Khoảng tin cậy

Nguyễn Thị Hồng Nhung

Ngày 25 tháng 11 năm 2016

Table of contents

Ước lượng điểm

- 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.
- Bài toán: tìm tham số θ
  - Chọn một mẫu ngẫu nhiên cỡ n: $X = (X_1, \dots, X_n)$.
  - Thống kê $\hat{\Theta} = h(X_1, \dots, X_n)$ gọi là một ước lượng điểm cho θ.
  - Với một mẫu thực nghiệm $x_1, \dots, x_n$, ta gọi $\hat{\theta} = h(x_1, \dots, x_n)$ là một giá trị ước lượng điểm cho θ.

Ước lượng điểm

Ví dụ 1

- *X = Chiều cao dân số trong một khu vực, $X \sim \mathcal{N}(\mu, \sigma^2)$. Phân phối của X phụ thuộc vào kỳ vọng μ và phương sai σ^2. Thống kê trung bình mẫu và phương sai mẫu*

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

  là những ước lượng điểm cho μ và σ^2
- *Với một mẫu thực nghiệm $x_1 = 150, x_2 = 155, x_3 = 167$, giá trị ước lượng điểm của μ và σ^2 lần lượt là $\bar{x} = 157.333, s^2 = 76.333$*

Ước lượng không chệch

Định nghĩa 1

Ước lượng điểm $\hat{\Theta}$ gọi là ước lượng không chệch (Unbiased estimator) cho tham số θ nếu

$$\mathbb{E}(\hat{\Theta}) = \theta. \qquad (1)$$

Nếu $\hat{\Theta}$ là ước lượng chệch của θ , độ sai khác

$$\mathbb{E}(\hat{\Theta}) - \theta$$

gọi là độ chệch của ước lượng, ký hiệu là Bias($\hat{\Theta}$) .

Ước lượng không chệch-Ví dụ

i. $\bar{X}$ là một ước lượng không chệch của μ

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sum_{i=1}^{n} \mathbb{E}X_i}{n} = \mu.$$

ii . S^2 là một ước lượng không chệch của σ^2

$$\mathbb{E}(S^2) = \sigma^2$$

iii . $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ là ước lượng chệch của σ^2

Ước lượng hiệu quả

Định nghĩa 2

Xét $\hat{\Theta}_1$ và $\hat{\Theta}_2$ là hai lượng không chệch của θ, $\hat{\Theta}_1$ gọi là ước lượng hiệu quả hơn $\hat{\Theta}_2$ nếu với cỡ mẫu n cho trước

$$\mathbb{V}ar(\hat{\Theta}_1) < \mathbb{V}ar(\hat{\Theta}_2).$$

Định lý 1

Trong một mẫu ngẫu nhiên cỡ n: $X_1, \ldots, X_n$ được chọn từ $X \sim \mathcal{N}(\mu, \sigma^2)$ thì $\bar{X}$ là ước lượng hiệu quả nhất cho μ.

Trung bình của bình phương sai số-MSE

- Trong một số trường hợp, ước lượng $\hat{\Theta}_2$ là ước lượng chệch (với độ chệch nhỏ), nhưng lại có phương sai nhỏ hơn các ước lượng không chệch $\hat{\Theta}_1$ khác. Khi đó, ta có thể muốn chọn $\hat{\Theta}_2$, mặc dù là ước lượng chệch nhưng nó có độ phân tán nhỏ hơn nhiều so với các ước lượng $\hat{\Theta}$ khác.
- Một độ đo kết hợp giữa độ chệch (Bias) và phương sai mẫu của một ước lượng là trung bình của bình phương sai số (Mean Squared Error - MSE)

$$MSE(\hat{\Theta}) = \mathbb{E}(\hat{\Theta} - \theta)^2 \quad (2)$$

- $$MSE(\hat{\Theta}) = \mathbb{V}ar(\hat{\Theta}) + (Bias(\hat{\Theta}))^2. \quad (3)$$

Trung bình của bình phương sai số (tt)

- Nếu $\hat{\Theta}$ là ước lượng không chệch: $MSE(\hat{\Theta}) = \mathbb{V}ar(\hat{\Theta})$
- Cho trước hai ước lượng $\hat{\Theta}_1$ và $\hat{\Theta}_2$, tiêu chuẩn MSE cho phép ta chon $\hat{\Theta}_2$ nếu, với cỡ mẫu n

$$MSE(\hat{\Theta}_2) < MSE(\hat{\Theta}_1)$$

- Hoặc $\mathbb{V}ar(\hat{\Theta}_1) - \mathbb{V}ar(\hat{\Theta}_2) > (Bias(\hat{\Theta}_2))^2 - (Bias(\hat{\Theta}_1))^2$.

