Bài 6: ƯỚC LƯỢNG THAM SỐ

Mục tiêu

- Giới thiệu một số khái niệm cơ bản liên quan đến bài toán ước lượng tham số của biến ngẫu nhiên: ước lượng điểm, ước lượng không chệch, ước lượng hiệu quả, ước lượng vững, … trình bày một số kiến thức về khái niệm ước lượng khoảng và đưa ra phương pháp ước lượng đối với một số tham số thống kê thường gặp nhất là kỳ vọng, phương sai và tỷ lệ.
- Kiến thức về ước lượng khoảng có ý nghĩa quan trọng chuẩn bị cho nội dung tiếp theo của bài toán kiểm định giả thuyết.

Thời lượng

- 8 tiết

Các kiến thức cần có

- Ước lượng điểm
- Khái niệm ước lượng điểm
- Ước lượng không lệch
- Ước lượng hiệu quả
- Ước lượng vững
- Ước lượng khoảng
- Khái niệm ước lượng khoảng
- Ước lượng khoảng cho kỳ vọng của biến ngẫu nhiên phân phối chuẩn
- Ước lượng khoảng cho phương sai của biến ngẫu nhiên phân phối chuẩn
- Ước lượng khoảng cho xác suất (tỷ lệ)

TÌNH HUỐNG KHỞI ĐỘNG BÀI

Tình huống

Để ước lượng phế phẩm của một dây chuyền sản xuất mới mua lại, công ty Thiên An kiểm tra ngẫu nhiên 100 sản phẩm do một nhà máy sản xuất thấy có 12 phế phẩm. Với độ tin cậy 95% , hãy ước lượng tỷ lệ phế phẩm của nhà máy đó. Nếu muốn độ chính xác là 0,03 thì phải lấy tối thiểu bao nhiêu sản phẩm?

Câu hỏi

1. Nhà sản xuất cần phải xem chất lượng của dây chuyền sản xuất. Vấn đề đặt ra là làm thế nào để nhà quản lý có thể ước lượng được tỷ lệ phế phẩm bình quân của dây chuyền?
2. Khoảng ước lượng cho tỷ lệ phế phẩm của nhà máy là bao nhiêu nếu giám đốc muốn độ tin cậy cho ước lượng đó là 95%?
3. Để khoảng ước lượng có độ chính xác cao (cỡ 0,03) thì cần phải tốn bao nhiêu tiền? Biết chi phí điều tra 01 mẫu mất 10000VNĐ

Trong bài này ta xét bài toán ước lượng tham số, một trong những bài toán quan trong và có nhiều ứng dụng của thống kê toán.

Bài toán:

Cho biến ngẫu nhiên X với tham số θ chưa biết, dựa vào thông tin mẫu (X_1, X_2, …, X_n) hãy ước lượng tham số θ.

6.1. Ước lượng điểm

6.1.1. Khái niệm

Thống kê (hàm đa biến) $\Theta^* = G(X_1, X_2, ..., X_n)$ dùng làm ước lượng cho tham số θ được gọi là *ước lượng điểm* cho θ.

Với mẫu cụ thể (x_1, x_2, …,x_n), giá trị của thống kê Θ^* là $\theta^* = G(x_1, x_2, ..., x_n)$, giá trị này có thể lấy làm giá trị ước lượng tương ứng cho θ.

CHÚ Ý

Thống kê là một hàm đa biến, còn mẫu ngẫu nhiên là một bộ các biến ngẫu nhiên. Khi gán các biến ngẫu nhiên đó vào vị trí các đối số tương ứng của hàm đa biến nói trên, ta thu được một biến ngẫu nhiên mới. Lúc đó thống kê trở thành một biến ngẫu nhiên và ta có thể lập các tham số của thống kê mới này, như kỳ vọng, phương sai, ... của thống kê đó.

Ví dụ 1:

Đối với biến ngẫu nhiên X, thống kê: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là một ước lượng điểm cho: $\theta = \mu = E(X)$. Giá trị cụ thể của ước lượng điểm này là $\overline{x}$.

Đối với một tham số cho trước, có rất nhiều thống kê có thể lấy làm ước lượng cho tham số đó (nói chung mọi hàm đa biến đều có thể được coi là ước lượng nào đó của tham số). Tuy nhiên, người ta thường quan tâm đến những ước lượng có những tính chất “Tốt”, “Phù hợp” (theo một nghĩa nào đấy) đối với tham số đang được quan tâm. “Không chệch”, “Hiệu quả” và “Vững” là những tính chất tốt thường được xét đến đối với các ước lượng tham số.

6.2. Ước lượng không chệch

Định nghĩa 1:

Thống kê Θ^* gọi là ước lượng không chệch cho tham số θ nếu

$$E(\Theta^*) = \theta.$$

Nếu khác đi ta nói Θ^* là một ước lượng chệch của θ.

Ví dụ 2:

Thống kê $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là một ước lượng không chệch cho tham số μ.

Thật vậy, ta có:

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

Ví dụ 3:

Ta có:

$$E(S^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

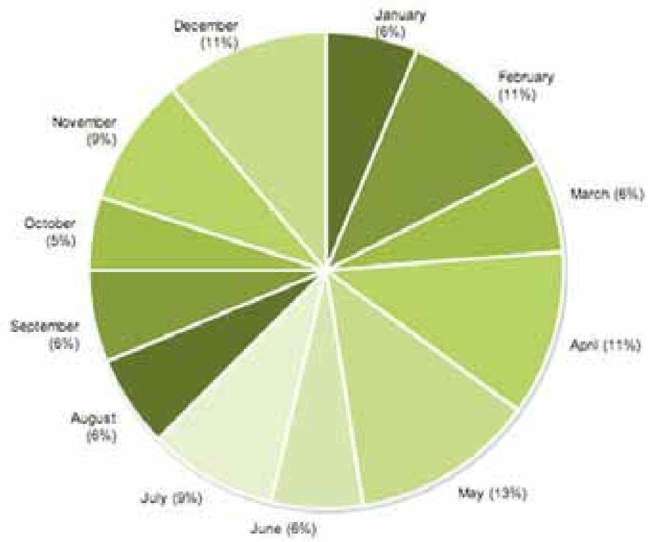
$$E(S'^2) = E\left[\frac{n}{n-1}S^2\right] = \frac{n}{n-1}E(S^2) = \sigma^2$$

Vậy S^2 là ước lượng chệch của σ^2 và S'^2 là ước lượng không chệch của σ^2.

