Chương 4: Thống kê - Ước lượng tham số

Lê Xuân Lý [1]

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

Hà Nội, tháng 8 năm 2014

[1] Email: lexuanly@gmail.com

Tổng thể

Khi nghiên cứu về một vấn đề người ta thường khảo sát trên một dấu hiệu nào đó, các dấu hiệu này được thể hiện trên nhiều phần tử.

Định nghĩa 1.1

Tập hợp các phần tử mang dấu hiệu ta quan tâm được gọi là tổng thể hay đám đông (population).



Một số lý do không thể khảo sát toàn bộ tổng thể

- **Giới hạn về thời gian, tài chính**: Ví dụ muốn khảo sát xem chiều cao của thanh niên VN hiện nay có tăng lên hay không ta phải khảo sát toàn bộ thanh niên VN (giả sử là 40 triệu người). Để khảo sát hết sẽ tốn nhiều thời gian và kinh phí. Ta có thể khảo sát một triệu thanh niên VN, từ chiều cao trung bình thu được ta suy ra chiều cao trung bình của người VN.
- **Phá vỡ tổng thể nghiên cứu**: Ví dụ ta cất vào kho $N = 10000$ hộp sản phẩmvà muốn biết tỷ lệ hộp hư sau 1 năm bảo quản. Ta phải kiểm tra từng hộp để xác định số hộp hư $M = 300$, tỷ lệ hộp hư trong kho là M/N. Một hộp sản phẩm sau khi kiểm tra thì mất phẩm chất, và vì vậy sau khi kiểm tra cả kho thì cũng "tiêu" luôn kho. Ta có thể lấy ngẫu nhiên $n = 100$ hộp ra kiểm tra, giả sử có $m = 9$ hộp bị hư. Tỷ lệ hộp hư 9% ta suy ra tỷ lệ hộp hư của cả kho.
- **Không xác định được chính xác tổng thể**: Ví dụ muốn khảo sát tỷ lệ người bị nhiễm HIV qua đường tiêm chích là bao nhiêu. Tổng thể lúc này là toàn bộ người bị nhiễm HIV, nhưng ta không thể xác định chính xác là bao nhiêu người (những người xét nghiệm thì bệnh viện biết, những người không xét nghiệm thì ...). Do đó ta chỉ biết một phần tổng thể. Ngoài ra số người bị nhiễm HIV mới và bị chết do HIV thay đổi liên tục nên tổng thể thay đổi liên tục.

Tập mẫu

Do đó người ta nghĩ ra cách thay vì khảo sát tổng thể, người ta chỉ cần chọn ra một tập nhỏ để khảo sát và đưa ra quyết định.

Định nghĩa 1.2

- *Tập mẫu là tập con của tổng thể và có tính chất tương tự như tổng thể.*
- *Số phần tử của tập mẫu được gọi là kích thước mẫu.*



Câu hỏi: Làm sao chọn được tập mẫu có tính chất tương tự như tổng thể để các kết luận của tập mẫu có thể dùng cho tổng thể ?

Một số cách chọn mẫu cơ bản

Một số cách chọn mẫu

- Chọn mẫu ngẫu nhiên có hoàn lại: Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó. Sau đó trả phần tử đó lại tổng thể trước khi lấy 1 phần tử khác. Tiếp tục như thế n lần ta thu được một mẫu có hoàn lại gồm n phần tử.
- Chọn mẫu ngẫu nhiên không hoàn lại: Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó rồi để qua một bên, không trả lại tổng thể. Sau đó lấy ngẫu nhiên 1 phần tử khác, tiếp tục như thế n lần ta thu được một mẫu không hoàn lại gồm n phần tử.
- Chọn mẫu phân nhóm: Đầu tiên ta chia tập nền thành các nhóm tương đối thuần nhất, từ mỗi nhóm đó chọn ra một mẫu ngẫu nhiên. Tập hợp tất cả mẫu đó cho ta một mẫu phân nhóm. Phương pháp này dùng khi trong tập nền có những sai khác lớn. Hạn chế là phụ thuộc vào việc chia nhóm.
- Chọn mẫu có suy luận: dựa trên ý kiến của chuyên gia về đối tượng nghiên cứu để chọn mẫu.

