6.1 PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

6.1.1 Khái niệm ước lượng điểm

Phương pháp ước lượng điểm chủ trương dùng một giá trị để thay cho giá trị của tham số θ chưa biết của tổng thể. Thông thường giá trị được chọn này là giá trị cụ thể của một thống kê $\boldsymbol{\Theta}$ nào đó của mẫu ngẫu nhiên

Với mẫu ngẫu nhiên $W=(X_1,X_2, \dots , X_n)$, thống kê ước lượng cho tham số θ có dạng

$$\boldsymbol{\Theta} = T\left(X_1, X_2, \dots, X_n\right)$$

Khi đó với mẫu cụ thể $w = (x_1,x_2, \dots , x_n)$, giá trị cụ thể của thống kê $\Theta = T(x_1,x_2, \dots , x_n)$, ước lượng cho tham số θ

6.1.2 Một số loại ước lượng điểm

6.1.2.1 Ước lượng không chệch (unbiased estimator)

Thống kê $\Theta = T(X_1, X_2, ..., X_n)$ được gọi là ước lượng không chệch của θ nếu với mọi giá trị của tham số θ

$$\mathrm{E}\left[\Theta\left(X_1, X_2, ..., X_n\right)\right] = \theta$$

6.1.2.2 Ước lượng hiệu quả (efficient estimator)

Ước lượng không chệch có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên gọi là ước lượng hiệu quả (hay ước lượng phương sai bé nhất)

Bất đẳng thức Cramer-Rao sau đây xác định một cận dưới các phương sai của các ước lượng không chệch, đó là một dấu hiệu để xét xem ước lượng không chệch có phải là ước lượng hiệu quả hay không

Dấu hiệu nghiên cứu là biến ngẫu nhiên X có hàm mật độ xác suất (hay hàm khối lượng xác suất) $f(x,\theta)$ và Θ là ước lượng không chệch bất kỳ của θ thì

$$\mathrm{D}\left(\Theta\right) \geq \frac{1}{n\,\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\theta)\right)}{\partial\theta}\right)^2}$$

Dựa vào bất đẳng thức Cramer-Rao ta có thể chứng minh được rằng trung bình mẫu là ước lượng hiệu quả của kỳ vọng μ của dấu hiệu nghiên cứu X của tổng thể có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

Ta có $\mathrm{D}\left(\overline{X}\right)=\dfrac{\sigma^2}{n}$

Mặt khác hàm mật độ của X có dạng $f(x,\mu)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\Rightarrow \ln f(x,\mu)=-\ln\left(\sigma\sqrt{2\pi}\right)-\frac{(x-\mu)^2}{2\sigma^2}\Rightarrow \frac{\partial \ln f(x,\mu)}{\partial\mu}=\frac{x-\mu}{\sigma^2}$$

$$n\mathrm{E}\left(\frac{\partial(\ln f(X,\mu))}{\partial\mu}\right)^2=n\mathrm{E}\left(\frac{X-\mu}{\sigma^2}\right)^2=\frac{n}{\sigma^4}\mathrm{E}\left(X-\mu\right)^2=\frac{n}{\sigma^4}\mathrm{D}\left(X\right)=\frac{n}{\sigma^2}$$

6.1.2.3 Ước lượng vững (consistent estimator)

Thống kê $\Theta = T(X_1, X_2, ..., X_n)$ được gọi là ước lượng vững của tham số θ của biến ngẫu nhiên gốc X nếu Θ hội tụ theo xác suất đến θ khi $n \to \infty$

Ta có các kết quả sau

- Trung bình mẫu là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.
- Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của tần suất p của tổng thể
- Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai của biến ngẫu nhiên gốc của tổng thể

6.2 PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY

Phương pháp ước lượng điểm nói trên có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng

Phương pháp ước lượng điểm cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu

Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy

Khoảng $[a; b]$ có hai đầu mút là hai thống kê

$$a = a(X_1, X_2, ..., X_n),\ b = b(X_1, X_2, ..., X_n)$$

phụ thuộc mẫu ngẫu nhiên $W=(X_1,X_2,\ldots,X_n)$ gọi là khoảng tin cậy của tham số θ với độ tin cậy β nếu

$$P\{a(X_1, X_2, ..., X_n) \le \theta \le b(X_1, X_2, ..., X_n)\} = \beta$$

Trong thực tế thường yêu cầu độ tin cậy β khá lớn, khi đó theo nguyên lý xác suất lớn biến cố

$$\{a(X_1,X_2,\ldots,X_n) \le \theta \le b(X_1,X_2,\ldots,X_n)\}$$

hầu như chắc chắn sẽ xảy ra trong một phép thử

Tiến hành một phép thử với mẫu ngẫu nhiên $W=(X_1,X_2,\ldots,X_n)$ ta thu được một mẫu cụ thể $w=(x_1,x_2,\ldots,x_n)$, tính được giá trị cụ thể $a=a(x_1,x_2,\ldots,x_n)$, $b=b(x_1,x_2,\ldots,x_n)$

Lúc đó có thể kết luận là: Qua mẫu cụ thể với độ tin cậy β tham số θ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng $[a; b]$, tức là $a \le \theta \le b$

6.2.1 Khoảng tin cậy của kỳ vọng biến ngẫu nhiên phân bố chuẩn

6.2.1.1. Trường hợp phương sai σ^2 đã biết

Khoảng tin cậy của tham số μ với độ tin cậy β có dạng

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

trong đó $\alpha = 1 - \beta$;

$U_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$

$\varepsilon = U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là ***độ chính xác của ước lượng***

Với độ tin cậy β không đổi, với độ chính xác ε_0 cho trước, *kích thước mẫu cần thiết* là số tự nhiên *n* nhỏ nhất thỏa mãn

$$n \geq \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2}$$

Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên phân bố chuẩn với độ lệch tiêu chuẩn 1 gram. Cần thử 25 sản phẩm loại này ta thu được kết quả

Trọng lượng (gram)	18	19	20	21
Số sản phẩm tương ứng	3	5	15	2

$\overline{x} = 19,64$. Với độ tin cậy $\beta = 0,95 \Rightarrow \alpha/2 = 0,025 \Rightarrow U_{\alpha/2} = 1,96$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1,96 \cdot \dfrac{1}{\sqrt{25}} = 0,392$

Khoảng tin cậy với độ tin cậy 95% của tham số μ là

$$19,248 \le \mu \le 20,032$$

Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 thì

$$n \ge \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} = \frac{1 \cdot 1,96^2}{0,3^2} = 42,68.\ \text{Chọn } n = 43$$

6.2.1.2 Trường hợp phương sai σ^2 chưa biết, $n \geq 30$

Khoảng tin cậy của tham số μ với độ tin cậy β

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Để xác định chiều cao trung bình của các cây bạch đàn trong khu rừng rộng trồng bạch đàn, ta tiến hành đo ngẫu nhiên 35 cây

Khoảng	r_i	x_i	$u_i = x_i$ - 8,25	$r_i u_i$	$r_i u_i^2$
6,5 - 7,0	2	6,75	-1,5	- 3	4,5
7,0 - 7,5	4	7,25	-1	- 4	4
7,5 - 8,0	10	7,75	- 0,5	- 5	2,5
8,0 - 8,5	11	8,25	0	0	0
8,5 - 9,0	5	8,75	0,5	2,5	1,25
9,0 - 9,5	3	9,25	1	3	3
$\sum$	35			- 6,5	15,25

$$\overline{u} = \frac{-6,5}{35} = -0,1857 \Rightarrow \overline{x} = 8,25 - 0,1857 \approx 8,06$$

$$s_X^2 = s_U^2 = \frac{1}{34}\left(15,25 - \frac{(-6,5)^2}{35}\right) = 0,413 \Rightarrow s = 0,64$$

Với độ tin cậy $\beta = 95\%, U_{\alpha/2} = 1,96$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\frac{s}{\sqrt{n}} = 1,96 \cdot \frac{0,64}{\sqrt{35}} = 0,21$

