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 trọng 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:

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

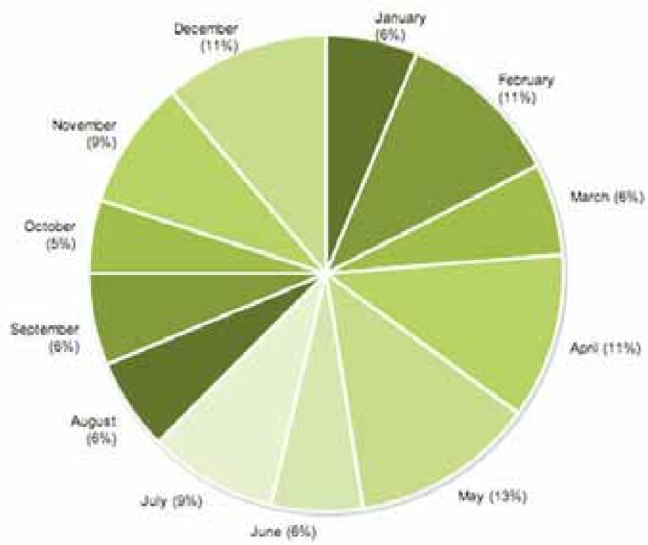
$$E(S'^2) = E\left[\frac{n}{n-1}S^2\right] = \frac{n}{n-1}E(S^2) = \sigma^2$$

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:

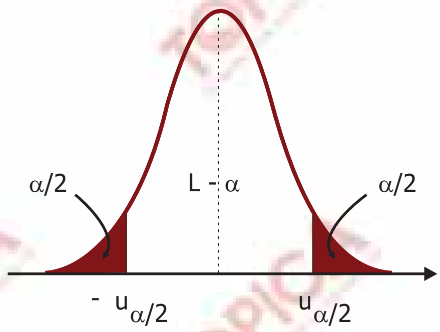
$$P\left\{-u_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < u_{\alpha/2}\right\} = 1-\alpha$$

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

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 $u_{\alpha/2}$.

CHÚ Ý

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



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:

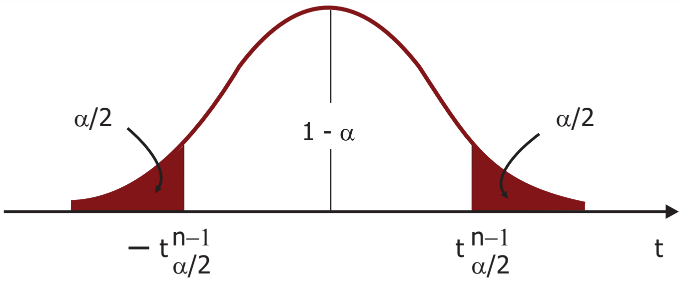
$$\mu \in (-\infty; 11,672 + \frac{0,2}{\sqrt{25}}2,33) = (-\infty; 11,765).$$

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:

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

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

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

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:

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

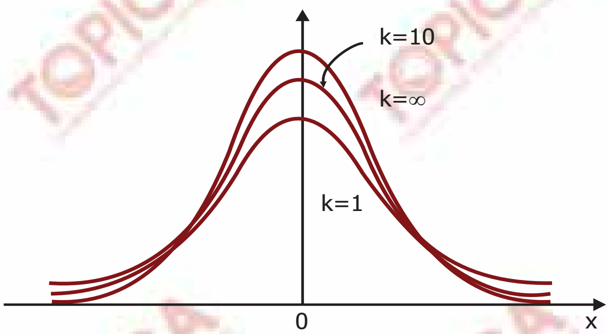
$$\mu \in \left(11,672 - \frac{0,137}{\sqrt{25}}1,71;\; +\infty) = (1,625;\; +\infty\right).$$

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:

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

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

Ta thấy rằng khi cỡ mẫu càng lớn thì độ sai lệch 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.