KHOẢNG TIN CẬY

Nguyễn Văn Thìn

BỘ MÔN THỐNG KÊ TOÁN HỌC
KHOA TOÁN - TIN HỌC
ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM

Tháng 2 năm 2016

Outline

Outline

Ước lượng khoảng

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

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 mẫu 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

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

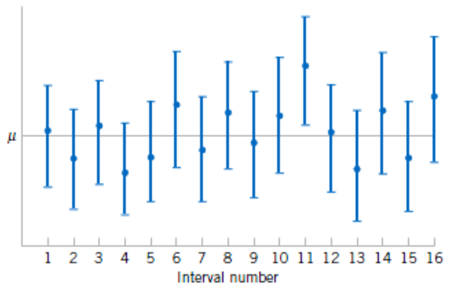
KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Khoảng tin cậy - Ý nghĩa

- Để cho ngắn gọn, ta thường dùng thuật ngữ "khoảng tin cậy $100(1-\alpha)\%$ cho tham số θ" thay vì "khoảng tin cậy cho tham số θ với độ tin cậy $100(1-\alpha)\%$".
- Thông thường, ta chọn độ tin cậy $1 - \alpha = 0.95$ (hoặc $0.9, 0.99$)
- Với mẫu thực nghiệm $\mathbf{x} = (x_1, \ldots, x_n)$, ta có khoảng tin cậy thực nghiệm cho tham số θ là $[l(\mathbf{x}), u(\mathbf{x})]$.
- **Ý nghĩa của độ tin cậy $(1-\alpha)$:** Cứ mỗi lần lấy mẫu, ta nhận được mẫu khác nhau và do đó khoảng tin cậy tìm được cũng khác nhau. Tuy nhiên, trong 100% lần lấy mẫu cỡ n thì
    - có $100(1-\alpha)\%$ lần giá trị tham số $\theta \in [l, u]$;
    - có $100\alpha\%$ lần giá trị tham số $\theta \notin [l, u]$.

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Khoảng tin cậy - Minh họa



Như vậy, khoảng tin cậy 95% cho tham số θ được tính từ mẫu thực nghiệm $\mathbf{x} = (x_1, \ldots, x_n)$ có thể chứa hoặc không chứa θ (ta không biết được). Tuy nhiên, ta tin tưởng 95% rằng khoảng này chứa tham số θ.

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Outline

Khoảng tin cậy cho trung bình - Phát biểu bài toán

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 tìm khoảng tin cậy cho μ với độ tin cậy $1-\alpha$ cho trước.

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.

Khoảng tin cậy cho trung bình - Cơ sở lý thuyết

Trường hợp kích thước mẫu $n \geq 30$ (hoặc $n < 30$ nhưng X có phân phối chuẩn), σ^2 đã biết

Ta có kết quả sau,

Mệnh đề 3

Trong trường hợp này, thống kê

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

có phân phối chuẩn $N(0,1)$.

Chứng minh.

Dễ dàng có được. □

Khoảng tin cậy cho trung bình - Cơ sở lý thuyết

Trường hợp kích thước mẫu $n \geq 30$ (hoặc $n < 30$ nhưng X có phân phối chuẩn), σ^2 đã biết

Định nghĩa 4

Phân vị mức γ của biến ngẫu nhiên X là giá trị q_γ sao cho

$$P(X \leq q_\gamma) = \gamma$$

Khi $X \sim N(0,1)$, ta thường ký hiệu z_γ thay cho q_γ và tìm z_γ bằng cách tra bảng. Dưới đây là một số giá trị z_γ thường gặp:

γ	0.95	0.975	0.98	0.985	0.99	0.995
z_γ	1.64	1.96	2.05	2.17	2.33	2.58

Khoảng tin cậy cho trung bình - Cơ sở lý thuyết

Trường hợp kích thước mẫu $n \geq 30$ (hoặc $n < 30$ nhưng X có phân phối chuẩn), σ^2 đã biết

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 $N(0,1)$.

- 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.

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 5

Trong trường hợp này, thống kê

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

có phân phối chuẩn $N(0,1)$.

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 $N(0,1)$.

- 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.

Khoảng tin cậy cho trung bình - Cơ sở lý thuyết

Trường hợp kích thước mẫu $n < 30$, σ^2 chưa biết, X tuân theo quy luật phân phối chuẩn.