Biểu diễn dữ liệu

Từ tổng thể ta trích ra tập mẫu có n phần tử. Ta có n số liệu.

Dạng liệt kê

Các số liệu thu được ta ghi lại thành dãy số liệu:

$$x_1, x_2, \ldots, x_n$$

Dạng rút gọn

Số liệu thu được có sự lặp đi lặp lại một sô giá trị thì ta có dạng rút gọn sau:

- Dạng tần số: $(n_1 + n_2 + \ldots + n_k = n)$

Giá trị	x_1	x_2	$\ldots$	x_k
Tần số	n_1	n_2	$\ldots$	n_k

- Dạng tần suất: $(p_k = n_k/n)$

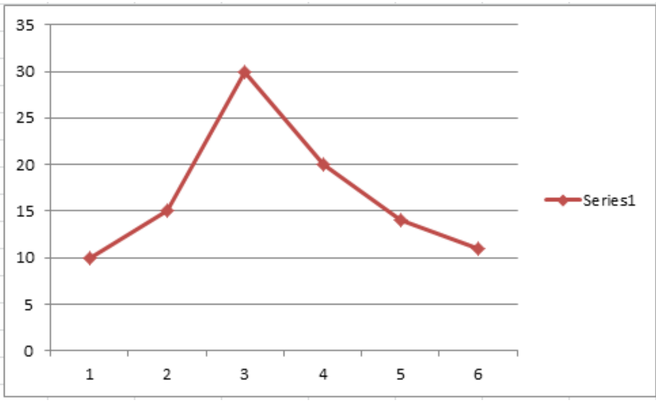
Giá trị	x_1	x_2	$\ldots$	x_k
Tần suất	p_1	p_2	$\ldots$	p_k

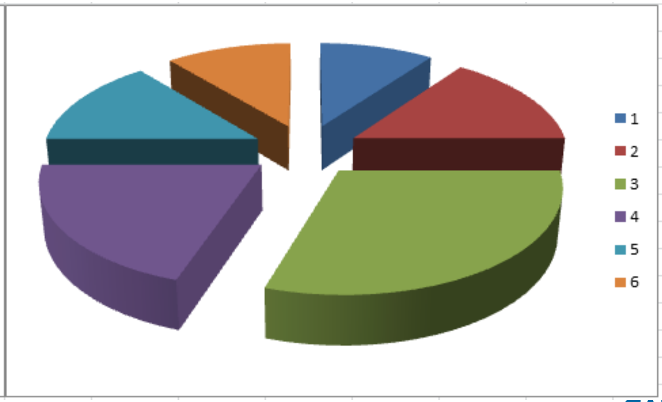
Biểu diễn dữ liệu

Ví dụ dạng rút gọn

Ta có bảng số liệu như sau:

Giá trị	1	2	3	4	5	6
Tần số	10	15	30	20	14	11
Tần suất	0.10	0.15	0.30	0.20	0.14	0.11





Biểu diễn dữ liệu

Dạng khoảng

Dữ liệu thu được nhận giá trị trong (a, b). Ta chia (a, b) thành k miền con bởi các điểm chia: $a_0 = a < a_1 < a_2 < ... < a_{k-1} < a_k = b$.

- Dạng tần số: $(n_1 + n_2 + \ldots + n_k = n)$

Giá trị	$(a_0 - a_1]$	$(a_1 - a_2]$	$\ldots$	$(a_{k-1} - a_k]$
Tần số	n_1	n_2	$\ldots$	n_k

- Dạng tần suất: $(p_k = n_k/n)$

Giá trị	$(a_0 - a_1]$	$(a_1 - a_2]$	$\ldots$	$(a_{k-1} - a_k]$
Tần suất	p_1	p_2	$\ldots$	p_k

- Một số vấn đề chú ý:
  - $k = 5 \to 15$.
  - Độ dài các khoảng thường chia bằng nhau.

Biểu diễn dữ liệu

Dạng khoảng

- Nếu độ dài các khoảng bằng nhau ta có thể chuyển về dạng rút gọn.

