Chương 4: Lý thuyết mẫu
Lý thuyết ước lượng

Nguyễn Thị Mộng Ngọc
University of Science, VNU - HCM
ngtmngoc@hcmus.edu.vn

Mẫu ngẫu nhiên

Mẫu ngẫu nhiên kích thước n là tập hợp của n biến ngẫu nhiên độc lập $X_1, X_2, \ldots, X_n$ được thành lập từ biến ngẫu nhiên X trong tổng thể nghiên cứu và có cùng quy luật phân phối xác suất với X.

Kí hiệu: $W = (X_1, X_2, \ldots, X_n)$

Thống kê

Một thống kê (statistic) là một hàm bất kì của các quan sát trong một mẫu ngẫu nhiên.

1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Các đặc trưng của mẫu ngẫu nhiên

Nếu $(X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n thì:

- **Trung bình mẫu:** $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- **Phương sai mẫu có hiệu chỉnh:**

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

- **Độ lệch chuẩn mẫu:** $S = \sqrt{S^2}$

đều là các thống kê.

1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Phân phối mẫu

Bởi vì thống kê là một mẫu ngẫu nhiên nên nó có phân phối xác suất.

Định nghĩa

Phân phối xác suất của một thống kê đgl một phân phối mẫu.

Ví dụ: Phân phối xác suất của $\bar{X}$ đgl phân phối mẫu của trung bình.

Nhận xét

Phân phối xác suất của một thống kê phụ thuộc vào phân phối của tổng thể, kích thước mẫu và phương pháp chọn mẫu.

Phân phối mẫu của trung bình và phương sai

Nếu tổng thể X có phân phối chuẩn $X \sim \mathcal{N}(\mu, \sigma^2)$ và $(X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên từ tổng thể trên thì

- $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- $\frac{(n-1)}{\sigma^2} S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$.
- $\frac{(\bar{X}-\mu)\sqrt{n}}{S} \sim t(n-1)$.
- $\bar{X}$ và S^2 là hai biến ngẫu nhiên độc lập.

Phân phối mẫu của trung bình và phương sai

Trường hợp tổng thể có phân phối xác suất chưa biết, từ định lí giới hạn trung tâm ta suy ra rằng

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Trong thực hành khi mẫu có kích thước đủ lớn ($n \geq 30$), ta có các phân phối xấp xỉ chuẩn sau:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1)$$

và

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx \mathcal{N}(0,1).$$

Sai số chuẩn (Standard Error) của trung bình

Sai số chuẩn (Standard Error) của trung bình, kí hiệu là $\sigma_{\bar{X}}$

$$\sigma_{\bar{X}} := \frac{\sigma}{\sqrt{n}}$$

Nhận xét:

- $\sigma_{\bar{X}}$ đo độ biến thiên của $\bar{X}$ xung quanh trung bình tổng thể μ.
- Sai số chuẩn càng nhỏ, ước lượng tham số từ tổng thể càng tốt và độ tin cậy cao.
- Độ biến thiên của tổng thể, σ, càng lớn thì sai số chuẩn, $\sigma_{\bar{X}}$, càng lớn.

Phân phối mẫu của tỉ lệ

Giả sử cần khảo sát đặc trưng $\mathcal{A}$ của tổng thể, khảo sát n phần tử và đặt

$$X_i = \begin{cases} 1 & \text{nếu thỏa } \mathcal{A} \\ 0 & \text{nếu khác} \end{cases}$$

thu được mẫu ngẫu nhiên $X_1, \ldots, X_n$ với $X_i \sim B(1, p)$, p là tỉ lệ phần tử thỏa đặc trưng $\mathcal{A}$.

Đặt $X = \sum_{i=1}^{n} X_i$ là số phần tử thỏa đặc trưng $\mathcal{A}$ trong mẫu khảo sát, thì $X \sim B(n, p)$.

