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

4.3. KHOẢNG TIN CẬY

Khái niệm

Xét một biến ngẫu nhiên X với θ là một tham số chưa biết của X (θ có thể là kỳ vọng, phương sai hoặc tỷ lệ...). Từ biến ngẫu nhiên này ta lập mẫu ngẫu nhiên kích thước n, $W_X = (X_1, X_2, \ldots, X_n)$. Khi đó, nếu định nghĩa được hai thống kê

$$\widehat{\Theta}_i = g_i(X_1, X_2, \ldots, X_n), \quad i = 1, 2,$$

sao cho, với một số γ lớn gần 1 cho trước, γ có thể là 0,95; 0,99; 0,995 ..., ta thu được hệ thức

$$P(\widehat{\Theta}_1 \leq \theta \leq \widehat{\Theta}_2) = \gamma,$$

thì, ta đã xây dựng được một ước lượng khoảng cho tham số θ và

- Khoảng ngẫu nhiên $(\widehat{\Theta}_1; \widehat{\Theta}_2)$ được gọi là khoảng tin cậy cho tham số θ.
- γ được gọi là độ tin cậy của ước lượng khoảng.
- $\widehat{\Theta}_2 - \widehat{\Theta}_1$ được gọi là độ dài của khoảng tin cậy.

Thực hiện một phép thử đối với mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ ta thu được mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, từ đó, tính được các giá trị của $\widehat{\Theta}_1$ và $\widehat{\Theta}_2$, ký hiệu là $\widehat{\theta}_1$ và $\widehat{\theta}_2$. Như vậy, có thể nhận định tham số θ nằm trong khoảng $(\widehat{\theta}_1, \widehat{\theta}_2)$.

Khái niệm

✍

- Các cận $\widehat{\Theta}_1$ và $\widehat{\Theta}_2$ phụ thuộc vào mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ nên chúng là các biến ngẫu nhiên.
- Đôi khi ta chỉ quan tâm đến giá trị lớn nhất hay giá trị nhỏ nhất của tham số. Ví dụ, xét hai kết luận sau:

  "Với xác suất 95% thì chiều cao trung bình của người Việt Nam nhỏ hơn 168 centimét".

  "Với xác suất 95%, chiều cao trung bình của sinh viên Đại học Bách khoa Hà Nội lớn hơn 165 centimét".

  Ở kết luận thứ nhất, với độ tin cậy 95% thì chiều cao trung bình của người Việt Nam nằm trong khoảng $(0; 168)$ và ta chỉ quan tâm đến giá trị lớn nhất. Ở kết luận thứ hai, với độ tin cậy 95% thì chiều cao trung bình của sinh viên Đại học Bách khoa Hà Nội thuộc khoảng $(165; +\infty)$ và ta chỉ quan tâm đến giá trị nhỏ nhất. Các khoảng tin cậy dạng này được gọi là khoảng tin cậy một phía.
- Khoảng tin cậy mà chỉ quan tâm đến cận trên được gọi là khoảng tin cậy trái.

  Khoảng tin cậy mà chỉ quan tâm tới cận dưới được gọi là khoảng tin cậy phải.

  Khoảng tin cậy mà ta quan tâm tới cả hai cận (trên và dưới) được gọi là khoảng tin cậy hai phía.

  Trong trường hợp các cận trên, cận dưới đối xứng qua ước lượng điểm của tham số thì ta có khoảng tin cậy đối xứng.

Bài toán

Bài toán 1

Giả sử biến ngẫu nhiên $X \sim \mathcal{N}(\mu;\sigma^2)$ với kỳ vọng $E(X)=\mu$ chưa biết cần ước lượng. Từ biến ngẫu nhiên X, lập mẫu ngẫu nhiên $W_X=(X_1,X_2,\ldots,X_n)$ kích thước n và xét bài toán tìm khoảng tin cậy cho μ dựa trên các quan sát $W_x=(x_1,x_2,\ldots,x_n)$.

✍ Ta xét các trường hợp sau đây

- Phương sai $V(X)=\sigma^2$ đã biết.
- Mẫu kích thước lớn.
- Phương sai $V(X)=\sigma^2$ chưa biết.

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai đã biết

- Vì $X \sim \mathcal{N}(\mu; \sigma^2)$, nên

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1). \tag{38}$$

- Ký hiệu $z_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ sao cho diện tích miền giới hạn bởi đường cong chuẩn, trục hoành và bên phải $z_{\alpha/2}$ là $\alpha/2$. Khi đó,

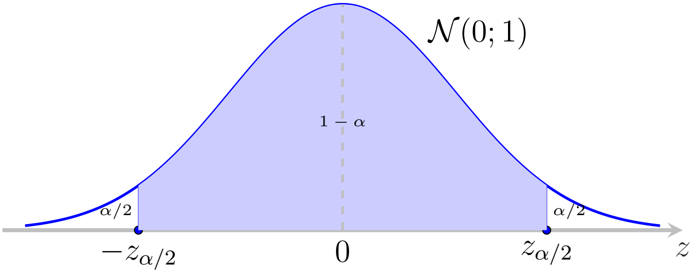
$$P\Big(-z_{\alpha/2} \le Z \le z_{\alpha/2}\Big) = 1 - \alpha \quad \text{hay} \quad P\Big(-z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\Big) = 1 - \alpha$$

(xem Hình 7).

- Suy ra,

$$P\Big(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha. \tag{39}$$

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai đã biết



Hình 7: $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai đã biết

Định lý 8

Nếu $\overline{X}$ là kỳ vọng mẫu của mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n lấy từ biến ngẫu nhiên $X \sim \mathcal{N}(\mu; \sigma^2)$ với σ^2 đã biết, thì khoảng tin cậy đối xứng cho μ với độ tin cậy $1 - \alpha$ được cho bởi

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \tag{40}$$

ở đây, $z_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân phối chuẩn được xác định từ hệ thức $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

Khi có mẫu cụ thể $W_X = (x_1, x_2, \ldots, x_n)$, ta tính được giá trị cụ thể $\overline{x}$ của $\overline{X}$ và nhận được khoảng tin cậy đối xứng cho μ là

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{41}$$

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai đã biết

Ví dụ 25

Trọng lượng (gam) của một loại sản phẩm là biến ngẫu nhiên X có phân phối chuẩn với độ lệch tiêu chuẩn là 6 gam. Cân thử 36 sản phẩm loại này ta thu được kết quả sau:

177	165	152	174	159	166	160	152	162	175	158	169
166	162	181	168	170	150	173	164	167	177	167	175
165	182	166	158	166	170	168	165	160	160	169	166

(a) Với độ tin cậy $1 - \alpha = 95\%$, hãy tìm khoảng tin cậy đối xứng cho trọng lượng trung bình của loại sản phẩm nói trên.

(b) Nếu yêu cầu độ tin cậy là 99% thì kết quả thế nào?

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai đã biết

Giải.

(a) Đây là bài toán ước lượng bằng khoảng tin cậy đối xứng cho kỳ vọng $E(X) = \mu$ của biến ngẫu nhiên gốc $X \sim \mathcal{N}(\mu; \sigma^2)$ với $\sigma = 6$.

- Với độ tin cậy $1 - \alpha = 0,95$, $\Phi(z_{\alpha/2}) = \Phi(z_{0,025}) = 0,975$. Suy ra $z_{0,025} = 1,96$.
- Từ số liệu đã cho, kích thước mẫu $n = 36$, trung bình mẫu $\overline{x} = 166,22$ (gam), áp dụng (41), suy ra khoảng tin cậy đối xứng cho trọng lượng trung bình của sản phẩm với độ tin cậy 95% là

$$166,22 - 1,96 \times \frac{6}{\sqrt{36}} \leq \mu \leq 166,22 + 1,96 \times \frac{6}{\sqrt{36}}$$

  hay

$$164,26 \ \leq \mu \leq \ 168,18.$$

- Như vậy, dựa trên dữ liệu mẫu, trọng lượng trung bình của loại sản phẩm nói trên từ 164,26 gam đến 168,18 gam.

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai đã biết

(b) Nếu yêu cầu độ tin cậy $1-\alpha = 99\%$, $z_{\alpha/2} = z_{0,005} = 2,58$.

Khoảng tin cậy cho trọng lượng trung bình của sản phẩm với độ tin cậy 99% là

$$166,22 - 2,58 \times \frac{6}{\sqrt{36}} \leq \mu \leq 166,22 + 2,58 \times \frac{6}{\sqrt{36}}$$

hay

$$163,64 \leq \mu \leq 168,80.$$

✍

- Trên cùng một mẫu dữ liệu, nếu độ tin cậy càng lớn thì độ dài của khoảng tin cậy càng lớn.
- Không thể viết $P(164,26 \leq \mu \leq 168,18) = 0,95$ vì độ tin cậy gắn với khoảng tin cậy ngẫu nhiên chứ không gắn với mẫu cụ thể. Hơn nữa vì μ là một hằng số nên nó chỉ có thể thuộc hoặc không thuộc khoảng (164,26; 168,18) nên $(164,26 \leq \mu \leq 168,18)$ không phải là sự kiện ngẫu nhiên.