Giá trị	x_1	x_2	$\dots$	x_k
Tần suất	p_1	p_2	$\dots$	p_k

Trong đó x_i là điểm đại diện cho $(a_{i-1}, a_i]$ thường được xác định là trung điểm của miền: $x_i = \frac{1}{2}(a_{i-1} + a_i)$

- Dạng rút gọn thường được thể hiện bằng đồ thị dạng đường hoặc dạng hình tròn.
- Dạng khoảng thường được thể hiện bằng đồ thị dạng hình cột.

Mẫu ngẫu nhiên

Tổng thể được đặc trưng bởi dấu hiệu nghiên cứu X là một biến ngẫu nhiên. Do đó khi nói về X là nói về tổng thể.
Từ tổng thể trích ra n phần tử làm một tập mẫu. Ta có 2 loại tập mẫu: *mẫu ngẫu nhiên* và *mẫu cụ thể*
Gọi X_i là biến ngẫu nhiên chỉ giá trị thu được của phần tử thứ $i, i = 1, 2, \dots, n$. Ta có $X_1, X_2, \dots, X_n$ là n biến ngẫu nhiên độc lập và có cùng phân phối với biến ngẫu nhiên X.

Định nghĩa 2.1

- *Mẫu ngẫu nhiên: là véctơ $W_X = (X_1, X_2, \dots, X_n)$, trong đó mỗi thành phần X_i là một biến ngẫu nhiên. Các biến ngẫu nhiên này độc lập và có cùng phân phối xác suất với X.*
- *Mẫu cụ thể: là véctơ $W_x = (x_1, x_2, \dots, x_n)$, trong đó mỗi thành phần x_i là một giá trị cụ thể.*
- *Với một mẫu ngẫu nhiên thì có nhiều mẫu cụ thể ứng với các lần lấy mẫu khác nhau.*

Mẫu ngẫu nhiên

Ví dụ 1

Một kệ chứa các đĩa nhạc với giá như sau:

Giá (ngàn đồng)	20	25	30	34	40
Số đĩa	35	10	25	17	13

Ta cần lấy 4 đĩa có hoàn lại để khảo sát.
Ta xét trong 2 trường hợp:

- *Xét về mặt định lượng: giá của từng đĩa là bao nhiêu?*
- *Xét về mặt định tính: đĩa đó có phải đĩa lậu không?*
  (Đĩa lậu là đĩa có giá dưới 25 ngàn đồng)

Mẫu ngẫu nhiên

Xét tổng thể về mặt định lượng

Lấy ngẫu nhiên một đĩa nhạc trong kệ. Gọi X là giá của đĩa nhạc này. Ta có bảng phân phối xác suất của X.

X	20	25	30	34	40
P	$0,35$	$0,10$	$0,25$	$0,17$	$0,13$

- Lấy ngẫu nhiên có hoàn lại 4 đĩa nhạc từ kệ.
  Gọi X_i là giá của đĩa nhạc thứ i lấy được, $i = 1, 2, 3, 4$.
- Ta thấy các biến X_i độc lập và có cùng phân phối xác suất với X.
- Ta có $W_X = (X_1, X_2, X_3, X_4)$ là một mẫu ngẫu nhiên.
- Bây giờ ta khảo sát giá cụ thể của 4 đĩa lấy ra, ta thấy:
  - Đĩa 1: giá 20 ngàn đồng
  - Đĩa 2: giá 30 ngàn đồng
  - Đĩa 3: giá 20 ngàn đồng
  - Đĩa 4: giá 40 ngàn đồng

  Lập $W_x = (x_1, x_2, x_3, x_4) = (20, 30, 20, 40)$, đây là mẫu cụ thể.

Mẫu ngẫu nhiên

Xét tổng thể về mặt định tính

Đĩa có giá dưới 25 ngàn đồng là đĩa "lậu". Lấy ngẫu nhiên một đĩa từ kệ.
Gọi X là số đĩa lậu lấy được.