Tỉ lệ mẫu $\hat{P}$ là một ước lượng của tỉ lệ p được xác định bởi

$$\hat{P} = \frac{X}{n}$$

Phân phối mẫu của tỉ lệ

Kì vọng và phương sai của $\hat{p}$ là

$$\mathbb{E}(\hat{P}) = p, \quad \mathbb{V}ar(\hat{P}) = \frac{p(1-p)}{n}$$

Theo định lí giới hạn trung tâm ta có

$$\frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1)$$

Vì vậy trong thực hành, khi
$np \geq 5, n(1-p) \geq 5$, thì $\hat{P} \approx N(p, \frac{p(1-p)}{n})$

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Bài toán ước lượng

Các tham số đặc trưng của tổng thể như trung bình, phương sai, tỷ lệ, ... được sử dụng rộng rãi trên nhiều lĩnh vực khác nhau. Tuy nhiên các tham số đặc trưng của tổng thể này thường chưa biết. Vì vậy cần ước lượng chúng bằng phương pháp mẫu.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Bài toán ước lượng

Giả sử biến ngẫu nhiên X có tham số θ chưa biết. Ước lượng tham số θ là dựa vào mẫu ngẫu nhiên $W = (X_1, X_2, \ldots, X_n)$ ta đưa ra thống kê $\hat{\theta}$ để ước lượng (dự đoán) θ.

Ví dụ:
○ **Ước lượng điểm**: chỉ ra $\hat{\theta} = \theta_0$ nào đó để ước lượng θ.
○ **Ước lượng khoảng**: chỉ ra một khoảng $(\hat{\theta}_L, \hat{\theta}_U)$ chứa θ sao cho $\mathbb{P}(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha$ cho trước, ($1 - \alpha$ đgl độ tin cậy của ước lượng).

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Ước lượng điểm - Ước lượng khoảng

- Một ước lượng điểm là một giá trị đơn.
- Một khoảng tin cậy cung cấp thông tin bổ sung về sự biến thiên của một ước lượng điểm tương ứng.

Ước lượng điểm

Định nghĩa

Một ước lượng điểm cho tham số tổng thể θ là một giá trị đơn $\hat{\theta}$ của một thống kê $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$.

Nhận xét

Thông thường giá trị được chọn này là giá trị cụ thể của một thống kê $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$ nào đó của mẫu ngẫu nhiên.

Ví dụ: Giá trị $\bar{x}$ của thống kê $\bar{X}$ được tính toán từ một mẫu ngẫu nhiên kích thước n là một ước lượng điểm của tham số trung bình tổng thể μ .

XSTK

N.T. M. Ngọc

1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Cùng một mẫu ngẫu nhiên ta có thể xây dựng được nhiều thống kê $\hat{\theta}$ khác nhau để ước lượng cho tham số tổng thể θ. Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau:

a. Ước lượng không chệch

Thống kê $\hat{\theta}$ của mẫu đgl ước lượng không chệch của tham số θ của biến ngẫu nhiên gốc X nếu:

$$\mathbb{E}(\hat{\theta}) = \theta$$

b. Ước lượng hiệu quả

Thống kê $\hat{\theta}$ của mẫu đgl ước lượng hiệu quả nhất của tham số θ của biến ngẫu nhiên gốc X nếu nó là ước lượng không chệch và có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng mẫu đó.

XSTK

N.T. M. Ngọc

1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

c. Ước lượng vững

Thống kê $\hat{\theta}$ của mẫu đgl ước lượng vững của tham số θ của biến ngẫu nhiên gốc X nếu: $\hat{\theta}$ hội tụ theo xác suất đến θ khi $n \to \infty$

d. Ước lượng đủ

Một ước lượng $\hat{\theta}$ đgl ước lượng đủ nếu nó chứa đựng toàn bộ các thông tin trong mẫu về tham số θ của ước lượng.

XSTK

N.T. M. Ngọc

1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Ước lượng khoảng

- Giả sử cần khảo sát một đặc tính X trên một tổng thể xác định.
- BNN X có phân phối $F(x;\theta)$, tham số θ chưa biết.
- Chọn một mẫu ngẫu nhiên cỡ n: $X = (X_1, \ldots, X_n)$.