6.2.1. Ước lượng hiệu quả

Định nghĩa 2:

Thống kê Θ^* được gọi là ước lượng hiệu quả cho tham số θ nếu $E(\Theta^*) = \theta$ và Θ^* có phương sai nhỏ nhất trong các ước lượng không chệch của θ.



6.2.2. Ước lượng vững

Định nghĩa 3

Thống kê Θ^* được gọi là ước lượng vững cho θ nếu:

$$\lim_{n\to\infty} P\{|\Theta^* - \theta| < \varepsilon\} = 1, \quad \forall \varepsilon > 0.$$

Ví dụ 4:

Theo Luật số lớn ta thấy thống kê

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

là ước lượng vững của kỳ vọng μ.

Trên đây là một số tính chất thường được xét đến khi đánh giá các thống kê dùng làm ước lượng cho một tham số. Trong thực hành, ngoài một số tham số đơn giản như kỳ vọng và phương sai, người ta còn quan tâm đến nhiều tham số khác và phải có những phương pháp thích hợp để tìm ra các ước lượng cho tham số cần quan tâm.

6.3. Ước lượng khoảng

Trong phần trên ta nói đến việc tìm ước lượng điểm cho tham số dựa vào dữ liệu mẫu. Tuy nhiên, vấn đề quan trọng là làm thế nào để đánh giá được chất lượng của một ước lượng thu được trong khi ước lượng điểm khó cho ta một kết luận chính xác về độ sai lệch giữa tham số và ước lượng điểm của nó. Trong mục này ta sẽ đưa ra một cách tiếp cận khác để ước lượng tham số đó là ước lượng khoảng. Phương pháp này được sử dụng rộng rãi khi tiến hành các phép kiểm định trong các lĩnh vực khoa học, kỹ thuật, kinh tế, …

6.3.1. Khái niệm

- Khoảng với hai đầu mút ngẫu nhiên:

$$(L;U) = \left(L(X_1, X_2, ..., X_n); U(X_1, X_2, ..., X_n)\right)$$

được gọi là ước lượng khoảng (hai phía) cho tham số θ với độ tin cậy $1-\alpha$ nếu:

$$P\{L(X_1, X_2, ..., X_n) < \theta < U(X_1, X_2, ..., X_n)\} = 1-\alpha$$

Khoảng $(L;+\infty)$ và $(-\infty;U)$ gọi là ước lượng một phía cho θ với độ tin cậy $1-\alpha$ nếu:

$$P\{L(X_1, X_2, ..., X_n) < \theta\} = P\{\theta < U(X_1, X_2, ..., X_n)\} = 1-\alpha.$$

Với mẫu cụ thể $(x_1, x_2, \ldots, x_n)$ giá trị của khoảng ước lượng cho θ là:

  - Khoảng ước lượng hai phía: $\theta \in (l;u) = \left(L(x_1, x_2, ..., x_n); U(x_1, x_2, ..., x_n)\right)$
  - Khoảng ước lượng phía trái: $\theta \in (l;+\infty) = \left(L(x_1, x_2, ..., x_n); +\infty\right)$
  - Khoảng ước lượng phía phải: $\theta \in (-\infty;u) = \left(-\infty; U(x_1, x_2, ..., x_n)\right)$

- Hiệu $u - l$ của khoảng ước lượng hai phía được gọi là độ chính xác của ước lượng.

6.3.2. Ước lượng khoảng cho kỳ vọng của biến ngẫu nhiên phân phối chuẩn

Cho biến ngẫu nhiên $X \sim N(\mu, \sigma^2)$ với tham số μ chưa biết và mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ có giá trị cụ thể $(x_1, x_2, \ldots, x_n)$.

6.3.2.1. Trường hợp σ^2 đã biết. Từ tính chất của phân phối chuẩn, ta có

$$\overline{X} \sim N(\mu, \sigma^2 / n) \;;\; \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1).$$

Với độ tin cậy $1-\alpha$ ta cần tìm điểm $u_{\alpha/2}$ sao cho:

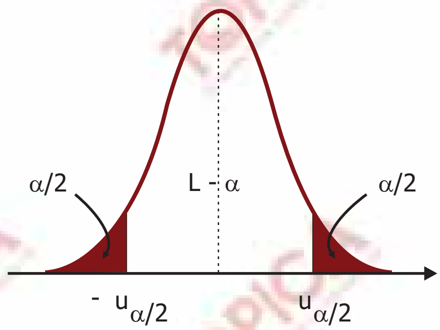
$$P\left\{-u_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma}\sqrt{n} < u_{\alpha/2}\right\} = 1-\alpha$$

$$P\left\{\overline{X} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right\} = 1-\alpha$$

CHÚ Ý

Độ tin cậy $1-\alpha$ thường được lấy lớn hơn 90%.

trong đó phân vị $u_{\alpha/2}$ thoả mãn $\Phi_0(u_{\alpha/2}) = 1-\alpha/2$. Tra bảng phân phối chuẩn ta tìm được $u_{\alpha/2}$.



