Chương 6. Lý thuyết ước lượng

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin
Đại học Quốc gia Thành phố Hồ Chí Minh

Mục tiêu

- Biết được khái niệm ước lượng điểm và ước lượng khoảng.
- Tính được khoảng ước lượng của trung bình tổng thể.
- Tính được khoảng ước lượng của tỉ lệ tổng thể.

6.1 Ước lượng điểm

Ước lượng điểm là gì?

- Các giá trị trung bình, phương sai, độ lệch chuẩn và trung vị của tổng thể được gọi là các **tham số** (parameters).
- Các giá trị trung bình, phương sai, độ lệch chuẩn và trung vị của mẫu được gọi là các **thống kê** (statistics).

Định nghĩa 6.1

1. Một ước lượng điểm là một giá trị dùng để ước lượng một tham số.
2. Một ước lượng khoảng là một khoảng giá trị dùng để ước lượng một tham số.

Ví dụ 6.2

- Nếu nói chiều cao trung bình của sinh viên nam trường Đại học Công nghệ Thông tin là 174 cm thì đó là một **giá trị ước lượng điểm.**
- Nếu nói chiều cao trung bình đó nằm trong khoảng từ 159 cm đến 169 cm hay 164 ± 5 cm thì ta đã có một ***ước lượng khoảng.***

6.1 Ước lượng điểm

Định nghĩa 6.3 **Ước lượng điểm** (point estimator) của tham số tổng thể là một biến ngẫu nhiên phụ thuộc vào thông tin mẫu, giá trị của nó cho ta một sự xấp xỉ của tham số chưa biết này. Một giá trị cụ thể của biến ngẫu nhiên đó được gọi là **giá trị ước lượng điểm** (point estimate).

Ví dụ 6.4

- Trung bình mẫu ngẫu nhiên $\overline{X}$ là ước lượng điểm (point estimator) của trung bình tổng thể μ.
- Độ lệch chuẩn mẫu ngẫu nhiên S^2 là ước lượng điểm của độ lệch chuẩn tổng thể σ^2.
- Tỷ lệ mẫu ngẫu nhiên $\hat{p}$ là ước lượng điểm của tỷ lệ tổng thể p.

Các giá trị cụ thể của $\overline{X}, S^2, \hat{p}$ được gọi là các giá trị ước lượng điểm (point estimate).

6.1 Ước lượng điểm

Ký hiệu

θ : tham số của tổng thể mà ta quan tâm

$\hat{\theta}$: thống kê mẫu hoặc ước lượng điểm của θ

Định nghĩa 6.5 Thống kê mẫu $\hat{\theta}$ được gọi là một ước lượng không lệch (unbiased estimator) của tham số tổng thể θ nếu

$$E(\hat{\theta}) = \theta.$$

Ví dụ 6.6 Trong chương 5, ta đã có:

- Trung bình mẫu ngẫu nhiên $\overline{X}$ là ước lượng không lệch của trung bình tổng thể μ vì $E(\overline{X}) = \mu$.
- Tỉ lệ mẫu ngẫu nhiên $\hat{p}$ ước lượng không lệch của tỷ lệ tổng thể p vì $E(\hat{p}) = p$.

6.2 Ước lượng khoảng

6.2 Ước lượng khoảng

- Vì một ước lượng điểm không thể cung cấp chính xác giá trị của tham số tổng thể nên ta thường dùng ước lượng khoảng (interval estimate).
- Ước lượng khoảng cung cấp thông tin về mức độ gần của ước lượng điểm do mẫu cung cấp với giá trị của tham số tổng thể.
- Ước lượng khoảng được xây dựng sao cho khi lấy mẫu lặp lại nhiều lần thì một tỷ lệ lớn các khoảng này sẽ bao quanh tham số tổng thể mà chúng ta đang quan tâm. Tỷ lệ này là độ tin cậy (confidence level) và khoảng được tạo ra được gọi là khoảng tin cậy (confidence interval).

6.2 Ước lượng khoảng

Định nghĩa 6.7

1. **Độ tin cậy** (confidence level), ký hiệu $1-\alpha$, của ước lượng khoảng của một tham số là xác suất khoảng ước lượng chứa tham số. Giả sử một số lượng lớn mẫu được chọn và quá trình ước lượng trên cùng một tham số được lặp lại.
2. **Khoảng tin cậy** (confidence interval) là một khoảng ước lượng cụ thể của một tham số tương ứng với độ tin cậy đã cho.