Định nghĩa 1

Một ước lượng khoảng (interval estimator) của một tham số θ là một cặp các thống kê $L(X_1, \ldots, X_n)$ và $U(X_1, \ldots, X_n)$ của một mẫu ngẫu nhiên thỏa $L(\mathbf{X}) \leq U(\mathbf{X})$, và $L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})$. Nếu một mẫu thực nghiệm $\mathbf{x} = (x_1, \ldots, x_n)$ được quan trắc, $[l(\mathbf{x}), u(\mathbf{x})]$ gọi là một khoảng ước lượng (interval estimate) cho θ.

Khoảng tin cậy

Định nghĩa 2

Xét vector ngẫu nhiên $\mathbf{X} = (X_1, \ldots, X_n)$ có hàm mật độ đồng thời phụ thuộc vào tham số $\theta \in \Theta$ và $L(\mathbf{X})$ và $U(\mathbf{X})$ là hai thống kê sao cho $L(\mathbf{X}) \leq U(\mathbf{X})$. Khi đó, khoảng ngẫu nhiên $[L(\mathbf{X}), U(\mathbf{X})]$ gọi là *khoảng tin cậy cho tham số θ với độ tin cậy* $100(1-\alpha)\%$ nếu

$$\mathbb{P}\{L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\} = 1 - \alpha \qquad (1)$$

Chú ý

- Đôi khi, "khoảng tin cậy cho tham số θ với độ tin cậy $100(1-\alpha)\%$" thường được viết gọn là "khoảng tin cậy $100(1-\alpha)\%$ cho tham số θ".
- Với mẫu thực nghiệm $\mathbf{x} = (x_1, \ldots, x_n)$, ta có khoảng tin cậy cụ thể cho tham số θ là
$$l(\mathbf{x}) \leq \theta \leq u(\mathbf{x})$$

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Ý nghĩa

Với 100 lần lấy mẫu cùng cỡ n từ tổng thể thì

- Có $100(1-\alpha)$ lần lấy, giá trị tham số $\theta \in [l, u]$ (hay nói cách khác, có $100(1-\alpha)\%$ số khoảng được tính toán theo cách này sẽ chứa giá trị thực của tham số θ).
- Có 100α lần lấy, giá trị tham số $\theta \notin [l, u]$.
- Khoảng tin cậy được tính theo cách này là $l \leq \theta \leq u$ với độ tin cậy $100(1-\alpha)\%$.
- $(1-\alpha)$ đgl hệ số tin cậy hay độ tin cậy.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Công thức tổng quát cho khoảng tin cậy

Công thức tổng quát cho mọi khoảng tin cậy:

Ước lượng điểm $\pm$ dung sai

trong đó, dung sai $=$ (nhân tố độ tin cậy)$\times$ sai số chuẩn.

Chú ý:

Giá trị của nhân tố độ tin cậy phụ thuộc vào độ tin cậy mong muốn.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Ước lượng trung bình của tổng thể

Bài toán

Cho tổng thể với trung bình μ với phương sai có thể đã biết hoặc chưa biết. Từ mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ hãy ước lượng μ với độ tin cậy $1-\alpha$.

Cách giải quyết

Ta chia bài toán thành 3 trường hợp (TH) sau:

- TH1 Kích thước mẫu $n \geq 30$ (hoặc $n < 30$ nhưng X có phân phối chuẩn), σ^2 đã biết
- TH2 Kích thước mẫu $n \geq 30$, σ^2 chưa biết
- TH3 Kích thước mẫu $n < 30$, σ^2 chưa biết, X tuân theo quy luật phân phối chuẩn.

TH 1: σ^2 đã biết và $n \geq 30$ (hoặc $n < 30$ nhưng tổng thể có phân phối chuẩn)

Mệnh đề 1

Trong trường hợp này, thống kê

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

có phân phối chuẩn $N(0,1)$.

TH 1: σ^2 đã biết và $n \geq 30$ (hoặc $n < 30$ nhưng tổng thể có phân phối chuẩn)

Với độ tin cậy $1-\alpha$, ta có

$$\mathbb{P}\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

hay

$$\mathbb{P}\left\{\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

với $z_{1-\frac{\alpha}{2}}$ là phân vị mức $1-\alpha/2$ của phân phối chuẩn hóa $N(0,1)$.