Mệnh đề 6

Trong trường hợp này, thống kê

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

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

Khoảng tin cậy cho trung bình - Cơ sở lý thuyết

Trường hợp kích thước mẫu $n < 30$, σ^2 chưa biết, X tuân theo quy luật phân phối chuẩn.

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Với độ tin cậy $1-\alpha$, ta có

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

hay

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

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

Khoảng tin cậy cho trung bình - Cơ sở lý thuyết

Trường hợp kích thước mẫu $n < 30$, σ^2 chưa biết, X tuân theo quy luật phân phối chuẩn.

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

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

$$\left[\bar{X} - t^{n-1}_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + t^{n-1}_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

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

$$\left[\bar{x} - t^{n-1}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + t^{n-1}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$

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

Khoảng tin cậy cho trung bình - Các bước thực hiện

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

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 $[\bar{x}-\epsilon, \bar{x}+\epsilon]$.

Khoảng tin cậy cho trung bình - Ví dụ

Ví dụ 7

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.

Khoảng tin cậy cho trung bình - Giải Ví dụ 7a

B1 Trung bình mẫu:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{1}{16}(1\times 0.8 + 1\times 1.0 + \ldots 1\times 2.5) = 1.625$$

Phương sai mẫu:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2$$
$$= \frac{1}{16-1}\left[1\times(0.8-1.625)^2 + \ldots + 1\times(2.5-1.625)^2\right] = 0.243$$

Khoảng tin cậy cho trung bình - Giải Ví dụ 7a (tt)

B2 Ta áp dụng TH1 vì $n < 30$, X có phân phối chuẩn và σ^2 đã biết.

B3 $1-\alpha = 0.96 \Rightarrow \alpha = 0.04 \Rightarrow z_{1-\alpha/2} = z_{0.98} = 2.05$.

B4 Dung sai:

$$\epsilon = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 2.05 \times \frac{0.63}{\sqrt{16}} = 0.323$$

KL KTC 96% cho mức lương trung bình hàng tháng của một công nhân là

$$[\bar{x}-\epsilon, \bar{x}+\epsilon] = [1.302, 1.948]$$

.

Khoảng tin cậy cho trung bình - Giải Ví dụ 7b

Áp dụng TH3 vì $n < 30$, X có phân phối chuẩn và σ^2 chưa biết.
Ta có, $1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow t^{n-1}_{1-\alpha/2} = t^{15}_{0.995} = 2.95$.
Dung sai:

$$\epsilon = t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}} = 2.95 \times \frac{\sqrt{0.243}}{\sqrt{16}} = 0.364$$

KL: KTC 99% cho mức lương trung bình là

$$[\bar{x}-\epsilon, \bar{x}+\epsilon] = [1.261, 1.989]$$

Khoảng tin cậy cho trung bình - Nhận xét

- So sánh KTC trong câu (a) và (b) ta nhận thấy rằng, KTC trong câu (b) rộng hơn KTC trong câu (a). Điều này là do (1) trong câu (a) ta biết thông tin về độ lệch chuẩn của tổng thể còn trong câu (b) thì không, (2) trong câu (b) cần tìm KTC có độ tin cậy cao hơn câu (a).
- Diễn giải kết quả câu (a):
  - Mức lương trung bình là một con số cố định. Nó có thể nằm trong khoảng $[1.302, 1.948]$ hoặc không.
  - **Phát biểu sai:** "Với xác suất 96%, mức lương trung bình thuộc khoảng $[1.302, 1.948]$".
  - **Phát biểu đúng:** "Ta tin chắc 96% rằng mức lương trung bình thuộc khoảng $[1.302, 1.948]$".
  - **Phát biểu đúng:** "Nếu một số lượng lớn các mẫu được thu thập và một khoảng tin cậy được tạo cho mỗi mẫu, thì xấp xỉ 96% các khoảng này sẽ chứa mức lương trung bình".

Outline

Khoảng tin cậy cho tỷ lệ - Phát biểu bài toán

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 $1 - \alpha$.