Sai số của ước lượng. Kích thước mẫu

- Giả sử θ là một tham số của tổng thể và $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$ là một hàm ước lượng điểm cho θ dựa trên mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n.
- Cho trước số $\varepsilon > 0$ và $1 - \alpha \in (0; 1)$, ta nói rằng $\widehat{\Theta}$ có sai số ε với độ tin cậy $1 - \alpha$ nếu

$$P\left(\left|\theta - \widehat{\Theta}\right| \leq \varepsilon\right) \geq 1 - \alpha. \tag{42}$$

- Nếu kích thước mẫu n càng lớn thì sai số càng nhỏ. Tuy nhiên kích thước mẫu càng lớn thì nhà nghiên cứu càng tốn nhiều thời gian, tiền bạc và công sức cho việc thu thập dữ liệu.
- Bài toán đặt ra là cần chọn kích thước mẫu tối thiểu là bao nhiêu để đủ đạt được sai số mong muốn.

Sai số của ước lượng. Kích thước mẫu

- Giả sử với độ tin cậy $1-\alpha$ cho trước, ta muốn có ước lượng cho giá trị trung bình μ với sai số không quá ε. Nếu $X \sim \mathcal{N}(\mu;\sigma^2)$ với σ^2 đã biết thì

$$P\left(|\overline{X}-\mu| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha. \tag{43}$$

- Do đó nếu n thỏa mãn $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \varepsilon$ hay tương đương với

$$n \geq \frac{\sigma^2 (z_{\alpha/2})^2}{\varepsilon^2} \tag{44}$$

  thì

$$P\left(|\overline{X}-\mu| \leq \varepsilon\right) \geq P\left(|\overline{X}-\mu| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

- Ta sẽ chọn kích thước mẫu n là số nguyên dương bé nhất thỏa mãn điều kiện (44) thì sẽ đạt được sai số mong muốn với độ tin cậy $1-\alpha$.

Sai số của ước lượng. Kích thước mẫu

✍

- Độ dài khoảng tin cậy đối xứng trong (41) là $2z_{\alpha/2}\sigma/\sqrt{n}$. Như vậy, nếu sử dụng $\overline{x}$ để ước lượng μ, thì sai số $|\overline{x}-\mu|$ nhỏ hơn hoặc bằng $\varepsilon = z_{\alpha/2}\sigma/\sqrt{n}$ với độ tin cậy $(1-\alpha)$.
- Với phương sai σ^2 đã biết không đổi và độ tin cậy $1-\alpha$ không đổi thì giá trị $z_{\alpha/2}$ không đổi, do đó sai số của ước lượng chỉ phụ thuộc vào kích thước mẫu n.

Ví dụ 26

Trong Ví dụ 25(a), sai số của ước lượng là

$$\varepsilon = 1,96 \times \frac{6}{\sqrt{36}} = 1,96.$$

Sai số của ước lượng. Kích thước mẫu

Ví dụ 27

Cần một kích thước mẫu là bao nhiêu nếu muốn ước lượng khoảng tin cậy cho μ trong Ví dụ 25 với độ tin cậy 95% và sai số của ước lượng là 1?

Giải.

- Ta có $\varepsilon = 1$, $\sigma = 6$, $z_{\alpha/2} = 1,96$. Từ (44),

$$n \geq \frac{6^2(1,96)^2}{1^2} = 138,2976.$$

- Nghĩa là ta phải cân ít nhất 139 sản phẩm. Tuy nhiên, vì ta đã cân 36 sản phẩm, nên thực tế cần cân bổ sung thêm 133 sản phẩm nữa.

Khoảng tin cậy một phía cho kỳ vọng, phương sai đã biết

- Ta có

$$P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_\alpha\right) = 1-\alpha \quad \text{hay} \quad P\left(\mu > \overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

- Tương tự,

$$P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > -z_\alpha\right) = 1-\alpha \quad \text{hay} \quad P\left(\mu < \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

- Từ đó ta có công thức xác định khoảng tin cậy một phía cho kỳ vọng $E(X) = \mu$ như sau.

Khoảng tin cậy một phía cho kỳ vọng, phương sai đã biết

Định lý 9

Giả sử $\overline{X}$ là trung bình mẫu của mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n lấy từ tổng thể phân phối chuẩn với phương sai là σ^2. Khi đó,

- Khoảng tin cậy trái cho μ với độ tin cậy $(1 - \alpha)$ là

$$-\infty < \mu \leq \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}. \tag{45}$$

- Khoảng tin cậy phải cho μ với độ tin cậy $(1 - \alpha)$ là

$$\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu < +\infty, \tag{46}$$

ở đây, z_α là giá trị tới hạn mức α của phân phối chuẩn được xác định từ hệ thức $\Phi(z_\alpha) = 1 - \alpha$.

Khoảng tin cậy một phía cho kỳ vọng, phương sai đã biết

Định lý 9 (tiếp theo)

Khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, ta tính được giá trị cụ thể $\overline{x}$ của $\overline{X}$ và nhận được khoảng tin cậy một phía cho μ ứng với mẫu này.

- Khoảng tin cậy trái cho μ là

$$-\infty < \mu \leq \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}. \tag{47}$$

- Khoảng tin cậy phải cho μ là

$$\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu < +\infty. \tag{48}$$

Khoảng tin cậy một phía cho kỳ vọng, phương sai đã biết

Ví dụ 28

Với số liệu trong Ví dụ 25, với độ tin cậy 99% hãy tính cận trên của khoảng tin cậy phải dựa vào số liệu đã cho.

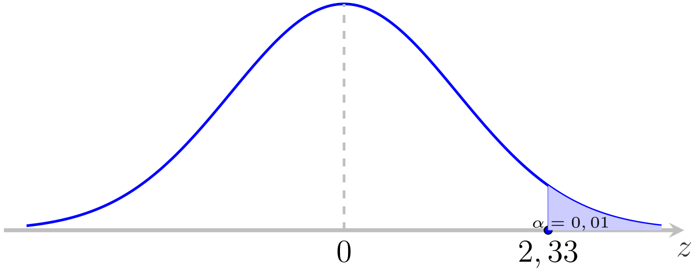
Giải.

- Đây là bài toán ước lượng khoảng cho kỳ vọng của tổng thể phân phối chuẩn, phương sai đã biết bằng khoảng tin cậy một phía (khoảng tin cậy phải).
- Với độ tin cậy $1-\alpha = 99\%$ suy ra $\Phi(z_\alpha) = \Phi(z_{0,01}) = 0,99$. Suy ra $z_{0,01} = 2,33$ (xem Hình 8).
- Giá trị bé nhất cho trọng lượng trung bình là

$$\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = 166,22 - 2,33\frac{6}{\sqrt{36}} = 163,89.$$

- Như vậy, với độ tin cậy 99%, ta tính được cận trên của khoảng tin cậy phải là 163,89 gam.

Khoảng tin cậy một phía cho kỳ vọng, phương sai đã biết



Hình 8: Giá trị z_α với độ tin cậy 99%

Trường hợp mẫu kích thước lớn

Ta sẽ xem xét khoảng tin cậy cho $E(X) = \mu$ trong trường hợp không cần giả thiết phân phối chuẩn của biến ngẫu nhiên gốc X.

- Giả sử mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n được thành lập từ biến ngẫu nhiên gốc X của tổng thể có kỳ vọng μ và phương sai σ^2 chưa biết.
- Nếu kích thước mẫu n lớn, thì theo Định lý giới hạn trung tâm,

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \text{ xấp xỉ phân phối chuẩn tắc } \mathcal{N}(0; 1).$$

- Do đó ta có thể sử dụng kết quả phần trên để xây dựng khoảng tin cậy cho μ. Tuy nhiên, vì σ chưa biết, khi n đủ lớn ta có thể thay σ bởi độ lệch tiêu chuẩn hiệu chỉnh S mà ít làm ảnh hưởng đến phân phối của Z.

