độ phân tán của số liệu là phương sai mẫu.

**Định nghĩa 4**

Phương sai mẫu, ký hiệu là $\widetilde{s}^2$, là trung bình của bình phương độ lệch giữa các giá trị mẫu với trung bình mẫu.

- Nếu mẫu cho dưới dạng liệt kê, thì

$$\widetilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{8}$$

- Nếu mẫu cho ở dạng rút gọn (1), thì

$$\widetilde{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2. \tag{9}$$

# Phương sai và độ lệch chuẩn mẫu

**Định nghĩa 4 (tiếp theo)**

Dạng tương đương của các công thức (8) và (9) lần lượt là

$$\widetilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big)^2, \tag{10}$$

và

$$\widetilde{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \Big(\frac{1}{n}\sum_{i=1}^{k} n_i x_i\Big)^2. \tag{11}$$

**Định nghĩa 5**

Độ lệch chuẩn mẫu, ký hiệu là $\widetilde{s}$, là căn bậc hai số học của phương sai mẫu,

$$\widetilde{s} = \sqrt{\widetilde{s}^2}. \tag{12}$$

# Ví dụ

### Ví dụ 10

Kiểm tra số lỗi trên sản phẩm đúc do một nhà máy sản xuất ta thu được dữ liệu sau đây:

| Số lỗi ở mỗi sản phẩm | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Số sản phẩm kiểm tra | 8 | 20 | 12 | 40 | 30 | 25 | 15 |

Hãy tính trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu của mẫu trên.

# Ví dụ

*Giải.*

- Áp dụng công thức (6), trung bình mẫu là

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i = \frac{499}{150} \approx 3,3267.$$

- Áp dụng công thức (8), phương sai mẫu là

$$\widetilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} n_i x_i\right)^2 = \frac{2073}{150} - \left(\frac{499}{150}\right)^2 \approx 2,7533.$$

- Độ lệch chuẩn mẫu là

$$\widetilde{s} = \sqrt{2,7533} \approx 1,6593.$$

# 4.1. MẪU NGẪU NHIÊN VÀ PHÂN PHỐI MẪU

# Phân phối của tổng thể

- Giả sử ta cần nghiên cứu dấu hiệu $\mathcal{X}$ nào đó của tổng thể. Ta có thể mô hình hóa $\mathcal{X}$ bởi một biến ngẫu nhiên, ký hiệu là $X$, bằng cách coi $X$ là giá trị của dấu hiệu $\mathcal{X}$ trên các phần tử của tổng thể.
- Phân phối xác suất của $X$ được gọi là phân phối xác suất của tổng thể.
- Giá trị $\mu = E(X)$ và $\sigma^2 = V(X)$ được gọi là kỳ vọng và phương sai của tổng thể.

# Mẫu ngẫu nhiên

**Định nghĩa 6**

Xét biến ngẫu nhiên $X$ từ tổng thể có phân phối xác suất $F_X(x)$. Bộ $(X_1, \dots, X_n)$ được gọi là một mẫu ngẫu nhiên có kích thước $n$, ký hiệu là $W_X = (X_1, X_2, \dots, X_n)$, nếu $X_1, \dots, X_n$ là các biến ngẫu nhiên độc lập và có cùng phân phối xác suất $F_X(x)$.

✍

- Trong định nghĩa trên, ta có thể coi $X_i$ là biến ngẫu nhiên chỉ giá trị của $X$ trên phần tử thứ $i$ trong mẫu.
- Do $X_1, \dots, X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối xác suất lên hàm phân phối xác suất đồng thời của chúng là:

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n).$$

- Các giá trị số $x_1, x_2, \dots, x_n$ được gọi là một mẫu cụ thể $W_x = (x_1, x_2, \dots, x_n)$ của mẫu ngẫu nhiên $W_X = (X_1, X_2, \dots, X_n)$.

