Chương 7
Ước lượng và kích thước mẫu

7-1 Giới thiệu

7-2 Ước lượng trung bình quần thể

7-3 Ước lượng tỉ lệ quần thể

Giới thiệu

- Trong chương 2 & 3, chúng ta đã sử dụng số liệu thống kê mô tả khi chúng ta tóm tắt dữ liệu bằng cách sử dụng các công cụ như đồ thị và số liệu thống kê chẳng hạn như giá trị trung bình và độ lệch chuẩn.
- Chương 6 chúng ta giới thiệu các critical value:

  z_α biểu thị điểm z với diện tích α ở bên phải.
- Nếu $\alpha = 0.025$, thì giá trị **critical value** là $z_{0.025} = 1.96$.
- Nghĩa là, critical value $z_{0.025} = 1.96$ có diện tích phần bên phải là 0.025

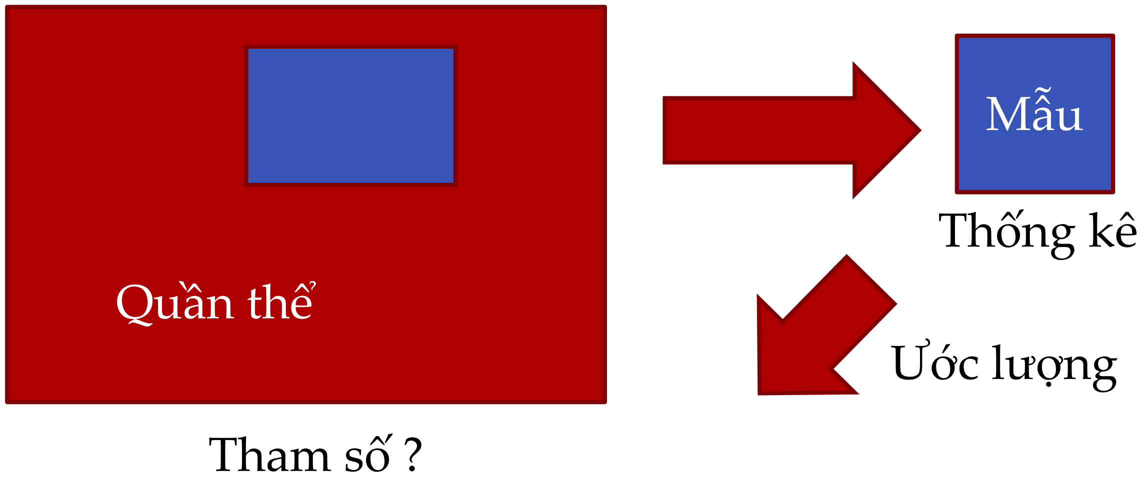
Giới thiệu

Chương này trình bày các khái niệm bắt đầu về thống kê suy diễn

- Hai vấn đề chính của thống kê suy diễn là
  - (1) sử dụng dữ liệu mẫu để ước lượng các giá trị của tham số quần thể.
  - (2) để kiểm định giả thuyết hoặc đưa ra các phát biểu về các tham số quần thể.
- Chúng ta giới thiệu các phương pháp để ước lượng các giá trị tham số trên quần thể quan trọng sau: trung bình và tỉ lệ.
- Chúng ta cũng trình bày các phương pháp xác định kích thước mẫu cần thiết để ước tính các thông số đó.

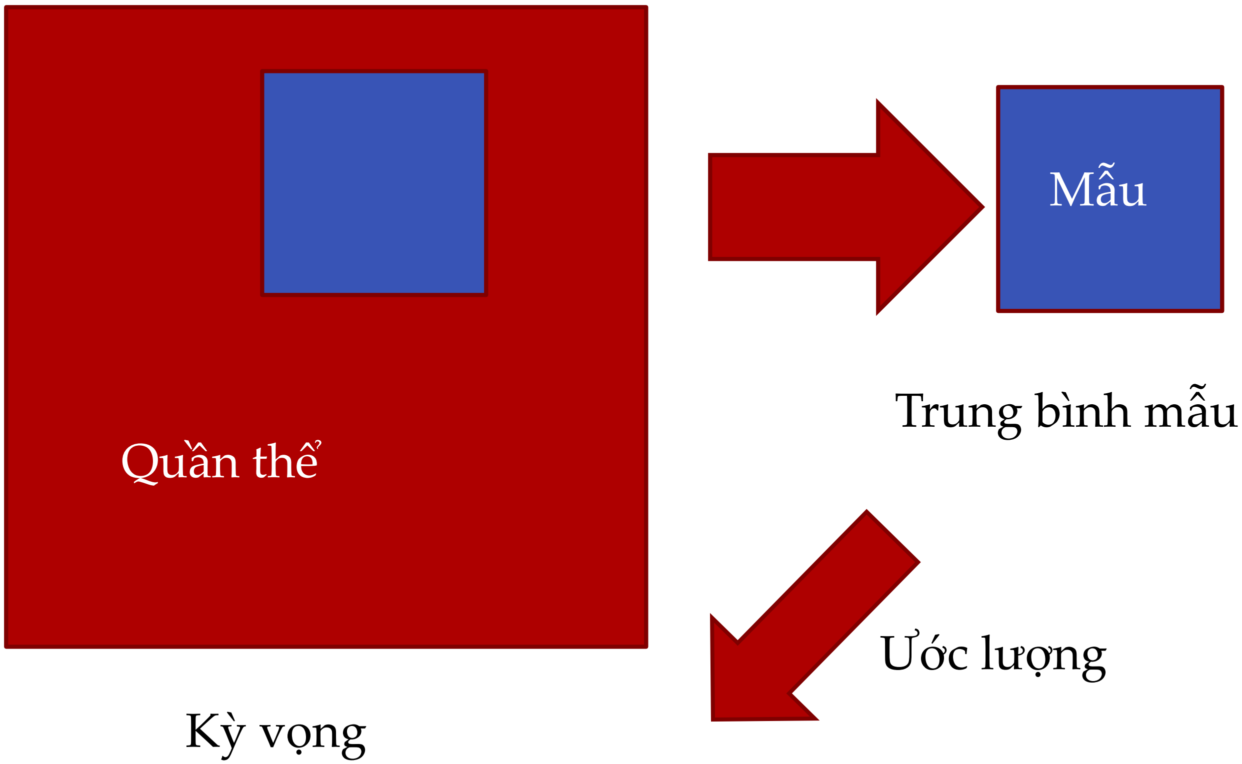
Giới thiệu

Ước lượng tham số: sử dụng thống kê mẫu để ước lượng cho tham số quần thể



- Ước lượng điểm (point estimation): xác định một giá trị số là giá trị ước lượng cho tham số quần thể
- Ước lượng khoảng (interval estimation): xác định một khoảng giá trị có nhiều khả năng chứa giá trị tham số quần thể

Ví dụ



Ví dụ ước lượng điểm

- Ví dụ: Biến ngẫu nhiên X phân phối chuẩn với kỳ vọng μ chưa biết. Trung bình mẫu $\overline{X}$