Bài toán. Gọi θ là tham số mà ta quan tâm. Tìm khoảng ước lượng của θ với độ tin cậy $1-\alpha$. Tức là, ta cần tìm một đoạn $[a, b]$ sao cho

$$P(a \leq \theta \leq b) = 1 - \alpha.$$

6.2.1 Ước lượng khoảng cho trung bình tổng thể khi biết σ

Cho $x_1, x_2, \ldots, x_n$ là một mẫu ngẫu nhiên có kích thước n được lấy từ một tổng thể có phân phối chuẩn $N(\mu, \sigma^2)$ trong đó **đã biết** σ. Giả sử ta cần tìm khoảng tin cậy $1 - \alpha$ của trung bình tổng thể.

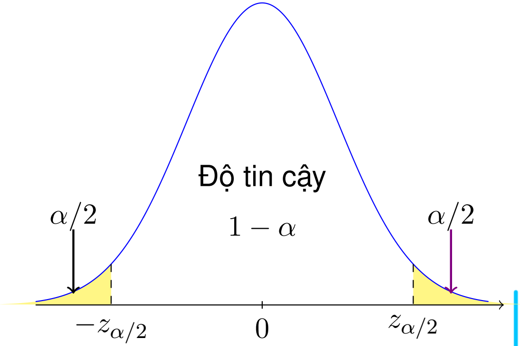
- Từ mẫu đã cho, ta tính được trung bình mẫu $\overline{x}$.
- Cần tìm ε sao cho

$$P(\overline{x} - \varepsilon \leq \mu \leq \overline{x} + \varepsilon) = 1 - \alpha.$$

$$P(-\varepsilon \leq \overline{x} - \mu \leq \varepsilon) = 1 - \alpha$$

$$P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} \leq \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

- Trong chương 5, ta đã biết biến ngẫu nhiên
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$
có phân phối chuẩn chuẩn tắc.
- Đặt $z_{\alpha/2} = \dfrac{\varepsilon}{\sigma/\sqrt{n}}$ và cần tìm $z_{\alpha/2}$ sao cho
$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$
- Với độ tin cậy $1 - \alpha$ thì giá trị $z_{\alpha/2}$ thỏa mãn
$$\Phi(z_{\alpha/2}) = 1 - \alpha/2.$$



- Với độ tin cậy $1-\alpha$ thì giá trị $z_{\alpha/2}$ thỏa mãn

$$\Phi(z_{\alpha/2}) = 1 - \alpha/2.$$

- Tìm được $z_{\alpha/2}$ bằng bảng phụ lục A4.
- Trung bình tổng thể sẽ thuộc khoảng (khoảng tin cậy của trung bình tổng thể với độ tin cậy $1-\alpha$)

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- Số $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là **sai số của ước lượng** hoặc **độ chính xác** của ước lượng.

Ví dụ 6.8 Giả sử rằng thời gian mua sắm của khách hàng tại một trung tâm thương mại có phân phối chuẩn với độ lệch chuẩn tổng thể là 20 phút. Chọn ngẫu nhiên 64 người đã mua sắm ở trung tâm đó. Người ta thấy rằng thời gian mua sắm trung bình của 64 người này là 75 phút. Tìm thời gian mua sắm trung bình của khách hàng tại trung tâm này với độ tin cậy 95%.

Giải.

- Ta có $\overline{x} = 75; n = 64$ và $\sigma = 20$;
- Độ tin cậy $1 - \alpha = 95\%$. Suy ra $\alpha = 0,05$ và $z_{\alpha/2} = 1,96$.
- Độ chính xác $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1,96\dfrac{20}{\sqrt{64}} = 4,9$.
- Khoảng tin cậy của trung bình tổng thể với độ tin cậy 95% là

$$[75 - 4,9; 75 + 4,9].$$

6.2.2 Ước lượng khoảng cho trung bình tổng thể khi chưa biết σ

Trường hợp 1: Kích thước mẫu $n \geq 30$.