# Mẫu ngẫu nhiên

**Ví dụ 11**

- Giả sử tuổi thọ $X$ (giờ) của loại pin A là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$. Chọn ngẫu nhiên $n = 10$ quả pin loại này và gọi $X_i$ là tuổi thọ của quả pin thứ $i$, $i = 1, 2, \ldots, 10$. Khi đó, $X_1, \ldots, X_{10}$ là độc lập và có cùng phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$. Tức là, ta có một mẫu ngẫu nhiên $W_X = (X_1, \ldots, X_{10})$ kích thước $n = 10$.
- Nếu ghi lại tuổi thọ của mỗi quả pin, ta thu được một mẫu cụ thể, chẳng hạn như sau:

$$W_x = (1000;\ 1001;\ 1002;\ 1003;\ 1004;\ 1005;\ 1006;\ 1007;\ 1008;\ 1009).$$

# Thống kê

Trong thống kê toán học, việc tổng hợp thông tin từ mẫu ngẫu nhiên $W_X = (X_1, X_2, \dots, X_n)$ thường dẫn đến việc xác định một hàm nào đó của các biến ngẫu nhiên $X_1, X_2, \dots, X_n$. Một hàm như thế được gọi là một thống kê.

**Định nghĩa 7**

Thống kê là một hàm của mẫu ngẫu nhiên $W_X = (X_1, X_2, \dots, X_n)$, ký hiệu là

$$\widehat{\Theta} = g(X_1, X_2, \dots, X_n). \tag{13}$$

✍

- Thống kê $\widehat{\Theta}$ là một hàm của các biến ngẫu nhiên $X_1, X_2, \dots, X_n$ nên nó cũng là một biến ngẫu nhiên.
- Nếu mẫu ngẫu nhiên có giá trị $W_x = (x_1, x_2, \dots, x_2)$, thì ta tính được giá trị cụ thể của $\widehat{\Theta}$, ký hiệu là $\widehat{\theta} = g(x_1, x_2, \dots, x_n)$ và gọi là giá trị quan sát của thống kê $\widehat{\Theta}$.

# Trung bình mẫu ngẫu nhiên

Cho mẫu ngẫu nhiên kích thước $n$, $W_X = (X_1, X_2, \ldots, X_n)$.

- Trung bình mẫu ngẫu nhiên là một thống kê, ký hiệu là $\overline{X}$, được định nghĩa bởi

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{14}$$

- Nếu biến ngẫu nhiên $X$ có $E(X) = \mu$ và $V(X) = \sigma^2$ thì thống kê $\overline{X}$ có kỳ vọng và phương sai được xác định bởi:

$$\mu_{\overline{X}} = \frac{1}{n}(\mu + \cdots + \mu) = \mu$$

và

$$\sigma^2_{\overline{X}} = \frac{1}{n^2}(\sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}.$$

# Trung bình mẫu ngẫu nhiên

✍

- Phương sai $V(\overline{X})$ của $\overline{X}$ nhỏ hơn phương sai $V(X)$ của $X$ là $n$ lần, nghĩa là các giá trị có thể có của $\overline{X}$ ổn định quanh kỳ vọng $\mu$ hơn các giá trị có thể có của $X$.
- Nếu một giá trị của mẫu ngẫu nhiên $W_X$ là $W_x = (x_1, x_2, \ldots, x_n)$ thì thống kê $\overline{X}$ nhận giá trị là trung bình mẫu

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

# Phương sai mẫu ngẫu nhiên

- Phương sai mẫu ngẫu nhiên là một thống kê, ký hiệu là $\widetilde{S}^2$, được định nghĩa bởi

$$\widetilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{15}$$

- Giá trị của $\widetilde{S}^2$ khi cho một mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, ký hiệu là $\widetilde{s}^2$, là phương sai mẫu

$$\widetilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

- Từ (15), ta tính được

$$E(\widetilde{S}^2) = \frac{n-1}{n}\sigma^2 \quad \text{với} \quad \sigma^2 = V(X). \tag{16}$$

# Phương sai mẫu ngẫu nhiên hiệu chỉnh

✍

- Để kỳ vọng của phương sai mẫu ngẫu nhiên $\widetilde{S}^2$ trùng với phương sai của biến ngẫu nhiên gốc $X$ ta cần một sự hiệu chỉnh. Từ (16) suy ra $E\left(\frac{n}{n-1}\widetilde{S}^2\right) = \sigma^2$.
- Đặt

$$S^2 = \frac{n}{n-1}\widetilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \tag{17}$$

thì $E(S^2) = \sigma^2$ và ta gọi $S^2$ là phương sai mẫu ngẫu nhiên đã hiệu chỉnh, hay phương sai mẫu ngẫu nhiên hiệu chỉnh.

