Chương 2 - ƯỚC LƯỢNG CÁC THAM SỐ

NỘI DUNG

Ước lượng trung bình tổng thể

– Ước lượng điểm trung bình tổng thể

– Ước lượng khoảng trung bình tổng thể

Ước lượng phương sai tổng thể

– Ước lượng điểm phương sai tổng thể

– Ước lượng khoảng phương sai tổng thể

Ước lượng khoảng xác suất các dấu hiệu định tính của một tổng thể

KHÁI NIỆM

Ước lượng điểm: là phương pháp dùng trị số của hàm ước lượng được tính toán ở mẫu để thay một cách gần đúng cho tham số tổng thể.

Ước lượng khoảng: là phương pháp mà tham số ước lượng của tổng thể nằm trong một khoảng với một xác suất (hay độ tin cậy) cho trước. (Khoảng này xác định được nhờ những kết quả khi nghiên cứu ở mẫu)

$$P(G_1 \le \theta \le G_2) = 1 - \alpha$$

trong đó:

- P : là xác suất của sự ước lượng;
- G1 & G2 : là giới hạn dưới và giới hạn trên của khoảng ước lượng (được xác định từ kết quả quan sát ở mẫu);
- 1 – α : là mức tin cậy của ước lượng, α thường chọn là 0,05; 0,01 hay 0,001 (mức sai lầm).

Hiệu số G2 – G1 được gọi là độ dài khoảng ước lượng và

$$\varepsilon = \frac{G_2 - G_1}{2}$$

ε gọi là sai số tới hạn của ước lượng (hay còn gọi là độ chính xác của ước lượng)

Sai số tương đối được tính

$$\varepsilon\% = \frac{\varepsilon}{\overline{\overline{X}}}.100$$

ƯỚC LƯỢNG TRUNG BÌNH TỔNG THỂ

Ước lượng điểm trung bình tổng thể

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

Ước lượng khoảng trung bình tổng thể

- Khi đã biết phương sai S^2 của tổng thể

$$U = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$$

trong đó

- X là trung bình mẫu.
- U phân phối chuẩn với độ tin cậy 1 – α
- μ : tham số trung bình
- σ : phương sai
- n : số mẫu

Tham số trung bình μ được tính theo công thức

$$P(\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Trong đó $u_{\alpha/2}$ là ????

Sai số tới hạn hay độ chính xác của ước lượng được tính

$$\varepsilon = u_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Khoảng tin cậy được tính

$$I = 2\varepsilon = 2u_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Tính dung lượng mẫu cần thiết để đạt được độ chính xác tương đối cho trước $\varepsilon_0(\%)$

$$\varepsilon(\%) = \frac{\varepsilon}{\bar{x}}.100 = u_{\alpha/2}\frac{\sigma}{\bar{x}\sqrt{n}}.100$$

$$= u_{\alpha/2}\frac{CV(\%)}{\sqrt{n}}$$

$$n_{\min} = \left(\frac{u_{\alpha/2}CV(\%)}{\varepsilon_0(\%)}\right)^2$$

Nếu $\alpha = 0{,}10$ thì $u_{\alpha/2} = 1{,}645$; $\alpha = 0{,}05$ thì $u_{\alpha/2} = 1{,}960$;
$\alpha = 0{,}02$ thì $u_{\alpha/2} = 2{,}326$; $\alpha = 0{,}01$ thì $u_{\alpha/2} = 2{,}576$

Khi chưa biết phương sai S^2 của tổng thể nhưng có dung lượng mẫu lớn (n > 30)

Khi dung lượng mẫu đủ lớn thì S = σ nên có thể thay thế S cho σ, khi đó việc ước lượng được tiến hành theo luật phân phối chuẩn theo công thức

$$P(\bar{X} - 1,96\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + 1,96\frac{S}{\sqrt{n}}) = 0,95$$

ở độ tin cậy 95% có thể phát biểu trung bình tổng thể μ nằm trong khoảng

$$(\bar{X} - 1,96\frac{S}{\sqrt{n}} ; \bar{X} + 1,96\frac{S}{\sqrt{n}})$$

$$\varepsilon = 1,96\frac{S}{\sqrt{n}}$$

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{x}}.100 = 1,96\frac{S}{\overline{x}\sqrt{n}}.100 = 1,96\frac{CV(\%)}{\sqrt{n}}$$

$$n_{\min} = \left(\frac{1,96CV(\%)}{\varepsilon_0(\%)}\right)^2$$

Tương tự với độ tin cậy 99%

$$P(\overline{x} - 2,58\frac{S}{\sqrt{n}} \le \mu \le \overline{x} + 2,58\frac{S}{\sqrt{n}}) = 0,99$$

