Chương 4:

LÍ THUYẾT ƯỚC LƯỢNG & KIỂM ĐỊNH GIẢ THIẾT

4.1. Bài toán ước lượng

– ***Ước lượng*** là *phỏng đoán* một giá trị chưa biết của tổng thể bằng cách dựa vào quan sát mẫu.

– Thông thường ta cần ước lượng giá trị trung bình, tỉ lệ, phương sai, độ lệch chuẩn,…

– Có 2 hình thức ước lượng:

a) ***Ước lượng điểm***: là đi tìm *một giá trị* gần đúng nhất với giá trị thực của tham số.

b) ***Ước lượng khoảng***: là đi tìm một *khoảng số thực* sao cho xác suất của tham số cần ước lượng rơi vào khoảng đó tương đối lớn.

4.1.1. Ước lượng điểm

– Biến ngẫu nhiên X cần nghiên cứu cho tổng thể, giả sử X có luật phân phối xác suất đã biết (chuẩn, nhị thức,…) nhưng còn phụ thuộc vào tham số θ chưa biết.

– Cần ước lượng tham số θ

– Ước lượng $\hat{\theta}$ của tham số θ gọi là *ước lượng tốt nhất* nếu thoả mãn 3 điều kiện sau:

+ *Không chệch:* $E(\hat{\theta}) = \theta$

+ *Bền vững:*

$$\lim_{n\to+\infty} P\left(\left|\hat{\theta}-\theta\right| < \varepsilon\right) = 1 \quad \text{hay} \quad \hat{\theta} \xrightarrow{n\to+\infty} \theta$$

+ *Có hiệu quả:*

$D(\hat{\theta})$ đạt giá trị nhỏ nhất

Một số kết quả

- $\overline{X}$ là ước lượng tốt nhất của kì vọng E(X).

- Phương sai mẫu hiệu chỉnh S_1^2 là một ước lượng không chệch của phương sai D(X) của X.

4.1.2. Ước lượng khoảng

1. Khoảng tin cậy (khoảng ước lượng), độ tin cậy

– Khoảng (c; d) gọi là ***khoảng tin cậy*** (khoảng ước lượng) của tham số θ, nếu:

$$P(\theta \in (c;d)) = 1 - \alpha = \gamma$$

với α khá bé (thường $0 < \alpha < 0,05$)

$1 - \alpha = \gamma$ (gần bằng 1) gọi là ***độ tin cậy***

Chú ý

- Theo định nghĩa độ tin cậy $\gamma = 1 - \alpha$ càng gần bằng 1 và khoảng tin cậy càng ngắn thì ước lượng càng chính xác.

– Trường hợp khoảng tin cậy đối xứng có dạng:

$$(\hat{\theta} - \varepsilon; \hat{\theta} + \varepsilon)$$

- ε gọi là độ chính xác hay sai số của ước lượng.
- 2ε gọi là bề rộng của khoảng.

2. Ước lượng khoảng của kì vọng

Bài toán:

X là BNN có phân phối chuẩn N(a, σ), nhưng kì vọng EX = a chưa biết. Tìm khoảng ước lượng đối xứng của kỳ vọng khi đã biết trung bình mẫu $\bar{x}$ với độ tin cậy γ = 1 - α

Giải:

Khoảng ước lượng của kì vọng (giá trị trung bình) có dạng: $(\bar{x} - \varepsilon, \bar{x} + \varepsilon)$

Trường hợp 1: Đã biết σ

Khi đó, ε= $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, nên khoảng ước lượng là:

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < a < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Với $z_{\alpha/2} = \phi^{-1}(1 - \frac{\alpha}{2})$

Trường hợp 2: Chưa biết σ

Khi đó, thay σ bởi độ lệch chuẩn mẫu hiệu chỉnh s_1

Nếu n > 30: Khoảng ước lượng:

$$\bar{x} - z_{\alpha/2}\frac{s_1}{\sqrt{n}} < a < \bar{x} + z_{\alpha/2}\frac{s_1}{\sqrt{n}}$$

Nếu n ≤ 30: Khoảng ước lượng:

$$\bar{x} - t_{n-1;\alpha/2}\frac{s_1}{\sqrt{n}} < a < \bar{x} + t_{n-1;\alpha/2}\frac{s_1}{\sqrt{n}}$$

Với $t_{n-1;\alpha/2}$ là phân phối Student

Ví dụ

Biết X là độ dài của một loại trục do một máy tiện làm ra là biến ngẫu nhiên có phân phối chuẩn N(a, 2).Từ một mẫu gồm 64 trục, người ta tính được trung bình mẫu $\bar{x} = 20{,}5\ cm$

Với độ tin cậy 95%, hãy tìm khoảng ước lượng cho độ dài trung bình của trụ, biết $z_{0,025}=1{,}96$.