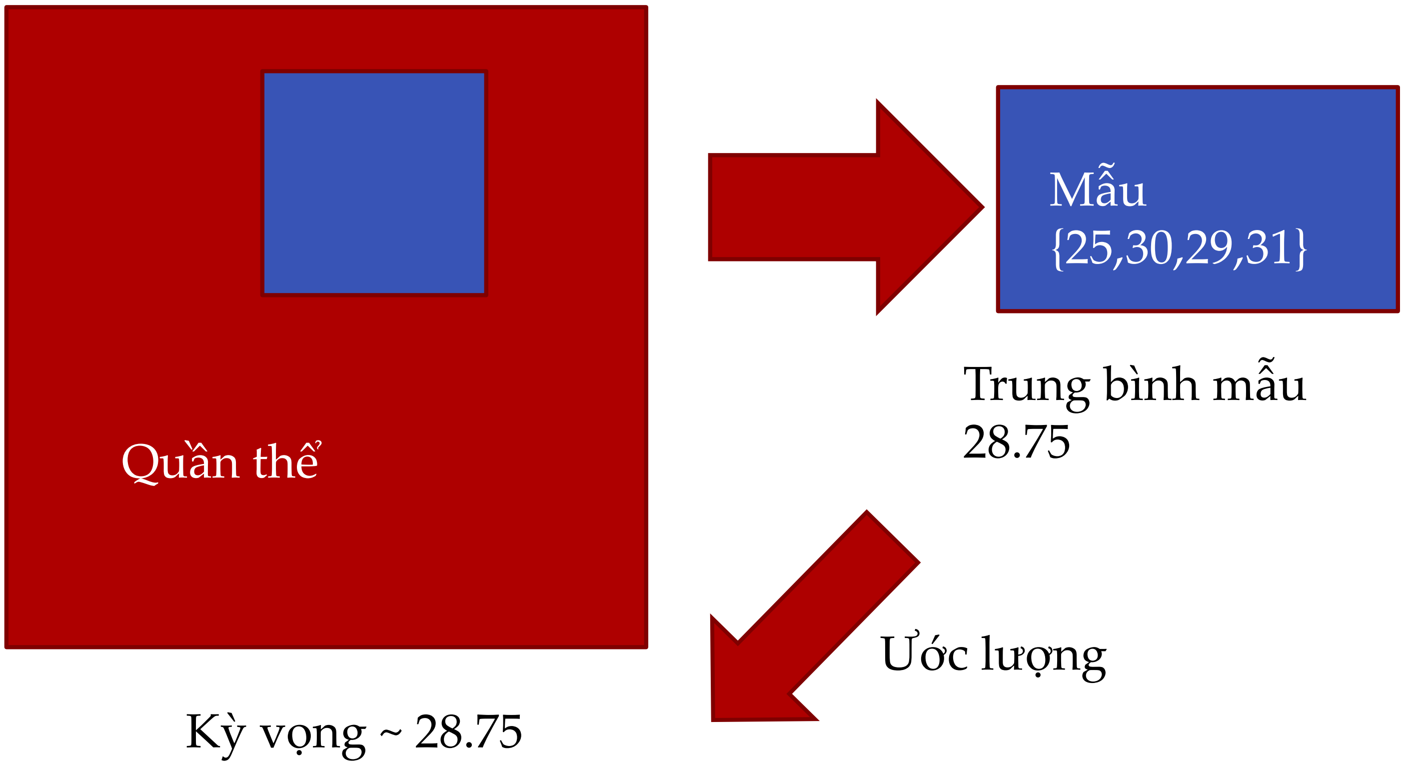
  là một bộ ước lượng điểm cho kỳ vọng μ.

  Lấy mẫu cụ thể, giá trị trung bình $\overline{x}$ là giá trị ước lượng điểm của μ.

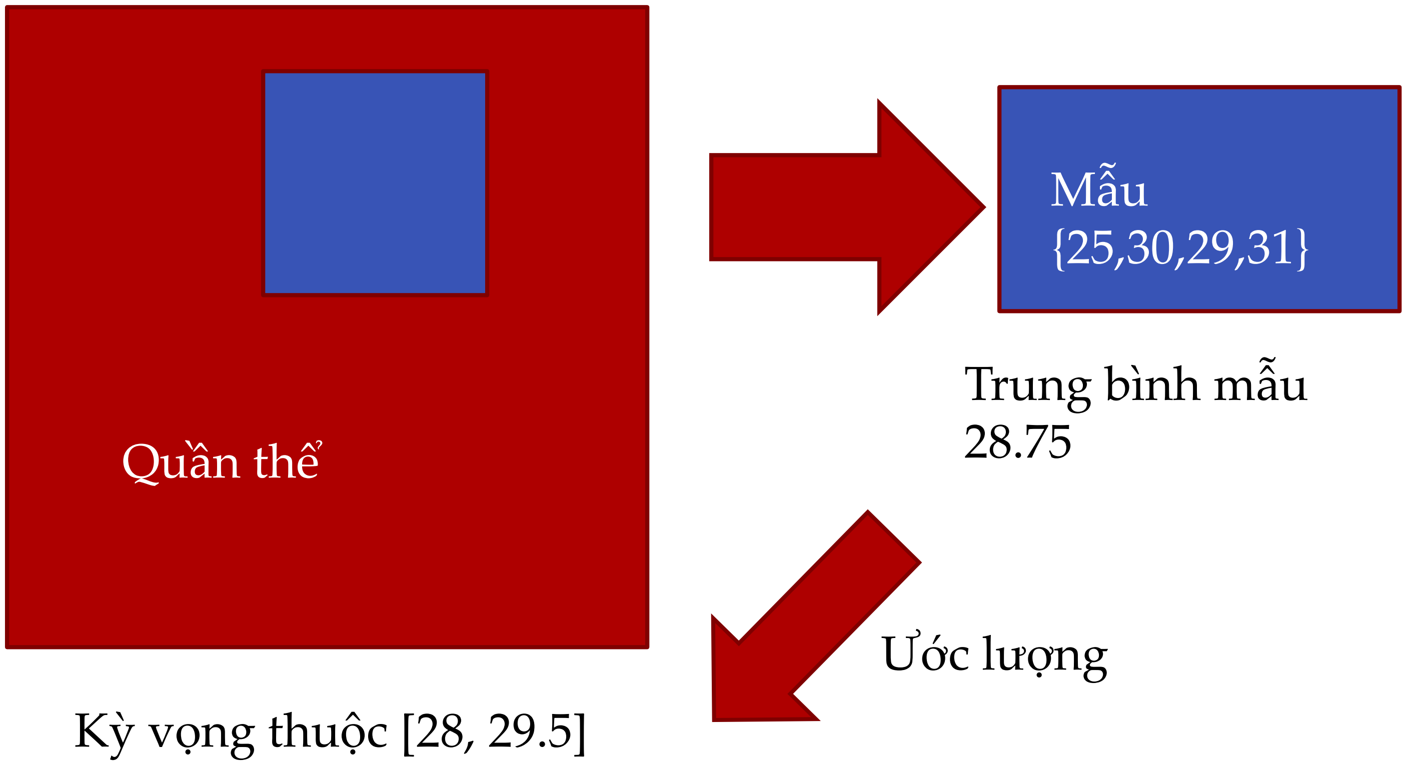
  Nếu mẫu là $x_1=25$, $x_2=30$, $x_3=29$, $x_4=31$ thì giá trị ước lượng điểm của μ là:

$$\overline{x} = \frac{25+30+29+31}{4} = 28.75$$

Ví dụ ước lượng điểm



Ví dụ ước lượng khoảng



Chương 7
Ước lượng và kích thước mẫu

7-1 Giới thiệu

7-2 Ước lượng trung bình quần thể

7-3 Ước lượng tỉ lệ quần thể

Khái niệm chính

1. Giá trị trung bình $\overline{x}$ của mẫu là ước lượng điểm tốt nhất của trung bình quần thể μ.
2. Sử dụng dữ liệu mẫu để xây dựng một khoảng ước lượng cho giá trị trung bình quần thể.
3. Xác định kích thước mẫu cần thiết để ước lượng trung bình quần thể.