TH 1: σ^2 đã biết và $n \geq 30$ (hoặc $n < 30$ nhưng tổng thể có phân phối chuẩn)

- Với mẫu ngẫu nhiên, khoảng tin cậy cho tham số μ với độ tin cậy $1-\alpha$ là

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

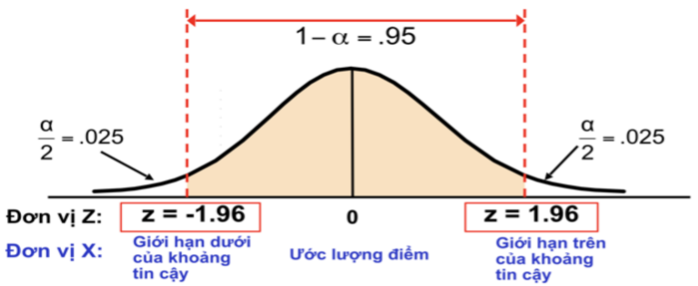
- Với mẫu thực nghiệm, khoảng tin cậy cho tham số μ với độ tin cậy $1-\alpha$ là

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

- Đại lượng $\epsilon = z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ được gọi là dung sai (sai số giới hạn) của khoảng tin cậy.

Tìm $z_{1-\alpha/2}$

- Xét một khoảng tin cậy 95% ($\alpha = 5\%$):



- Tra bảng phân phối chuẩn hóa Z: $z_{1-\alpha/2} = z_{0.975} = 1.96$.
- Cách kí hiệu khác: $z_{\alpha/2}$ (phân vị trên - upper percentile).

TH 1: σ^2 đã biết và $n \geq 30$ (hoặc $n < 30$ nhưng tổng thể có phân phối chuẩn)

Ví dụ: Hàm lượng kẽm trung bình thu hồi được từ một mẫu các giá trị đo kẽm tại 36 điểm đo khác nhau được xác định là 2.6g/ml. Xác định các khoảng tin cậy 95% và 99% cho mật độ kẽm trung bình ở sông. Giả thiết độ lệch tiêu chuẩn tổng thể là 0.3.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

TH 2: σ^2 chưa biết và $n \geq 30$.

Ta có thể dùng ước lượng của $Var(X)$ là S^2 để thay thế cho σ^2. Định lí giới hạn trung tâm nói rằng

Mệnh đề 2

Trong trường hợp này, thống kê

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

có phân phối chuẩn $N(0, 1)$.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

TH 2: σ^2 chưa biết và $n \geq 30$.

Với độ tin cậy $1 - \alpha$, ta có

$$\mathbb{P}\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right\} = 1 - \alpha$$

hay

$$\mathbb{P}\left\{\bar{X} - z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right\} = 1 - \alpha$$

với $z_{1-\frac{\alpha}{2}}$ là phân vị mức $1 - \alpha/2$ của phân phối chuẩn hóa $N(0, 1)$.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

TH 2: σ^2 chưa biết và $n \geq 30$.

- Với mẫu ngẫu nhiên, khoảng tin cậy cho tham số μ với độ tin cậy $1 - \alpha$ là
$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$
- Với mẫu thực nghiệm, khoảng tin cậy cho tham số μ với độ tin cậy $1 - \alpha$ là
$$\left[\bar{x} - z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$
- Đại lượng $\epsilon = z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$ được gọi là dung sai (sai số giới hạn) của khoảng tin cậy.

TH 3: σ^2 chưa biết; $n < 30$ và X có phân phối chuẩn.

Mệnh đề 3

Trong trường hợp này, thống kê

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

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

TH 3: σ^2 chưa biết; $n < 30$ và X có phân phối chuẩn.

Với độ tin cậy $1-\alpha$, ta có

$$\mathbb{P}\left\{-t_{1-\frac{\alpha}{2}}^{n-1} \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{1-\frac{\alpha}{2}}^{n-1}\right\} = 1-\alpha$$

hay

$$\mathbb{P}\left\{\bar{X} - t_{1-\frac{\alpha}{2}}^{n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{1-\frac{\alpha}{2}}^{n-1}\frac{S}{\sqrt{n}}\right\} = 1-\alpha$$

với $t_{1-\frac{\alpha}{2}}^{n-1}$ là phân vị mức $1-\frac{\alpha}{2}$ của luật phân phối Student với $n-1$ bậc tự do.