Giải:

- Các đặc trưng mẫu: n=64, $\bar{x} = 20{,}5$

- Độ tin cậy: $1-\alpha=0{,}95 \Rightarrow \alpha=0{,}05$ và $z_{\alpha/2}=z_{0{,}025}=1{,}96$

- Sai số ước lượng: $\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{2}{\sqrt{64}} = 0{,}49$

Vậy khoảng ước lượng của độ dài trung bình của trụ là:
$(\bar{x} - \varepsilon, \bar{x} + \varepsilon) = (20{,}01; 20{,}99)$

Ví dụ:

Khối lượng (kg) của một thiết bị có phân phối chuẩn $N(\mu;\sigma^2)$ *với* $\sigma = 0.2$ *(kg). Chọn ngẫu nhiên 25 thiết bị người ta tính được trung bình mẫu* $\overline{x} = 65,1$ *(kg). Với độ tin cậy 95% hãy tìm khoảng tin cậy (đối xứng) cho khối lượng trung bình của thiết bị này. Cho biết* $z_{0,025} = 1,96$.

Giải.

- Các số đặc trưng mẫu: $n = 25; \bar{x} = 65,1$.
- Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05; z_{\alpha/2} = z_{0,025} = 1,96$.
- Sai số ước lượng:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1,96 * \frac{0,2}{\sqrt{25}} = 0,0784$$

- Khoảng tin cậy cho khối lượng trung bình μ của thiết bị này:

$$\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon \Leftrightarrow 65,02 < \mu < 65,18$$

Ví dụ

Đo đường kính của 100 trục máy do 1 nhà máy sản xuất được bảng số liệu sau:

Đường kính (cm)	9,75	9,8	9,85	9,9
Số trục máy	5	37	42	16

Hãy ước lượng khoảng trung bình của trục máy với độ tin cậy 95%, biết $z_{0,025}=1,96$

Chú ý: Xác định độ tin cậy và kích thước mẫu

- Độ tin cậy: $1-\alpha = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{S_1}\right)$

- Kích thước mẫu: $n = \left[\left(\frac{\Phi^{-1}\left(\frac{\gamma}{2}\right)S_1}{\varepsilon}\right)^2\right] + 1$

Ví dụ

Đo đường kính của 100 chi tiết do một máy sản xuất ta thu được bảng số liệu:

Đường kính (cm)	9,75	9,80	9,85	9,90
Số chi tiết	5	37	42	16

a.Với độ chính xác 0,006 hãy xác định độ tin cậy?

b. Muốn độ chính xác 0,003; độ tin cậy 95% thì cần kiểm tra bao nhiêu chi tiết.

2. Ước lượng khoảng của tỉ lệ

Bài toán: Xét dấu hiệu A trong tổng thể có tỉ lệ là p (chưa biết). Để khảo sát giá trị p ta tiến hành điều tra mẫu có kích thước n. Giả sử trong mẫu này có k phần tử mang dấu hiệu A. Hãy tìm khoảng ước lượng cho p.

Giải:

Khoảng ước lượng cho p là:

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

$$Trong\ \text{đó}, \hat{p} = \frac{k}{n} : Tỉ\ lệ\ mẫu$$

Ví dụ:

Gieo 400 hạt đậu xanh thì có 50 hạt không nảy mầm. Với độ tin cậy 98%, hãy tìm khoảng ước lượng (tin cậy) cho tỉ lệ hạt nảy mầm, biết $z_{0,01}=2,326$, $z_{0,025}=1,96$.