Trung bình của bình phương sai số-MSE (tt)

- Nếu cả $\hat{\Theta}_1$ và $\hat{\Theta}_2$ là ước lượng không chệch, tiêu chuẩn MSE trở thành tiêu chuẩn so sánh dựa trên phương sai mẫu.
- Tiêu chuẩn MSE tương đương với việc so sánh tỷ số

$$Eff(\hat{\Theta}_1, \hat{\Theta}_2) = \frac{MSE(\hat{\Theta}_2)}{MSE(\hat{\Theta}_1)} \tag{4}$$

và chọn $\hat{\Theta}_2$ nếu $Eff(\hat{\Theta}_1, \hat{\Theta}_2) < 1$.

Sai số chuẩn - Standard Error

Định nghĩa 3

Sai số chuẩn (SE) của một ước lượng $\hat{\Theta}$ chính là độ lệch tiêu chuẩn của nó, cho bởi

$$SE(\hat{\Theta}) = \sqrt{\mathbb{V}ar(\hat{\Theta})} \tag{5}$$

Ký hiệu khác: $\hat{\sigma}_{\hat{\Theta}}$

Ước lượng điểm, Các tiêu chuẩn ước lượng | Các tiêu chuẩn ước lượng

Ví dụ 2

Tham số	Ước lượng T	$\mathbb{V}ar$	$SE(T)$
μ	$\bar{X}$	$\frac{\sigma^2}{n}$	$\frac{S}{\sqrt{n}}$
p	$\hat{p}$	$\frac{p(1-p)}{n}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
σ^2	S^2	$\frac{2\sigma^4}{n-1}$	$S^2\sqrt{\frac{2}{n-1}}$

Ước lượng bền vững

Định nghĩa 4

Gọi $\hat{\Theta}_n = h(X_1, \ldots, X_n)$ là một ước lượng điểm của tham số θ. Ước lượng $\hat{\Theta}_n$ gọi là bền vững(consistency) nếu $\hat{\Theta}_n \to^{\mathbb{P}} \theta$, tức là

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\Theta}_n - \theta| \le \epsilon\right) = 1, \ \forall \epsilon > 0.$$

Định lý 2

Nếu $\mathbb{E}(\hat{\Theta}_n) \to \theta$ và $D(\hat{\Theta}_n) \to 0$ khi $n \to \infty$ thì $\hat{\theta}_n$ là ước lượng vững.

Ví dụ 3

1. *S^2 là ước lượng vững của σ^2.*
2. *Với $X \sim B(n, p)$, $\hat{p} = \frac{\bar{X}}{n}$ là ước lượng vững cho p.*

Giới thiệu

- Giả sử cần khảo sát 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: $X = (X_1, \ldots, X_n)$.

Định nghĩa 5

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 $L(X) \le U(X)$ và $L(X) \le \theta \le U(X)$. Nếu một mẫu thực nghiệm $x = (x_1, \ldots, x_n)$ được quan trắc, $[l(x), u(x)]$ gọi là một khoảng ước lượng (interval estimator) cho θ.

Khoảng tin cậy

Định nghĩa 6

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

$$\mathbb{P}\left(L(X) \leq \theta \leq U(X)\right) = 1 - \alpha. \tag{6}$$

- Với mẫu thực nghiệm $x = (x_1, \ldots, x_n)$ ta có khoảng tin cậy cụ thể cho tham số θ là $l(x) \leq \theta \leq u(x)$.

Khoảng tin cậy

- Ý nghĩa: với 100% lần lấy mẫu cỡ n thì
  - i có $100(1-\alpha)\%$ lần giá trị tham số $\theta \in [l, u]$;
  - ii có $100\alpha\%$ lần giá trị tham số $\theta \notin [l, u]$.

Khoảng tin cậy cho kỳ vọng

Bài toán 1

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

Trường hợp biết phương sai

Các giả định

- Mẫu ngẫu nhiên được chọn từ tổng thể có phân phối chuẩn, tức là $X_1, \ldots, X_n \sim^{i.i.d} \mathcal{N}(\mu, \sigma^2)$.
- Phương sai σ^2 của tổng thể đã biết.