Khoảng tin cậy đối xứng cho μ với mẫu kích thước n lớn

Định lý 10

Khi n đủ lớn thì

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \text{ xấp xỉ phân phối chuẩn tắc } \mathcal{N}(0;1). \tag{49}$$

Khi đó,

$$\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}} \tag{50}$$

là khoảng tin cậy mẫu lớn cho μ với độ tin cậy $1-\alpha$.
Khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ ta tính được các giá trị cụ thể $\overline{x}$ của $\overline{X}$ và s của S và xác định được khoảng tin cậy đối xứng cho μ là

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}. \tag{51}$$

Khoảng tin cậy đối xứng cho μ với mẫu kích thước n lớn

✍ Nói chung kích thước mẫu n nên ít nhất là 40 để sử dụng kết quả này một cách đáng tin cậy. Định lý giới hạn trung tâm thường áp dụng cho $n \geq 30$, nhưng kích thước mẫu lớn hơn được khuyến nghị ở đây vì việc thay thế σ bằng S dẫn đến sự thay đổi trong kết quả của Z ở (49).

Khoảng tin cậy đối xứng cho μ với mẫu kích thước n lớn

Ví dụ 29

Để ước lượng trọng lượng trung bình của một loại sản phẩm do một máy sản xuất, người ta cân thử ngẫu nhiên 100 sản phẩm loại này và thu được kết quả sau:

Trọng lượng (gam)	40 – 42	42 – 44	44 – 46	46 – 48	48 – 50	50 – 52
Số sản phẩm	7	13	25	35	15	5

Hãy ước lượng trọng lượng trung bình của loại sản phẩm nói trên bằng khoảng tin cậy đối xứng với độ tin cậy 95%.

Khoảng tin cậy đối xứng cho μ với mẫu kích thước n lớn

Giải. Gọi X là trọng lượng của loại sản phẩm trên, trọng lượng trung bình của sản phẩm là $E(X) = \mu$ chưa biết, cần ước lượng. Đây là bài toán ước lượng khoảng của kỳ vọng của biến ngẫu nhiên chưa biết phân phối xác suất và phương sai chưa biết, kích thước mẫu $n = 100$ đủ lớn. Sử dụng Định lý 49 như sau:

- Chọn thống kê $Z = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$. Vì $n = 100$ đủ lớn, nên Z xấp xỉ $\mathcal{N}(0;1)$.
- Áp dụng khoảng tin cậy đối xứng (49) cho μ, với $z_{\alpha/2} = z_{0,025} = 1,96$, $\overline{x} = 46,06$ và $s = 2,48$ ta nhận được
$$46,06 - 1,96 \times \frac{2,48}{\sqrt{100}} \le \mu \le 46,06 + 1,96 \times \frac{2,48}{\sqrt{100}}$$
hay
$$45,573 \le \mu \le 46,546.$$
- Với độ tin cậy 95%, trọng lượng trung bình của loại sản phẩm trên là từ 45,573 gam đến 46,546 gam.

Khoảng tin cậy một phía cho μ với mẫu kích thước n lớn

✍ Với các lập luận như đã trình bày ở trên, khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, ta tính được $\overline{x}$ và s và nhận được các khoảng tin cậy trái và phải cho μ lần lượt là:

$$-\infty < \mu \leq \overline{x} + z_\alpha \frac{s}{\sqrt{n}} \tag{52}$$

và

$$\overline{x} - z_\alpha \frac{s}{\sqrt{n}} \leq \mu < +\infty. \tag{53}$$

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai chưa biết

- Từ X lập một mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n. Như đã biết,

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

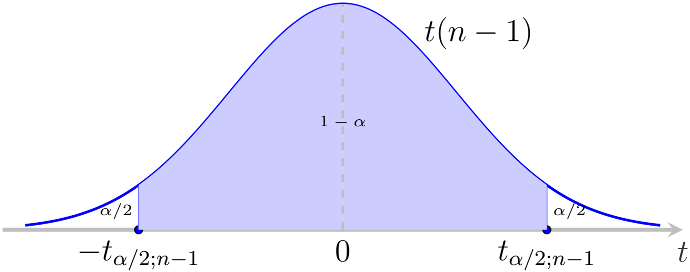
- Tương tự cách làm với trường hợp σ đã biết, thay σ bởi S và phân phối chuẩn bởi phân phối Student, ta có (xem Hình 9),

$$P(-t_{\alpha/2;n-1} \leq T \leq t_{\alpha/2;n-1}) = 1 - \alpha,$$

  hay

$$P\left(-t_{\alpha/2;n-1} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2;n-1}\right) = 1 - \alpha.$$

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai chưa biết



Hình 9: $P(-t_{\alpha/2;n-1} \leq T \leq t_{\alpha/2;n-1}) = 1-\alpha$

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai chưa biết

- Từ đó,

$$P\left(\overline{X} - t_{\alpha/2;n-1}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha. \quad (54)$$

- Khi có mẫu cụ thể kích thước n, ta tính được $\overline{x}$ và s và từ đó tìm được khoảng tin cậy cho μ với độ tin cậy $1-\alpha$.

Khoảng tin cậy đối xứng cho kỳ vọng, phương sai chưa biết

Định lý 11

Nếu $\overline{X}$ và S là trung bình mẫu và độ lệch chuẩn hiệu chỉnh mẫu của một mẫu ngẫu nhiên kích thước n, $W_X = (X_1, X_2, \ldots, X_n)$, được chọn từ biến ngẫu nhiên X có phân phối chuẩn với phương sai σ^2 chưa biết, thì khoảng tin cậy đối xứng cho μ với độ tin cậy $1-\alpha$ là

$$\overline{X} - t_{\alpha/2;n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}, \tag{55}$$

ở đây, $t_{\alpha/2;n-1}$ là giá trị tới hạn mức $\alpha/2$ của phân phối Student với $n-1$ bậc tự do.
Khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ kích thước n, ta tính được $\overline{x}$ của $\overline{X}$ và s của S và xác định được khoảng tin cậy đối xứng cho μ là

$$\overline{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}. \tag{56}$$

Chọn kích thước mẫu

- Tương tự như mục trước, ta có

$$P\left(|\overline{X}-\mu| \le t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

- Do đó, nếu n thỏa mãn

$$t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} \le \varepsilon$$

hay tương đương với

$$n \ge \frac{s^2(t_{\alpha/2;n-1})^2}{\varepsilon^2} \tag{57}$$

thì ta sẽ có

$$P\left(|\overline{X}-\mu| \le \varepsilon\right) \ge 1-\alpha.$$

Chọn kích thước mẫu

- Ta không thể tìm n thỏa mãn (57) vì cả s và $t_{\alpha/2;n-1}$ đều phụ thuộc vào n.
- Tuy nhiên, khi số bậc tự do lớn thì $t_{\alpha/2;n-1}$ và $z_{\alpha/2}$ gần như nhau, vì vậy ta có thể chọn n thỏa mãn

$$n \geq \frac{s^2 (z_{\alpha/2})^2}{\varepsilon^2}. \tag{58}$$

- Như vậy, nếu tìm được ước lượng cho s thì ta sẽ tìm được n từ bất đẳng thức này.
- Người ta thường lấy một mẫu sơ bộ kích thước n đủ lớn ($n \geq 30$) để tính s rồi sau đó tìm n từ bất đẳng thức (58).

Chọn kích thước mẫu

Ví dụ 30

Ta muốn xây dựng một ước lượng với độ tin cậy 95% và sai số 2 dặm cho vận tốc trung bình của ô tô trên đường cao tốc. Một mẫu điều tra sơ bộ với cỡ mẫu $n = 50$ cho ta $s = 9$ dặm. Hỏi cần lấy mẫu cỡ tối thiểu là bao nhiêu để đạt được sai số và độ tin cậy đã đặt ra.

Giải.

- Ta có $\varepsilon = 2$, $s = 9$, $1 - \alpha = 95\%$ nên $z_{\alpha/2} = 1,96$. Từ bất đẳng thức (58) ta có

$$n \geq \frac{s^2(z_{\alpha/2})^2}{\varepsilon^2} = \frac{9^2(1,96)^2}{2^2} = 77,79.$$

- Nghĩa là ta phải lấy cỡ mẫu ít nhất là 78. Tuy nhiên vì ta đã có mẫu sơ bộ với cỡ 50, nên thực ra ta chỉ cần bổ sung thêm 28 quan sát nữa.

Chọn kích thước mẫu

✍

- Việc lấy mẫu sơ bộ để tìm s là hợp lý vì ta biết rằng s hội tụ đến σ khi $n \to \infty$, do đó khi n lớn thì các giá trị của s khá "ổn định" và "gần" σ.
- Trong ví dụ trên, phương sai của tổng thể dữ liệu được ước lượng thông qua độ phân tán của mẫu sơ bộ. Ta cũng có thể ước lượng phương sai của tổng thể dữ liệu theo những cách khác.