Hình 1: Đồ thị phân phối chuẩn và các phân vị xác định khoảng tin cậy

Với mẫu cụ thể $(x_1, x_2, \ldots, x_n)$ ta có khoảng ước lượng (hai phía) cho μ là:

$$\mu \in (\overline{x} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2}; \quad \overline{x} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}).$$

Tương tự ta có các khoảng ước lượng một phía của μ là:

- Ước lượng giá trị tối thiểu:

  $$\mu \in (\overline{x} - \frac{\sigma}{\sqrt{n}}u_{\alpha}; \; +\infty)$$

  trong đó $\Phi_0(u_\alpha) = 1-\alpha$, tra bảng phân phối chuẩn ta tìm được u_α.

CHÚ Ý

Ngoài cách tra bảng, ta dùng lệnh trong Excel: normsinv (1-α/2). Tham khảo phần phụ lục

- Ước lượng giá trị tối đa:

  $$\mu \in (-\infty; \; \overline{x} + \frac{\sigma}{\sqrt{n}}u_{\alpha})$$

Ví dụ 5:

Điều tra thu nhập (triệu/năm) hàng năm của 25 hộ gia đình trong vùng ta có bảng số liệu:

Thu nhập	11,5	11,6	11,7	11,8	11,9	12
Số hộ	5	8	4	6	1	1

Hãy ước lượng mức thu nhập trung bình trong vùng với độ tin cậy 95%, biết rằng thu nhập là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn $\sigma = 0,2$.

Giải:

Gọi X là biến ngẫu nhiên thu nhập hộ gia đình trong vùng, ta có:

$$X \sim N(\mu; 0,2^2).$$

Từ đó $\overline{x} = 11,672$, $\Phi_0(u_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0,975$; $u_{0,025} = 1,96$. Vậy khoảng ước lượng của thu nhập trung bình μ là:

$$\mu \in (11,672 - \frac{0,2}{\sqrt{25}}1,96;\ 11,672 + \frac{0,2}{\sqrt{25}}1,96) = (11,594;\ 11,75).$$

Ví dụ 6:

(Xét Ví dụ 5) Hãy ước lượng giá trị tối thiểu và giá trị tối đa của mức thu nhập trung bình trong vùng với độ tin cậy 99%.

Giải:

Ta có độ tin cậy $1 - \alpha = 99\%$, $\alpha = 0,01$, tra bảng ta có: $u_\alpha = u_{0,01} = 2,33$.

Ước lượng giá trị tối thiểu:

$$\mu \in (11,672 - \frac{0,2}{\sqrt{25}}2,33;\ +\infty) = (11,579;\ +\infty)$$

Ước lượng giá trị tối đa:

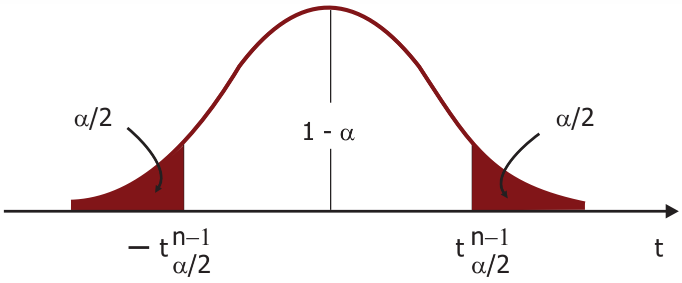
$$\mu \in (-\infty; 11,672 + \frac{0,2}{\sqrt{25}}2,33) = (-\infty; 11,765).$$

6.3.2.2. Trường hợp σ^2 chưa biết

Ta có thống kê $T = \frac{\overline{X} - \mu}{S'}\sqrt{n}$ có phân phối Student với n – 1 bậc tự do. Với độ tin cậy $1 - \alpha$ ta tìm được điểm phân vị $t_{\alpha/2}^{n-1}$ sao cho:

$$P\left\{-t_{\alpha/2}^{n-1} < \frac{\overline{X}-\mu}{S'}\sqrt{n} < t_{\alpha/2}^{n-1}\right\} = 1-\alpha;$$

$$\left\{\overline{X} - \frac{S'}{\sqrt{n}}t_{\alpha/2}^{n-1} < \mu < \overline{X} + \frac{S'}{\sqrt{n}}t_{\alpha/2}^{n-1}\right\} = 1-\alpha,$$



Hình 2: Đồ thị phân phối Student và các phân vị xác định khoảng tin cậy

trong đó phân vị $t_{\alpha/2}^{n-1}$ được tìm từ bảng phân phối Student.

Vậy với mẫu cụ thể ta có khoảng ước lượng cho μ:

$$\mu \in \left(\overline{x} - \frac{s'}{\sqrt{n}}t_{\alpha/2}^{n-1};\quad \overline{x} + \frac{s'}{\sqrt{n}}t_{\alpha/2}^{n-1}\right).$$

Tương tự ta có các khoảng ước lượng một phía là:

- Ước lượng giá trị tối thiểu:

$$\mu \in \left(\overline{x} - \frac{s'}{\sqrt{n}}t_{\alpha}^{n-1};\ +\infty\right)$$

  phân vị t_{α}^{n-1} được tìm từ bảng phân phối Student

- Ước lượng giá trị tối đa:

$$\mu \in \left(-\infty;\ \overline{x} + \frac{s'}{\sqrt{n}}t_{\alpha}^{n-1}\right).$$

CHÚ Ý

Ngoài cách tra bảng, ta dùng lệnh trong Excel: tinv(α,n-1). Tham khảo phần phụ lục

Ví dụ 7:

Điều tra thu nhập (triệu/năm) hàng năm của 25 hộ gia đình trong vùng ta có bảng số liệu:

Thu nhập	11,5	11,6	11,7	11,8	11,9	12
Số hộ	5	8	4	6	1	1

Hãy ước lượng mức thu nhập trung bình trong vùng với độ tin cậy 95%, biết rằng thu nhập là biến ngẫu nhiên có phân phối chuẩn.