Giải.

a. - Tỉ lệ mẫu: $\hat{p} = \dfrac{k}{n} = \dfrac{400 - 50}{400} = 0,875.$

- Độ tin cậy: $1 - \alpha = 0,98 \Rightarrow \alpha = 0,02; z_{\alpha/2} = z_{0,01} = 2,326$

- Sai số: $\varepsilon = z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 2,326 * \sqrt{\dfrac{0,875 * (1 - 0,875)}{400}} = 0,0385$

- Khoảng tin cậy cho tỉ lệ hạt nảy mầm p:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,8365 < p < 0,9135$$

Ví dụ:

Hãy ước lượng tỷ lệ chính phẩm của một nhà máy bằng khoảng tin cậy đối xứng với độ tin cậy 0,95 biết rằng kiểm tra 100 sản phẩm của nhà máy thì thấy có 10 phế phẩm.

i) Khoảng tin cậy cho số lượng cá thể

Từ khoảng tin cậy cho tỉ lệ ta có thể suy ra khoảng tin cậy cho số lượng cá thể mang dấu hiệu nghiên cứu hoặc số lượng cá thể của tổng thể. Cụ thể:

- Gọi K và N lần lượt là số lượng cá thể mang dấu hiệu nghiên cứu và số lượng cá thể của tổng thể. Khi đó, tỉ lệ dấu hiệu nghiên cứu:

$$p = \frac{K}{N} \quad (1)$$

- Dựa trên mẫu điều tra, ta ước lượng được khoảng tin cậy cho tỉ lệ p:

$$p_1 < p < p_2 \quad (2)$$

Từ (1) và (2) suy ra khoảng tin cậy cho giá trị K hoặc N.

Ví dụ:

Trong một cuộc bầu cử ở một địa phương có 10000 cử tri. Người ta phỏng vấn ngẫu nhiên 300 cử tri thì thấy có 156 người ủng hộ cho ứng cử viên A. Với độ tin cậy 95% tìm khoảng tin cậy (đối xứng) cho số lượng cử tri ở địa phương này ủng hộ cho ứng cử viên A.

Cho biết: $z_{0,025}=1,96$

Giải. Gọi K là số lượng cử tri ở địa phương này ủng hộ ứng cử viên A. Ta có tỉ lệ ủng hộ ứng cử viên A: $p = K/10000$.

Tỉ lệ mẫu: $\hat{p} = k/n = 156/300 = 0,52$

Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05; z_{\alpha/2} = z_{0,025} = 1,96$

Sai số:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96 * \sqrt{\frac{0,52 * 0,48}{300}} = 0,0565$$

Khoảng tin cậy cho tỉ lệ ủng hộ:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,4635 < p < 0,5765$$

Từ đó, suy ra:

$$0,4635 < K/10000 < 0,5765 \Leftrightarrow 4635 < K < 5765$$

Ví dụ:

Để ước lượng số lượng cá có trong hồ, người ta làm như sau. Bắt ngẫu nhiên 500 con cá, sau đó đánh dấu vào các con cá đã được bắt và thả chúng xuống hồ. Sau đó bắt ngẫu nhiên 200 con cá để kiểm tra thì thấy có 30 con có đánh dấu. Với độ tin cậy 90% tìm khoảng tin cậy đối xứng cho số lượng cá có trong hồ.

Biết $z_{0,05}$=1,645, $z_{0,025}$=1,96

Giải. Gọi N là số lượng cá trong hồ. Ta có tỉ lệ cá được đánh dấu: $p = 500/N$.
Tỉ lệ mẫu: $\hat{p} = k/n = 30/200 = 0,15$
Độ tin cậy: $1 - \alpha = 0,9 \Rightarrow \alpha = 0,1; z_{\alpha/2} = z_{0,05} = 1,645$
Sai số:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,645 * \sqrt{\frac{0,15 * 0,85}{200}} = 0,0415$$

Khoảng tin cậy cho tỉ lệ cá được đánh dấu:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,1085 < p < 0,1915$$

Từ đó, suy ra:

$$0,1085 < 500/N < 0,1915 \Leftrightarrow 2610,9 < N < 4608,3$$

Bài tập

Bài 1:

Trọng lượng của một loại trứng gà được cho bởi bảng sô liệu sau:

X-Trọng lượng (g)	25-30	30-35	35-40	40-45	45-50
Số quả	15	17	40	18	10

Bằng khoảng tin cậy đối xứng hãy ước lượng trọng lượng trung bình của loại trứng gà này với độ tin cậy 95%. Cho biết trọng lượng trứng gà là đại lượng ngẫu nhiên có phân phối chuẩn.