X	0	1
P	$0,65$	$0,35$

- Lấy ngẫu nhiên có hoàn lại 4 đĩa nhạc từ kệ.
  Gọi X_i là só đĩa lậu lấy được khi lấy một đĩa lần thứ i, $i = 1, 2, 3, 4$.
- Ta thấy các biến X_i độc lập và có cùng phân phối xác suất với X.
- Ta có $W_X = (X_1, X_2, X_3, X_4)$ là một mẫu ngẫu nhiên.
- Bây giờ ta khảo sát giá cụ thể của 4 đĩa lấy ra, ta thấy:
  • Đĩa 1: giá 20 ngàn đồng $\to x_1 = 1$
  • Đĩa 2: giá 30 ngàn đồng $\to x_2 = 0$
  • Đĩa 3: giá 20 ngàn đồng $\to x_3 = 1$
  • Đĩa 4: giá 40 ngàn đồng $\to x_4 = 0$
  Lập $W_x = (x_1, x_2, x_3, x_4) = (1, 0, 1, 0)$, đây là mẫu cụ thể.

Các đặc trưng mẫu

Thống kê

Cho $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên.
Biến ngẫu nhiên $Y = g(X_1, X_2, ..., X_n)$ (với g là một hàm nào đó) được gọi là một thống kê

Các tham số đặc trưng

- *Xét tổng thể về mặt định lượng*: tổng thể được đặc trưng bởi dấu hiệu nghiên cứu X, (X là biến ngẫu nhiên). Ta có:
  • Trung bình tổng thể: $EX = \mu$
  • Phương sai tổng thể: $VX = \sigma^2$
  • Độ lệch chuẩn của tổng thể: σ.
- *Xét tổng thể về mặt định tính*: tổng thể có kích thướcN, trong đó có M phần tử có tính chất A. Khi đó $p = M/N$ gọi là tỷ lệ xảy ra A của tổng thể.

Các đặc trưng mẫu

Trung bình mẫu

Cho $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên.

- Thống kê - **Trung bình mẫu ngẫu nhiên**:

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

- Mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì $\overline{X}$ nhận giá trị:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\overline{x}$ được gọi là **trung bình mẫu**.

Nếu mẫu dạng rút gọn thì: $\overline{x} = \frac{1}{k}\sum\limits_{i=1}^{n} x_i n_i$

Các đặc trưng mẫu

Phương sai mẫu(chưa hiệu chỉnh)

Cho $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên.

- Thống kê - **Phương sai mẫu ngẫu nhiên**:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2$$

- Mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì S^2 nhận giá trị:

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

S^2 được gọi là **Phương sai mẫu (chưa hiệu chỉnh)**.

- Vấn đề: $E(S^2) = \dfrac{n-1}{n}\sigma^2$

Các đặc trưng mẫu

Phương sai mẫu hiệu chỉnh

Ta phải hiệu chỉnh đi để thu được giá trị thay thế σ^2 tốt hơn.

- Thống kê - **Phương sai mẫu ngẫu nhiên hiệu chỉnh**:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

- Mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì s^2 nhận giá trị:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

  s^2 được gọi là **Phương sai mẫu hiệu chỉnh**.
- s được gọi là **độ lệch chuẩn mẫu hiệu chỉnh**.

Ước lượng điểm

Vấn đề

Cho biến ngẫu nhiên gốc X có phân phối xác suất đã biết nhưng chưa biết tham số θ nào đó.
Mẫu số liệu thu thập được của X là: $(x_1, x_2, ..., x_n)$.
Khi đó $\overline{\theta} = g(x_1, x_2, ..., x_n)$ được gọi là một ước lượng điểm của θ
Muốn biết ước lượng này tốt hay xấu ta phải so sánh với θ.