Khái niệm chính

Trung bình mẫu $\bar{x}$ là ước lượng điểm tốt nhất của trung bình quần thể μ.

Khoảng ước lượng cho trung bình μ

- Khoảng ước lượng cho trung bình μ của quần thể là: $\bar{X} \pm E$ hay

$$\bar{X} - E < \mu < \bar{X} + E$$

(trong đó $\bar{X}$ là trung bình của

tập mẫu có kích thước n,

E gọi là biên độ lỗi) .

- Vấn đề là: E được tính như thế nào?

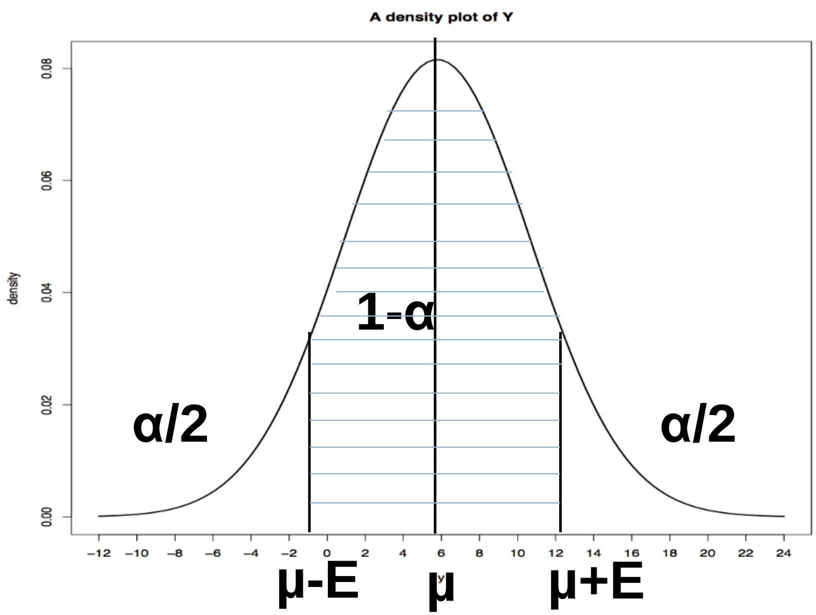
Khoảng ước lượng cho trung bình μ

- E được tính dựa vào định lý giới hạn trung tâm.
- Định lý giới hạn trung tâm: khi ta lấy mẫu ngẫu nhiên, kích thước tập mẫu càng lớn thì phân phối xác suất của đặc trưng trung bình của tập mẫu càng gần với phân phối chuẩn.

Khoảng ước lượng cho trung bình μ

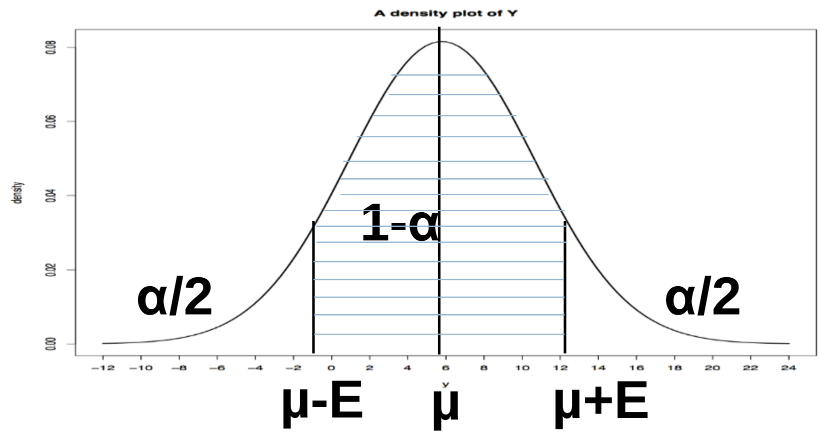
- Theo định lý giới hạn trung tâm, nếu n đủ lớn, $\overline{X}$ có phân phối chuẩn với kỳ vọng là μ, phương sai là $\frac{\sigma^2}{n}$
- Trong đó μ là trung bình của quần thể và σ^2 là phương sai của quần thể

Phân phối chuẩn của $\overline{X}$



Khoảng ước lượng cho trung bình μ

Phân phối chuẩn của $\overline{X}$



- $P(\mu - E < \overline{\boldsymbol{X}} < \mu + E) = 1 - \alpha$
- $P(\mu - E < \overline{\boldsymbol{X}} < \mu + E) = P(\overline{\boldsymbol{X}} < \mu + E) - P(\overline{\boldsymbol{X}} < \mu - E)$
- $= P(\overline{\boldsymbol{X}} - E < \mu) - P(\overline{\boldsymbol{X}} + E < \mu)$
- $= 1 - P(\overline{\boldsymbol{X}} - E > \mu) - (1 - P(\overline{\boldsymbol{X}} + E > \mu))$
- $= P(\overline{\boldsymbol{X}} + E > \mu) - P(\overline{\boldsymbol{X}} + E < \mu)$
- $= P(\mu < \overline{\boldsymbol{X}} + E) - P(\mu > \overline{\boldsymbol{X}} - E)$
- Suy ra: **$P(\overline{X} - E < \mu < \overline{X} + E) = 1 - \alpha$**

Định nghĩa

Độ tin cậy (confidence level) là xác suất 1 - α (thường được biểu thị bằng giá trị phần trăm tương đương) mà khoảng tin cậy thực sự chứa tham số quần thể, giả định rằng quá trình ước lượng được lặp lại một số lượng lớn lần. Độ tin cậy cũng được gọi là bậc tin cậy (degree of confidence), hoặc hệ số tin cậy (confidence coefficient).

Các lựa chọn phổ biến nhất là 90%, 95%, or 99%.

$(\alpha = 0.10)$, $(\alpha = 0.05)$, $(\alpha = 0.01)$

Diễn giải khoảng tin cậy

Chúng ta phải cẩn thận để diễn giải các khoảng tin cậy một cách chính xác.