Giải:

Gọi X là thu nhập của hộ gia đình trong vùng, lúc đó $X \sim N(\mu;\sigma^2)$, đây là trường hợp σ chưa biết. Ta có:

$$\bar{x} = 11,672, \quad s'^2 = 0,0188, \quad s' = 0,137.$$

$$1-\alpha = 0,95 \;\; ; \;\; \alpha = 0,05$$

$$t_{\alpha/2}^{n-1} = t_{0,025}^{24} = 2,06.$$

Vậy khoảng ước lượng cho thu nhập trung bình là:

$$\mu \in (11,62; \;\; 11,73).$$

Tương tự ta có các khoảng ướng lượng một phía

Ước lượng giá trị tối thiểu:

$$t_{\alpha}^{n-1} = t_{0,05}^{24} = 1,71$$

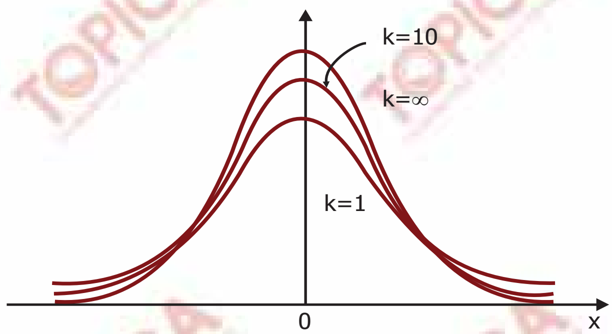
$$\mu \in \left(11,672 - \frac{0,137}{\sqrt{25}}1,71; +\infty) = (1,625; +\infty\right).$$

CHÚ Ý

Nếu cỡ mẫu n đủ lớn $n \geq 30$ thì thống kê T có thể xấp xỉ phân phối chuẩn N(0;1). Vậy ta có các khoảng ước lượng cho μ tương tự như trường hợp σ đã biết

Ước lượng giá trị tối đa:

$$\mu \in (-\infty; \; 11,672 + \frac{0,137}{\sqrt{25}}1,71) = (-\infty; \; 11,719).$$



Hình 3: Đồ thị phân phối Student với các bậc tự do khác nhau

6.3.2.3. Xác định cỡ mẫu

Ước lượng khoảng hai phía cho μ trong trường hợp σ đã biết là:

$$\mu \in (\bar{x} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2}; \; \bar{x} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2})$$

$$\Rightarrow \varepsilon = |\bar{x} - \mu| < \frac{\sigma}{\sqrt{n}}u_{\alpha/2} \qquad (*)$$

Ta thấy rằng khi cỡ mẫu càng lớn thì độ sai lêch giữa μ và $\overline{x}$ càng nhỏ, ta gọi ε là độ chính xác của ước lượng. Trong (*) ta thấy rằng nếu cho trước độ chính xác của ước lượng là ε_0 thì cỡ mẫu tối thiểu là:

$$n_0 = [(\frac{\sigma}{\varepsilon_0} u_{\alpha/2})^2] + 1.$$

trong đó [ ] là ký hiệu phần nguyên.

Ví dụ 8:

(xét Ví dụ 5) nếu cho trước độ chính xác là 0,05 thì cần phải lấy bao nhiêu mẫu điều tra. Ta có $\varepsilon_0 = 0,05;\ \sigma = 0,2;\ u_{0,025} = 1,96$.

Vậy cỡ mẫu tối thiểu cần phải lấy là:

$$n_0 = [(\frac{0,2}{0,05} 1,96)^2] + 1 = [61,465] + 1 = 62.$$

Tương tự trong trường hợp σ chưa biết ta có:

$$\varepsilon = |\overline{x} - \mu| < \frac{s'}{\sqrt{n}} t_{\alpha/2}^{n-1}.$$

Vậy nếu cho trước độ chính xác ε_0 thì cỡ mẫu tối thiểu là:

$$n_0 = [(\frac{s'}{\varepsilon_0} t_{\alpha/2}^{n-1})^2] + 1.$$

Ví dụ 9:

Xét Ví dụ 7, hãy xác định cỡ mẫu nếu biết độ chính xác là 0.05.

Ta có: $\varepsilon_0 = 0,05;\ s' = 0,137;\ t_{0,025}^{24} = 2,06$.

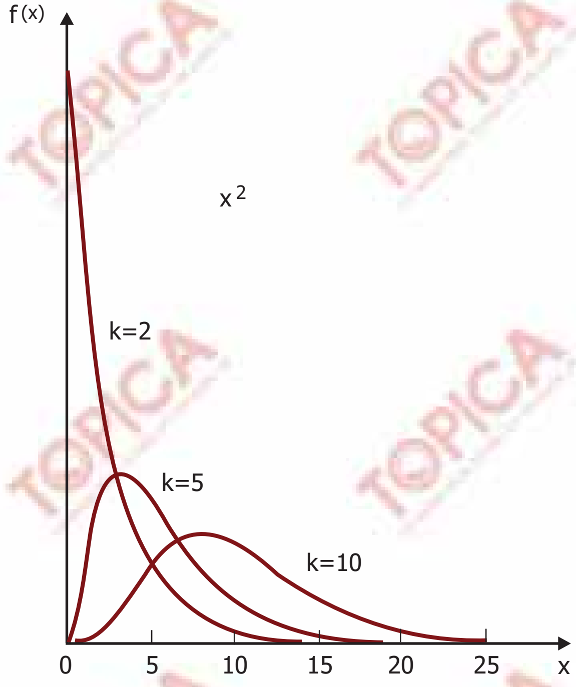
Vậy: $n_0 = [(\frac{0,137}{0,05} \times 2,06)^2] + 1 = [31,85] + 1 = 32.$

6.3.3. Ước lượng khoảng cho phương sai của biến ngẫu nhiên phân phối chuẩn

Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu;\sigma^2)$ và mẫu ngẫu nhiên $(X_1, X_2,\ldots,X_n)$ có giá trị mẫu $(x_1, x_2,\ldots,x_n)$. Khi đó thống kê:

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

có phân phối khi–bình phương với n–1 bậc tự do.