- Giá trị của $S^2$ khi cho một mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ là

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{18}$$

và gọi là phương sai mẫu đã hiệu chỉnh, hay phương sai mẫu hiệu chỉnh.

# Phương sai mẫu ngẫu nhiên hiệu chỉnh

### Ví dụ 12

So sánh giá thành của một loại sản phẩm tại 4 siêu thị được chọn ngẫu nhiên ở thành phố Hà Nội cho thấy các mức tăng so với tháng trước là 12, 15, 17 và 20 nghìn đồng cho một kilôgam. Tìm phương sai mẫu hiệu chỉnh của mẫu giá tăng này.

*Giải.*

- Tính trung bình mẫu
$$\overline{x} = \frac{1}{4}(12 + 15 + 17 + 20) = 16 \text{ nghìn đồng.}$$
- Áp dụng công thức (18), phương sai mẫu hiệu chỉnh là
$$s^2 = \frac{1}{3}\sum_{i=1}^{4}(x_i - 16)^2 = \frac{1}{3}\left[(12-16)^2 + (15-16)^2 + (17-16)^2 + (20-16)^2\right] = \frac{34}{3}.$$

# Độ lệch chuẩn mẫu ngẫu nhiên

- Độ lệch chuẩn mẫu ngẫu nhiên ký hiệu là $\widetilde{S}$ và được xác định bởi

$$\widetilde{S} = \sqrt{\widetilde{S}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{19}$$

- Độ lệch chuẩn mẫu ngẫu nhiên hiệu chỉnh ký hiệu là $S$ và được xác định bởi

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{20}$$

# Độ lệch chuẩn mẫu ngẫu nhiên

- Giá trị của $\widetilde{S}$ và $S$ khi cho một mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ lần lượt là

$$\widetilde{s} = \sqrt{\widetilde{s}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{21}$$

  và

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{22}$$

  và được gọi là độ lệch chuẩn mẫu và độ lệch chuẩn mẫu hiệu chỉnh tương ứng.

# Sử dụng máy tính để tính các đặc trưng mẫu

✍ Ta có thể sử dụng máy tính bỏ túi, chẳng hạn máy CASIO FX570VN PLUS, để tính $\overline{x}$ và $s$ theo các bước sau.

1. **Bước 1:** Chuyển đổi máy tính về chương trình thống kê **MODE** $\rightarrow$ **3** $\rightarrow$ **AC**
2. **Bước 2:** Bật chức năng cột tần số/tần suất **SHIFT** $\rightarrow$ **MODE** $\rightarrow$ **Mũi tên đi xuống** $\rightarrow$ **4(STAT)** $\rightarrow$ **1(ON)**
3. **Bước 3:** Bật chế độ màn hình để nhập dữ liệu, Nhập số liệu **SHIFT** $\rightarrow$ **1** $\rightarrow$ **1(TYPE)** $\rightarrow$ **1(1-VAR)** Chú ý nhập xong số liệu thì bấm **AC** để thoát.
4. **Bước 4:** Xem kết quả:
    - Trung bình mẫu $\overline{x}$: **SHIFT** $\rightarrow$ **1** $\rightarrow$ **4(VAR)** $\rightarrow$ **2**
    - Độ lệch tiêu chuẩn mẫu hiệu chỉnh $s$: **SHIFT** $\rightarrow$ **1** $\rightarrow$ **4** $\rightarrow$ **4**

# Tần suất mẫu ngẫu nhiên

- Ký hiệu $p$ là tỷ lệ cá thể trong tổng thể mang một dấu hiệu A định tính nào đó.
- Ta có thể mô hình hóa sự kiện một cá thể có hoặc không có dấu hiệu A bởi một biến ngẫu nhiên $X$ được định nghĩa như sau:
  $X = 0$ nếu cá thể không có dấu hiệu A;
  $X = 1$ nếu có thể có dấu hiệu A.
- Khi đó $X \sim \mathcal{B}(p)$, tức $X$ là biến ngẫu nhiên có phân phối Bernoulli với tham số $p$.