1. $\overline{x}, s$ là trung bình và độ lệch chuẩn của một mẫu đang có.
2. $\overline{X}, S$ là trung bình và độ lệch chuẩn của một mẫu ngẫu nhiên.
3. Biến ngẫu nhiên $Z = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ được xem là có phân phối chuẩn chuẩn tắc $Z \sim N(0;1)$.
4. Trong phụ lục A4, tìm $z_{\alpha/2}$.
5. Khoảng tin cậy của μ với độ tin cậy $1 - \alpha$ là

$$\left[\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right].$$

6.2.2 Ước lượng khoảng cho trung bình tổng thể khi chưa biết σ

Trường hợp 2: Kích thước mẫu $n < 30$ và tổng thể có phân phối chuẩn

1. Biến ngẫu nhiên $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ có phân phối Student với bậc tự do $n - 1$.
2. Trong phụ lục A5, dòng $n - 1$, tìm $t_{\alpha/2}$ thỏa mãn $P(t > t_{\alpha/2}) = \dfrac{\alpha}{2}$.
3. Khoảng tin cậy của μ với độ tin cậy $1 - \alpha$ là

$$\left[\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right].$$

Ví dụ 6.9 Theo một thống kê cho thấy số thu nhập của 7 công nhân trong năm 2021 của một công ty được cho như sau (đơn vị triệu đồng)

109,2   118   121,8   126,2   143,2   168,8   198,6

Giả sử thu nhập trong năm 2021 của công ty có phân phối chuẩn. Tính khoảng thu nhập trung bình các công nhân của công ty này với độ tin cậy 99%.

Giải.

- Trung bình mẫu: $\overline{x} = 140,829$
- Độ lệch chuẩn mẫu: $s = 32,205$
- Tìm $t_{\alpha/2}$ với độ tin cậy $1 - \alpha = 0,99$ và bậc tự do 6. Ta có $t_{\alpha/2} = 3,707$.
- Khoảng tin cậy cần tìm

$$\left[\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right] \text{ hay } [95,706; 185,952]$$

Ví dụ 6.10 Kiểm tra tuổi thọ (tính bằng giờ) của 50 bóng đèn do nhà máy A sản xuất, người ta được bảng số liệu sau

Tuổi thọ	(3350;3450]	(3450;3550]	(3550;3650]	(3650;3750]
Số bóng đèn	10	20	12	8

a. Ước tính tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất với độ tin cậy 97%.

b. Dựa vào mẫu trên để ước lượng tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất có độ chính xác 29 giờ thì phải đảm bảo độ tin cậy là bao nhiêu?

Tuổi thọ	(3350;3450]	(3450;3550]	(3550;3650]	(3650;3750]
Số bóng đèn	10	20	12	8

	A	B
	Tuoi	Sobong
	3400	10
	3500	20
	3600	12
	3700	8

```
> d <- read.table("bongden.csv", header=TRUE, sep=",")
> d2 <- rep(d$Tuoi, d$Sobong)
> multi.fun <- function(x){c(mean=mean(x), var=var(x), median=median(x),
+  sd=sd(x))}
> multi.fun(d2)
      mean        var     median         sd
3536.00000 9697.95918 3500.00000   98.47822
> |
```

Tuổi thọ	(3350;3450]	(3450;3550]	(3550;3650]	(3650;3750]
Số bóng đèn	10	20	12	8

Giải. a. Ước tính tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất với độ tin cậy 97%. (Kích thước mẫu bằng 50 và chưa biết độ lệch chuẩn tổng thể)

- Trung bình mẫu: $\overline{x} = 3536$ (giờ)
- Độ lệch chuẩn mẫu: $s = 98,478$ (giờ)
- Độ tin cậy $1 - \alpha = 0,97$. Suy ra $\Phi(z_{\alpha/2}) = 1 - \alpha/2 = 0,985$. Do đó $z_{\alpha/2} = 2,17$.
- Độ chính xác:

$$z_{\alpha/2}\frac{s}{\sqrt{n}} = 2,17\frac{98,478}{\sqrt{50}} = 30,221.$$

- Khoảng tin cậy của tuổi thọ trung của bóng đèn với độ tin cậy 97% là

$$[3505,779; 3566,221].$$

b. Dựa vào mẫu trên để ước lượng tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất có độ chính xác 29 giờ thì phải đảm bảo độ tin cậy là bao nhiêu?