Hình 4: Đồ thị phân phối khi–bình phương với các bậc tự do khác nhau

Vậy với độ tin cậy $1-\alpha$ ta có:

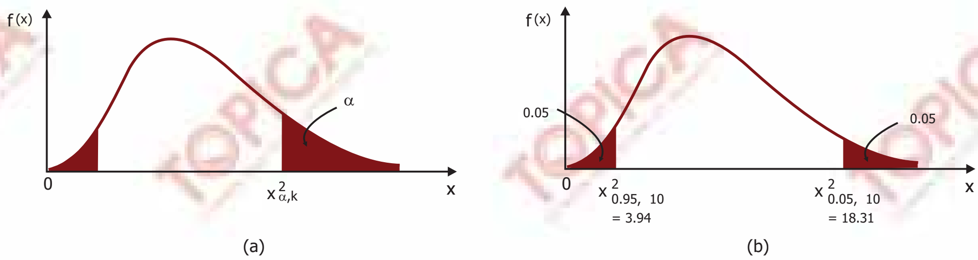
$$P\left\{\chi^2_{1-\alpha/2,n-1} < \chi^2 = \frac{(n-1)S'^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha$$

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

trong đó các điểm phân vị $\chi^2_{\alpha,k}$ được xác định bởi:

$$P\left\{\chi^2 > \chi^2_{\alpha,k}\right\} = \alpha$$

giá trị phân vị $\chi^2_{\alpha,k}$ được tìm từ bảng phân phối khi–bình phương.



Hình 5: Đồ thị phân phối Student với các bậc tự do khác nhau

Vậy với mẫu cụ thể ta có khoảng ước lượng hai phía của σ^2:

$$\sigma^2 \in \left(\frac{(n-1)s'^2}{\chi^2_{\alpha/2,n-1}};\ \frac{(n-1)s'^2}{\chi^2_{1-\alpha/2,n-1}}\right).$$

Tương tự ta có các ước lượng một phía của σ^2:

- Ước lượng giá trị tối thiểu:

$$\sigma^2 \in \left(\frac{(n-1)s'^2}{\chi^2_{\alpha,n-1}};\ +\infty\right)$$

- Ước lượng giá trị tối đa:

$$\sigma^2 \in \left(0;\ \frac{(n-1)s'^2}{\chi^2_{1-\alpha,n-1}}\right).$$

CHÚ Ý

Ngoài cách tra bảng, ta dùng lệnh trong Excel: chiinv(p,n-1). Tham khảo phần phụ lục

Ví dụ 10:

Kiểm tra ngẫu nhiên 20 bao gạo do một máy đóng bao tự động đóng, ta có phương sai hiệu chỉnh $s'^2 = 0,0153(kg)^2$. Hãy tìm ước lượng khoảng tối đa cho độ chính xác của trọng lượng các bao gạo với độ tin cậy 95%. Biết rằng trọng lượng các bao gạo do máy tự động đóng là biến ngẫu nhiên có phân phối chuẩn.

Giải:

Gọi X là trọng lượng bao gạo, $X \sim N(\mu;\sigma^2)$.

Ta có $s'^2 = 0,0153$, $1-\alpha = 0,95 \Rightarrow \alpha = 0,05$. Tra bảng phân phối khi bình phương ta có: $\chi^2_{0,95,19} = 10,117$.

Vậy ta có:

$$\sigma^2 \in \left(0;\ \frac{(20-1)s'^2}{\chi^2_{0,95,19}}\right) = \left(0;\ \frac{19\times 0,0153}{10,117}\right)$$

$$\Rightarrow \sigma^2 \in (0;\ 0,17).$$

CHÚ Ý

Nội dung của các mục 5.3.2 và 5.3.3 được trình bày với giả thiết biến ngẫu nhiên gốc có phân phối chuẩn. Với hiệu lực của Định lý Giới hạn trung tâm, giả thiết về tính phân phối chuẩn của biến ngẫu nhiên gốc có thể bỏ qua, nếu cỡ mẫu n "đủ lớn". Tuy nhiên, khái niệm "đủ lớn" của cỡ mẫu n phụ thuộc rất nhiều vào dạng phân phối của biến ngẫu nhiên gốc. Chẳng hạn nếu biến ngẫu nhiên gốc có phân bố lên tục và đối xứng thì với $n \geq 6$, các kết quả của mục 5.3.2 đã đúng, còn nếu biến ngẫu nhiên gốc có phân phối rời rạc và không đối xứng thì các kết quả đó vẫn chưa đúng thậm chí khi $n = 300$.