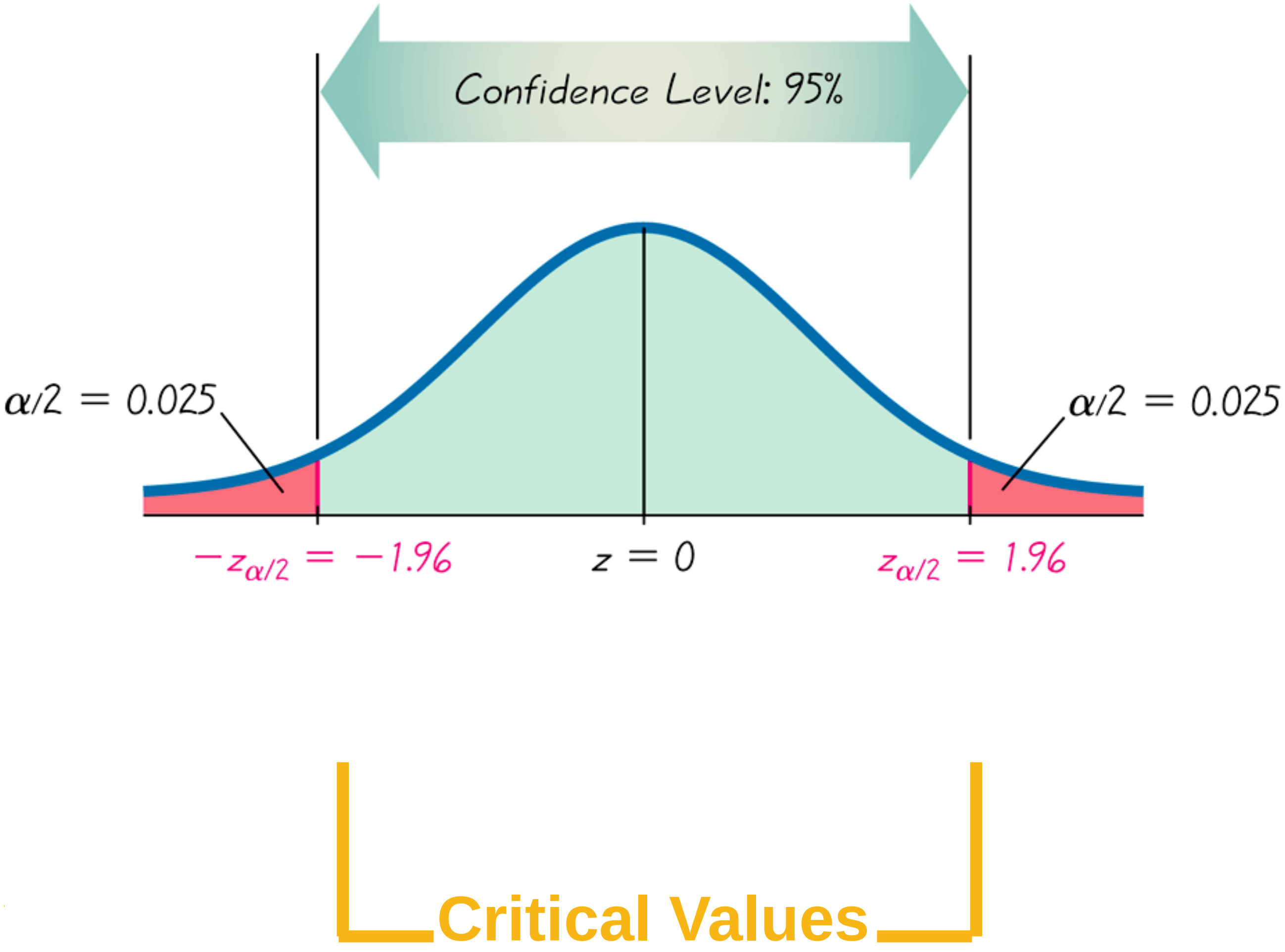
Ví dụ: khoảng tin cậy $0.828 < p < 0.872$.

"Chúng ta tin tưởng 95%(1 - α) rằng khoảng giá trị từ 0.828 đến 0.872 chứa giá trị thực của tỷ lệ quần thể p"

Điều này có nghĩa là nếu chúng ta chọn nhiều mẫu khác nhau có kích thước 1007 và xây dựng các khoảng tin cậy tương ứng, 95% trong số chúng thực sự chứa giá trị của tỷ lệ quần thể p.

(Lưu ý rằng trong cách giải thích đúng này, mức 95% là tỷ lệ thành công của quá trình được sử dụng để ước lượng tỷ lệ.)

Finding $z_{\alpha/2}$ for a 95% Confidence Level



Ví dụ

- Xét quần thể gồm các bao đựng gạo nhỏ. Người ta cần ước lượng trung bình khối lượng của các bao là bao nhiêu gram. Giả sử người ta lấy mẫu gồm 35 bao và cân bao gạo đó, sau đó tính trung bình khối lượng 35 bao. Chọn độ tin cậy là 1-α=0.95. Hãy xây dựng khoảng ước lượng cho trung bình khối lượng của toàn bộ các bao đựng gạo. Giả sử trung bình khối lượng 35 bao là 362.3 gam, độ lệch chuẩn của quần thể là 5 gram

- Gọi μ là trung bình khối lượng của toàn bộ các bao đựng gạo trong quần thể.
- $\bar{X}$ là khối lượng trung bình của 35 bao đựng gạo trong tập mẫu.
- Ta có $\bar{X}$=362.3 gram. Độ tin cậy 1-α=0.95
- Khoảng ước lượng cho trung bình μ của quần thể: $\bar{X} \pm E$ hay
- $\bar{X} - E < \mu < \bar{X} + E$ trong đó $\bar{X}$=362.3, E là giá trị sao cho

$$P(\mu - E < \bar{X} < \mu + E) = P(\bar{X} - E < \mu < \bar{X} + E) = 1 - \alpha$$

Ví dụ

- Xét quần thể gồm các bao đựng gạo nhỏ. Người ta cần ước lượng trung bình khối lượng của các bao là bao nhiêu gram. Giả sử người ta lấy mẫu gồm 35 bao và cân bao gạo đó, sau đó tính trung bình khối lượng 35 bao. Chọn độ tin cậy là 1-α=0.95. Hãy xây dựng khoảng ước lượng cho trung bình khối lượng của toàn bộ các bao đựng gạo. Giả sử trung bình khối lượng 35 bao là 362.3 gam, độ lệch chuẩn của quần thể là 5 gam

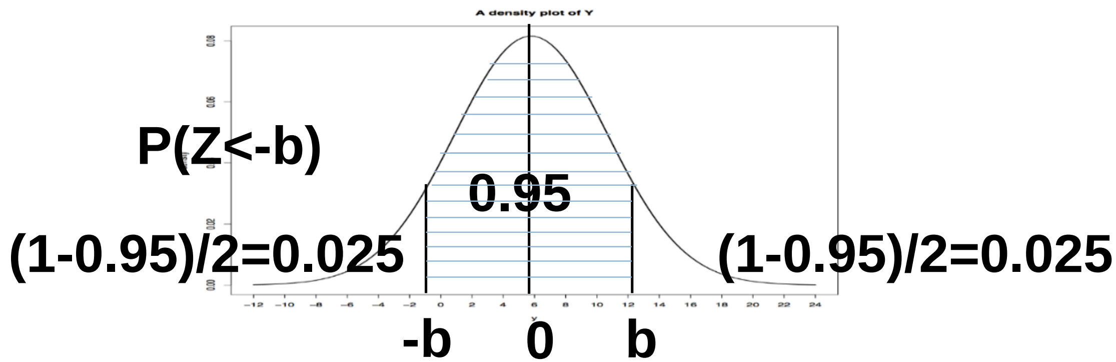
$$\mathbf{P(\mu\text{-}E< \bar{X} <\mu+E)= P(\bar{X}-E<\mu<\bar{X}+E)=1\text{-}\alpha}$$