Chọn kích thước mẫu

Ví dụ 31

Ta muốn ước lượng số năm đi học của những người trưởng thành trong một vùng với độ tin cậy 95% và sai số không quá một năm thì cần lấy mẫu tối thiểu là bao nhiêu?

Giải.

- Ta có $\varepsilon = 1$, $1 - \alpha = 95\%$ nên $z_{\alpha/2} = 1,96$.
- Để ước lượng phương sai của tổng thể dữ liệu, ta nhận thấy rằng số năm học dao động từ 0 đến 18. Nếu phân phối của số năm đi học là phân phối chuẩn thì hầu hết các giá trị quan sát sẽ thuộc khoảng $(\mu - 3\sigma; \mu + 3\sigma)$. Khoảng này có độ dài 6σ do đó ta dùng ước lượng $6\sigma = 18$ hay $\sigma = 3$. Từ đó ta tìm được

$$n \geq \frac{3^2(1,96)^2}{1^2} = 34,5744$$

- Ta cũng có thể đoán rằng số năm học không thể quá 24 và do đó độ lệch tiêu chuẩn σ không thể quá 4 để có ước lượng thô về σ từ đó tìm n.

Khoảng tin cậy một phía cho kỳ vọng, phương sai chưa biết

Bằng cách xây dựng tương tự như trên, khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ kích thước n, ta tính được giá trị cụ thể $\overline{x}$ của $\overline{X}$ và s của S và xây dựng được các khoảng tin cậy một phía cho μ.

- Khoảng tin cậy trái cho μ là

$$-\infty < \mu \leq \overline{x} + t_{\alpha;n-1}\frac{s}{\sqrt{n}}; \tag{59}$$

- Khoảng tin cậy phải cho μ là

$$\overline{x} - t_{\alpha;n-1}\frac{s}{\sqrt{n}} \leq \mu < +\infty, \tag{60}$$

trong đó, $t_{\alpha;n-1}$ là giá trị tới hạn mức α của phân phối Student với $n-1$ bậc tự do.

Khoảng tin cậy một phía cho kỳ vọng, phương sai chưa biết

Ví dụ 32

Để kiểm tra sự chính xác của hệ thống đóng gói tự động các bao gạo khi xuất khẩu tại một nhà máy, người ta đã chọn ngẫu nhiên 16 bao và tính được trọng lượng trung bình là 49,75 kg và độ lệch tiêu chuẩn hiệu chỉnh mẫu là 0,5 kg. Giả thiết rằng trọng lượng của những bao gạo có phân phối chuẩn. Với độ tin cậy 95% có thể khẳng định trọng lượng trung bình của bao gạo cao nhất là bao nhiêu? Dựa trên kết quả thu được có thể khẳng định về trung bình các bao gạo đã bị đóng thiếu hay không nếu biết rằng trọng lượng chuẩn mỗi bao là 50 kg.

Khoảng tin cậy một phía cho kỳ vọng, phương sai chưa biết

Giải. Gọi X là trọng lượng các bao gạo được đóng gói tự động, $X \sim \mathcal{N}(\mu; \sigma^2)$ với σ^2 chưa biết. Đây là bài toán ước lượng bằng khoảng tin cậy một phía cho kỳ vọng $E(X) = \mu$ của của biến ngẫu nhiên có phân phối chuẩn, phương sai chưa biết, mẫu kích thước $n = 16 < 30$.

- Với độ tin cậy 95% suy ra $t_{\alpha;n-1} = t_{0,05;15} = 1,75$.
- Áp dụng (59), khoảng tin cậy cần tìm là
$$0 < \mu \leq 49,75 + 1,75 \times \frac{0,5}{\sqrt{16}}$$
hay
$$0 < \mu \leq 49,97.$$
- Với độ tin cậy 95% giá trị lớn nhất cho trọng lượng trung bình của các bao gạo là 49,97 kg. Ta có thể khẳng định về trung bình các bao gạo đã bị đóng thiếu.

4.3. KHOẢNG TIN CẬY

Khoảng tin cậy cho phương sai

Bài toán 2

Giả sử biến ngẫu nhiên $X \sim \mathcal{N}(\mu; \sigma^2)$ với phương sai $V(X) = \sigma^2$ chưa biết. Lập một mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n. Bài toán đặt ra là hãy tìm khoảng tin cậy cho phương sai σ^2 dựa trên số liệu quan sát $W_x = (x_1, x_2, \ldots, x_n)$.

Khoảng tin cậy hai phía

- Theo Định lý 22 (Chương 2), thống kê

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

Đại lượng này gọi là đại lượng xoay, nó giúp ta có thể tra cứu các giá trị tới hạn từ bảng phân phối Khi-bình phương.

- Khi đó, ta có thể viết (xem Hình 10)

$$P(\chi^2_{1-\alpha/2;n-1} \leq \chi^2 \leq \chi^2_{\alpha/2;n-1}) = 1-\alpha.$$

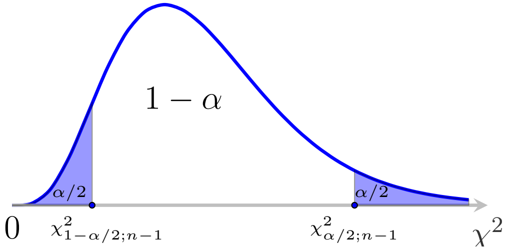
Hay

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

- Từ đây suy ra

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2;,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}\right) = 1-\alpha. \quad (61)$$

Khoảng tin cậy hai phía



Hình 10: $P(\chi^2_{1-\alpha/2;n-1} \le \chi^2 \le \chi^2_{\alpha/2;n-1}) = 1-\alpha$

Khoảng tin cậy hai phía

Định lý 12

Nếu S^2 là phương sai hiệu chỉnh mẫu của mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n chọn từ biến ngẫu nhiên X phân phối chuẩn với phương sai σ^2 chưa biết, thì khoảng tin cậy cho σ^2 với độ tin cậy $1-\alpha$ là

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}, \tag{62}$$

ở đây, $\chi^2_{\alpha/2;n-1}$ và $\chi^2_{1-\alpha/2;n-1}$ là các giá trị tới hạn mức $\alpha/2$ và $1-\alpha/2$ tương ứng của phân phối Khi-bình phương với $n-1$ bậc tự do.

Khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ kích thước n, ta tính được giá trị s^2 của S^2 và xác định được khoảng tin cậy hai phía cho phương sai là

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;n-1}}. \tag{63}$$

Khoảng tin cậy hai phía

Ví dụ 33

Trọng lượng của một loại sản phẩm tuân theo luật phân phối chuẩn. Cân thử trọng lượng của 10 sản phẩm ta được kết quả tương ứng (gam)

$$46,4;\ 46,1;\ 45,8;\ 47,0;\ 46,1;\ 45,9;\ 45,8;\ 46,9;\ 45,2;\ 46,0$$

Với độ tin cậy 95% hãy tìm khoảng tin cậy cho phương sai của trọng lượng sản phẩm.

Khoảng tin cậy hai phía

Giải. Gọi X là trọng lượng sản phẩm, $X \sim \mathcal{N}(\mu; \sigma^2)$. Phương sai của trọng lượng sản phẩm là $V(X) = \sigma^2$ chưa biết, cần ước lượng. Đây là bài toán ước lượng khoảng của phương sai của biến ngẫu nhiên phân phối chuẩn.

- Với $\alpha = 0,05$ và số bậc tự do $n-1 = 9$ ta nhận được $\chi^2_{\alpha/2;n-1} = \chi^2_{0,025;9} = 19,023$ và $\chi^2_{1-\alpha/2;n-1} = \chi^2_{0,975;9} = 2,7$.
- Từ dữ liệu bài ra, tính

$$s^2 = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)} = \frac{10 \times 21273,12 - (461,2)^2}{10 \times 9} = 0,2862.$$

- Do đó, khoảng tin cậy cho σ^2 với độ tin cậy 95% là

$$\frac{9 \times 0,2862}{19,023} \leq \sigma^2 \leq \frac{9 \times 0,2862}{2,7}$$

hay

$$0,1354 \leq \sigma^2 \leq 0,9540.$$

Khoảng tin cậy một phía

Tương tự như trên, khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ kích thước n, ta tính được giá trị s^2 của S^2 và xác định được khoảng tin cậy một phía cho phương sai.

- Khoảng tin cậy trái cho phương sai là

$$0 < \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\alpha;n-1}} \tag{64}$$

- Khoảng tin cậy phải cho phương sai là

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha;n-1}} \leq \sigma^2 < +\infty, \tag{65}$$

ở đây $\chi^2_{\alpha;n-1}$ và $\chi^2_{1-\alpha;n-1}$ là các giá trị tới hạn mức α và $1-\alpha$ tương ứng của phân phối Khi-bình phương với $n-1$ bậc tự do.