6.3.4. Ước lượng khoảng cho xác suất (tỷ lệ)

Cho biến cố A với xác suất xảy ra p chưa biết, thực hiện n lần thử của biến cố A, gọi m là số lần A xuất hiện. Ta có tần suất xuất hiện biến cố A là $f = \frac{m}{n}$. Theo lý thuyết xác suất, ta thấy thống kê:

$$U = \frac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$$

có phân phối xấp xỉ phân phối chuẩn N(0; 1) khi cỡ mẫu n đủ lớn. Với độ tin cậy $1-\alpha$ ta có:

$$P\left\{-u_{\alpha/2} < U = \frac{f-p}{\sqrt{f(1-f)}}\sqrt{n} < u_{\alpha/2}\right\} = 1-\alpha$$

$$P\left\{f - \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha/2} < p < f + \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha/2}\right\} = 1-\alpha.$$

Vậy ta có khoảng ước lượng hai phía của p là:

$$p \in (f - \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha/2};\ f + \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha/2}),$$

trong đó phân vị $u_{\alpha/2}$ tìm từ bảng phân phối chuẩn.

Tương tự ta xác định khoảng ước lượng một phía của p như sau:

- Ước lượng giá trị tối thiểu:

$$p > f - \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha}.$$

- Ước lượng giá trị tối đa:

$$p < f + \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha}.$$

Ví dụ 11:

Kiểm tra ngẫu nhiên 100 sản phẩm do một nhà máy sản xuất thấy có 12 phế phẩm. Với độ tin cậy 95% , hãy ước lượng tỷ lệ phế phẩm của nhà máy đó.

Giải:

Gọi p là tỷ lệ phế phẩm của nhà máy, ta có n = 100; m = 12; f = m/n = 12/100 = 0,12. Vậy: $1-\alpha = 0,95 \Rightarrow \alpha/2 = 0,025$, tra bảng phân phối chuẩn ta có $u_{0,025} = 1,96$. Từ đó khoảng ước lượng của p là:

$$p \in \left(0,12 - \frac{\sqrt{0,12\times0,88}}{\sqrt{100}}1,96;\ 0,12 + \frac{\sqrt{0,12\times0,88}}{\sqrt{100}}1,96\right)$$

$$\Rightarrow p \in (0,056;\ 0,184).$$

- Xác định cỡ mẫu

  Trong thực hành, trước khi tiến hành thu thập số liệu, người ta cần xác định cỡ mẫu tối thiểu đối với mỗi nghiên cứu để đáp ứng được mục tiêu nghiên cứu đồng thời thỏa mãn các điều kiện thực tế. Một trong những cách xác định cỡ mẫu có thể trình bày như sau:

Xét khoảng ước lượng hai phía cho p, ta có:

$$\varepsilon = |p - f| < \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{\alpha/2}.$$

Khi n càng lớn thì độ sai lêch giữa p và tần suất f càng nhỏ, ε được gọi là độ chính xác của ước lượng. Nếu cho trước độ chính xác là ε_0 khi đó cỡ mẫu tối thiểu cần có là

$$n_0 = \left[\left(\frac{\sqrt{f(1-f)}}{\varepsilon_0} u_{\alpha/2} \right)^2 \right] + 1.$$

Ví dụ 12:

Xét bài toán trong Ví dụ 11. Nếu muốn độ chính xác là 0,03 thì phải lấy tối thiểu bao nhiêu sản phẩm để kiểm tra?

Giải:

Ta có độ chính xác $\varepsilon_0 = 0,03$; $u_{0,025} = 1,96$.

Vậy số sản phẩm tối thiểu cần lấy ra kiểm tra

$$n_0 = \left[\left(\frac{\sqrt{0,12 \times 0,88}}{0,03} \times 1,19 \right)^2 \right] + 1 = [450,75 + 1]$$

$$\Rightarrow n_0 = 451.$$

TÓM LƯỢC CUỐI BÀI

Trong bài này các bạn cần nắm vững khái niệm về bài toán ước lượng tham số và phương pháp tìm ước lượng khoảng cho các tham số kỳ vọng, phương sai và khoảng ước lượng cho tỷ lệ của tổng thể, cách xác định cỡ mẫu khi cho biết trước độ chính xác của ước lượng. Chú ý phần ước lượng khoảng cho kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trong các trường hợp phương sai đã biết và chưa biết và cách tìm các giá trị phân vị bằng cách tra bảng hoặc sử dụng các lệnh trong phần mềm Excel.

BÀI TẬP TRẮC NGHIỆM

1. Cho mẫu ngẫu nhiên (x_1, x_2, …, x_{25}) rút ra từ biến ngẫu nhiên X có:

s' = 2,4 và $\sum_{k=1}^{25} x_k^2 = 165$.

Ta có trung bình mẫu $\overline{x}$ là:

a. 1,036

b. 1,035

c. 1,034

d. 1,033

e. 1,045

2. Cho (X_1, X_2, …,X_n) là mẫu ngẫu nhiên rút ra từ biến ngẫu nhiên X có phân phối đều trên đoạn $[0;\theta]$, ước lượng điểm của tham số θ là $\theta = 2\overline{X}$.

Khi đó phương sai của $\hat{\theta}$ là:

a. 0

b. $\dfrac{\theta^2}{12}$

c. $\dfrac{\theta^2}{3}$

d. $\dfrac{\theta^2}{12n}$

e. $\dfrac{\theta^2}{3n}$

3. Cho X là biến ngẫu nhiên có phân phối chuẩn $N(\mu,\sigma^2)$, một mẫu ngẫu nhiên (x_1, x_2,..,x_{25}) của X có các giá trị sau: $\sum_{k=1}^{25} x_k = 175$; $\sum_{k=1}^{25} x_k^2 = 1550$.