Giải. b. Ta có độ chính xác bằng 29 giờ, tức là

$$z_{\alpha/2}\frac{s}{\sqrt{n}} = 29$$

Suy ra

$$z_{\alpha/2} = 29\frac{\sqrt{n}}{s} = 2,08$$

Do đó

$$\Phi(z_{\alpha/2}) = \Phi(2,08) = 1 - \frac{\alpha}{2}.$$

Tìm trong phụ lục A4, ta có $\Phi(2,08) = 0,9812$ và do đó $\alpha = 0,0376.$
Như vậy, độ tin cậy là $1 - \alpha = 0,9624.$

6.2.3 Ước lượng tỉ lệ của tổng thể

Bài toán. Thăm dò ý kiến của 100 người dân được chọn ngẫu nhiên tại một thành phố cho thấy có 80% trong số này ủng hộ ứng viên A. Với độ tin cậy 98%, hãy ước lượng tỉ lệ của tất cả các người dân ủng hộ ứng viên A tại thành phố này.

- p : tỉ lệ tổng thể (tỉ lệ phần tử có tính chất $\mathcal{P}$ trong tổng thể)
- $\hat{p}$: tỉ lệ mẫu (tỉ lệ phần tử có tính chất $\mathcal{P}$ trong mẫu)
- Nếu $np, n(1-p) \geq 5$ thì tỉ lệ mẫu ngẫu nhiên $\hat{p}$ có phân phối chuẩn $N(p; \sigma_{\hat{p}}^2)$ với $\sigma_{\hat{p}}^2 = \dfrac{p(1-p)}{n}$.
- Biến ngẫu nhiên $\dfrac{\hat{p}-p}{\sigma_{\hat{p}}}$ có phân phối chuẩn chuẩn tắc.

Từ một mẫu đã chọn, khoảng ước lượng của p có dạng $[\hat{p}-\varepsilon; \hat{p}+\varepsilon]$. Khi đó

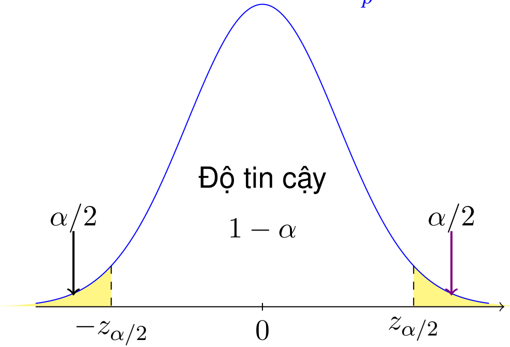
$$P(\hat{p}-\varepsilon \le p \le \hat{p}+\varepsilon) = 1-\alpha$$

$$P(-\varepsilon \le \hat{p}-p \le \varepsilon) = 1-\alpha$$

$$P\left(\frac{-\varepsilon}{\sigma_{\hat{p}}} \le \frac{\hat{p}-p}{\sigma_{\hat{p}}} \le \frac{\varepsilon}{\sigma_{\hat{p}}}\right) = 1-\alpha.$$

Đặt $z_{\alpha/2} = \dfrac{\varepsilon}{\sigma_{\hat{p}}}$

Phân phối của $\dfrac{\hat{p}-p}{\sigma_{\hat{p}}}$



- Vì p chưa biết (ta cần ước lượng) nên khi n đủ lớn, ta có thể thay p trong $\sigma_{\hat{p}}$ bởi giá trị của tỉ lệ mẫu $\hat{p}$.
- Với độ tin cậy $1-\alpha$, khoảng tin cậy chứa tỉ lệ tổng thể là

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

trong đó $\Phi(z_{\alpha/2}) = 1 - \dfrac{\alpha}{2}$ (xem phụ lục A4).

- Độ chính xác (sai số) là $z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

Ví dụ 6.11 Thăm dò ý kiến của 100 người dân được chọn ngẫu nhiên tại một thành phố cho thấy có 80% trong số này ủng hộ ứng viên A. Với độ tin cậy 98%, hãy ước lượng tỉ lệ của tất cả các người dân ủng hộ ứng viên A tại thành phố này.

Giải. Theo đề bài

- Tỉ lệ mẫu $\hat{p} = 0,8$
- Kích thước mẫu $n = 100$
- Độ tin cậy $1 - \alpha = 0,98$ suy ra $1 - \dfrac{\alpha}{2} = 0,99$ Do đó $z_{\alpha/2} = 2,33$
- Độ chính xác (sai số) là

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2,33\frac{0,4}{10} = 0,0932.$$

- Khoảng tin cậy $[0,7068; 0,8932]$.