Khoảng tin cậy cho độ lệch chuẩn

✍ Lấy căn bậc hai các cận trong các khoảng tin cậy cho phương sai ta sẽ thu được các khoảng tin cậy cho các độ lệch chuẩn.

4.3. KHOẢNG TIN CẬY

Bài toán

Bài toán 3

Ta xét một tổng thể mà mỗi cá thể có hoặc không có dấu hiệu A nào đó. Gọi p là tỷ lệ cá thể có dấu hiệu A trong tổng thể, thông thường p chưa biết. Dựa trên một mẫu được chọn ngẫu nhiên, hãy tìm khoảng tin cậy cho p với độ tin cậy $1 - \alpha$ cho trước.

Xấp xỉ phân phối chuẩn cho tỷ lệ

- Xây dựng một mẫu ngẫu nhiên kích thước n từ một tổng thể kích thước lớn (có thể là vô hạn) và gọi X $(\leq n)$ là số cá thể có dấu hiệu A trong n cá thể được chọn. Khi đó, $\widehat{P} = X/n$ là một ước lượng điểm cho p. Nếu p không quá gần 0 hoặc 1 thì ta có thể xây dựng ước lượng khoảng cho p dựa trên phân phối mẫu của $\widehat{P}$.
- Chú ý rằng n và p là các tham số của phân phối nhị thức và $\widehat{P}$ là thống kê xấp xỉ phân phối chuẩn với $E(\widehat{P}) = p$ và $V(\widehat{P}) = p(1-p)/n$ nếu n đủ lớn. Thông thường, để áp dụng được xấp xỉ này, ta yêu cầu np và $n(1-p)$ lớn hơn hoặc bằng 5.

Xấp xỉ phân phối chuẩn cho tỷ lệ

Định lý 13

Nếu n đủ lớn thì phân phối của thống kê

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\widehat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \tag{66}$$

xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0;1)$.

Khoảng tin cậy đối xứng

- Để xây dựng khoảng tin cậy cho tỷ lệ p, từ Định lý 13,

$$P\left(z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) \approx 1 - \alpha.$$

- Do đó

$$P\left(z_{\alpha/2} \leq \frac{\widehat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \approx 1 - \alpha,$$

hay

$$P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \approx 1 - \alpha. \tag{67}$$

Khoảng tin cậy đối xứng

✍

- Đại lượng $\sqrt{p(1-p)/n}$ trong công thức (67) được gọi là sai số tiêu chuẩn của hàm ước lượng điểm $\widehat{P}$.
- Ta thấy cận trên và cận dưới của khoảng tin cậy trong công thức (67) chứa tham số p chưa biết. Tuy nhiên, ta có thể dùng $\widehat{P}$ để thay thế cho p và nhận được

$$P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}} \leq p \leq \widehat{P} + z_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}}\right) \approx 1-\alpha. \tag{68}$$

Khoảng tin cậy đối xứng

Định lý 14

Nếu $\widehat{P}$ là tỷ lệ các quan sát trong một mẫu ngẫu nhiên kích thước n, thì xấp xỉ khoảng tin cậy cho tỷ lệ p của tổng thể với độ tin cậy $1-\alpha$ là

$$\widehat{P} - z_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}} \le p \le \widehat{P} + z_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}}, \tag{69}$$

ở đây, $z_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân phối chuẩn.

Khi có mẫu cụ thể, ta tính được giá trị cụ thể $\widehat{p}$ của $\widehat{P}$ và nhận được khoảng tin cậy đối xứng của tỷ lệ p:

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \le p \le \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}. \tag{70}$$

✍ Vì p chưa biết nên ta không kiểm tra được điều kiện $np \ge 5$ và $n(1-p) \ge 5$. Trong thực hành ta dùng các điều kiện $n\widehat{p} \ge 5$ và $n(1-\widehat{p}) \ge 5$.

Khoảng tin cậy đối xứng

Ví dụ 34

Điều tra nhu cầu tiêu dùng loại hàng A trong 100 hộ gia đình ở khu dân cư B thấy 60 hộ gia đình có nhu cầu loại hàng trên. Với độ tin cậy $1-\alpha = 95\%$ hãy tìm khoảng tin cậy đối xứng của tỷ lệ hộ gia đình có nhu cầu loại hàng đó. Giải thích kết quả thu được.

Giải. Gọi p là tỷ lệ hộ gia đình ở khu dân cư B có nhu cầu mặt hàng A. Kiểm tra điều kiện $n\widehat{p} = 100 \times 0,6 = 60 > 5$ và $n(1-\widehat{p}) = 100 \times 0,4 = 40 > 5$. Đây là bài toán ước lượng khoảng cho tỷ lệ trường hợp mẫu kích thước n đủ lớn.

Khoảng tin cậy đối xứng

- Áp dụng khoảng tin cậy đối xứng (70), trong đó $z_{\alpha/2} = z_{0,025} = 1,96$.
- Với $n = 100$, $x = 60$, $\widehat{p} = x/n = 0,6$, suy ra khoảng tin cậy cần tìm là

$$0,6 - 1,96\sqrt{\frac{0,6 \times 0,4}{100}} \leq p \leq 0,6 - 1,96\sqrt{\frac{0,6 \times 0,4}{100}}$$

  hay

$$0,504 \leq p \leq 0,696.$$

- Vậy tỷ lệ hộ gia đình ở khu dân cư B có nhu cầu loại hàng A là từ 50,4% đến 69,6% với độ tin cậy 95%.

Vì mọi giá trị trong khoảng tin cậy này đều nhỏ hơn 70% nên ta cũng có thể khẳng định rằng có ít hơn 70% hộ gia đình có nhu cầu mặt hàng A.

Chọn kích thước mẫu

- Ta có

$$P\left(|\widehat{P}-p| \leq z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha. \tag{71}$$

- Do đó, nếu n thỏa mãn

$$z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \varepsilon$$

hay tương đương với

$$n \geq \frac{(z_{\alpha/2})^2 p(1-p)}{\varepsilon^2} \tag{72}$$

thì

$$P\left(|\widehat{P}-p| \leq \varepsilon\right) \geq 1-\alpha.$$

- Như vậy ta cần lấy n là số nguyên dương nhỏ nhất thỏa mãn (72).

Chọn kích thước mẫu

- Tuy nhiên, vì giá trị của p chưa biết, nên vế phải của (72) chưa xác định. Có hai cách để khắc phục tình trạng này.
  (a) Cách thứ nhất là lấy một mẫu sơ bộ kích thước k để thu được tần suất $\widehat{p}_k$ và lấy $\widehat{p}_k$ làm ước lượng ban đầu cho p. Khi đó bất đẳng thức (72) trở thành

$$n \geq \frac{(z_{\alpha/2})^2 \widehat{p}_k (1 - \widehat{p}_k)}{\varepsilon^2} \tag{73}$$

  với điều kiện

$$k\widehat{p}_k \geq 5 \quad \text{và} \quad k(1 - \widehat{p}_k) \geq 5. \tag{74}$$

  Ta sẽ lấy n là số nguyên dương nhỏ nhất thỏa mãn (73) và (74).

Chọn kích thước mẫu

(b) Cách thứ hai, sử dụng bất đẳng thức Cauchy $p(1-p) \leq \frac{1}{4}$ ta nhận được

$$z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq \frac{u_{1-\alpha/2}}{2\sqrt{n}}.$$

Nếu ta chọn n thỏa mãn điều kiện

$$n \geq \frac{\left(u_{1-\alpha/2}\right)^2}{4\varepsilon^2} \tag{75}$$

thì n thỏa mãn điều kiện (72). Vậy ta sẽ lấy n là số nguyên dương nhỏ nhất thỏa mãn (75).

Chọn kích thước mẫu

Ví dụ 35

Một kỹ sư muốn ước lượng tỷ lệ sản phẩm loại A của một nhà máy với độ tin cậy 90% và sai số không quá 0,02. Hỏi kỹ sư đó phải lấy một mẫu kích thước n bằng bao nhiêu?

Giải.

- Nếu làm theo cách thứ nhất, trước hết người kỹ sư lấy một mẫu kích thước $n = 1000$ sản phẩm kiểm tra thấy có 640 sản phẩm loại A. Khoảng tin cậy của tỷ lệ sản phẩm loại A của nhà máy dựa trên mẫu điều tra này theo (70) là

$$\left(0,64 - 1,645\sqrt{\frac{0,64(1-0,64)}{1000}} \le p \le 0,64 + 1,645\sqrt{\frac{0,64(1-0,64)}{1000}}\right)$$

hay

$$(0,64 - 0,25 \le p \le 0,64 + 0,25).$$

Chọn kích thước mẫu

- Sai số của ước lượng là 0,25 lớn hơn 0,02. Vậy cần lấy mẫu có kích thước lớn hơn nữa. Kích thước mẫu n phải thỏa mãn (73), tức là
$$n \geq \frac{(1,645)^2(0,64)(0,36)}{(0,02)^2} = 1558,67.$$
Vậy $n = 1559$.
- Nếu sử dụng cách thứ hai thì phải chọn n thỏa mãn (75), tức là
$$n \geq \frac{(1,645)^2}{(4)(0,02)^2} = 1691,266.$$
Do đó $n = 1692$.