Khi đó ước lượng khoảng hai phía cho trung bình μ với độ tin cậy 95% là:

a. $7 \pm 0,7360 t_{0,025}^{(24)}$

b. $7 \pm 0,7360 t_{0,025}^{(25)}$

c. $7 \pm 0,5417 t_{0,025}^{(24)}$

d. $7 \pm 0,5417 u_{0,025}$

4. Cho biến ngẫu nhiên $X \sim N(\mu;4)$ và một mẫu ngẫu nhiên cỡ n = 35, $\overline{x} = 15, s^{'} = 2,5$, với độ tin cậy 95%, khi đó ước lượng khoảng hai phía phương sai của X là:

a. $(0,280;0,125)$
b. (0,290; 0,125)
c. 2
d. (0,280; 0,135)
e. (0,290; 0,135)

5. Cho biến ngẫu nhiên $X \sim N(\mu;\sigma^2)$ và một mẫu ngẫu nhiên cỡ n = 25, $s' = 1,5$, với độ tin cậy 95%, khi đó ước lượng khoảng hai phía phương sai của X là:

a. (0,045; 0,101)
b. (1,157; 2,596)
c. (1,045; 2,341)
d. (1,111; 2,493)
e. (1,115; 2,494)

6. Lấy ngẫu nhiên n sản phẩm trong kho và kiểm tra thấy có k phế phẩm. Với độ tin cậy 98% khoảng ước lượng cho tỷ lệ phế phẩm p là: (0,481; 0,519), khi đó tần suất mẫu về tỷ lệ phế phẩm là:

a. 0,51
b. 0,52
c. 0,53
d. 0,5
e. 0,54

7. Trong thời gian của một chuyến tàu chạy từ A đến B là biến ngẫu nhiên có phân phối chuẩn. Quan sát 25 chuyến tàu từ A đến B ta thu được số liệu:

Thời gian (h)	15 - 18,5	15,5 - 16	16 - 16,5	16,5 - 17
Số chuyến	2	8	13	2

Với độ tin cậy 95%, hãy ước lượng thời gian trung bình tàu chạy từ A đến B?

8. Trong một báo cáo về tỷ lệ ủng hộ ứng cử viên A cho chức tổng thống người ta đã thăm dò 1500 cử tri khẳng định rằng tỷ lệ đó nằm trong khoảng (0,519; 0,581). Khi đó thì số người đã ủng hộ cho ứng cử viên A là bao nhiêu?

a. 850
b. 840
c. 830
d. 825
e. 820

9. Với số liệu đã cho như trong bài tập 8, người ta đã khẳng định tỷ lệ số người ủng hộ ứng cử viên A với độ tin cậy là bao nhiêu?

a. 99%
b. 98%
c. 97%
d. 96%
e. 95%

10. Trong một cuộc điều tra về chiều cao của nam thanh niên Việt nam trong những năm gần đây, người ta đã kết luận rằng chiều cao trung bình trong khoảng (165,84; 172,82). Hãy cho biết trung bình mẫu điều tra là bao nhiêu, biết rằng chiều cao của nam thanh niên là biến ngẫu nhiên có phân phối chuẩn.

a. 169,13
b. 169,23
c. 169,33
d. 169,43
e. 169,73

11. Xét các số liệu cho trong bài tập 10, biết rằng phương sai mẫu hiệu chỉnh là 87,47 và người ta đã điều tra một mẫu cỡ n = 30 nam thanh niên. Vậy kết luận đã khẳng định về chiều cao trung bình của nam thanh niên Việt nam với độ tin cậy là bao nhiêu?

a. 0,95
b. 0,96
c. 0,97
d. 0,98
e. 0,9

12. Cho biến ngẫu nhiên $X \sim N(\mu;16)$ và một mẫu ngẫu nhiên cỡ n, với độ tin cậy 95% khoảng ước lượng cho kỳ vọng là (15,63; 17,59). Khi đó cỡ mẫu n phải là:

a. 64
b. 65
c. 66
d. 74
e. không xác định được

13. Gieo thử 400 hạt giống thì thấy rằng có 20 hạt không nảy mầm. Khi đó ước lượng khoảng đối xứng cho tỷ lệ hạt giống nảy mầm với độ tin cậy 95% là:

a. (0,925; 0,976)
b. (0,929; 0,971)
c. (0,929; 0,984)
d. (0,925; 0,984)

14. Cho biến ngẫu nhiên $X \sim N(\mu;9)$ với mẫu ngẫu nhiên cỡ n = 99 thì ước lượng khoảng cho μ với độ tin cậy 95% là $10 \pm \frac{3}{\sqrt{99}} u_{0,025}$. Nếu ta thêm quan sát thứ x_{100} = 11 thì ước lượng khoảng cho μ bây giờ sẽ là:

a. $10 \pm 0,3u_{0,025}$

b. $10,01 \pm 0,3u_{0,025}$

c. $10 \pm 0,299u_{0,025}$

d. $10,01 \pm 0,299u_{0,025}$

e. không có câu trả lời

15. Giả sử X là biến ngẫu nhiên có phân phối đều trên $[0,\theta], \overline{X}$ là trung bình mẫu của mẫu ngẫu nhiên cỡ n rút ra từ X. Giả sử θ có ước lượng điểm là $\hat{\theta} = \overline{X}$ khi đó $\hat{\theta}$ có phương sai là:

a. $\dfrac{\theta^2}{12n}$

b. $\dfrac{\theta^2}{12n} + \dfrac{\theta^2}{4}$

c. $\dfrac{\theta^2}{3n}$

d. $\dfrac{\theta^2}{12}$

e. $\dfrac{\theta^2}{12} + \dfrac{\theta^2}{4n}$

16. Cho biến ngẫu nhiên $X \sim N(\mu, \mu^2)$ khi đó với độ tin cậy 95% khoảng ước lượng cho μ dựa trên mẫu ngẫu nhiên (X_1, X_2, …, X_n) là:

a. $\overline{X} \pm (\mu)u_{\alpha/2}$

b. $\overline{X} \pm (\dfrac{\mu}{\sqrt{n}})u_{\alpha/2}$

c. $\overline{X} \pm \left(\dfrac{|\overline{X}|}{\sqrt{n}}\right)u_{\alpha/2}$

d. $\overline{X} \pm \left(\dfrac{|\overline{X}|}{n}\right)u_{\alpha/2}$

e. $\overline{X} \pm \left(\dfrac{|\overline{X}|}{\sqrt{n}}\right)u_{\alpha/2}$

BÀI TẬP

1. Điều tra 30 sinh viên về điểm thi môn xác suất - thống kê thu được số liệu sau:

Điểm thi	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10
Số sinh viên	5	3	10	8	4

Biết rằng điểm thi của sinh viên là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 0,5.

a. Hãy ước lượng điểm thi trung bình của sinh viên với độ tin cậy 95%.

b. Nếu muốn độ chính xác của ước lượng là 0,2 thì cần phải điều tra bao nhiêu sinh viên?

2. Xét số liệu trong bài tập 1 nếu biết rằng điểm thi của sinh viên là biến ngẫu nhiên có phân phối chuẩn.

a. Hãy ước lượng điểm thi của sinh viên với độ tin cậy là 90%.

b. Nếu muốn độ chính xác của ước lượng là 0,1 thì cần điều tra bao nhiêu sinh viên?

3. Để ước lượng lượng gỗ trong một khu rừng người ta tiến hành khai thác 50 cây gỗ và đo được số liệu sau, với độ tin cậy 95%

Khối lượng gỗ (m^3)	20	40	60	80
Số cây	10	15	17	8

a. Tìm ước lượng không chệch cho trung bình và phương sai

b. Ước lượng trữ lượng gỗ trung bình.

c. Ước lượng gỗ tối đa.

Biết rằng khối lượng gỗ là biến ngẫu nhiên có phân phối chuẩn.

4. Lượng điện sử dụng của các hộ gia đình là biến ngẫu nhiên có phân phối chuẩn, điều tra 30 hộ gia đình thu được số liệu sau:

Lượng điện tiêu thụ (kw/h)	150-160	160-170	170-180	180-190
Số hộ gia đình	5	11	10	4

Hãy tìm ước lượng không chệch của :

a. Lượng điện tiêu thụ trung bình.

b. Phương sai của lượng điện tiêu thụ.

5. Xét số liệu đã cho trong bài tập 4, với độ tin cậy 95%. Hãy ước lượng lượng điện tiêu thụ trung bình hàng tháng của các hộ gia đình.

6. Xét số liệu trong bài tập 4, với độ tin cậy 90%. Hãy ước lượng độ phân tán của lượng điện tiêu thụ.

7. Thời gian của một chuyến tàu chạy từ A đến B là biến ngẫu nhiên có phân phối chuẩn. Quan sát 25 chuyến tàu từ A đến B ta thu được số liệu:

Thời gian (h)	15 - 18,5	15,5 - 16	16 - 16,5	16,5 - 17
Số chuyến	2	8	13	2

Với độ tin cậy 95%, hãy ước lượng thời gian trung bình tàu chạy từ A đến.

8. Xét số liệu trong bài tập 7, chuyến tàu chạy từ A đến B gọi là đạt chuẩn nếu thời gian chạy nhỏ hơn 16 h. Hãy ước lượng tỷ lệ số chuyến đạt chuẩn với độ tin cậy 95%.

9. Gieo thử 400 hạt giống thì thấy rằng có 20 hạt không nảy mầm. Hãy ước lượng tỷ lệ hạt giống nảy mầm với độ tin cậy 95%.

10. Trong đợt vận động bầu cử tổng thống người ta đã phỏng vấn 1600 cử tri thì được biết 960 người sẽ ủng hộ ứng cử viên A. Với độ tin cậy 90% hãy ước lượng tỷ lệ tối đa số phiếu ủng hộ ứng cử viên A.

11. Theo dõi mức tiêu hao năng lượng của một loại nhiên liệu ta thu được số liệu sau:

Lượng tiêu hao (lít)	11-13	13-15	15-17	17-19	19-21
	8	5	13	10	9

Giả sử mức tiêu hao nhiên liệu của động cơ là biến ngẫu nhiên có phân phối chuẩ. Với độ tin cậy là 93% hãy ước lượng mức tiêu hao tối đa của loại động cơ trên.

12. Theo dõi thời gian gia công một chi tiết máy của 25 công nhân ta thu được số liệu sau:

Thời gian (phút)	16-18	18 -20	20-22	22-24	24-26
Số công nhân	2	6	7	6	4

Hãy ước lượng tối đa phương sai của thời gian gia công chi tiết với độ tin cậy 97%, biết rằng thời gian gia công chi tiết là biến ngẫu nhiên có phân phối chuẩn.

13. Xét số liệu trong bài tập số 12. Công nhân được gọi là lao động giỏi nếu thời gian gia công chi tiết nhỏ hơn 20 phút. Hãy ước lượng tối đa tỷ lệ công nhân lao động giỏi với độ tin cậy là 96%.

14. Kiểm tra 36 gói đường loại 500g của một máy đóng gói tự động thu được số liệu:

Trọng lượng (g)	495	497	499	501	503
Số gói	4	8	10	9	5

Giả thiết trọng lượng gói đường là biến ngẫu nhiên có phân phối chuẩn.

a. Hãy ước lượng trọng lượng trung bình các gói đường với độ tin cậy 92 %

b. Nếu muốn độ chính xác của ước lượng là 0,3% thì cần cân thử bao nhiêu gói đường?

15. Xét số liệu trong bài tập 14. Hãy ước lượng tỷ lệ các gói đường có trọng lượng trong khoảng [79;501) với độ tin cậy là 98%.