Khoảng tin cậy cho chiều cao trung bình μ của các cây bạch đàn

$$7,87 \le \mu \le 8,29$$

6.2.1.3 Trường hợp phương sai σ^2 chưa biết, $n < 30$

Khoảng tin cậy của tham số μ với độ tin cậy β

$$\left[\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}};\ \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right]$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\alpha/2$ của phân bố Student $n-1$ bậc tự do

$\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}$ được gọi là ***độ chính xác của ước lượng***

Với độ tin cậy β không đổi, với độ chính xác ε_0 cho trước, *kích thước mẫu cần thiết* là số tự nhiên *n* nhỏ nhất thỏa mãn

$$n \geq \left(\frac{S \cdot t_{\alpha/2}(n-1)}{\varepsilon_0}\right)^2$$

Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Gieo thử giống này trên 16 mảnh vườn thí nghiệm thu được như sau (đơn vị kg/ha):

172, 173, 173,174, 174, 175, 176, 166, 166, 167, 165, 173, 171, 170, 171, 170

Từ các số liệu trên ta tính được: $\overline{x} = 171;\ s = 3,4254$

Với độ tin cậy $\beta = 95\%$

$$\alpha = 0,05;\ \frac{\alpha}{2} = 0,025 \Rightarrow t_{\alpha/2}(n-1) = t_{0,025}(15) = 2,131$$

Độ chính xác $\varepsilon = t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}} = 2,131 \cdot \dfrac{3,4254}{\sqrt{16}} = 1,885$

Vậy khoảng tin cậy với độ tin cậy $\beta = 95\%$ cho năng suất trung bình của loại hạt giống này $169,115 \le \mu \le 172,885$

6.2.2 Khoảng tin cậy cho xác suất *p*

Ta cần nghiên cứu một dấu hiệu định tính A nào đó của tổng thể

Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0

Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli với tham số p

Kỳ vọng $\mathrm{E}X = p$ và phương sai $\mathrm{D}X = p(1-p)$

Lấy mẫu ngẫu nhiên $W=(X_1,X_2,\ldots,X_n)$

Tần suất mẫu $f = \dfrac{X_1 + \cdots + X_n}{n}$

Ta có thể xem thống kê

$U = \dfrac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}}$ xấp xỉ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn

Khoảng tin cậy cho tần suất p của tổng thể với độ tin cậy β

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

Với điều kiện n đủ lớn $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$

Độ chính xác của khoảng tin cậy $\varepsilon = U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}$

Với độ tin cậy β không đổi, với độ chính xác ε_0 cho trước, *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn

$$n \geq f(1-f)\left(\frac{U_{\alpha/2}}{\varepsilon_0}\right)^2$$

Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 1600 cử tri, được biết có 960 người trong số đó sẽ bỏ phiếu cho ứng cử viên A

Với độ tin cậy $\beta = 95\%$ hãy ước lượng ứng cử viên A sẽ chiếm được tối thiểu bao nhiêu % số phiếu bầu

Từ mẫu cụ thể trên ta có $f = \dfrac{960}{1600} = 0,6$

Thỏa mãn điều kiện n đủ lớn $\begin{cases} nf = 960 > 10 \\ n(1-f) = 640 > 10 \end{cases}$

Độ chính xác của ước lượng $\varepsilon = 1,96\sqrt{\dfrac{0,6 \cdot 0,4}{1600}} = 0,024$

Khoảng tin cậy 95% $0,576 \le p \le 0,624$

Vậy tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A

6.2.3 Ước lượng phương sai của biến ngẫu nhiên có phân bố chuẩn

Giả sử trong tổng thể biến ngẫu nhiên gốc X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nhưng chưa biết phương sai σ^2 của nó

Lấy mẫu ngẫu nhiên $W=(X_1,X_2,\ldots,X_n)$

Ta sẽ chọn thống kê thích hợp để ước lượng cho tham số σ^2 trong các trường hợp sau

6.2.3.1. Đã biết kỳ vọng μ

Chọn thống kê $$T = \frac{nS^{*2}}{\sigma^2} \qquad T \sim \chi^2(n)$$

Với độ tin cậy β, với cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$

có thể tìm hai giá trị tới hạn của $\boldsymbol{T}$ mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n)$ $\chi^2_{\alpha_2}(n)$

$$P\left\{T > \chi^2_{1-\alpha_1}(n)\right\} = 1-\alpha_1 \text{ và } P\left\{T > \chi^2_{\alpha_2}(n)\right\} = \alpha_2$$