Khoảng tin cậy cho kỳ vọng
Trường hợp biết phương sai

Xây dựng khoảng tin cậy

- Chọn mẫu ngẫu nhiên cỡ n: $X_1, \ldots, X_n \sim^{i.i.d} \mathcal{N}(\mu, \sigma^2)$
- Thống kê trung bình mẫu $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Phân phối mẫu của $\bar{X}$: $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- Đặt

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{7}$$

thì $Z \sim \mathcal{N}(0, 1)$.

Khoảng tin cậy cho kỳ vọng
Trường hợp biết phương sai

Xây dựng khoảng tin cậy

Với độ tin cậy $100(1-\alpha)\%$, ta có

$$\mathbb{P}\left(\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right\}\right) = 1 - \alpha. \tag{8}$$

hay

$$\mathbb{P}\left(\left\{\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\}\right) = 1 - \alpha \tag{9}$$

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

Khoảng tin cậy cho kỳ vọng
Trường hợp biết phương sai

Định nghĩa 7

Nếu $\bar{x}$ là trung bình mẫu của một mẫu ngẫu nhiên cỡ n được chọn từ một tổng thể có phương sai σ^2 đã biết, khoảng tin cậy $100(1-\alpha)\%$ cho kỳ vọng μ được xác định như sau

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (10)$$

với $z_{1-\frac{\alpha}{2}}$ là phân vị mức $1-\frac{\alpha}{2}$ của $Z \sim \mathcal{N}(0,1)$

Khoảng tin cậy cho kỳ vọng
Trường hợp biết phương sai

Độ chính xác và cỡ mẫu

- $\epsilon = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ gọi là độ chính xác (hay sai số) của ước lượng.
- Chiều dài khoảng tin cậy 2ϵ
- Cho trước sai số và độ tin cậy $100(1-\alpha)\%$, từ đó suy ra công thức tính cỡ mẫu

$$n = \left(\frac{\sigma * z_{1-\frac{\alpha}{2}}}{\epsilon} \right)^2 \quad (11)$$

Khoảng tin cậy cho kỳ vọng
Trường hợp biết phương sai

Ví dụ 4

Đường kính của một ống piston trong động cơ xe máy có phân phối chuẩn với độ lệch chuẩn $\sigma = 0.001$ *mm. Một mẫu ngẫu nhiên gồm 15 ống piston có đường kính trung bình* $\bar{x} = 74.036$ *mm.*

1. *Lập KTC 95% cho đường kính trung bình của piston.*
2. *Lập KTC 99% cho đường kính trung bình.*

Trường hợp biết phương sai

Ví dụ 5

Đo chỉ số IQ của các sinh viên trong 1 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 $\sigma = 10.50$.

i *Vẽ đồ thị stem-leaf cho dữ liệu trên.*
ii *Lập khoảng tin cậy 95% cho chỉ số IQ trung bình.*
iii *Lập khoảng tin cậy 99% cho chỉ số IQ trung bình.*

Trường hợp không biết phương sai, mẫu nhỏ ($n \leq 30$)

Các giả định

- Mẫu ngẫu nhiên được chọn từ tổng thể có phân phối chuẩn
- Phương sai σ^2 của tổng thể không biết; ta có thể dùng phương sai mẫu S^2 để thay thế.
- Cỡ mẫu nhỏ : $n \leq 30$.

Trường hợp không biết phương sai, mẫu nhỏ ($n < 30$)

Xây dựng khoảng tin cậy

- Chọn mẫu ngẫu nhiên cỡ n: $X_1, \ldots, X_n \sim^{i.i.d} \mathcal{N}(\mu, \sigma^2)$
- Thống kê trung bình mẫu và phương sai mẫu

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

- Thay σ bởi S trong công thức (7) thu được biến ngẫu nhiên

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

Trường hợp không biết phương sai, mẫu nhỏ $(n < 30)$

- Phân phối Student -t .

Định nghĩa 8

Xét $X = (X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \sigma^2)$ với μ, σ^2 không biết. Biến ngẫu nhiên $T = \frac{\bar{X}-\mu}{S\sqrt{n}}$ có phân phối Student với $(n-1)$ bậc tự do.

Hàm mật độ của T có dạng

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})\sqrt{k\pi}\left(\frac{t^2}{k}+1\right)^{\frac{k+1}{2}}}, -\infty < t < \infty.$$

Trường hợp không biết phương sai, mẫu nhỏ $(n < 30)$

- Gọi t_α^n là phân vị mức α của biến ngẫu nhiên T có phân phối Student với n bậc tự do.
- t_α^n được xác định như sau

$$\mathbb{P}(T < t_\alpha^n) = \alpha \qquad (12)$$

- Tìm t_α^n : tra bảng Student.