Chọn kích thước mẫu

Ví dụ 36

Một cuộc thăm dò dư luận tại thành phố A được tiến hành để hỏi xem có nên thực hiện một dự án quan trọng với kinh phí lớn trong thành phố hay không. Tỷ lệ người ủng hộ việc thực hiện dự án trên mẫu thăm dò cho ta ước lượng tỷ lệ người ủng hộ việc thực hiện dự án trên toàn bộ số dân của thành phố. Nếu ước lượng tỷ lệ này với độ tin cậy 95% và sai số không quá 0,03 thì phải lấy kích thước mẫu bao nhiêu?

Chọn kích thước mẫu

Giải.

- Nếu sử dụng cách thứ hai, ta phải chọn n thỏa mãn (75) hay
$$n \geq \frac{(1,96)^2}{(4)(0,03)^2} = 1067,111.$$
Do đó $n = 1068$.
- Điều thú vị nhất trong tính toán trên là kích thước mẫu tối thiểu n chỉ phụ thuộc vào độ chính xác và độ tin cậy mà ta mong muốn, chứ không phụ thuộc vào tổng số dân trên thành phố.
- Nếu hỏi ý kiến 1068 người, thì ta đạt được độ tin cậy là 95% và độ chính xác 0,03 mà không phụ thuộc vào tổng số dân trên thành phố là 5 triệu, là 50 triệu hay 500 triệu!

Chọn kích thước mẫu

✍

- Cái khó khăn ở đây là không phải thu thập ý kiến của càng nhiều người càng tốt. Vấn đề lớn là phải đảm bảo đây là mẫu ngẫu nhiên.
- Nếu p khá gần 0,5 thì sự khác nhau giữa số n tìm được theo hai cách sẽ không nhiều lắm. Nhưng nếu p khá gần 0 hoặc 1 thì sự khác biệt sẽ rất lớn. Do đó, nếu cảm thấy tỷ lệ p rất bé hoặc rất lớn thì nên dùng cách thứ nhất.

Khoảng tin cậy một phía

Bằng cách xây dựng tương tự như trên, khi có mẫu cụ thể, ta tính được giá trị cụ thể $\widehat{p}$ của $\widehat{P}$ và xác định được khoảng tin cậy một phía cho tỷ lệ p.

- Khoảng tin cậy trái cho tỷ lệ p là

$$0 < p \leq \widehat{p} + z_\alpha \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}. \tag{76}$$

- Khoảng tin cậy phải cho tỷ lệ p là

$$\widehat{p} - z_\alpha \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p < 1. \tag{77}$$

4.3. KHOẢNG TIN CẬY

Bài tập Mục 4.3

Bài 28

Công thức khoảng ước lượng đối xứng với độ tin cậy $1-\alpha$ cho kỳ vọng của biến ngẫu nhiên $X \sim N(\mu;\sigma^2)$ khi σ^2 đã biết là:

A. $[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$

B. $(-\infty; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$

C. $[\bar{x} - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}; +\infty)$

D. $(-\infty; +\infty)$

Bài 29

Công thức khoảng ước lượng đối xứng với độ tin cậy $1-\alpha$ cho kỳ vọng của biến ngẫu nhiên $X \sim N(\mu;\sigma^2)$ khi σ^2 chưa biết là:

A. $[\bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}; \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}]$

B. $(-\infty; \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}]$

C. $[\bar{x} - z_{\alpha/2;n-1}\frac{s}{\sqrt{n}}; +\infty)$

D. $(-\infty; +\infty)$

Bài 30

Công thức khoảng ước lượng đối xứng với độ tin cậy $1-\alpha$ cho phương sai của biến ngẫu nhiên $X \sim N(\mu;\sigma^2)$ khi μ chưa biết là:

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

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

C. $\left[\frac{nS^2}{\chi^2_{\alpha/2;n-1}}; \frac{nS^2}{\chi^2_{1-\alpha/2;n-1}}\right]$

D. $\left[-\infty; \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}\right]$

Bài 31

Công thức khoảng ước lượng đối xứng với độ tin cậy $1-\alpha$ cho tỷ lệ là:

A. $[\bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}; \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}]$

B. $[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$

C. $[\widehat{p} - z_{\alpha/2}\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}; \widehat{p} + z_{\alpha/2}\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}]$

D. $[\widehat{p} - z_{\alpha/2}\frac{\sqrt{\widehat{p}(1+\widehat{p})}}{\sqrt{n}}; \widehat{p} + z_{\alpha/2}\frac{\sqrt{\widehat{p}(1+\widehat{p})}}{\sqrt{n}}]$

Bài 32

Công thức ước lượng giá trị tối thiểu với độ tin cậy $1-\alpha$ cho kỳ vọng của biến ngẫu nhiên $X \sim N(\mu;\sigma^2)$ khi σ^2 chưa biết là:

A. $[\bar{x} - t_{\alpha;n-1}\frac{s}{\sqrt{n}};+\infty)$

B. $[\bar{x} + t_{\alpha;n-1}\frac{s}{\sqrt{n}};+\infty)$

C. $(-\infty;\bar{x} - t_{\alpha;n-1}\frac{s}{\sqrt{n}}]$

D. $(-\infty;\bar{x} + t_{\alpha;n-1}\frac{s}{\sqrt{n}}]$

Bài 33

Công thức ước lượng giá trị tối đa với độ tin cậy $1-\alpha$ cho giá trị tỷ lệ p là:

A. $[\widehat{p} - z_\alpha\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}};+\infty)$

B. $[\widehat{p} + z_\alpha\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}};+\infty)$

C. $(-\infty;\widehat{p} - z_\alpha\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}]$

D. $(-\infty;\widehat{p} + z_\alpha\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}]$

Bài 34

Tuổi thọ của một loại bóng đèn do một dây chuyền công nghệ sản xuất ra có độ lệch chuẩn là 305 giờ. Người ta lấy ngẫu nhiên ra 45 bóng đèn loại này thấy tuổi thọ trung bình là 2150 giờ. Với độ tin cậy 95% hãy ước lượng tuổi thọ trung bình của loại bóng đèn nói trên.

Bài 35

Một kỹ sư cho biết trọng lượng tạp chất trong một sản phẩm tuân theo luật phân phối chuẩn với độ lệch chuẩn bằng 3,8 gam. Một mẫu ngẫu nhiên gồm 9 sản phẩm được tiến hành kiểm tra và thấy lượng tạp chất như sau (đơn vị tính là gam):

$$18{,}2 \quad 13{,}7 \quad 15{,}9 \quad 17{,}4 \quad 21{,}8 \quad 16{,}6 \quad 12{,}3 \quad 18{,}8 \quad 16{,}2$$

(a) Tìm khoảng tin cậy cho trọng lượng trung bình tạp chất của sản phẩm với độ tin cậy 99%.

(b) Không cần tính toán, nếu độ tin cậy 95% thì khoảng ước lượng trung bình sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

Bài 36

Giả sử chiều dài của một chi tiết sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn với độ lệch chuẩn là 0,2m. Người ta sản xuất thử nghiệm 35 sản phẩm loại này và tính được chiều dài trung bình là 25m. Với độ tin cậy 95% hãy ước lượng khoảng cho chiều dài trung bình của chi tiết sản phẩm đang được thử nghiệm.

Bài 37

Để xác định trọng lượng trung bình của các bao gạo được đóng gói bằng máy tự động, người ta chọn ngẫu nhiên ra 20 bao gạo và thấy trung bình mẫu là 49,2kg và độ lệch chuẩn mẫu hiệu chỉnh là 1,8kg. Biết rằng trọng lượng các bao gạo xấp xỉ phân phối chuẩn. Hãy tìm khoảng tin cậy cho trọng lượng trung bình của một bao gạo với độ tin cậy 99%.

Bài 38

Thời gian đợi phục vụ tại một cửa hàng ăn nhanh là biến ngẫu nhiên xấp xỉ phân phối chuẩn. Người ta khảo sát 16 người thì thấy thời gian đợi trung bình là 4 phút và độ lệch chuẩn mẫu hiệu chỉnh là 1,8 phút. Với độ tin cậy 99% hãy tìm khoảng tin cậy cho thời gian chờ đợi trung bình của một khách hàng tại cửa hàng ăn nhanh này.