Như vậy có từ 70,68% đến 89,32% người dân ủng hộ ứng viên A.

Theo bất đẳng thức Cauchy, ta có độ chính xác (sai số)

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

Do đó, sai số tối đa trong ước lượng tỉ lệ tổng thể là $\dfrac{z_{\alpha/2}}{2\sqrt{n}}$.

Ví dụ 6.12 Lấy ngẫu nhiên 344 sản phẩm trong một kho hàng để kiểm tra thì thấy có 261 sản phẩm tốt.

a. Dựa vào mẫu trên, để ước tính tỉ lệ sản phẩm bị lỗi có độ chính xác là 0,059 thì độ tin cậy bằng bao nhiêu?

b. Nếu muốn ước lượng tỉ lệ sản phẩm bị lỗi với độ chính xác nhỏ hơn $0,02$ với độ tin cậy 95% thì cần kiểm tra ít nhất bao nhiêu sản phẩm?

Giải. a. Theo đề bài

- Tỉ lệ mẫu $\hat{p} = \frac{261}{344} = 0,759$
- Kích thước mẫu $n = 344$
- Độ chính xác là

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = z_{\alpha/2}\sqrt{\frac{0,759(1-0,759)}{344}} = 0,059.$$

- Suy ra $z_{\alpha/2} = 2,58$. Do đó độ tin cậy là 99%.

b. Nếu muốn ước lượng tỉ lệ sản phẩm bị lỗi với độ chính xác nhỏ hơn 0,02 với độ tin cậy 95% thì cần kiểm tra ít nhất bao nhiêu sản phẩm?

Giải. b. Độ chính xác nhỏ hơn $0,02$

$$\frac{z_{\alpha/2}}{2\sqrt{n}} < 0,02 \text{ hay } n > \frac{z_{\alpha/2}^2}{4.0,02^2}$$

Vì độ tin cậy $1-\alpha = 0,95$ nên $z_{\alpha/2} = 1,96$.

Do đó

$$n > \frac{1,96^2}{4.0,02^2} = 2401.$$

Như vậy, cần kiểm tra ít nhất 2402 sản phẩm.

Bài tập

Bài 6.1 Chọn ngẫu nhiên 30 người để kiểm tra thời gian sử dụng chiếc điện thoại di động đầu tiên. Người ta thấy rằng thời gian sử dụng trung bình của 30 người này là 5,6 năm. Giả sử thời gian sử dụng chiếc điện thoại di động đầu tiên có phân phối chuẩn với độ lệch chuẩn là 0,8 năm. Tính khoảng thời gian trung bình sử dụng chiếc điện thoại đầu tiên với độ tin cậy 99%. (Đáp số: [5,22; 5,98])

Bài 6.2 Để ước tính trọng lượng trung bình của một loại sản phẩm, người ta chọn 26 sản phẩm và thu được kết quả như sau

Trọng lượng (gam)	190	195	198	200	204	205
Số sản phẩm	5	4	2	8	6	1

Giả sử trọng lượng sản phẩm có phân phối chuẩn. Hãy ước lượng trọng lượng trung bình của sản phẩm trên với độ tin cậy 95%. (Đáp số: [196,21; 200,33])

Bài tập

Bài 6.3 Một mẫu ngẫu nhiên gồm 100 cửa hàng mua máy in hiệu T cho thấy rằng 59 cửa hàng có kế hoạch gia tăng việc mua hàng của mình trong năm tới. Hãy ước lượng tỷ lệ các cửa hàng trong tổng thể tất cả các cửa hàng mua máy in hiệu T mà có kế hoạch gia tăng việc mua hàng của mình trong năm tới với độ tin cậy 95%.

Bài 6.4 Chọn ngẫu nhiên 500 laptop trong một kho chứa thì thấy có 27 laptop hiệu UIT.

a. Dựa vào mẫu trên, để ước lượng tỉ lệ laptop hiệu UIT trong toàn bộ kho có độ chính xác 0,0177 thì phải đảm bảo độ tin cậy bằng bao nhiêu?

b. Dựa vào mẫu trên, nếu muốn có độ chính xác của ước lượng tỉ lệ laptop UIT nhỏ hơn 0,01 với độ tin cậy 95% thì cần chọn ít nhất bao nhiêu laptop nữa?