Trường hợp không biết phương sai, mẫu nhỏ $(n < 30)$

Xây dựng khoảng tin cậy

Với độ tin cậy $100(1-\alpha)\%$ và $T = \frac{\bar{X}-\mu}{S\sqrt{n}}$ ta có

$$\mathbb{P}\left(\left\{-t^{n-1}_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t^{n-1}_{1-\frac{\alpha}{2}}\right\}\right) = 1-\alpha \qquad (13)$$

hay

$$\mathbb{P}\left(\left\{\bar{X} - t^{n-1}_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t^{n-1}_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right\}\right) = 1-\alpha \qquad (14)$$

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

Trường hợp không biết phương sai, mẫu nhỏ $(n \leq 30)$

Định nghĩa 9

Nếu $\bar{x}$ và s lần lượt là trung bình mẫu và độ lệch tiêu mẫu của một mẫu ngẫu nhiên cỡ n được chọn từ một tổng thể có phân phối chuẩn với kỳ vọng μ và phương sai σ^2 không biết, khoảng tin cậy $100(1-\alpha)\%$ cho kỳ vọng μ được xác định như sau

$$\bar{x} - t^{n-1}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t^{n-1}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

với $t^{n-1}_{1-\frac{\alpha}{2}}$ là phân vị mức $1-\frac{\alpha}{2}$ của $T \sim t(n-1)$.

Trường hợp không biết phương sai, mẫu lớn $(n > 30)$

Các giả định

- Mẫu ngẫu nhiên được chọn từ một tổng thể với kỳ vọng μ và phương sai σ^2 không biết, sử dụng phương sai mẫu S^2 thay thế cho σ^2.
- Cỡ mẫu: $n > 30$

Trường hợp không biết phương sai, mẫu lớn $(n > 30)$

Khi cỡ mẫu lớn, đại lượng ngẫu nhiên

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

sẽ xấp xỉ với phân phối chuẩn hóa $\mathcal{N}(0, 1)$ theo định lý giới hạn trung tâm. Do đó, khoảng tin cậy cho kỳ vọng μ với độ tin cậy $100(1 - \alpha)\%$ cho bởi

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \tag{15}$$

với $z_{1-\frac{\alpha}{2}}$ là phân vị mức $1 - \frac{\alpha}{2}$ của phân phối chuẩn hóa.

Ví dụ 6

Cadmium (Cd), một kim loại nặng, là chất độc đối với các loài động vật. Tuy nhiên, các loài nấm lại có khả năng hấp thụ Cd với hàm lượng cao. Chính phủ một số nước ra quy định giới hạn hàm lượng Cd tối đa trong rau quả khô là 0.5(*ppm*)*. Trong một nghiên cứu về hàm lượng Cd trong loài nấm Boletus pinocola do M. Melgar và các cộng sự thực hiện đăng trên tạp chí Journal of Enviroment Science and Health, cho số liệu về hàm lượng Cd trong một mẫu 12 cây nấm như sau:*

$$0.24\ 0.59\ 0.62\ 0.16\ 0.77\ 1.33$$
$$0.92\ 0.19\ 0.33\ 0.25\ 0.59\ 0.32$$

Lập KTC 99% *cho hàm lượng Cd trung bình trong nấm Boletus pinocola. Nếu muốn sai số ước lượng* $\epsilon = 0.1$ *thì khảo sát tối thiểu bao nhiêu cây nấm?*

Khoảng tin cậy cho kỳ vọng | Trường hợp không biết phương sai

Ví dụ 7

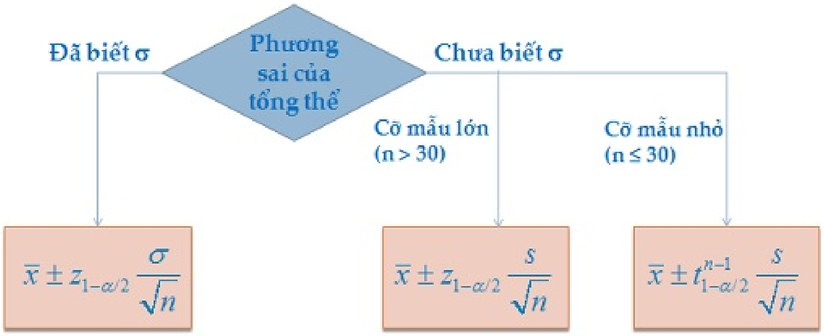
Biết lương tháng (Đv: triệu đồng) của thanh niên trong độ tuổi $25 - 35$ *ở một khu vực có phân phối chuẩn. Khảo sát 50 thanh niên.*