Do đó $$P\left\{\chi^2_{1-\alpha_1}(n) < T < \chi^2_{\alpha_2}(n)\right\} = 1-(\alpha_1+\alpha_2) = \beta$$

$$P\left\{\ldots\right\} = \beta$$

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha_2}(n)}; \frac{nS^{*2}}{\chi^2_{1-\alpha_1}(n)}\right)$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng

- Nếu $\alpha_1 = \alpha_2 = \alpha/2$ khoảng tin cậy đối xứng có dạng

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha/2}(n)} ; \frac{nS^{*2}}{\chi^2_{1-\alpha/2}(n)} \right)$$

- Nếu $\alpha_1 = 0,\ \alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha}(n)} ; +\infty \right)$$

- Nếu $\alpha_1 = \alpha,\ \alpha_2 = 0$ khoảng tin cậy bên trái của σ^2 có dạng

$$\left(0 ; \frac{nS^{*2}}{\chi^2_{1-\alpha}(n)} \right)$$

Ví dụ 6.9: Mức hao phí nguyên liệu cho 1 đơn vị sản phẩm là biến ngẫu nhiên có phân bố chuẩn với trung bình là 20 gam. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử 25 sản phẩm và thu được kết quả sau

Hao phí nguyên liệu (gam)	19,5	20,0	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy $\beta = 90\%$,

Tìm khoảng tin cậy của σ^2 nếu $\alpha_1 = \alpha_2 = \alpha/2 = 0,05$

Tra bảng $\chi^2(n)$ ta có:

$$\chi^2_{\alpha/2}(n) = \chi^2_{0,05}(25) = 37,65;\ \chi^2_{1-\alpha/2}(n) = \chi^2_{0,95}(25) = 14,61$$

x_i	r_i	$x_i-\mu$	$(x_i-\mu)^2$	$r_i(x_i-\mu)^2$
19,5	5	– 0,5	0,25	1,25
20,0	18	0,0	0,00	0,00
20,5	2	0,5	0,25	0,50
Σ	25			1,75

$s^{*2} = 1,75/25 = 0,07$

Vậy với độ tin cậy 90%, qua mẫu cụ thể này, khoảng tin cậy của σ^2 là

$$\left(\frac{25\cdot 0,07}{37,65};\frac{25\cdot 0,07}{14,61}\right) \quad \text{hay} \quad 0,0464 < \sigma^2 < 0,1198$$

6.2.3.2. Chưa biết kỳ vọng μ

Chọn thống kê $$T = \frac{(n-1)S^2}{\sigma^2} \qquad T \sim \chi^2(n-1)$$

Với độ tin cậy β, với cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$

có thể tìm hai giá trị tới hạn của $\boldsymbol{T}$ mức α_1, α_2 là

$$\chi^2_{1-\alpha_1}(n-1) \qquad \chi^2_{\alpha_2}(n-1)$$

$$P\left\{T > \chi^2_{1-\alpha_1}(n-1)\right\} = 1-\alpha_1 \quad \text{và} \quad P\left\{T > \chi^2_{\alpha_2}(n-1)\right\} = \alpha_2$$

Do đó $$P\left\{\chi^2_{1-\alpha_1}(n-1) < T < \chi^2_{\alpha_2}(n-1)\right\} = 1-(\alpha_1+\alpha_2) = \beta$$

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right\} = \beta$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)};\frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right)$$

Tùy theo cách chọn cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$ ta có các khoảng tin cậy của σ^2

- Nếu $\alpha_1 = \alpha_2 = \alpha/2$ khoảng tin cậy đối xứng có dạng

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

- Nếu $\alpha_1 = 0,\ \alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)} ; +\infty \right)$$

- Nếu $\alpha_1 = \alpha,\ \alpha_2 = 0$ khoảng tin cậy bên trái của σ^2 có dạng

$$\left(0 ; \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)} \right)$$