# Tần suất mẫu ngẫu nhiên

## Định nghĩa 8

Từ tổng thể, lấy một mẫu ngẫu nhiên kích thước $n$, $W_X = (X_1, X_2, \ldots, X_n)$. Số cá thể có dấu hiệu A trong mẫu là $X = \sum_{i=1}^{n} X_i$. Tần suất mẫu ngẫu nhiên là một thống kê được ký hiệu và xác định bởi

$$\widehat{P} = \frac{X}{n}. \tag{23}$$

✍

- Dễ thấy,

$$E(\widehat{P}) = p \quad \text{và} \quad V(\widehat{P}) = \frac{p(1-p)}{n}. \tag{24}$$

- Khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, giá trị quan sát của $\widehat{P}$ là

$$\widehat{p} = \frac{x}{n}, \tag{25}$$

  ở đây, $x$ là số cá thể có dấu hiệu A trong mẫu.
- $\widehat{p}$ được gọi là tần suất mẫu.

# Tần suất mẫu ngẫu nhiên

**Ví dụ 13**

Phỏng vấn ngẫu nhiên 1600 cử tri thấy 960 người ủng hộ cho ứng cử viên A. Khi đó, tỷ lệ cử tri ủng hộ cho ứng cử viên A trong mẫu này là

$$\widehat{p} = \frac{x}{n} = \frac{960}{1600} = 0,6.$$

# 4.1. MẪU NGẪU NHIÊN VÀ PHÂN PHỐI MẪU

# Phân phối mẫu

✍ Vì thống kê $\widehat{\Theta}$ là một biến ngẫu nhiên nên nó có phân phối xác suất.

**Định nghĩa 9**

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

# Phân phối mẫu của $\overline{X}$

✍ Phân phối xác suất của $\overline{X}$ là một phân phối mẫu và được gọi là phân phối của trung bình mẫu ngẫu nhiên.

**Định lý 1**

Nếu biến ngẫu nhiên $X$ có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$ thì trung bình mẫu $\overline{X}$ của mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2/n)$.

**Ví dụ 14**

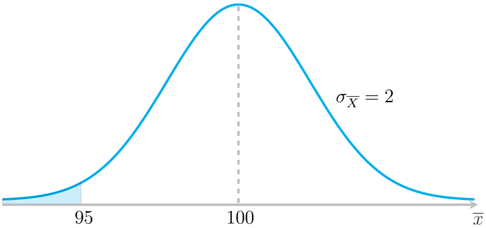
Giả sử điện trở $X$ $(\Omega)$ do một công ty sản xuất là biến ngẫu nhiên có phân phối chuẩn với trung bình 100 $\Omega$ và độ lệch chuẩn là 8 $\Omega$. Tính xác suất để một mẫu ngẫu nhiên gồm 16 điện trở sẽ có điện trở trung bình nhỏ hơn 95 $\Omega$.

# Phân phối mẫu của $\overline{X}$

*Giải.*

- Theo Định lý 1 thì $\overline{X}$ có phân phối chuẩn (Hình 3) với $\mu_{\overline{X}} = 100$ và $\sigma_{\overline{X}} = \sqrt{\frac{64}{16}} = 2$.
- Do đó,

$$P(\overline{X} < 95) = \Phi\left(\frac{95-100}{2}\right) = \Phi(-2,5) = 1 - \Phi(2,5) \approx 1 - 0,99379 = 0,00621.$$




**Hình 3:** Phân phối của $\overline{X}$ trong Ví dụ 14

# Định lý giới hạn trung tâm

**Định lý 2**

Nếu mẫu ngẫu nhiên kích thước $n$, $W_X = (X_1, X_2, \ldots, X_n)$, được xây dựng từ biến ngẫu nhiên $X$ có kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$ hữu hạn và $\overline{X}$ là trung bình của mẫu ngẫu nhiên, thì giới hạn (theo nghĩa phân phối) của phân phối xác suất của