Khoảng tin cậy cho tỷ lệ - Cở sở lý thuyết

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

Khoảng tin cậy cho tỷ lệ - Cở sở lý thuyết

Mệnh đề 8

Thống kê

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

và

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

Khoảng tin cậy cho tỷ lệ - Cở sở lý thuyết

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Do đó, khi kích thước mẫu đủ lớn,

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

hay

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

Khoảng tin cậy cho tỷ lệ - Cở sở lý thuyết

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

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

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

Khoảng tin cậy cho tỷ lệ - Các bước thực hiện

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

B1 Tìm tỉ lệ mẫu: $\hat{p}$.

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

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 $[\hat{p} - \epsilon, \hat{p} + \epsilon]$.

Khoảng tin cậy cho tỷ lệ - Ví dụ

Ví dụ 9

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.

Khoảng tin cậy cho phương sai và độ lệch chuẩn

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Bài toán

Cho tổng thể X có phân phối chuẩn với phương sai σ^2 chưa biết. Từ mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ hãy tìm khoảng tin cậy cho σ^2 và σ với độ tin cậy $1-\alpha$ cho trước.

Nhắc lại phân phối mẫu của $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ như sau,

Mệnh đề 10

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Khoảng tin cậy cho phương sai và độ lệch chuẩn

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Do đó,

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

trong đó, $\chi^2_\gamma(k)$ là phân vị mức γ của phân phối Chi bình phương với k bậc tự do.
Hay

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

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

Khoảng tin cậy cho phương sai và độ lệch chuẩn

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Vậy

- Với mẫu ngẫu nhiên, khoảng tin cậy $100(1-\alpha)\%$ cho σ^2 là
$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right]$$
- Với mẫu thực nghiệm, khoảng tin cậy $100(1-\alpha)\%$ cho σ^2 là
$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right]$$

Khoảng tin cậy cho phương sai và độ lệch chuẩn

Vậy

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

$$\left[\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}\right]$$

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

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}}\right]$$

Outline

Xác định kích thước mẫu.

- **Vấn đề**
  Trước khi lấy mẫu, với độ tin cậy cho trước, ta mong muốn khoảng tin cậy tìm được phải có chiều dài không vượt quá một giá trị nào đó. Hỏi khi đó ta phải lấy mẫu có kích cỡ tối thiểu là bao nhiêu?

- **Bài toán**
  Tìm giá trị nhỏ nhất của n sao cho $\epsilon \leq \epsilon_0$, với ϵ_0 và α cho trước.

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 $\epsilon \leq \epsilon_0$ ta cần chọn

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

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:

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

Xác định kích thước mẫu
Khi ước lượng tỷ lệ tổng thể

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

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

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

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

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

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

Outline

Xác định độ tin cậy.

- **Vấn đề**
  Bây giờ cố định kích thước mẫu, ta mong muốn khoảng tin cậy tìm được phải có chiều dài xác định. Hỏi khi đó độ tin cậy của khoảng là bao nhiêu?

- **Bài toán**
  Tìm $1-\alpha$ khi biết n và ϵ.

Xác định độ tin cậy
Khi ước lượng trung bình tổng thể

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

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

ta suy ra

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

sau khi xác định được $z_{1-\alpha/2}$ ta suy ra độ tin cậy $1-\alpha$ (tra bảng).

b. Nếu chưa biết $\mathbb{V}ar(X) = \sigma^2$, khi đó ta căn cứ vào mẫu đã cho để tính s. Từ đó xác định $z_{1-\alpha/2}$ theo công thức

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

Rồi suy tiếp ra độ tin cậy $1-\alpha$ như đã làm ở trên.

Xác định độ tin cậy

Khi ước lượng tỷ lệ tổng thể

Từ công thức

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

ta suy ra

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

Từ đây ta suy ngược ra $1-\alpha$ như đã làm ở trên.

KHOẢNG TIN CẬY
Nguyễn Văn Thìn
Giới thiệu
Khoảng tin cậy cho trung bình
Khoảng tin cậy cho tỷ lệ.
Xác định kích thước mẫu
Xác định độ tin cậy

Ví dụ 11

Theo dõi 1000 bệnh nhân ung thư phổi thấy có 823 bệnh nhận chết trong vòng 10 năm.

a. Lập KTC 95% cho tỷ lệ bệnh nhân chết vì ung thư phổi.
b. Nếu muốn sai số bé hơn 0.03 thì phải theo dõi tối thiểu bao nhiêu bệnh nhân trong 10 năm?