- Cần tính E. Quay lại bài toán phân phối chuẩn.
- $\bar{X}$ có phân phối chuẩn với kỳ vọng là μ, phương sai là $\frac{\sigma^2}{n}$. Suy ra độ lệch chuẩn là $\frac{\sigma}{\sqrt{n}}$

Ví dụ

$P(\mu - E < \overline{X} < \mu + E) = P(\overline{X} - E < \mu < \overline{X} + E) = 1 - \alpha$

- Chuyển đổi về phân phối Z theo công thức $Z = \dfrac{X - \mu}{\sigma}$.
- Suy ra $X = Z * \sigma_{\bar{x}} + \mu$ thay vào công thức trên ta có
- $P(\mu - E < Z * \sigma_{\bar{x}} + \mu < \mu + E) = P\left(\frac{-E}{\sigma_{\bar{x}}} < Z < \frac{E}{\sigma_{\bar{x}}}\right)$
- Thay $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (σ là độ lệch chuẩn của trung bình quần thể)
- $P\left(\frac{-E}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{E}{\frac{\sigma}{\sqrt{n}}}\right) = P(-b < Z < b) = 1 - \alpha$ (trong đó Z có phân phối chuẩn chính tắc, $b = \frac{-E}{\frac{\sigma}{\sqrt{n}}}$)
- $P(Z < -b) = \alpha/2$
- $\rightarrow b = z_{\alpha/2}$
- Tra bảng Z $\rightarrow b = z_{\alpha/2} = 1.95$

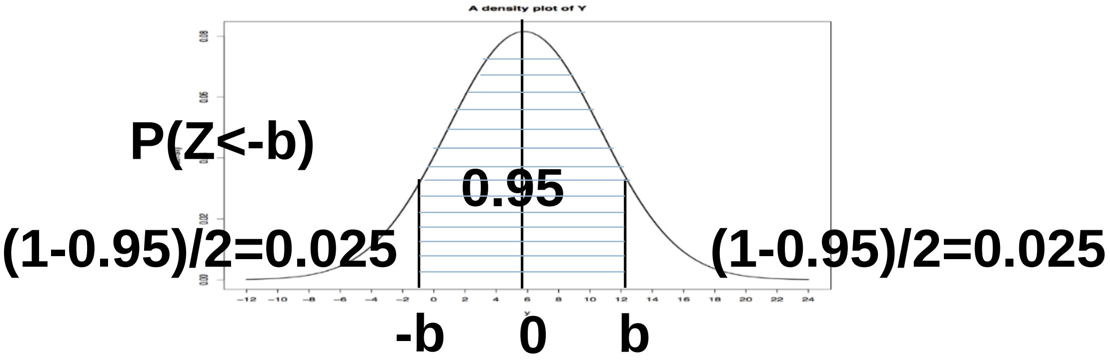


Ví dụ

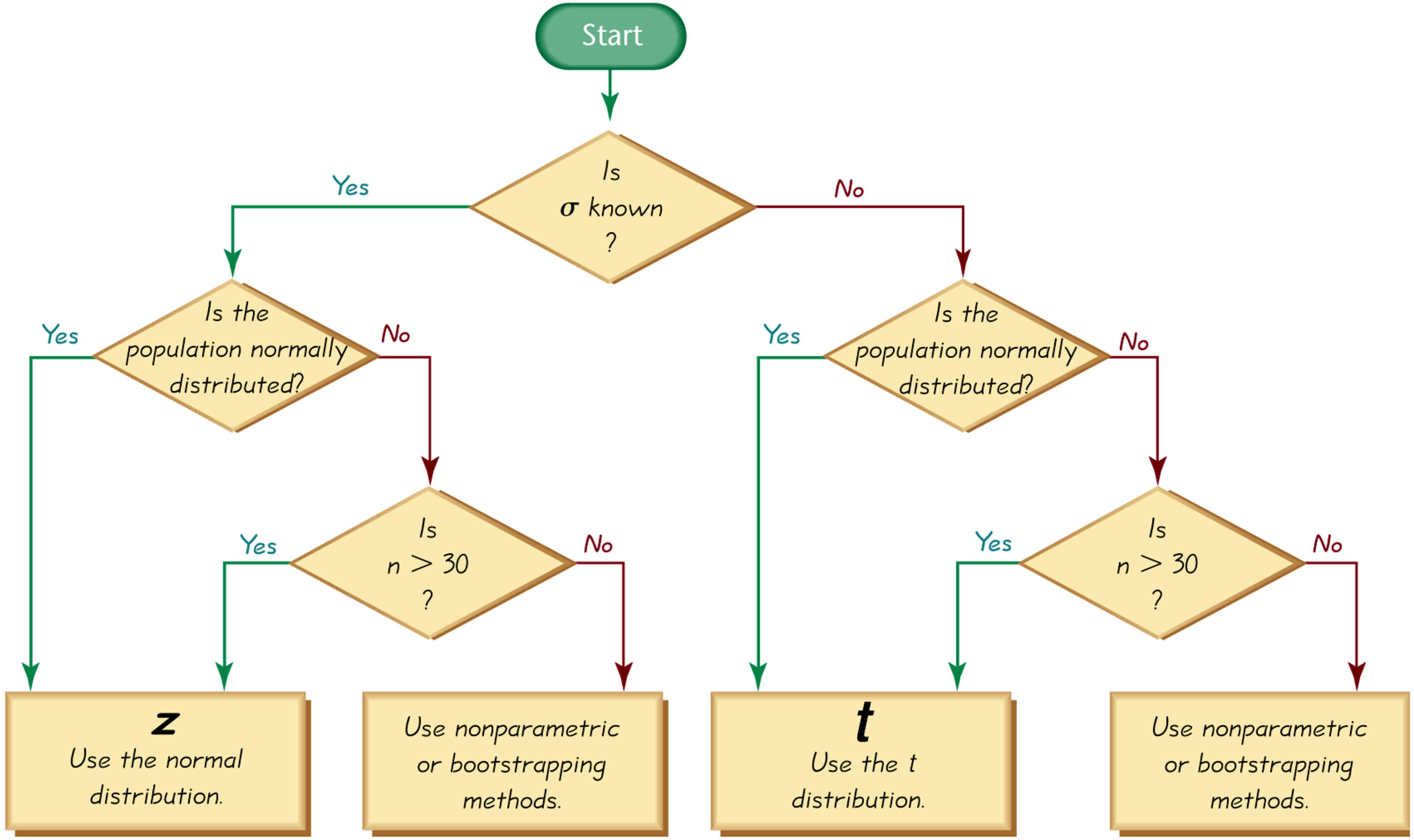
$$P\left(\frac{-E}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{E}{\frac{\sigma}{\sqrt{n}}}\right) = P(-b<Z<b) = 1-\alpha$$