TH 3: σ^2 chưa biết; $n < 30$ và X có phân phối chuẩn.

- Với mẫu ngẫu nhiên, khoảng tin cậy cho tham số μ với độ tin cậy $1-\alpha$ là
$$\left[\bar{X} - t_{1-\frac{\alpha}{2}}^{n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}}^{n-1}\frac{S}{\sqrt{n}}\right]$$
- Với mẫu thực nghiệm, khoảng tin cậy cho tham số μ với độ tin cậy $1-\alpha$ là
$$\left[\bar{x} - t_{1-\frac{\alpha}{2}}^{n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}}^{n-1}\frac{s}{\sqrt{n}}\right]$$
- Đại lượng $\epsilon = t_{1-\frac{\alpha}{2}}^{n-1}\frac{s}{\sqrt{n}}$ được gọi là dung sai (sai số giới hạn) của khoảng tin cậy.

TH 3: σ^2 chưa biết; $n < 30$ và X có phân phối chuẩn.

Ví dụ: Các hàm lượng của 7 container axit sulfuric là $9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6$ lít. Tìm khoảng tin cậy 95% cho giá trị trung bình của tất cả các container đó, giả sử có phân phối chuẩn ước lượng.

Ước lượng trung bình của tổng thể

Tóm lại,

Các bước thực hiện

B1 Tìm trung bình mẫu $\bar{x}$ và phương sai mẫu s^2.

B2 Xác định trường hợp áp dụng:

- TH1 $n \geq 30$ (hoặc $n < 30$, X có phân phối chuẩn) và σ^2 đã biết.
- TH2 $n \geq 30$, σ^2 chưa biết.
- TH3 $n < 30$, X có phân phối chuẩn, và σ^2 chưa biết.

B3 Tìm phân vị: $z_{1-\alpha/2}$ nếu là TH1 và TH2; hoặc $t^{n-1}_{1-\alpha/2}$ nếu là TH3.

B4 Tìm dung sai:

$$\epsilon = \begin{cases} z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} & \text{nếu TH1} \\ z_{1-\alpha/2}\frac{s}{\sqrt{n}} & \text{nếu TH2} \\ t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}} & \text{nếu TH3} \end{cases}$$

KL Khoảng tin cậy $100(1-\alpha)\%$ cho trung bình của tổng thể là $[\bar{x}-\epsilon, \bar{x}+\epsilon]$.

Ước lượng trung bình tổng thể

Ví dụ:

Biết lương tháng của công nhân (Đv: triệu đồng) trong một nhà máy có phân phối chuẩn. Chọn ngẫu nhiên 16 công nhân khảo sát.

Lương tháng	0.8	1.0	1.2	1.3	1.5	1.7	2	2.3	2.5
Số công nhân	1	1	2	2	2	3	2	2	1

a. Giả sử $\sigma = 0.63$, tìm KTC 96% cho mức lương trung bình hàng tháng của một công nhân.

b. Lập KTC 99% cho mức lương trung bình.

Ước lượng tỷ lệ của tổng thể

Bài toán

Cho tổng thể X, trong đó tỷ lệ cá thể mang đặc tính $\mathcal{A}$ nào đó trong tổng thể là p. Từ mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ hãy tìm khoảng tin cậy cho p với độ tin cậy $1-\alpha$.

- Gọi Y là số phần tử thỏa tính chất $\mathcal{A}$ trong n phần tử khảo sát, thì $Y \sim B(n, p)$.
- Đặt

$$\hat{P} = \frac{Y}{n} \tag{2}$$

- Thống kê $\hat{P}$ có kỳ vọng và phương sai lần lượt là

$$\mathbb{E}(\hat{P}) = \mu_{\hat{P}} = p, \quad \mathbb{V}\mathrm{ar}(\hat{P}) = \sigma^2_{\hat{P}} = \frac{p(1-p)}{n}$$

Ước lượng tỷ lệ của tổng thể

Mệnh đề 4

Thống kê

$$Z = \frac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0,1) \quad (3)$$

và

$$W = \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \rightsquigarrow \mathcal{N}(0,1) \quad (4)$$