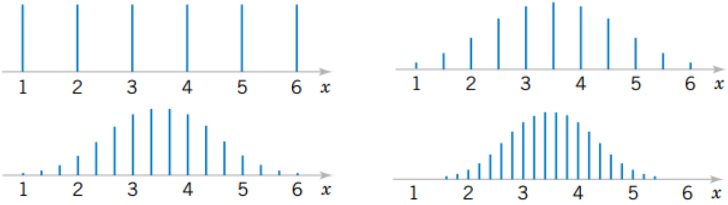
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{26}$$

khi $n \to \infty$, là phân phối chuẩn tắc.

# Định lý giới hạn trung tâm

✍

- Sự chính xác của xấp xỉ chuẩn của $\overline{X}$ phụ thuộc vào kích thước mẫu $n$.
- Trong thực hành, nếu $n \geq 30$, thống kê $Z$ trong (26) sẽ xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0;1)$.




**Hình 4:** Phân phối của số chấm trung bình khi tung 1, 2, 3 hoặc 5 con xúc sắc

# Định lý giới hạn trung tâm

**Ví dụ 15**

Cho $X$ là biến ngẫu nhiên liên tục có phân phối đều trên đoạn $[2; 4]$. Tìm phân phối xác suất của trung bình mẫu của mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ với kích thước $n = 40$ được xây dựng từ biến ngẫu nhiên $X$.

*Giải.*

- Kỳ vọng và phương sai của $X$ lần lượt là

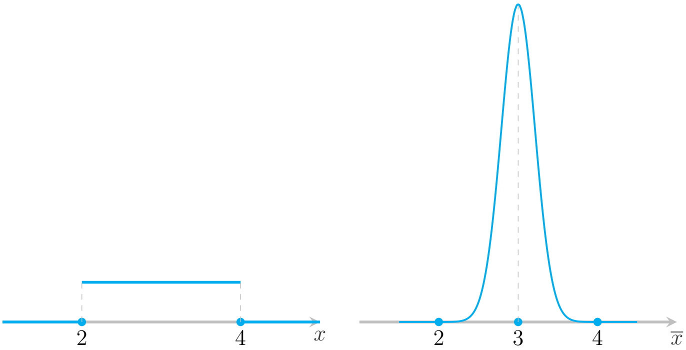
$$\mu_X = \frac{2+4}{2} = 3 \quad \text{và} \quad \sigma_X^2 = \frac{(4-2)^2}{12} = \frac{1}{3}.$$

- Theo Định lý 2, $\overline{X}$ có phân phối xấp xỉ phân phối chuẩn với

$$\mu_{\overline{X}} = 3 \quad \text{và} \quad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} = \frac{1}{3 \times 40} = \frac{1}{120}.$$

Phân phối xác suất của $X$ và phân phối mẫu của $\overline{X}$ cho trong Hình 5.

# Định lý giới hạn trung tâm




**Hình 5:** Phân phối của $X$ và $\overline{X}$ trong Ví dụ 15

# Phân phối mẫu của $\overline{X}_1 - \overline{X}_2$

- Giả sử $X_1$ và $X_2$ là hai biến ngẫu nhiên tương ứng với hai tổng thể khác nhau và $X_1$ và $X_2$ có phân phối chuẩn $\mathcal{N}(\mu_1;\sigma_1^2)$ và $\mathcal{N}(\mu_2;\sigma_2^2)$.
- Xét hai mẫu ngẫu nhiên độc lập $W_{X_1} = (X_{11}, X_{12}, \ldots, X_{1n_1})$ và $W_{X_2} = (X_{21}, X_{22}, \ldots, X_{2n_2})$ kích thước $n_1$ và $n_2$ tương ứng. Khi đó, $\overline{X}_1 - \overline{X}_2$ là biến ngẫu nhiên có phân phối chuẩn với kỳ vọng và phương sai là

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2 \tag{27}$$

và

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}. \tag{28}$$

✍ Nếu $X_1$ và $X_2$ không có phân phối chuẩn và nếu $n_1 \geq 30$ và $n_2 \geq 30$ thì ta có thể sử dụng Định lý giới hạn trung tâm, $\overline{X}_1$ và $\overline{X}_2$ có phân phối xấp xỉ phân phối chuẩn. Khi đó, $\overline{X}_1 - \overline{X}_2$ là biến ngẫu nhiên có phân phối xấp xỉ phân phối chuẩn với kỳ vọng và phương sai được xác định bởi các công thức (27) và (28).