Bài 39

Một mẫu ngẫu nhiên gồm 16 thùng hàng được chọn ra từ tất cả các thùng hàng được sản xuất bởi nhà máy trong một tháng. Trọng lượng của 16 thùng hàng lần lượt như sau (đơn vị tính là kg):

18,6	18,4	19,2	19,8	19,4	19,5	18,9	19,4
19,7	20,1	20,2	20,1	18,6	18,4	19,2	19,8

Tìm khoảng tin cậy cho trọng lượng trung bình tổng thể của tất cả các thùng hàng của nhà máy với độ tin cậy 95%, biết rằng trọng lượng thùng hàng được chọn ngẫu nhiên là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Bài 40

Để định mức thời gian gia công một chi tiết máy, người ta theo dõi ngẫu nhiên quá trình gia công 35 chi tiết máy và thu được số liệu:

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

Giả sử thời gian gia công chi tiết máy là biến ngẫu nhiên tuân theo luật phân phối chuẩn. Với độ tin cậy 95% hãy ước lượng khoảng tin cậy cho thời gian gia công trung bình một chi tiết máy nói trên.

Bài 41

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

Với độ tin cậy 99% hãy tìm khoảng ước lượng đối xứng của áp lực trung bình của các thùng trên. Biết rằng áp lực là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Bài 42

Một bài báo trong Nuclear Engineering International (tháng 2 năm 1988, trang 33) mô tả một số đặc điểm của các thanh nhiên liệu được sử dụng trong một lò phản ứng hạt nhân của một công ty điện lực ở Na Uy. Người ta đo tỷ lệ làm giàu của 12 thanh và có được dữ liệu sau:

$$2{,}94 \quad 3{,}00 \quad 2{,}90 \quad 2{,}90 \quad 2{,}75 \quad 2{,}95 \quad 2{,}75 \quad 3{,}00 \quad 2{,}95 \quad 2{,}82 \quad 2{,}81 \quad 3{,}05$$

Giả sử tỷ lệ làm giàu của các thanh nhiên liệu tuân theo luật phân phối chuẩn. Hãy ước lượng khoảng cho tỷ lệ làm giàu trung bình của các thanh nhiên liệu với độ tin cậy 95%.

Bài 43

Người ta chọn ngẫu nhiên ra 49 sinh viên của một trường đại học và thấy chiều cao trung bình mẫu là 163cm và độ lệch chuẩn mẫu hiệu chỉnh là 12cm. Hãy tìm khoảng ước lượng với độ tin cậy 99% cho chiều cao trung bình của sinh viên của trường đó.

Bài 44

Một trường đại học tiến hành một nghiên cứu xem trung bình một sinh viên tiêu hết bao nhiêu tiền điện thoại trong một tháng. Họ điều tra 60 sinh viên và cho thấy số tiền trung bình mẫu là 95 nghìn và độ lệch chuẩn mẫu hiệu chỉnh là 36 nghìn. Hãy ước lượng khoảng với độ tin cậy 95% cho số tiền điện thoại trung bình trong một tháng của mỗi sinh viên.

Bài 45

Người ta điều tra 35 người nghiện thuốc lá được chọn ngẫu nhiên từ số lượng người nghiện hút thuốc lá của một thành phố thấy số điếu thuốc hút trong 5 ngày của họ là:

31	37	48	40	59	97	98	87	80	68	64	45
48	62	74	76	79	85	83	81	93	82	85	79
34	57	95	49	59	63	48	79	50	55	63	

Hãy tìm khoảng ước lượng cho số điếu thuốc hút trung bình trong 5 ngày của những người nghiện thuốc lá của thành phố đó với độ tin cậy 99%.

Bài 46

Để điều tra tiền điện phải trả trong một tháng của một hộ dân cư ở phường A, người ta kiểm tra ngẫu nhiên 200 hộ gia đình ở phường này và được kết quả sau:

Số tiền (nghìn đồng)	[80,180)	[180,280)	[280,380)	[380,480)	[480,580)	[580,680)	[680,78
Số hộ gia đình	14	25	43	46	39	23	10

Ước lượng khoảng cho số tiền trung bình một hộ dân phải trả ở phường đó với độ tin cậy 95%.

Bài 47

Để ước lượng số lượng xăng hao phí trên một tuyến đường của một hãng xe khách, người ta tiến hành chạy thử nghiệm 55 lần liên tiếp trên tuyến đường này và có được số liệu:

Lượng xăng hao phí	10,5-11	11-11,5	11,5-12	12-12,5	12,5-13	13-13,5
Tần số	5	12	15	13	6	4

Hãy ước lượng lượng xăng hao phí trung bình cho một xe với độ tin cậy 95%.

Bài 48

Để xác định giá trung bình đối với một loại hàng hóa trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng thu được số liệu sau:

Giá (nghìn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng	5	8	13	14	30	11	8	6	4	1

Với độ tin cậy 95% hãy ước lượng giá trung bình của loại hàng đó tại thời điểm đang xét. Biết rằng giá hàng hóa là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Bài 49

Để ước lượng cho tỷ lệ những cây bạch đàn có chiều cao đạt chuẩn phục vụ cho việc khai thác ở một nông trường lâm nghiệp, người ta tiến hành đo ngẫu nhiên chiều cao của 135 cây và thấy có 36 cây cao từ 7,5 m trở lên. Hãy ước lượng khoảng cho tỷ lệ các cây bạch đàn có chiều cao trên 7,5 m với độ tin cậy 95%.

Bài 50

Để ước lượng số cá có trong hồ người ta bắt từ hồ lên 100 con đánh dấu rồi thả lại vào hồ. Sau đó người ta bắt lên 300 con thì thấy có 32 con bị đánh dấu. Hãy ước lượng khoảng cho số cá có trong hồ với độ tin cậy 99%.

Bài 51

Để điều tra thị phần xe máy, người ta chọn ngẫu nhiên ra 450 người mua xe máy trong một tháng ở các địa bàn ở một thành phố thì có 275 người mua xe Honda. Tìm khoảng tin cậy cho tỷ lệ người mua xe Honda với độ tin cậy 95%.

Bài 52

Kiểm tra ngẫu nhiên 400 sản phẩm do một hệ thống máy mới sản xuất thì thấy có 387 chính phẩm. Hãy ước lượng tỷ lệ chính phẩm tối thiểu của hệ thống máy mới với độ tin cậy 95%.

Bài 53

Mở thử 200 hộp của kho đồ hộp thấy có 10 hộp bị biến chất. Với độ tin cậy 95% hãy ước lượng tỷ lệ hộp bị biến chất tối đa của kho.

Bài 54

Cần phải lập một mẫu ngẫu nhiên với kích thước là bao nhiêu để tỷ lệ phế phẩm của mẫu là 0,2 và độ dài khoảng tin cậy đối xứng là 0,05 và độ tin cậy của ước lượng là 95%.

Bài 55

Làm cách nào để ước lượng số thú hiếm trong một khu rừng với độ tin cậy 95%.

Bài 56

Để nghiên cứu về thời gian xem ti vi của một thanh niên từ 18 đến 35 tuổi trong vòng một tuần, người ta tiến hành khảo sát trên 40 người và cho ta bảng số liệu sau:

39	02	43	35	15	54	23	21	25	07	24	33	17	
23	24	43	11	15	17	15	19	06	43	35	25	37	
15	14	08	11	29	12	13	25	15	28	24	06	16	7

Hãy tìm khoảng ước lượng cho thời gian xem ti vi trung bình của thanh niên trong độ tuổi trên trong vòng một tuần với độ tin cậy 99%.

Bài 57

Thử nghiệm 560 bóng đèn điện tử do một nhà máy sản xuất thì thấy 10 bóng có lỗi kỹ thuật. Hãy tìm ước lượng cho tỷ lệ bóng có lỗi kỹ thuật tối đa với độ tin cậy 95%.

Bài 58

Một kỹ sư nghiên cứu về cường độ nén của bê tông đang được thử nghiệm. Anh ta tiến hành kiểm tra 12 mẫu vật và có được các dữ liệu sau đây:

$$2216 \quad 2234 \quad 2225 \quad 2301 \quad 2278 \quad 2255 \quad 2249 \quad 2204 \quad 2286 \quad 2263 \quad 2275 \quad 2295$$

Giả sử cường độ nén của bê tông đang thử nghiệm tuân theo luật phân phối chuẩn.

(a) Hãy ước lượng khoảng với độ tin cậy 95% cho cường độ nén trung bình của bê tông đang được thử nghiệm.

