XÁC SUẤT - THỐNG KÊ

CHƯƠNG 4: LÝ THUYẾT ƯỚC LƯỢNG

TS. Phan Thị Hường

Trường Đại học Bách Khoa TP HCM
Khoa Khoa học ứng dụng, bộ môn Toán ứng dụng
Email: huongphan@hcmut.edu.vn

TP. HCM — 2020.

NỘI DUNG

1. MỘT SỐ KHÁI NIỆM CƠ BẢN

NỘI DUNG

1. MỘT SỐ KHÁI NIỆM CƠ BẢN

2. SỬ DỤNG CÁC THỐNG KÊ MẪU ĐỂ MÔ TẢ DỮ LIỆU.

NỘI DUNG

1. MỘT SỐ KHÁI NIỆM CƠ BẢN
2. SỬ DỤNG CÁC THỐNG KÊ MẪU ĐỂ MÔ TẢ DỮ LIỆU.
3. ƯỚC LƯỢNG KHOẢNG

NGÀNH KHOA HỌC DỮ LIỆU

Thống kê là một ngành khoa học dữ liệu mà trong lĩnh vực này chúng ta cần giải quyết các vấn đề liên qua đến thu thập dữ liệu, biểu diễn dữ liệu, phân tích dữ liệu và sử dụng dữ liệu để đưa ra những quyết định, giải quyết các vấn đề, hoặc thiết kế sản phẩm, hay quy trình sản xuất, ...

NGÀNH KHOA HỌC DỮ LIỆU

Thống kê là một ngành khoa học dữ liệu mà trong lĩnh vực này chúng ta cần giải quyết các vấn đề liên qua đến thu thập dữ liệu, biểu diễn dữ liệu, phân tích dữ liệu và sử dụng dữ liệu để đưa ra những quyết định, giải quyết các vấn đề, hoặc thiết kế sản phẩm, hay quy trình sản xuất, ...

PHÂN TÍCH SỰ BIẾN THIÊN

Các phương pháp thống kê cho phép chúng ta mô tả và hiểu về sự biến thiên.

Tổng thể và mẫu

- **Tổng thể (population):** tập hợp tất cả những phần tử mang đặc trưng quan tâm hay cần nghiên cứu.

TỔNG THỂ VÀ MẪU

- **Tổng thể (population):** tập hợp tất cả những phần tử mang đặc trưng quan tâm hay cần nghiên cứu.
- **Mẫu (sample):** là một tập con được chọn ra từ tổng thể.

TỔNG THỂ VÀ MẪU

- **Tổng thể (population):** tập hợp tất cả những phần tử mang đặc trưng quan tâm hay cần nghiên cứu.
- **Mẫu (sample):** là một tập con được chọn ra từ tổng thể.
- **Tham số (parameter):** là một đặc trưng cụ thể của một tổng thể.

TỔNG THỂ VÀ MẪU

- **Tổng thể (population):** tập hợp tất cả những phần tử mang đặc trưng quan tâm hay cần nghiên cứu.
- **Mẫu (sample):** là một tập con được chọn ra từ tổng thể.
- **Tham số (parameter):** là một đặc trưng cụ thể của một tổng thể.
- **Thống kê (statistic):** là một đặc trưng cụ thể của một mẫu.

TỔNG THỂ VÀ MẪU

- **Tổng thể (population):** tập hợp tất cả những phần tử mang đặc trưng quan tâm hay cần nghiên cứu.
- **Mẫu (sample):** là một tập con được chọn ra từ tổng thể.
- **Tham số (parameter):** là một đặc trưng cụ thể của một tổng thể.
- **Thống kê (statistic):** là một đặc trưng cụ thể của một mẫu.