# Phân phối mẫu của $\overline{X}_1 - \overline{X}_2$

### Định lý 3

Giả sử hai mẫu ngẫu nhiên độc lập $W_{X_1} = (X_{11}, X_{12}, \ldots, X_{1n_1})$ và $W_{X_2} = (X_{21}, X_{22}, \ldots, X_{2n_2})$ được xây dựng từ hai biến ngẫu nhiên $X_1$ và $X_2$ độc lập có kỳ vọng $E(X_1) = \mu_1$, $E(X_2) = \mu_2$ và phương sai $V(X_1) = \sigma_1^2$, $V(X_2) = \sigma_2^2$. Khi đó, phân phối xác suất của

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{29}$$

xấp xỉ phân phối chuẩn tắc, nếu các điều kiện của Định lý giới hạn trung tâm được thỏa mãn. Ngoài ra, nếu $X_1$ và $X_2$ có phân phối chuẩn $\mathcal{N}(\mu_1; \sigma_1^2)$ và $\mathcal{N}(\mu_2; \sigma_2^2)$ thì thống kê $Z$ trong (29) có phân phối chuẩn tắc $\mathcal{N}(0; 1)$.

# Phân phối Khi-bình phương với phương sai mẫu hiệu chỉnh

**Định lý 4**

Nếu $S^2$ là phương sai hiệu chỉnh của mẫu ngẫu nhiên kích thước $n$, $W_X = (X_1, X_2, \dots, X_n)$, được thành lập từ biến ngẫu nhiên $X$ có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$, thì thống kê

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \tag{30}$$

có phân phối Khi-bình phương với $n-1$ bậc tự do.

# Phân phối Student với thống kê mẫu

**Định lý 5**

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước $n$ được xây dựng từ biến ngẫu nhiên $X$ có phân phối chuẩn với kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$ chưa biết. Khi đó, thống kê

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{31}$$

có phân phối Student với $n - 1$ bậc tự do.

# Phân phối Fisher với hai phương sai mẫu hiệu chỉnh

Giả sử hai mẫu ngẫu nhiên kích thước $n_1$ và $n_2$ được chọn từ hai tổng thể có phân phối chuẩn với phương sai $\sigma_1^2$ và $\sigma_2^2$. Từ Định lý 4,

$$\chi_1^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \quad \text{và} \quad \chi_2^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2}$$

là các biến ngẫu nhiên có phân phối Khi-bình phương với số bậc tự do tương ứng là $n_1 - 1$ và $n_2 - 1$. Hơn nữa, vì các mẫu được chọn ngẫu nhiên, nên chúng độc lập. Do đó, sử dụng Định lý 22 (Chương 2) với $\chi_1^2 = U$ và $\chi_2^2 = V$ ta nhận được kết quả dưới đây.

# Phân phối Fisher với hai phương sai hiệu chỉnh mẫu

**Định lý 6**

Nếu $S_1^2$ và $S_2^2$ là hai phương sai hiệu chỉnh của hai mẫu ngẫu nhiên độc lập được chọn từ hai tổng thể có phân phối chuẩn với phương sai $\sigma_1^2$ và $\sigma_2^2$ tương ứng, thì

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

có phân phối Fisher với $\nu_1 = n_1 - 1$ và $\nu_2 = n_2 - 1$ bậc tự do.