(b) Hãy ước lượng khoảng tin cậy phải cho cường độ nén trung bình của bê tông đang được thử nghiệm với độ tin cậy 99%.

Bài 59

Nghiên cứu về năng suất của loại hoa màu A, người ta kiểm tra năng suất của 64 điểm trồng loại hoa màu này thu được bảng số liệu

Năng suất (tạ/ha)	40–45	45–50	50–55	55-60	60–65	65–70
Số điểm	2	5	15	30	8	4

(a) Hãy ước lượng năng suất trung bình của loại hoa màu A với độ tin cậy 95%; Nếu muốn sai số của ước lượng giảm đi 2 lần thì cần kiểm tra bao nhiêu điểm để đảm bảo yêu cầu nêu trên?

(b) Biết rằng trên toàn miền Bắc có 10.000 điểm trồng loại hoa màu A. Hãy cho biết có khoảng bao nhiêu điểm đạt năng suất trên 60 tạ/ha? Hãy kết luận với độ tin cậy 99%.

(c) Hãy cho biết tỷ lệ những điểm có năng suất trên 60 tạ/ha của loại hoa màu A tối thiểu là bao nhiêu? Hãy kết luận với độ tin cậy 95%?

Bài 60

Trọng lượng những viên gạch trong một quá trình sản xuất gạch được giả sử là tuân theo luật phân phối chuẩn. Một mẫu ngẫu nhiên gồm 25 viên gạch vừa sản xuất ra trong ngày có trọng lượng trung bình 2,45 kg và độ lệch chuẩn mẫu hiệu chỉnh là 0,15 kg.

(a) Tìm khoảng tin cậy của trọng lượng trung bình của tất cả các viên gạch trong ngày với độ tin cậy 99%.

(b) Không cần tính toán, với độ tin cậy 95% thì khoảng tin cậy trung bình sẽ rộng hơn, hẹp hơn hay bằng với kết quả ý (a)?

(c) Một mẫu ngẫu nhiên gồm 20 viên gạch sẽ được chọn ra trong ngày mai. Không cần tính toán, với độ tin cậy 99% thì khoảng tin cậy cho trọng lượng trung bình của tất cả các viên gạch sản xuất ra trong ngày mai sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

(d) Sự thật rằng, độ lệch chuẩn mẫu của các viên gạch sản xuất trong ngày mai là 0,10kg. Không cần tính toán, với độ tin cậy 99% thì khoảng tin cậy cho trọng lượng trung bình của tất cả các viên gạch sản xuất ra trong ngày mai sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

Bài 61

Một trường đại học lớn đang quan tâm về lượng thời gian sinh viên tự nghiên cứu mỗi tuần. Người ta tiến hành khảo sát một mẫu ngẫu nhiên gồm 16 sinh viên, dữ liệu cho thấy thời gian nghiên cứu trung bình của một sinh viên là 15,26 giờ/tuần và độ lệch chuẩn hiệu chỉnh là 6,43 giờ. Giả sử thời gian nghiên cứu của sinh viên của trường đại học trên là tuân theo luật phân phối chuẩn.

(a) Tìm khoảng tin cậy cho lượng thời gian tự nghiên cứu trung bình mỗi tuần cho tất cả sinh viên trường đại học này với độ tin cậy 95%.

(b) Không cần tính toán, khoảng tin cậy của trung bình tổng thể khi ước lượng sẽ rộng hơn hay hẹp hơn với ba điều kiện sau:

b1. Mẫu gồm 30 sinh viên được chọn ra, với tất cả các điều kiện khác giống như ý (a)?

b2. Độ lệch chuẩn mẫu là 4,15 giờ, tất cả các điều kiện khác giống như ý (a)?

b3. Độ tin cậy 99%, tất cả các điều kiện khác giống như ý (a)?

Bài 62

Xét mẫu bộ dữ liệu gồm 60 quan sát sau:
1.14, 1.11, 1.19, 1.00, 2.10, 1.07, 1.10, 1.24, 1.21, 1.08, 2.05, 1.19, 1.01, 1.36, 1.61, 1.22, 1.68, 1.40, 1.52, 1.41, 1.72, 1.38, 1.94, 2.02, 2.10, 1.13, 1.07, 1.30, 1.22, 1.50, 1.17, 1.24, 1.01, 1.11, 1.95, 1.14, 1.33, 1.81, 1.36, 1.25, 5.19, 1.02, 2.71, 1.87, 1.24, 1.20, 1.02, 3.06, 1.22, 1.20, 1.71, 1.25, 1.10, 1.43, 2.29, 1.31, 1.72, 1.85, 1.37, 1.09
Từ phân bố của dữ liệu, ta có thể mô hình hoá phân bố của tổng thể với hàm mật độ xác suất như sau:

$$f(x;\theta) = \begin{cases} (\theta+1)x^{-\theta-2} & \text{nếu } x \geq 1 \\ 0 & \text{nếu } x < 1 \end{cases}$$

trong đó $\theta > 0$ là tham số của mô hình.

(a) Hãy tính giá trị kỳ vọng của phân phối trên.

(b) Hãy tính ước lượng mô men của tham số θ.

(c) Hãy tính ước lượng hợp lý cực đại của tham số θ.

(d) Khi kích thước mẫu n đủ lớn thì thống kê $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ có phân phối xấp xỉ chuẩn tắc. Áp dụng định lý giới hạn này, hãy xây dựng công thức khoảng ước lượng cho tham số θ với độ tin cậy 90%.

Bài 63

Xét bộ dữ liệu sau về số khách hàng vào một cửa hàng trong 1 giờ sau một chiến dịch marketing (bao gồm 100 quan sát):
20, 8, 19, 13, 18, 13, 8, 13, 19, 23, 18, 11, 16, 17, 13, 16, 19, 12, 11, 18, 13, 19, 16, 10, 15, 19, 17, 20, 15, 16, 16, 13, 19, 19, 5, 13, 11, 13, 14, 17, 10, 16, 12, 11, 18, 13, 16, 13, 11, 9, 14, 8, 16, 12, 20, 20, 13, 10, 16, 13, 19, 10, 7, 12, 20, 12, 14, 17, 19, 19, 20, 16, 15, 12, 14, 9, 10, 9, 13, 15, 18, 15, 21, 11, 16, 16, 19, 19, 14, 11, 13, 15, 9, 13, 17, 19, 18, 16, 18, 12.
Giả sử số khách hàng vào cửa hàng trong 1 giờ có phân bố Poisson với tham số $\lambda > 0$. Hãy xây dựng 2 công thức khoảng ước lượng cho tham số λ với độ tin cậy 90% (trong trường hợp kích thước mẫu đủ lớn) bằng cách áp dụng hai định lý giới hạn sau:

(a) Thống kê $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ có phân phối xấp xỉ chuẩn tắc.

(b) Thống kê $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ có phân phối xấp xỉ chuẩn tắc.

Bài 64

Thời gian (X) một loại hạt phóng xạ phân rã có thể mô hình hoá bởi phân bố mũ với tham số $\lambda > 0$. Xét bộ dữ liệu sau về thời gian loại hạt phóng xạ trên phân rã (bao gồm 100 quan sát):
28, 1.5, 0.9, 5.6, 3.8, 3.4, 3.7, 4.3, 0.1, 14.7, 4.4, 7, 1.8, 11.9, 5.4, 3.7, 4.1, 1.6, 2.1, 5.6, 4.4, 0.9, 2.7, 15.2, 0.8, 1, 1.6, 0.1, 7.7, 11, 1.4, 6.8, 7.4, 2, 1.6, 6.9, 9.6, 0.5, 0.4, 11.5, 0.6, 3.8, 1, 1.9, 2.1, 9.6, 4.6, 8.4, 3.2, 5.4, 13.5, 0.9, 0.7, 5.4, 3.8, 0.2, 2.4, 1.5, 7.8, 0.3, 12.6, 1.1, 10.1, 8, 10.8, 2.4, 5, 2.4, 1, 3.6, 0.2, 1.8, 0.1, 4.4, 4.9, 1, 6.4, 26.5, 6.8, 0.2, 10.2, 2, 3.1, 0.7, 6.8, 0.1, 1.6, 3.3, 5.9, 0.6, 3.9, 1.4, 2.9, 9.2, 12.4, 11.1, 0.2, 0.6, 4.8, 3.2
Hãy xây dựng 2 công thức khoảng ước lượng cho tham số λ với độ tin cậy 90% (trong trường hợp kích thước mẫu đủ lớn) bằng cách áp dụng hai định lý giới hạn sau:

(a) Thống kê $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ có phân phối xấp xỉ chuẩn tắc.

(b) Thống kê $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ có phân phối xấp xỉ chuẩn tắc.