Tổng thể và mẫu

- **Tổng thể (population):** tập hợp tất cả những phần tử mang đặc trưng quan tâm hay cần nghiên cứu.
- **Mẫu (sample):** là một tập con được chọn ra từ tổng thể.
- **Tham số (parameter):** là một đặc trưng cụ thể của một tổng thể.
- **Thống kê (statistic):** là một đặc trưng cụ thể của một mẫu.

Thống kê mô tả và thống kê suy diễn

Hai nhánh chính của thống kê:

- Thống kê mô tả: Thu thập, sắp xếp, biểu diễn, tổng hợp dữ liệu.
- Thống kê suy diễn: Phân tích và đưa ra kết luận dữ liệu.

Quy trình xử lý số liệu tổng quát trong các ngành kỹ thuật:

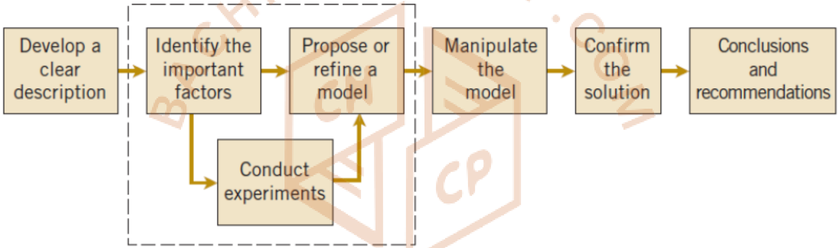


Figure 1-1 The engineering method

THỐNG KÊ MẪU

- Với một mẫu cỡ n, trung bình mẫu

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Thống kê mẫu

- Với một mẫu cỡ n, trung bình mẫu

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- Với một mẫu cỡ n, phương sai mẫu

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

và độ lệch chuẩn mẫu là s.

Thống kê mẫu

- Với một mẫu cỡ n, trung bình mẫu

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- Với một mẫu cỡ n, phương sai mẫu

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

  và độ lệch chuẩn mẫu là s.

- Trung vị tổng thể (median) là một giá trị x_0 thoả cả 2 tính chất

$$\mathbb{P}(X \leq x_0) \geq 0.5 \text{ và } \mathbb{P}(X \geq x_0) \geq 0.5$$

THỐNG KÊ MẪU

- Với một mẫu cỡ n, trung bình mẫu

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- Với một mẫu cỡ n, phương sai mẫu

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

  và độ lệch chuẩn mẫu là s.

- Trung vị tổng thể (median) là một giá trị x_0 thoả cả 2 tính chất

$$\mathbb{P}(X \leq x_0) \geq 0.5 \text{ và } \mathbb{P}(X \geq x_0) \geq 0.5$$

- Trung vị mẫu: Trong một tập dữ liệu được sắp xếp theo thứ tự tăng dần, trung vị là giá trị "chính giữa"của dữ liệu (50% bên trên, 50% bên dưới).

CÁCH TÌM TRUNG VỊ MẪU

- Vị trí của trung vị: sắp xếp dữ liệu theo thứ tự tăng dần, gọi i là vị trí của trung vị

$$i = \frac{n+1}{2}$$

+ Nếu i chẵn, trung vị = X_i,
+ Nếu i lẻ, trung vị = $\dfrac{X_{[i]} + X_{[i]+1}}{2}$, với $[i]$ là phần nguyên của i.

- Mức phân vị $100p\%$ tổng thể là giá trị x_0 thoả cả 2 tính chất

$$\mathbb{P}(X \leq x_0) \geq p \text{ và } \mathbb{P}(X \geq x_0) \geq 1 - p$$

- Mức phân vị $100p\%$ mẫu: là giá trị x_0 thoả cả 2 tính chất sau:
  1. Có ít nhất $100p\%$ dữ liệu nhỏ hơn hoặc bằng giá trị x_0.
  2. Có ít nhất $100p\%$ dữ liệu lớn hơn hoặc bằng x_0.

  Nếu có hai giá trị cùng thoả cả hai tính chất này thì mức phân vị $100p\%$ của mẫu bằng trung bình của 2 giá trị này.