Trung bình tổng thể μ nằm trong khoảng

$$(\overline{x} - 2,58\frac{S}{\sqrt{n}}; \overline{x} + 2,58\frac{S}{\sqrt{n}})$$

$$\varepsilon = 2,58\frac{S}{\sqrt{n}}$$

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{x}}.100 = 2,58\frac{S}{\overline{x}\sqrt{n}}.100 = 2,58\frac{CV(\%)}{\sqrt{n}}$$

$$n_{\min} = \left(\frac{2,58CV(\%)}{\varepsilon_0(\%)}\right)^2$$

Nếu mẫu được chọn n > 0,1N. Để bảo đảm độ chính xác của ước lượng thì sai số tới hạn sẽ nhân thêm hệ số điều chỉnh

$$\sqrt{\frac{N-n}{N-1}}$$

Tham số trung bình μ sẽ là

$$P\left(\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}\le\mu\le\bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}\right)=1-\alpha$$

Ví dụ: Ước lượng trung bình năng suất cá thể và khoảng tin cậy của tổ hợp bông lai F_1 S02-13/TM1 trồng tại Đại học Nông Lâm Tp. HCM, 2008 theo số liệu bảng sau:

Bảng 2.1: Năng suất cá thể 50 cây (g/cây)

84,4	73,5	84,5	93,5	75,2	74,3	93,5	82,1	68,6	85,4
66,8	57,7	79,4	91,0	129,0	74,6	61,4	91,4	99,5	82,7
95,2	88,3	28,4	77,0	81,2	39,7	86,4	51,5	51,2	80,5
101,0	77,7	90,0	92,9	80,8	67,2	57,1	57,3	34,2	79,5
80,3	88,0	61,1	63,8	101,0	70,2	95,1	97,0	50,3	73,7

Yêu cầu

- Tính trung bình; phương sai; và độ lệch chuẩn
- Ước lượng khoảng trung bình tổng thể ở độ tin cậy 95%
- Tính sai số tới hạn và sai số tương đối
- Tính số mẫu điều tra tối thiểu khi muốn sai số tương đối không vượt quá 5%.

Giải

Trung bình = 76.92

Phương sai = 351.68

Độ lệch chuẩn = 18.75

- Ước lượng khoảng trung bình ở độ tin cậy 95%

$$P(76,92-1,96\frac{18,75}{\sqrt{50}}\leq \mu \leq 76,92+1,96\frac{18,75}{\sqrt{50}})=0,95$$

$$= P(76,92-5,20\leq \mu \leq 76,92+5,20)=0,95$$

$$= P(71,72\leq \mu \leq 82,12)=0,95$$

- Tính sai số tới hạn và sai số tương đối

  Sai số tới hạn sai ε = 5,20 g/cây, sai số tương đối ε(%) là 6,76

- Để sai số tương đối $\varepsilon_0(\%)$ cho trước không vượt quá 5% thì số mẫu điều tra tối thiểu phải đạt

$$n_{\min} = \left(\frac{2CV(\%)}{\varepsilon_0(\%)} \right)^2 = \left(\frac{2 \times 24,38}{5} \right)^2$$

$= 91$ cây

Khi chưa biết phương sai S^2 của tổng thể và có dung lượng mẫu nhỏ (n < 30)

- Trong trường hợp mẫu nhỏ việc ước lượng được tiến hành theo luật phân phối Student
- Công thức ước lượng khoảng số trung bình tổng thể theo luật Student

$$P\left(\bar{x} - t_{\alpha_2}^{(n-1)} \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha_1}^{(n-1)} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

khoảng tin cậy của μ sẽ là

$$\left(\overline{x} - t_{\alpha_2}^{(n-1)}\frac{S}{\sqrt{n}};\overline{x} + t_{\alpha_1}^{(n-1)}\frac{S}{\sqrt{n}}\right)$$

Khi ước lượng khoảng tin cậy đối xứng α1 = α2 = α/2 thì khoảng tin cậy của μ sẽ la

$$P(\overline{x} - t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}} \le \mu \le \overline{x} + t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}}) = 1-\alpha$$

Sai số tới hạn ε la

$$\varepsilon = t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}}$$

khoảng tin cậy sẽ là

$$I = 2\varepsilon = 2t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}},$$

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{X}}.100 = t_{\alpha/2}^{(n-1)}\frac{S}{\overline{X}\sqrt{n}}.100 = t_{\alpha/2}^{(n-1)}\frac{CV(\%)}{\sqrt{n}}$$

Dung lượng mẫu cần thiết để đạt được độ chính xác $\varepsilon_0(\%)$ là

$$n_{\min} = \left(t_{\alpha/2}^{(n-1)}\frac{CV(\%)}{\varepsilon_0(\%)}\right)^2$$

khi n > 0,1N, để bảo đảm độ chính xác của ước lượng thì sai số tới hạn sẽ nhân thêm hệ số điều chỉnh

$$\sqrt{\frac{N-n}{N-1}}$$

Khi ước lượng độ tin cậy của tổng thể

$$P\left(\bar{x} - t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{x} + t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

Dựa vào đây để tính toán ε, $\varepsilon(\%)$, từ đó sẽ xác định được dung lượng mẫu cần thiết để đạt được độ chính xác tương đối cho trước $\varepsilon_0(\%)$.