# 4.1. MẪU NGẪU NHIÊN VÀ PHÂN PHỐI MẪU

# Bài tập

## Bài 1

Một hãng sản xuất chip máy tính quảng cáo rằng dưới 5% sản phẩm của họ bị lỗi. Kiểm tra ngẫu nhiên 1000 con chip do hãng sản xuất thì phát hiện thấy $3,5\%$ số chip bị lỗi. Hãy cho biết:

**(a)** Tổng thể muốn nghiên cứu là gì?

**(b)** Mẫu thu thập được là gì?

**(c)** Tham số quan tâm trong nghiên cứu này là gì?

**(d)** Giá trị thống kê trong nghiên cứu là gì?

**(e)** Giá trị 5% chỉ tham số hay thống kê?

**(f)** Giá trị $3,5\%$ chỉ tham số hay thống kê?

## Bài 2

Bảng dữ liệu sau đây liệt kê tỷ lệ phần trăm nam giới và nữ giới trong năm nhóm tuổi không có bảo hiểm y tế tại Hoa Kỳ vào tháng 9 năm 2008.

| Nhóm tuổi | Nam | Nữ |
|---|---|---|
| Dưới 18 | 8,5 | 8,5 |
| 18-24 | 32,3 | 24,9 |
| 25-34 | 30,4 | 21,4 |
| 35-44 | 21,3 | 17,1 |
| 45-64 | 13,5 | 13,0 |

Hãy sử dụng kỹ thuật biểu đồ hình cột và biểu đồ hình quạt để trình bày những số liệu này.

# Bài tập

## Bài 3

Quan sát thời gian khô (đơn vị: giờ) của 15 mẫu sơn Latex, chúng ta thu được số liệu như sau:

$$3,4; 2,5; 4,8; 2,9; 3,6; 2,8; 3,3; 5,6; 3,7; 2,8; 4,4; 4,0; 5,2; 3,0; 4,8$$

**(a)** Hãy biểu diễn tập dữ liệu trên dưới dạng bảng tần số với 4 khoảng trong đó cận trái của khoảng đầu tiên là 2.

**(b)** Hãy vẽ biểu đồ tổ chức đồ cho tập dữ liệu trên dựa trên cách biểu diễn ở ý (a).

# Bài tập

## Bài 4

Quan sát thời gian khô (đơn vị: giờ) của 15 mẫu sơn Latex, chúng ta thu được số liệu như sau:

$$3,4; 2,5; 4,8; 2,9; 3,6; 2,8; 3,3; 5,6; 3,7; 2,8; 4,4; 4,0; 5,2; 3,0; 4,8$$

**(a)** Hãy tính các đặc trưng mẫu đo trung tâm của dữ liệu: trung bình mẫu, trung vị mẫu.

**(b)** Hãy tính các đặc trưng mẫu đo độ phân tán của dữ liệu: phương sai mẫu và độ lệch chuẩn mẫu hiệu chỉnh.

# Bài tập

## Bài 5

Theo dõi thời gian gia công 34 chi tiết máy ta thu được số liệu như sau:

| Thời gian (phút) | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 |
|---|---|---|---|---|---|---|
| Số chi tiết máy | 3 | 4 | 10 | 9 | 5 | 3 |

**(a)** Hãy tính các độ đo trung tâm của dữ liệu: trung bình mẫu, trung vị mẫu.

**(b)** Hãy tính các độ đo mức độ phân tán của dữ liệu: phương sai mẫu và độ lệch chuẩn mẫu hiệu chỉnh.

**(c)** Hãy vẽ biểu đồ tổ chức đồ cho tập dữ liệu.

# Bài tập

## Bài 6

Đo áp lực $X$ (tính bằng kg/cm$^2$) của 18 thùng chứa ta được bảng kết quả sau:

| Áp lực (kg/cm$^2$) | 19,6 | 19,5 | 19,9 | 20,0 | 19,8 | 20,5 | 21,0 | 18,5 | 19,7 |
|---|---|---|---|---|---|---|---|---|---|
| Số thùng | 1 | 2 | 2 | 4 | 2 | 3 | 2 | 1 | 1 |

**(a)** Hãy tính các độ đo trung tâm của dữ liệu: trung bình mẫu, trung vị mẫu.

**(b)** Hãy tính các độ đo mức độ phân tán của dữ liệu: phương sai mẫu và độ lệch chuẩn mẫu hiệu chỉnh.

# Bài tập

## Bài 7

Giả sử chiều dài của một chi tiết máy là biến ngẫu nhiên có phân phối chuẩn với trung bình là 25 centimét độ lệch chuẩn là 0,2 centimét. Đo chiều dài của 10 chi tiết máy.