Ghi chú 2.1

- *Mức phân vị* 25% *mẫu được gọi là mức phân vị thứ nhất, ký hiệu* $Q1$.
- *Mức phân vị* 50% *mẫu được gọi là trung vị, ký hiệu* $Q2$.
- *Mức phân vị* 75% *mẫu được gọi là mức phân vị thứ 3, ký hiệu* $Q3$.
- *Các giá trị* $Q1$, $Q2$, $Q3$ *được gọi là các điểm tứ phân vị, và định nghĩa các khoảng tứ phân vị của dữ liệu.*

Ghi chú 2.1

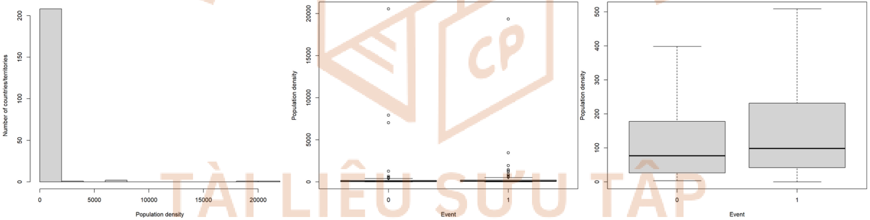
- *Mức phân vị* 25% *mẫu được gọi là mức phân vị thứ nhất, ký hiệu* $Q1$.
- *Mức phân vị* 50% *mẫu được gọi là trung vị, ký hiệu* $Q2$.
- *Mức phân vị* 75% *mẫu được gọi là mức phân vị thứ 3, ký hiệu* $Q3$.
- *Các giá trị Q1, Q2, Q3 được gọi là các **điểm tứ phân vị**, và định nghĩa các khoảng tứ phân vị của dữ liệu.*

Cách tìm mức phân vị $100p\%$ mẫu

1. Sắp xếp dữ liệu theo thứ tự tăng dần.
2. Nếu np là một số **không nguyên**, mức phân vị $100p\%$ của mẫu là giá trị tại vị trí $\lfloor np \rfloor$, trong đó $\lfloor np \rfloor$ là số nguyên nhỏ nhất lớn hơn np.
3. Nếu np là **số nguyên**, thì mức phân vị $100p\%$ của mẫu là giá trị trung bình của 2 giá trị tại vị trí np và vị trí $np+1$.

ĐỊNH NGHĨA 2.1

Điểm ngoại lai là những giá trị trong dữ liệu có giá trị khác biệt rõ nét so với các giá trị còn lại.



XÁC ĐỊNH ĐIỂM NGOẠI LAI BẰNG NGUYÊN TẮC TUKEY FENCES

1. Tính miền phân vị: $IQR = Q3 - Q1$, trong đó $Q1$ là điểm phân vị thứ 1 và $Q3$ là điểm phân vị thứ 3 của dữ liệu.
2. Một giá trị được gọi là ngoại lai (outlier) nếu nó cách điểm tứ phân vị gần nhất một khoảng hơn 1.5 lần miền phân vị, tức là

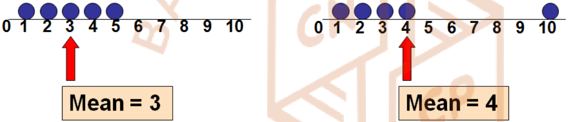
$$x > Q_3 + 1.5IQR \qquad \text{hoặc} \qquad x < Q_1 - 1.5IQR.$$

3. Một điểm ngoại lệ được gọi là xa (extreme outlier) nếu nó cách điểm tứ phân vị gần nhất một khoảng nhiều hơn 3 miền phân vị, tức là

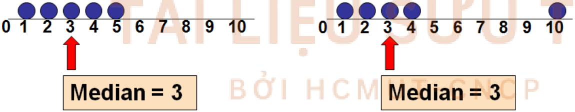
$$x > Q_3 + 3IQR \qquad \text{hoặc} \qquad x < Q_1 - 3IQR$$