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Do đó, khi kích thước mẫu đủ lớn,

$$\mathbb{P}\left\{-z_{1-\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \leq z_{1-\alpha/2}\right\} = 1 - \alpha \quad (5)$$

hay

$$\mathbb{P}\left\{\hat{P} - z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right\} = 1 - \alpha \quad (6)$$

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Vậy

- Với mẫu ngẫu nhiên, khoảng tin cậy $100(1-\alpha)\%$ cho p là

$$\left[\hat{P} - z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P} + z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right]$$

- Với mẫu thực nghiệm, khoảng tin cậy $100(1-\alpha)\%$ cho p là

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Ước lượng tỷ lệ của tổng thể

Trong thực hành, ta thực hiện theo các bước sau đây:

Các bước thực hiện

B1 Tìm tỉ lệ mẫu: $\hat{p}$.

B2 Kiểm tra điều kiện: $n\hat{p} \geq 5$ và $n(1-\hat{p}) \geq 5$.

B3 Tìm phân vị: $z_{1-\alpha/2}$ bằng cách tra bảng.

B4 Tìm dung sai: $\epsilon = z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

KL Khoảng tin cậy $100(1-\alpha)\%$ cho tỷ lệ của tổng thể là $[\hat{p} - \epsilon, \hat{p} + \epsilon]$.

Ví dụ

Biết lương tháng của công nhân (Đv: triệu đồng) trong một nhà máy có phân phối chuẩn. Chọn ngẫu nhiên 16 công nhân khảo sát.

Lương tháng	0.8	1.0	1.2	1.3	1.5	1.7	2	2.3	2.5
Số công nhân	1	1	2	2	2	3	2	2	1

Công nhân gọi là có thu nhập cao nếu lương tháng từ 2 triệu đồng trở lên. Hãy lập khoảng tin cậy 95% cho tỉ lệ công nhân có thu nhập cao.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Xác định kích thước mẫu.

Nhận xét

- Chất lượng của ước lượng phản ánh qua độ tin cậy $1-\alpha$ và dung sai ϵ.
- Độ tin cậy càng cao và dung sai càng nhỏ thì ước lượng đó càng tốt.
- Tuy nhiên, dung sai ϵ lại phụ thuộc vào kích thước mẫu n và độ tin cậy $1-\alpha$.

Câu hỏi

Với độ tin cậy $1-\alpha$, nếu ta muốn dung sai ϵ đạt được ở một mức nào đó cho trước thì kích thước mẫu n tối thiểu là bao nhiêu?

Bài toán

Tìm giá trị nhỏ nhất của n sao cho $\epsilon \le \epsilon_0$, với ϵ_0 và α cho trước.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Xác định kích thước mẫu

Khi ước lượng trung bình tổng thể

a. Nếu biết $\mathbb{V}ar(X) = \sigma^2$, từ công thức

$$\epsilon = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Để $\epsilon \le \epsilon_0$ ta cần chọn

$$n \ge (z_{1-\alpha/2})^2\frac{\sigma^2}{\epsilon_0^2}$$

b. Nếu chưa biết σ^2, ta căn cứ vào mẫu đã cho để tính s^2. Từ đó ta xác định được kích thước mẫu tối thiểu:

$$n \ge (z_{1-\alpha/2})^2\frac{s^2}{\epsilon_0^2}$$

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Xác định kích thước mẫu

Khi ước lượng tỷ lệ tổng thể

a. Khi đã biết $\hat{p}$, để $\epsilon \le \epsilon_0$ thì từ công thức

$$\epsilon = z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow n \ge (z_{1-\alpha/2})^2\frac{\hat{p}(1-\hat{p})}{\epsilon_0^2}$$

b. Khi chưa biết $\hat{p}$, ta có $\epsilon = z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Do $\hat{p}(1-\hat{p})$ đạt giá trị cực đại 0.25 khi $\hat{p} = 0.5$ nên

$$\epsilon \le z_{1-\alpha/2}\sqrt{\frac{0.25}{n}}$$

Do đó, để $\epsilon \le \epsilon_0$ ta chọn n sao cho $z_{1-\alpha/2}\sqrt{\frac{0.25}{n}} \le \epsilon_0$ tức là

$$n \ge \frac{0.25(z_{1-\alpha/2})^2}{\epsilon_0^2}$$

Ví dụ

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?*

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Bài toán: Ước lượng phương sai tổng thể