ƯỚC LƯỢNG PHƯƠNG SAI TỔNG THỂ

Ước lượng điểm phương sai tổng thể

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

Ước lượng khoảng phương sai tổng thể

Việc ước lượng khoảng phương sai tổng thể được tiến hành theo luật phân phối "khi bình phương"

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

khoảng tin cậy

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

Khi ước lượng khoảng tin cậy đối xứng thì
$\alpha 1 = \alpha 2 = \alpha/2$ thi

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

Ví dụ: Hãy ước lượng phương sai tổng thể về năng suất (cá thể) của tổ hợp bông lai F_1S02-13/TM1 trồng tại Đại học Nông Lâm Tp. HCM, 2008 theo số liệu Bảng 2.2 với S = 18,76 với độ tin cậy 0,95.

Ở đây $\alpha = 1 - 0,95 = 0,05$; $\alpha/2 = 0,025$; $1 - \alpha/2 = 0,975$.

Tra giá trị c^2 trong phần mềm Excel, ta có $\chi^{2(49)}_{0,025} = 70,222$ và $\chi^{2(49)}_{0.975} = 31,555$. Thay các giá trị này vào công thức ta có:

$$\left(\frac{49\times18,76^2}{70,222};\frac{49\times18,76^2}{31,555}\right) = (546,24\ ;\ 245,46)$$

$$P(245,46 \leq \sigma^2 \leq 546,24) = 0,95$$

Như vậy ở mức tin cậy 0,95 phương sai tổng thể năng suất cá thể của tổ hợp bông lai F1 S02-13/TM1 nằm trong khoảng từ 245,46 đến 546,24

ƯỚC LƯỢNG KHOẢNG XÁC SUẤT CÁC DẤU HIỆU ĐỊNH TÍNH CỦA TỔNG THỂ

Công thức ước lượng xác suất p tổng thể

$$P\left(p_m - u_{\alpha_2}\sqrt{\frac{p_m(1-p_m)}{n}} \le p \le p_m + u_{\alpha_1}\sqrt{\frac{p_m(1-p_m)}{n}}\right) = 1-\alpha$$

khoảng tin cậy của p là

$$\left(p_m - u_{\alpha_2}\sqrt{\frac{p_m(1-p_m)}{n}}; p_m + u_{\alpha_1}\sqrt{\frac{p_m(1-p_m)}{n}}\right)$$

p_m là xác suất mẫu

u_{a1} và u_{a2} hai giá trị tới hạn chuẩn

Khi ước lượng khoảng tin cậy đối xứng thì $\alpha_1 = \alpha_2 = \alpha/2$ thì

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

Với $\alpha = 0,05$, $u_{\alpha/2} = 1,96$; $\alpha = 0,01$, $u_{\alpha/2} = 2,58$ và

$\alpha = 0,001$, $u_{\alpha/2} = 3,29$

Sai số tới hạn của ước lượng

$$\varepsilon = u_{\alpha/2}\sqrt{\frac{p_m(1-p_m)}{n}}$$

Dung lượng mẫu cần thiết để đạt được độ chính xác cho trước ε_0 là

$$n_{\min} = (u_{\alpha/2})^2 \frac{p_m(1-p_m)}{\varepsilon_0^{\,2}}$$

Ví dụ: Để kiểm nghiệm tỷ lệ nảy mầm một giống bắp lai, người ta đã tiến hành thử 4 mẫu, mỗi mẫu 100 hạt. Kết quả như sau:

Mẫu thử	1	2	3	4	Tổng
Số hạt nảy mầm	93	89	87	96	365

Hãy ước lượng khoảng xác suất nảy mầm của lô hạt giống và số lượng hạt cần thử để đạt sai số không vượt quá 3% với độ tin cậy 95%.

Giải: $n = 400$, $pm = 365/400 = 0{,}913$, $a = 1 – 0{,}95 = 0{,}05$, $u_{0,025} = 1{,}96$ và $\varepsilon_0 = 0{,}03$

Với dung lượng mẫu đủ lớn, theo công thức trên ta có

$$P\left(0{,}885 \le p \le 0{,}940\right) = 0{,}95$$

Số hạt cần thử: $n_{\min} = (u_{\alpha/2})^2 \dfrac{p_m(1-p_m)}{\varepsilon_0{}^2}$

$$= 1{,}96^2 \frac{0{,}913(1-0{,}913)}{0{,}03^2} = 340{,}8 \approx 341$$