- $\frac{-E}{\frac{\sigma}{\sqrt{n}}} = -z_{\alpha/2} \rightarrow E = z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} = 1.95 * \frac{5}{\sqrt{35}} = 1.65$
- Vậy khoảng ước lượng cho trung bình μ quần thể các bao gạo với độ tin cậy là 0.95 là $\bar{X} \pm E$ hay

$$362.3 - 1.65 < \mu < 362.3 + 1.65$$



Chọn phân phối phù hợp



Chọn phân phối phù hợp

Sử dụng phân phối chuẩn

σ đã biết và quần thể có phân phối chuẩn hoặc $n > 30$

$$\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

Sử dụng phân phối t

σ chưa biết và quần thể có phân phối chuẩn hoặc $n > 30$

$$\bar{x} - t_{\alpha/2} * \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2} * \frac{s}{\sqrt{n}}$$

Sử dụng phương pháp phi tham số

Quần thể không có phân phối chuẩn và $n \le 30$

Phân phối t

Nếu phân phối của quần thể là phân phối chuẩn thì phân phối của

là một phân phối *t* cho tất cả các mẫu có kích thước. Nó thường được gọi là phân phối t và được sử dụng để tìm các critical values được biểu thị bằng .

Định nghĩa

Số bậc tự do của tập hợp dữ liệu mẫu là số các giá trị mẫu có thể khác nhau sau khi có ràng buộc cụ thể trên tất cả các giá trị dữ liệu.

Bậc tự do thường viết tắt là df.

df= n – 1

Đối với các phương pháp trong phần này

Ký hiệu

μ = trung bình quần thể

$\bar{X}$ = trung bình mẫu

s = độ lệch chuẩn của mẫu

n = kích thước mẫu

E = biên độ lỗi

$t_{\alpha/2}$ = critical value t tách một diện tích α / 2 ở đuôi phải của phân bố t.

Khoảng tin cậy cho ước lượng của μ (Với σ không biết)

Trong đó

df = *n* – 1

tra trong bảng A-3.

Ví dụ

Một phát biểu phổ biến là tỏi làm giảm mức cholesterol. Trong một thử nghiệm về hiệu quả của tỏi, 49 đối tượng được điều trị bằng liều tỏi sống, và mức cholesterol của chúng được đo trước và sau khi điều trị.

Những thay đổi về nồng độ cholesterol LDL (mg / dL) có giá trị trung bình là 0,4 và độ lệch chuẩn là 21,0.

Sử dụng số liệu thống kê mẫu n = 49, $\bar{x}$ = 0,4 và s = 21,0 để xây dựng ước lượng khoảng tin cậy 95% của thay đổi ròng trung bình trong cholesterol về nồng độ sau khi điều trị bằng tỏi.

Khoảng tin cậy cho thấy hiệu quả của tỏi trong việc giảm cholesterol về nồng độ là gì?

Ví dụ (tt)

Yêu cầu được thỏa mãn: mẫu ngẫu nhiên đơn giản và n = 49 (tức là n> 30).

95% hàm ý là $\alpha = 0.05$.
Với $n = 49$, df = 49 – 1 = 48
df gần nhất 50, vì vậy = 2.009

Sử dụng = 2.009, $s = 21.0$ và $n = 49$, biên độ lỗi là:

$$E = t_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2.009 \cdot \frac{21.0}{\sqrt{49}} = 6.027$$

Ví dụ (tt)

Xây dựng khoảng tin cậy: $\bar{x} = 0.4, E = 6.027$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$0.4 - 6.027 < \mu < 0.4 + 6.027$$

$$-5.6 < \mu < 6.4$$

Chúng tôi tin tưởng 95% rằng khoảng giá trị -5.6 và 6.4 thực sự chứa giá trị μ, trung bình của những thay đổi về nồng độ cholesterol của quần thể.

Bởi vì giới hạn khoảng tin cậy chứa giá trị là 0, rất có thể là giá trị trung bình của những thay đổi trong nồng độ cholesterol bằng 0, cho thấy rằng việc điều trị bằng tỏi không ảnh hưởng đến nồng độ cholesterol.

Các thuộc tính quan trọng của phân phối t

1. Phân phối t khác nhau đối với các cỡ mẫu khác nhau. (Xem slide sau cho các trường hợp n = 3 và n = 12.)

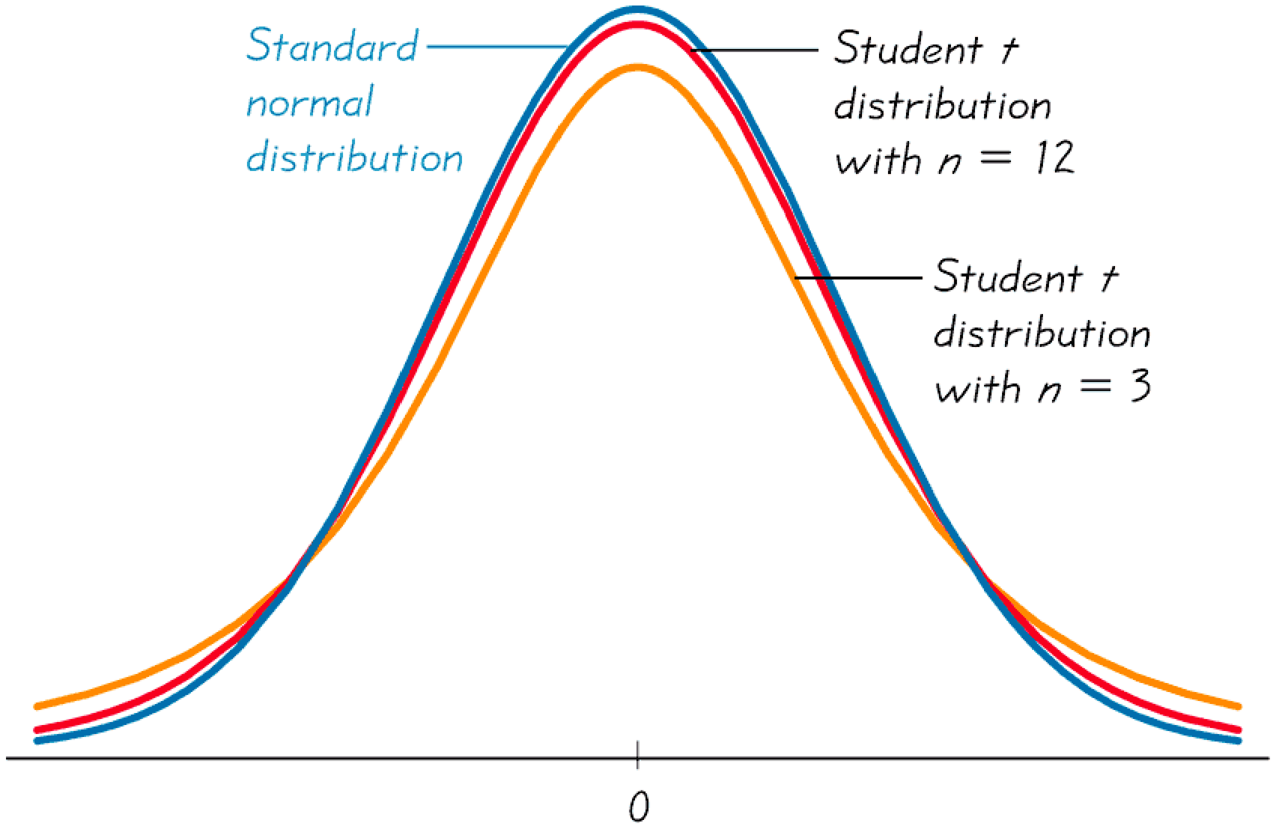
2. Phân phối t có hình dạng chuông đối xứng chung giống như phân bố chuẩn nhưng nó phản ánh sự biến thiên lớn hơn (với phân bố rộng hơn) được mong đợi với các mẫu nhỏ.