SO SÁNH TRUNG BÌNH VÀ TRUNG VỊ

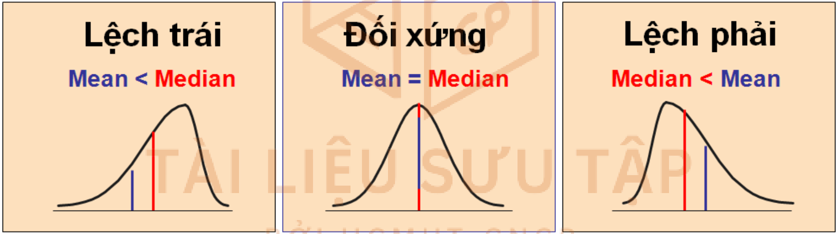
- Trung bình bị ảnh hưởng lớn bởi các điểm ngoại lai, trung vị thì không bị ảnh hưởng bởi các điểm ngoại lai.



$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3 \qquad \frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$



- Trung bình thường được sử dụng để đo giá trị trung tâm của dữ liệu, nếu các điểm ngoại lai không tồn tại hoặc sau khi loại bỏ các điểm ngoại lai.
- Vị trí của trung vị và trung bình ảnh hưởng bởi hình dạng của phân phối:



VÍ DỤ 2.1

Kết quả đo huyết áp của 10 *bệnh nhân như sau:*

$$120, 130, 120, 150, 130, 170, 180, 160, 170, 150$$

(A) *Tính trung bình mẫu và phương sai mẫu.*
(B) *Tính các điểm tứ phân vị của dữ liệu.*
(C) *Tìm các điểm ngoại lai của dữ liệu.*

KHÁI NIỆM ƯỚC LƯỢNG ĐIỂM

ĐỊNH NGHĨA 3.1

- *Họ các BNN $X_1, X_2, X_3, ..., X_n$ được gọi là một mẫu ngẫu nhiên với cỡ mẫu n nếu các BNN này là độc lập và có cùng phân phối.*
- *Nếu X là một BNN với hàm mật độ xác suất $f(x)$, phụ thuộc vào tham số θ, và nếu X_1, X_2, ..., X_n là một mẫu ngẫu nhiên từ X, thống ke mẫu $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ được gọi là một ước lượng điểm của θ.*
- *Chú ý rằng $\hat{\Theta}$ là một BNN nhiên.*
- *Giá trị $\hat{\Theta}$ tính trên một mẫu thực nghiệm $x_1, x_2, ;x_n$ được gọi là giá trị ước lượng điểm ký hiệu là θ.*

GHI CHÚ 3.1

Các thống kê mẫu: trung bình mẫu, phương sai mẫu, trung vị mẫu, ... nếu được tính từ các mẫu ngẫu nhiên thì đều là những giá trị ước lượng điểm

KHOẢNG TIN CẬY

- Đối với bài toán ước lượng điểm, một câu hỏi được đặt ra là: làm sao xác định được độ không chắc chắn của một ước lượng điểm cho một tham số của tổng thể?
- Để trả lời cho câu hỏi trên, ta sử dụng ước lượng khoảng tin cậy (confidence interval estimate)

Khoảng tin cậy và độ tin cậy

Định nghĩa 3.2

Cho $0 < \alpha < 1$, *một khoảng* $[l, u]$ *được gọi là một khoảng tin cậy* $100 \times (1-\alpha)\%$ *cho tham số* θ *nếu*

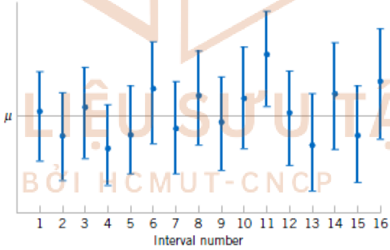
$$P(L \leq \theta \leq U) = 1 - \alpha.$$