**(a)** Hãy tìm phân phối xác suất của chiều dài trung bình của 10 chi tiết máy được đo.

**(b)** Tính xác suất để chiều dài trung bình của 10 chi tiết máy được đo lớn hơn 25,06 centimét.

## Bài 8

Giả sử tuổi thọ (đơn vị: năm) của một loại thiết bị là biến ngẫu nhiên có phân phối mũ với tham số $\lambda = 0,2$. Quan sát tuổi thọ của 50 thiết bị.

**(a)** Hãy tìm phân phối xác suất xấp xỉ của tuổi thọ trung bình của 50 thiết bị được quan sát.

**(b)** Tính xác suất để tuổi thọ trung bình của 50 thiết bị được quan sát nhỏ hơn 4,9 năm.

# Bài tập

## Bài 9

Giả sử tỷ lệ khách hàng yêu thích sản phẩm của một công ty công nghệ là 55%. Phỏng vấn 100 khách hàng, gọi $X$ là số khách hàng trả lời yêu thích sản phẩm của công ty.

**(a)** Hãy tìm phân phối xác suất của $X$.

**(b)** Hãy tìm phân phối xác suất xấp xỉ của tỷ lệ khách hàng yêu thích sản phẩm của công ty trong 100 người được phỏng vấn.

**(c)** Tính xác suất để tỷ lệ khách hàng yêu thích sản phẩm của công ty trong 100 người được phỏng vấn lớn hơn 50%

# Bài tập

## Bài 10

Giả sử tỷ lệ thiết bị bị lỗi trong dây chuyển công nghệ cũ là $1\%$ và trong dây chuyền công nghệ mới là $0,5\%$. Quan sát ngẫu nhiên 100 thiết bị được sản xuất theo dây chuyền công nghệ cũ và 120 thiết bị được sản xuất theo dây chuyền công nghệ mới.

**(a)** Hãy tìm phân phối xác suất xấp xỉ của hiệu 2 tỷ lệ thiết bị bị lỗi trong 2 mẫu trên.

**(b)** Tính xác suất để tỷ lệ thiết bị bị lỗi trong mẫu sản xuất theo công nghệ cũ cao hơn trong mẫu sản xuất theo công nghệ mới.

# Bài tập

## Bài 11

Giả sử thu nhập ($X$) của sinh viên tốt nghiệp Ngành Công nghệ thông tin (CNTT), Đại học Bách khoa Hà Nội (HUST) trung bình là 15000 USD/năm và độ lệch chuẩn là 1500 USD; còn thu nhập ($Y$) của sinh viên tốt nghiệp Ngành CNTT Đại học Quốc gia Hà Nội (VNU) trung bình là 13000 USD/năm và độ lệch chuẩn là 1200 USD.

**(a)** Giả sử $X$ và $Y$ có phân phối chuẩn. Quan sát thu nhập của 10 sinh viên tốt nghiệp Ngành CNTT của HUST và 12 sinh viên Ngành CNTT của VNU. Hãy tính xác suất để thu nhập trung bình của 10 sinh viên HUST cao hơn 12 sinh viên VNU ít nhất 1500 USD.

**(b)** Giả sử không biết $X$ và $Y$ có phân phối gì. Quan sát thu nhập của 50 sinh viên tốt nghiệp Ngành CNTT của HUST và 60 sinh viên của VNU. Hãy tính xác suất để thu nhập trung bình của 50 sinh viên của HUST cao hơn 60 sinh viên của VNU ít nhất 1500 USD.

# Bài tập

**Bài 12**

**(a)** Giả sử $T$ có phân phối Student với 12 bậc tự do. Hãy tìm các hằng số $a, b$ sao cho $P(T > a) = 0,025$ và $P(-1,78 < T < b) = 0,9$.

**(b)** Giả sử $T$ có phân phối Fisher với 10 và 12 bậc tự do. Hãy tìm các hằng số $a, b$ sao cho $P(F > a) = 0,025$ và $P(-1,78 < F < b) = 0,9$.