3. Phân phối t có giá trị trung bình là t = 0 (giống như phân bố chuẩn có giá trị trung bình là 0).

4. Độ lệch chuẩn của phân phối t thay đổi theo cỡ mẫu và lớn hơn 1 (không giống như phân bố chuẩn, có $\sigma = 1$).

5. Khi cỡ mẫu n lớn hơn, phân phối t sẽ gần hơn với phân bố chuẩn.

Phân phối t với $n = 3$ và $n = 12$



Ví dụ

Người ta đã chết vì tai nạn tàu thuyền và máy bay vì ước tính lỗi thời của trọng lượng trung bình của đàn ông đã được sử dụng.

Trong những thập kỷ gần đây, trọng lượng trung bình của đàn ông đã tăng lên đáng kể, vì vậy chúng tôi cần cập nhật ước tính của mình về phương tiện đó để tàu thuyền, máy bay, thang máy và các thiết bị khác không bị quá tải nguy hiểm.

Sử dụng trọng lượng của đàn ông từ một mẫu ngẫu nhiên, chúng tôi thu thập các số liệu thống kê mẫu này cho mẫu ngẫu nhiên đơn giản:

n = 40 và $\bar{x}$ = 172,55 lb.

Nghiên cứu từ một số nguồn khác cho thấy rằng quần thể trọng lượng của đàn ông có độ lệch chuẩn được đưa ra bởi σ = 26 lb.

Ví dụ

a. Tìm ước lượng điểm tốt nhất về trọng lượng trung bình của quần thể của tất cả nam giới.

b. Xây dựng ước lượng khoảng tin cậy 95% của trọng lượng trung bình của tất cả nam giới.

c. Kết quả đề xuất về trọng lượng trung bình 166,3 lb được sử dụng để xác định khả năng chở khách an toàn của tàu cá vào năm 1960 (như được đưa ra trong khuyến cáo an toàn của Bộ Giao thông và An toàn Quốc gia M-04-04) như thế nào?

Ví dụ

a. Trung bình mẫu 172,55 lb là ước lượng điểm tốt nhất về trọng lượng trung bình của quần thể của tất cả nam giới.

b. Khoảng tin cậy 95% ngụ ý rằng α = 0,05, vì vậy $z_{\alpha/2}$= 1.96.

Tính biên độ lỗi.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{26}{\sqrt{40}} = 8.0574835$$

Xây dựng khoảng tin cậy.

$$\bar{x} - E < \mu < \bar{x} - E$$

$$172.55 - 8.0574835 < \mu < 172.55 + 8.0574835$$

$$164.49 < \mu < 180.61$$

Ví dụ

c. Dựa trên khoảng tin cậy, có thể trọng lượng trung bình 166,3 lb được sử dụng trong năm 1960 có thể là trọng lượng trung bình của nam giới ngày nay.

Tuy nhiên, ước lượng điểm tốt nhất là 172,55 lb cho thấy trọng lượng trung bình của nam giới bây giờ lớn hơn đáng kể so với 166,3 lb.

Kích thước mẫu

Giả sử chúng tôi muốn thu thập dữ liệu mẫu để ước tính trung bình quần thể.

Câu hỏi đặt ra là phải lấy mẫu kích thước bao nhiêu?

Tìm kích thước mẫu để ước lượng trung bình quần thể

= trung bình quần thể

= độ lệch chuẩn của quần thể

$\bar{x}$ = trung bình mẫu

= biên độ lỗi mong đợi

Quy tắc làm tròn cho kích thước mẫu n

Nếu cỡ mẫu được tính n không phải là số nguyên, hãy làm tròn giá trị của n đến số nguyên lớn hơn tiếp theo.

Tìm kích thước mẫu n khi σ không xác định

1. Sử dụng quy tắc sau để ước tính độ lệch chuẩn:

2. Bắt đầu quá trình thu thập mẫu mà không biết σ, sử dụng một vài giá trị đầu tiên, tính toán độ lệch chuẩn mẫu và sử dụng nó thay cho σ. Giá trị ước tính của σ sau đó có thể được cải thiện khi thu được nhiều dữ liệu mẫu hơn và kích thước mẫu có thể được tinh chỉnh cho phù hợp.

3. Ước lượng giá trị của σ bằng cách sử dụng kết quả của một số nghiên cứu trước đó khác.

Ví dụ

Giả sử rằng chúng ta muốn ước tính điểm số IQ trung bình của quần thể sinh viên ngành thống kê. Cần lấy mẫu bao nhiêu sinh viên để làm bài kiểm tra IQ để tin tưởng 95% rằng giá trị ước lượng chứa giá trị trung bình thực sự μ của quần thể với sai số là 3 điểm IQ.

$\alpha = 0.05$

$\alpha/2 = 0.025$

$z_{\alpha/2} = 1.96$

$E = 3$

$\sigma = 15$

Với một mẫu ngẫu nhiên đơn giản chỉ có 97 sinh viên thống kê, chúng ta tin tưởng 95% rằng giá trị ước lượng chứa giá trị trung bình thực sự μ của quần thể với sai số là 3 điểm IQ.

Chương 7
Ước lượng và kích thước mẫu

7-1 Giới thiệu

7-2 Ước lượng trung bình quần thể

7-3 Ước lượng tỉ lệ quần thể

Khái niệm chính

1. Tỉ lệ mẫu là ước lượng điểm tốt nhất của tỉ lệ quần thể.
2. Sử dụng tỉ lệ mẫu để xây dựng khoảng tin cậy để ước lượng giá trị đúng của tỉ lệ quần thể và chúng ta cũng nên biết cách diễn dịch ý nghĩa về khoảng tin cậy.
3. Xác định kích thước mẫu cần thiết để ước lượng tỉ lệ quần thể.

Khái niệm chính

Tỉ lệ mẫu $\hat{p}$ là ước lượng điểm tốt nhất của ước lượng tỉ lệ quần thể p.