Trong đó l *và* u *lần lượt các ước lượng điểm của* L *và* U. *Đại lượng* $\gamma = 1 - \alpha$ *được gọi là* *độ tin cậy (confidence level)* *của khoảng này.*

KHOẢNG TIN CẬY VÀ ĐỘ TIN CẬY

Ý nghĩa:

- Khoảng tin cậy được tính theo cách này được biểu diễn là: $l \le \theta \le u$ với độ tin cậy $100 \times (1-\alpha)\%$.
- Nếu lặp lại nhiều lần việc lấy mẫu từ một tổng thể, thì sẽ có $100 \times (1-\alpha)\%$ số khoảng được tính toán theo cách này sẽ chứa giá trị thực của tham số θ.



Khoảng tin cậy một phía

Tương tự ta định nghĩa các khoảng tin cậy một phía như sau:

Định nghĩa 3.3

- *Khoảng tin cậy trái (left-sided confident interval) của tham số* θ *là khoảng* $[l, \infty]$ *với độ tin cậy* $100 \times (1-\alpha)\%$ *nếu*

$$P(L < \theta) = \gamma,$$

trong đó l *là một ước lượng điểm của* L.

- *Khoảng tin cậy phải (right-sided confident interval) của tham số* θ *là khoảng* $[-\infty, u]$ *với độ tin cậy* $100 \times (1-\alpha)\%$ *nếu*

$$P(\theta < U) = \gamma,$$

trong đó u *là một ước lượng điểm của* U.

KHOẢNG TIN CẬY: CÔNG THỨC TỔNG QUÁT

- Công thức tổng quát cho mọi khoảng tin cậy hai phía là:

$$\boxed{\text{Ước lượng điểm} \pm (\text{nhân tố tin cậy})(\text{sai số chuẩn})}$$

- Công thức tổng quát cho mọi khoảng tin cậy trái là:

$$\boxed{\text{Ước lượng điểm} - (\text{nhân tố tin cậy})(\text{sai số chuẩn})}$$

- Công thức tổng quát cho mọi khoảng tin cậy phải là:

$$\boxed{\text{Ước lượng điểm} + (\text{nhân tố tin cậy})(\text{sai số chuẩn})}$$

- Giá trị của nhân tố độ tin cậy (reliability factor) phụ thuộc vào độ tin cậy γ mong muốn.

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH BIẾT σ^2

- **Các giả định:**

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH BIẾT σ^2

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$.

Khoảng tin cậy cho kỳ vọng: TH biết σ^2

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$.
  - Biết phương sai σ^2.

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH BIẾT σ^2

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1,\dots,X_n$ được chọn từ tổng thể có phân phối chuẩn, $X_i \sim \mathcal{N}(\mu,\sigma^2)$, $i = 1,\dots,n$.
  - Biết phương sai σ^2.
- **Công thức xác định khoảng tin cậy :**

KTC hai phía: $\boxed{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$,

KTC trái: $\boxed{\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}} \le \mu}$,

KTC phải: $\boxed{\mu \le \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}}$,

trong đó $\bar{x}$ là trung bình mẫu, σ là độ lệch tiêu chuẩn của tổng thể (đã biết) và $z_{\alpha/2}$ là điểm phân vị (trên) mức $\alpha/2$ của phân phối chuẩn hóa $\mathcal{N}(0,1)$.

SAI SỐ/ĐỘ CHÍNH XÁC

- Sai số của KTC hai phía với độ tin cậy $\gamma = 1 - \alpha$

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

SAI SỐ/ĐỘ CHÍNH XÁC

- Sai số của KTC hai phía với độ tin cậy $\gamma = 1 - \alpha$

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Câu hỏi: Với độ tin cậy $100(1-\alpha)\%$ Cần chọn cỡ mẫu bao nhiêu để thu được sai số E^* như mong muốn?