Bài 2:

Điều tra doanh số hàng tháng của 100 hộ kinh doanh một loại hàng, có bảng số liệu:

Doanh số (X-triệu đồng)	11,5	11,6	11,7	11,8	11,9	12,0
Số hộ tương ứng	10	15	20	30	15	10

Bằng khoảng tin cậy đối xứng hãy ước lượng doanh số trung bình hàng tháng của các hộ kinh doanh mặt hàng này với độ tin cậy 95%. Giả thiết doanh số là ĐLNN tuân theo qui luật phân phối chuẩn.

4.2. BÀI TOÁN KIỂM ĐỊNH GIẢ THIẾT

Ở phần trước ta đã nghiên cứu các tham số đặc trưng của tổng thể trên cơ sở thông tin của mẫu bằng *phương pháp ước lượng*. Phần này tiếp tục nghiên cứu dấu hiệu của tổng thể bằng phương pháp khác là *kiểm định giả thiết thống kê.*

Đứng trước một tổng thể, ta thường có một giả thiết nào đó. Bằng cách sử dụng các kết quả nghiên cứu từ mẫu để khẳng định hay bác bỏ giả thiết ban đầu gọi là kiểm định giả thiết thống kê.

Các giả thiết thống kê thường được kí hiệu là: H

Các loại sai lầm của kiểm định giả thiết:

– Loại 1: Bác bỏ H trong khi H đúng.

– Loại 2: Chấp nhận H trong khi H sai.

1. Phương pháp kiểm định:

Chấp nhận xảy ra sai lầm loại 1 với xác suất không vượt quá α (α gọi là mức ý nghĩa). Với mức ý nghĩa đó, ta sẽ chấp nhận giả thiết H nếu xác suất xảy ra sai lầm loại 2 nhỏ nhất.

2. Các loại kiểm định giả thiết về tham số t của biến ngẫu nhiên:

– Kiểm định 2 phía đối với tham số t:

Giả thiết H: "$t = t_o$" với đối thiết H_1: "$t \neq t_o$"

– Kiểm định phía phải đối với tham số t:

Giả thiết H: "$t = t_o$" với đối thiết H_1: "t > to"

– Kiểm định phía trái đối với tham số t:

Giả thiết H: "$t = t_o$" với đối thiết H_1: "$t < t_o$"

3. Kiểm định giả thiết về giá trị trung bình

Bài toán:

Giả sử tổng thể có trung bình (kỳ vọng) là a (giá trị của a chưa biết). Mẫu có kích thước n, trung bình mẫu $\overline{X}$, độ lệch chuẩn mẫu hiệu chỉnh S_1.

Hãy kiểm định giả thiết H: $a = a_0$ với mức ý nghĩa α.

(a_0 là giá trị cụ thể nào đó đã biết).

Kiểm định giả thiết H: a = a_0 với đối thiết H_1 : a $\neq$ a_0

Bước 1: Tính $Z_\alpha = \Phi^{-1}\left(\frac{1-\alpha}{2}\right)$

Bước 2: Tính

$$Z_0 = \frac{|\overline{X} - a_0|}{S_1} \cdot \sqrt{n}$$

Bước 3: Kết luận

-Nếu $Z_0 \leq Z_\alpha$ thì chấp nhận H

- Nếu $Z_0 > Z_\alpha$ thì bác bỏ H và chấp nhận H_1 với mức ý nghĩa α.

Ví dụ:

Điểm trung bình môn toán năm học trước là 5,72. Năm học này, theo dõi 100 sinh viên thu được bảng số liệu sau:

Điểm	3	4	5	6	7	8	9
Số SV	3	5	27	43	12	6	4

Hãy kiểm định giả thiết: Phải chăng điểm trung bình môn toán năm nay đã có sự thay đổi so với năm trước với mức ý nghĩa 1%?