Ví dụ

Trung tâm nghiên cứu Pew đã tiến hành một cuộc khảo sát 1007 người lớn và nhận thấy rằng 85% người trong số họ biết Twitter là gì.

Ước lượng điểm tốt nhất của p, tỷ lệ quần thể, là tỷ lệ mẫu:

$$\hat{p} = 0.85$$

Khoảng ước lượng cho tỉ lệ p

Sử dụng tỉ lệ mẫu để ước lượng giá trị đúng của tỉ lệ quần thể:

- Theo chương 6, phân phối tỉ lệ mẫu của biến X là phân phối nhị thức vì thỏa các điều kiện sau:
  - Số lần thí nghiệm của tiến trình ngẫu nhiên đang xét là cố định
  - Hậu quả của thí nghiệm chỉ có thể được phân thành 2 lớp (thành công hay thất bại)
  - Xác suất thành công trong mọi lần thí nghiệm là như nhau
  - Các lần thí nghiệm là độc lập nhau
  - X = số lần thí nghiệm thành công trong n lần thí nghiệm
- Trong pp nhị thức, tính xác suất khi số phép thử lớn (ví dụ như 100) là gần như không thể.
- Phân phối chuẩn có thể được dùng để xấp xỉ phân phối nhị thức khi n lớn.

Khoảng ước lượng cho tỉ lệ p

- Điều kiện để phân phối chuẩn có thể được dùng để xấp xỉ phân phối nhị thức khi n lớn:
  - $np \geq 5$ & $n(1 - p) \geq 5$
  - *n càng lớn thì xấp xỉ càng tốt*.
- Khi phân phối nhị thức xấp xỉ phân phối chuẩn, phân phối của tỉ lệ mẫu có:

$$\mu_{\hat{p}} = \hat{p} \qquad \sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{\boldsymbol{n}}}$$

$$\hat{p} - E \leq p \leq \hat{p} + E$$

$$\hat{p} - z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{\boldsymbol{n}}} \leq p \leq \hat{p} + z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{\boldsymbol{n}}}$$

Quy tắc làm tròn cho ước lượng khoảng tin cậy của p

Làm tròn giới hạn khoảng tin cậy cho p đến **ba chữ số có nghĩa**

Ví dụ

Trong một cuộc thăm dò ý kiến của Trung tâm nghiên cứu Pew của 1007 người lớn được lựa chọn ngẫu nhiên cho thấy 85% người được hỏi biết rằng Twitter là gì. Kết quả mẫu là n = 1007 và $\hat{p} = 0.70$.

a. Tìm biên độ lỗi E tương ứng với mức tin cậy 95%.

b. Ước lượng khoảng tin cậy của tỷ lệ quần thể p với mức tin cậy là 95%.

c. Dựa trên kết quả, chúng ta có thể kết luận một cách chắc chắc rằng hơn 75% người lớn biết Twitter là gì không?

d. Giả sử bạn là phóng viên báo, viết một phát biểu ngắn gọn mô tả chính xác kết quả trong đó bao gồm tất cả thông tin liên quan.

Ví dụ (tt)

Kiểm tra yêu cầu: mẫu ngẫu nhiên đơn giản; số thử nghiệm cố định, 1007; các thử nghiệm độc lập; hai kết quả cho mỗi thử nghiệm; xác suất vẫn không đổi. Lưu ý: số lần thành công và thất bại đều là ít nhất 5.

a) Sử dụng công thức để tìm biên độ của lỗi

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(0.85)(0.15)}{1007}}$$

$$E = 0.0220545$$

Ví dụ (tt)

b) Khoảng tin cậy 95%:

$$\hat{p} - E < p < \hat{p} + E$$

$$0.85 - 0.0220545 < p < 0.85 + 0.0220545$$

$$0.828 < p < 0.872$$

Ví dụ (tt)

c) Dựa trên khoảng tin cậy thu được trong phần (b), có hơn 75% người lớn biết Twitter là gì.

Bởi vì các giới hạn của 0.828 và 0.872 có khả năng chứa tỷ lệ quần thể thực, tỷ lệ quần thể là một giá trị lớn hơn 0,75.

Ví dụ (tt)

d) Dưới đây là một phát biểu tóm tắt kết quả:

85% người lớn ở Hoa Kỳ biết Twitter là gì. Tỷ lệ này dựa trên cuộc thăm dò ý kiến của Trung tâm nghiên cứu Pew với 1007 người lớn được chọn ngẫu nhiên.

Về lý thuyết, trong 95% các cuộc thăm dò như vậy, tỷ lệ phần trăm nên khác nhau không quá 2,2% theo một trong hai hướng từ tỷ lệ phần trăm sẽ được tìm thấy bằng cách phỏng vấn tất cả người lớn ở Hoa Kỳ.

Kích thước mẫu

Giả sử chúng tôi muốn thu thập dữ liệu mẫu để ước tính tỷ lệ quần thể.

Câu hỏi đặt ra là phải lấy mẫu kích thước bao nhiêu?

Xác định kích thước mẫu

(solve for *n* by algebra)

Kích thước mẫu để ước tính tỷ lệ p

Khi ước lượng tỉ lệ đã biết:

Khi ước lượng tỉ lệ chưa biết:

Quy tắc làm tròn để xác định kích thước mẫu

Nếu cỡ mẫu n được tính không phải là số nguyên, hãy làm tròn giá trị của n đến số nguyên lớn hơn tiếp theo.

Ví dụ:

Nhiều công ty quan tâm đến việc biết phần trăm người lớn mua quần áo trực tuyến.

Cần phải lấy mẫu bao nhiêu người lớn để khảo sát để tin tưởng rằng 95% giá trị tỷ lệ thực của quần thể có biên độ lỗi là 3%?

a. Sử dụng kết quả gần đây từ Cục điều tra dân số: 66% người lớn mua quần áo trực tuyến.

b. Giả sử rằng chúng ta không có thông tin nào trước đó cho thấy giá trị xác suất của tỷ lệ.

Ví dụ (tt)

a) Sử dụng $\hat{p} = 0.66$ and $\hat{q} = 1 - \hat{p} = 0.34$

$\alpha = 0.05$ so $z_{\alpha/2} = 1.96$

$E = 0.03$

$$n = \frac{\left(z_{\alpha/2}\right)^2 \hat{p}\hat{q}}{E^2} = \frac{(1.96)^2 (0.66)(0.34)}{(0.03)^2} = 957.839 = 958$$

Để tin tưởng rằng 95% giá trị tỷ lệ thực của quần thể có biên độ lỗi là 3%, cần lấy mẫu ngẫu nhiên đơn giản là 958 người lớn.

Ví dụ (tt)

b) Sử dụng $\alpha = 0.05$ so $z_{\alpha/2} = 1.96$

$E = 0.03$

$$n = \frac{\left(z_{\alpha/2}\right)^2 \cdot 0.25}{E^2}$$

$$= \frac{(1.96)^2 \cdot 0.25}{(0.03)^2}$$

$$= 1067.1111$$

$$= 1068$$

Để tin tưởng rằng 95% giá trị tỷ lệ thực của quần thể có biên độ lỗi là 3%, cần lấy mẫu ngẫu nhiên đơn giản là 1068 người lớn.