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E^*}\right)^2$$

VÍ DỤ 3.1

Đo chỉ số IQ của các sinh viên trong một trường đại học, khảo sát 18 *sinh viên thu được kết quả sau:*

130	122	119	142	136	127
120	152	141	132	127	118
150	141	133	137	129	142

Biết rằng chỉ số IQ của sinh viên tuân theo phân phối chuẩn với $\sigma = 10.5$.

(B) *Lập khoảng tin cậy* 95% *cho chỉ số IQ trung bình.*

(C) *Với độ tin cậy* 95%, *nếu muốn sai số ước lượng bằng* 3.5 *thì phải khảo sát thêm bao nhiêu sinh viên nữa?*

PHÂN PHỐI STUDENT

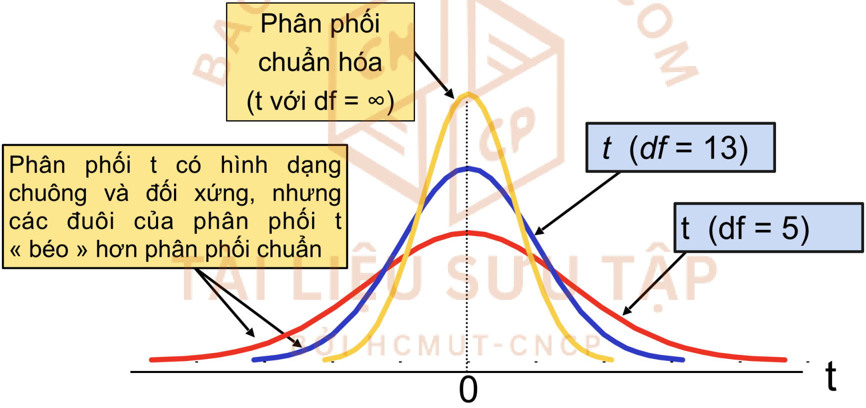
- Xét mẫu ngẫu nhiên $X_1,\dots,X_n$ được chọn từ tổng thể có phân phối chuẩn có trung bình μ,
- trung bình mẫu $\bar{X}$ và độ lệch chuẩn S.
- Khi đó, biến ngẫu nhiên

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}},$$

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

PHÂN PHỐI STUDENT

- $T \to Z$ khi n tăng.



KHOẢNG TIN CẬY CHO KỲ VỌNG: TH KHÔNG BIẾT σ^2

- **Các giả định:**

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH KHÔNG BIẾT σ^2

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$.

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH KHÔNG BIẾT σ^2

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$.
  - Phương sai σ^2 không biết.

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH KHÔNG BIẾT σ^2

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$.
  - Phương sai σ^2 không biết.
- **Công thức xác định khoảng tin cậy :**

KTC hai phía: $\boxed{\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}}$,

KTC trái: $\boxed{\bar{x} - t_{n-1,\alpha}\frac{s}{\sqrt{n}} \leq \mu}$,

KTC phải: $\boxed{\mu \leq \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}}}$,

trong đó $\bar{x}$ là trung bình mẫu, s là độ lệch tiêu chuẩn mẫu và $t_{n-1,\alpha/2}$ là điểm phân vị (trên) mức $\alpha/2$ của phân phối Student.

Ví dụ 3.2

Một mẫu ngẫu nhiên cỡ $n = 20$ *được chọn từ tổng thể có phân phối chuẩn có* $\bar{x} = 50$ *và* $s = 8$.

- (A) *Hãy lập khoảng tin cậy* 95% *cho* μ.
- (B) *Hãy lập khoảng tin cậy phải* 95% *cho* μ.

KHOẢNG TIN CẬY CHO KỲ VỌNG: TH MẪU LỚN.