- **Các giả định:**
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn với kỳ vọng μ và phương sai σ^2.
  - **Công thức tính khoảng tin cậy:** khoảng tin cậy $100(1-\alpha)\%$ cho σ^2 có dạng

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}},$$

trong đó $\chi^2_{\alpha/2,n-1}$ và $\chi^2_{1-\alpha/2,n-1}$ lần lượt là phân vị trên và phân vị dưới mức $\alpha/2$ và $1-\alpha/2$ của biến ngẫu nhiên Chi bình phương với $n-1$ bậc tự do.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Phân phối Chi bình phương

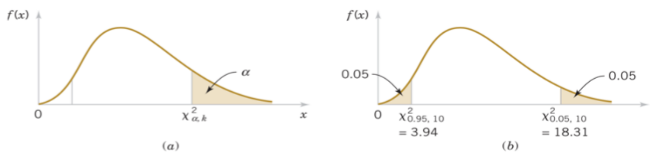
Xét $X_1, X_2, \ldots, X_n$ là một mẫu ngẫu nhiên được chọn từ một tổng thể có phân phối chuẩn với kỳ vọng μ và phương sai σ^2. Xét S^2 là phương sai mẫu, thì biến ngẫu nhiên

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

có phân phối Chi-bình phương (χ^2) với $n-1$ bậc tự do.

XSTK
N.T. M. Ngọc
1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Phân phối Chi bình phương



- Phân vị trên (upper percentile) mức α: $\mathbb{P}(X^2 > \chi^2_{\alpha,n-1}) = \alpha$ (hình (a)).
- Phân vị dưới (lower percentile) mức $1-\alpha$: $\mathbb{P}(X^2 > \chi^2_{1-\alpha,n-1}) = 1-\alpha$ (hình (b)).

Khoảng tin cậy cho phương sai

- Cho độ tin cậy $100(1-\alpha)\%$, bởi vì $X^2 = (n-1)S^2/\sigma^2$ có phân phối Chi bình phương với $n-1$ bậc tự do nên ta có

$$\mathbb{P}\left(\chi^2_{1-\alpha/2,n-1} \le X^2 \le \chi^2_{\alpha/2,n-1}\right) = 1-\alpha.$$

- Do vậy ta có

$$\mathbb{P}\left(\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right) = 1-\alpha.$$

- Ta thu được

$$\mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha.$$

XSTK

N.T. M. Ngọc

1. Lý thuyết mẫu
1.1 Mẫu ngẫu nhiên
1.2 Các đặc trưng của mẫu ngẫu nhiên
1.3 Phân phối mẫu
2. Lý thuyết ước lượng
2.1 Bài toán ước lượng tham số thống kê
2.2 Ước lượng điểm
2.3 Ước lượng khoảng
2.4 Ước lượng trung bình tổng thể
2.5 Ước lượng tỷ lệ tổng thể
2.6 Xác định kích thước mẫu
2.7 Ước lượng phương sai tổng thể

Ví dụ

Trong một nhà máy sản xuất kẹo, dây chuyền tự động được lập trình để đóng gói những bịch kẹo có trọng lượng là 52 g, độ lệch chuẩn cho phép là ± 1 g. Một kỹ sư kiểm tra chất lượng có nghi vấn rằng máy đóng bịch tự động hoạt động không tốt, và trọng lượng một số bịch kẹo do dây chuyền đóng gói có trọng lượng nhỏ hơn hoặc lớn hơn nhiều so với quy định. Để kiểm tra, kỹ sư này chọn ngẫu nhiên 10 bịch kẹo trong 1 lô hàng, và tính được phương sai mẫu bằng 4.2 g. Hãy tìm khoảng tin cậy 95% cho độ lệch chuẩn và cho kết luận xem máy đóng bịch có hoạt động tốt hay không?