Lương tháng	1.8	2.5	3.2	3.9	4.6	5.3	6.0	6.7	7.4
Số thanh niên	3	3	8	9	11	7	5	2	2

a. *Lập khoảng tin cậy* 95% *cho lương tháng của thanh niên trong khu vực này.*
b. *Nếu muốn sai số ước lượng* $\epsilon = 0.10$ *mà vẫn giữ cỡ mẫu* $n = 50$ *thì độ tin cậy là bao nhiêu ?*

Nguyễn Thị Hồng Nhung | Khoảng tin cậy | Ngày 25 tháng 11 năm 2016 | 31 / 43

Khoảng tin cậy cho kỳ vọng-Tóm tắt



Ước lượng tỷ lệ của tổng thể

Bài toán 2

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 $(X_1, \ldots, X_n)$ hãy tìm khoảng tin cậy cho p với độ tin cậy $(1-\alpha)$.

Ước lượng tỷ lệ của tổng thể

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

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

- Biến ngẫu nhiên $\hat{P}$ có kỳ vọng và phương sai lần lượt là

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

Khoảng tin cậy cho tỷ lệ

Ước lượng tỷ lệ của tổng thể

Mệnh đề 1

Thống kê

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

và

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

Ước lượng tỷ lệ của tổng thể

Do đó, với độ tin cậy $100(1-\alpha)\%$

$$\mathbb{P}\left(\left\{-z_{1-\frac{\alpha}{2}} \le \frac{\hat{P}-p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \le z_{1-\frac{\alpha}{2}}\right\}\right) = 1-\alpha \qquad (16)$$

hay

$$\mathbb{P}\left\{\hat{P} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \le p \le \hat{P} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\}\right\} = 1-\alpha \qquad (17)$$

Ước lượng tỷ lệ của tổng thể

Vậy

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

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

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

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

Khoảng tin cậy cho tỷ lệ

Nhận xét 1

- *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 càng tốt.*
- *Tuy nhiên, dung sai ϵ lại phụ thuộc vào kích thước mẫu n và độ tin cậy $(1-\alpha)$.*

Câu hỏi 1

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 ?

Khi ước lượng trung bình tổng thể

Bài toán 3

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

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

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

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

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

Khi ước lượng trung bình tổng thể

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

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

Khi ước lượng tỷ lệ tổng thể

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

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

b Khi chưa biết $\hat{p}$, ta có $\epsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
Do $\hat{p}(1-\hat{p})$ đạt giá trị cực đại 0.25 khi $\hat{p} = 0.5$ nên $\epsilon \leq z_{1-\frac{\alpha}{2}} \sqrt{\frac{0.25}{n}}$. Do đó, để $\epsilon \leq \epsilon_0$ ta chọn n sao cho $z_{1-\frac{\alpha}{2}} \sqrt{\frac{0.25}{n}} \leq \epsilon_0$, tức là

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

Xác định độ tin cậy

Câu hỏi 2

Khi ước lượng các số đặc trưng của tổng thể bằng các số liệu quan sát của một mẫu kích thước n, nếu ta muốn dung sai ϵ đủ nhỏ thì độ tin cậy $(1-\alpha)$ sẽ là bao nhiêu ?

Bài toán 4

Tìm $(1-\alpha)$ khi biết n và ϵ.

Khi ước lượng trung bình tổng thể

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

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

ta suy ra $z_{1-\frac{\alpha}{2}} = \frac{\epsilon\sqrt{n}}{\sigma}$. sau khi xác định được $z_{1-\frac{\alpha}{2}}$ ta suy ra độ tin cậy $(1-\alpha)$.

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 $z_{1-\frac{\alpha}{2}}$ theo công thức $z_{1-\frac{\alpha}{2}} = \frac{\epsilon\sqrt{n}}{s}$ Sau đó suy ra độ tin cậy $(1-\alpha)$ như ở trên.

Khi ước lượng tỷ lệ tổng thể

Từ công thức

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

ta suy ra

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

Từ đó ta suy ra $(1-\alpha)$ như ở trên.

Ví dụ 8

Theo dõi 1000 bệnh nhân ung thư phổi thấy có 823 bệnh nhân 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 ?*