- **Các giả định:**
  - Cỡ mẫu lớn.
  - Không biết phân phối của tổng thể, hoặc tổng thể không có phân phối chuẩn.
- **Công thức xác định khoảng tin cậy (σ không biết) :**

KTC hai phía: $\boxed{\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \lessapprox \mu \lessapprox \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}}$,

KTC trái: $\boxed{\bar{x} - z_{\alpha}\frac{s}{\sqrt{n}} \lessapprox \mu}$,

KTC phải: $\boxed{\mu \lessapprox \bar{x} + z_{\alpha}\frac{s}{\sqrt{n}}}$.

Chú ý: Nếu σ đã biết thì thay s trong các công thức trên bằng σ.

Ví dụ 3.3

Người ta đo nồng độ thuỷ ngân của một mẫu gồm 100 *con cá ở một hồ địa phương (ppm) và ghi nhận nồng độ thuỷ ngân trung bình là* 0.75 *(ppm) với độ lệch chuẩn là* 0.05 *(ppm).*

(A) *Hãy ước lượng* 90% *KTC cho nồng độ thuỷ ngân trung bình của các con cá được nuôi ở hồ này.*

(B) *Hãy ước lượng* 90% *KTC trái cho nồng độ thuỷ ngân trung bình của các con cá được nuôi ở hồ này.*

KHOẢNG TIN CẬY CHO TỶ LỆ

MỆNH ĐỀ 3.1

Giả sử rằng chúng ta cần ước lượng cho tỷ lệ phần tử thỏa tính chất $\mathscr{A}$ trong một tổng thể, ký hiệu là p. Gọi X là số phần tử thỏa tính chất $\mathscr{A}$ trong một mẫu gồm n phần tử. Nếu n đủ lớn, thì

$$\frac{\hat{P} - P}{\sqrt{P(1-P)/n}} \simeq N(0,1)$$

trong đó $\hat{P} = \dfrac{X}{n}$ *là tỷ lệ mẫu.*

KHOẢNG TIN CẬY CHO TỶ LỆ

- **Giả định:** cỡ mẫu đủ lớn.
- **Công thức xác định khoảng tin cậy :**

KTC hai phía: $\boxed{\hat{p} - z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \lessapprox \mu \lessapprox z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} + \hat{p}}$

KTC trái: $\boxed{\hat{p} - z_{\alpha}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \lessapprox \hat{p}}$

KTC phải: $\boxed{\hat{p} \lessapprox \hat{p} + z_{\alpha}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}}$

KHOẢNG TIN CẬY CHO TỶ LỆ

- Độ chính xác (sai số) của khoảng tin cậy:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}.$$

- Câu hỏi: Với độ tin cậy $100(1-\alpha)\%$, chọn cỡ mẫu n bằng bao nhiêu để có được sai số E^* (cho trước) như mong muốn?

KHOẢNG TIN CẬY CHO TỶ LỆ

- Độ chính xác (sai số) của khoảng tin cậy:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}.$$

- Câu hỏi: Với độ tin cậy $100(1-\alpha)\%$, chọn cỡ mẫu n bằng bao nhiêu để có được sai số E^* (cho trước) như mong muốn?
  $\Rightarrow$ Vì $p(1-p) \le 0.25, \quad 0 < p < 1$, ta có

$$E \le E^* \Leftrightarrow n \ge \left(\frac{z_{\alpha/2}}{E^*}\right)^2 \times 0.25.$$

trong đó E^* là sai số cho trước .

VÍ DỤ 3.4

Trong một nhà máy, ở khâu kiểm tra chất lượng sản phẩm, người ta lấy ngẫu nhiên 100 sản phẩm trong một lô hàng thì phát hiện được 20 sản phẩm kém chất lượng.

- *Hãy tìm KTC 95% cho tỷ lệ sản phẩm kém chất lượng của mỗi lô hàng.*
- *Với độ tin cậy 99%, nếu muốn độ chính xác bằng 0.04 thì phải kiểm tra bao nhiêu sản phẩm?*