Ước lượng không chệch

Thống kê $\overline{\theta}$ được gọi là ước lượng không chệch của θ nếu thoả mãn: $E\overline{\theta} = \theta$

Kết quả

Cho biến ngẫu nhiên X có $EX = \mu, VX = \sigma^2$. Mẫu số liệu quan sát $(x_1, x_2, ..., x_n)$.
Khi đó ta có kết quả:

- Ước lượng không chệch cho μ là: $\overline{x}$
- Ước lượng không chệch cho σ^2 là: s^2

Xác định ước lượng điểm

Ví dụ 2

Điều tra năng suất lúa trên diện tích 100 hécta trồng lúa của một vùng, ta thu được bảng số liệu sau:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Số ha có năng suất tương ứng	10	20	30	15	10	10	5

a. Tính trung bình mẫu, độ lệch chuẩn mẫu hiệu chỉnh.
b. Những thửa ruộng có năng suất từ 48 tạ/ha trở lên là những thửa ruộng có năng suất cao. Tính trung bình mẫu, độ lệch chuẩn mẫu hiệu chỉnh của những thửa ruộng có năng suất cao.

$\overline{x} = 46; s = 3,30$

Xác định ước lượng điểm

Ví dụ 3

Quan sát tuổi thọ của một số người ta có bảng số liệu sau:

Tuổi(năm)	20-30	30-40	40-50	50-60
Số người	5	14	25	6

Tính trung bình mẫu, độ lệch chuẩn mẫu hiệu chỉnh của biến ngẫu nhiên X chỉ tuổi thọ của con người.

$\overline{x} = 41,4; s = 8,271$

Ước lượng khoảng

- Cho biến ngẫu nhiên X có $EX = \mu, VX = \sigma^2$.
  Mẫu cụ thể của X là $(x_1, x_2, ..., x_n)$
  Chú ý: nếu cỡ mẫu $n \leq 30$ thì ta phải thêm điều kiện $X \sim N(\mu, \sigma^2)$.
- Bài toán đặt ra là tìm khoảng ước lượng cho μ với xác suất xảy ra bằng $(1-\alpha)$ cho trước. Điều đó tương đương với việc tim khoảng (a, b) sao cho:

$$P(a < \mu < b) = 1 - \alpha$$

• (a, b) được gọi là khoảng tin cậy (hoặc khoảng ước lượng) của μ.
• $(1-\alpha)$ được gọi là độ tin cậy.

Ước lượng khoảng cho kỳ vọng

Trường hợp 1: σ đã biết

- Chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0; 1)$
- Xét cặp số không âm α_1, α_2 thoả mãn: $\alpha_1 + \alpha_2 = \alpha$ và các phân vị chuẩn tắc $u_{\alpha_1}, u_{1-\alpha_2}$:
  • $P(Z < u_{\alpha_1}) = \alpha_1$. Do tính chất của phân phối chuẩn tắc: $u_{\alpha_1} = -u_{1-\alpha_1}$
  • $P(Z < u_{1-\alpha_2}) = 1 - \alpha_2$
  Suy ra $P(-u_{1-\alpha_1} < Z < u_{1-\alpha_2}) = P(u_{\alpha_1} < Z < u_{1-\alpha_2})$
  $= P(Z < u_{1-\alpha_2}) - P(Z < u_{\alpha_1}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$
- $1 - \alpha = P(-u_{1-\alpha_1} < Z < u_{1-\alpha_2}) = P(-u_{1-\alpha_1} < \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} < u_{1-\alpha_2})$
  $= P(\overline{X} - u_{1-\alpha_2}\dfrac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{1-\alpha_1}\dfrac{\sigma}{\sqrt{n}})$
- Từ mẫu cụ thể $(x_1, x_2, .., x_n)$, ta có khoảng ước lượng cho μ với độ tin cậy $1 - \alpha$ là:

$$(\overline{x} - u_{1-\alpha_2}\frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\alpha_1}\frac{\sigma}{\sqrt{n}})$$

Như vậy có vô số khoảng ước lượng cho μ.

Ước lượng khoảng cho kỳ vọng

Trường hợp 1: σ đã biết

- Khoảng ước lượng đối xứng ($\alpha_1 = \alpha_2 = \alpha/2$):

$$\boxed{(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})}, \text{hàm laplace: } \phi(u_{1-\frac{\alpha}{2}}) = 0,5 - \frac{\alpha}{2}$$

  trong đó $\epsilon = u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ gọi là *độ chính xác của ước lượng*.

  Chú ý: Khoảng đối xứng là khoảng ước lượng có độ dài ngắn nhất.
- Khoảng ước lượng một phía ($\alpha_1 = \alpha; \alpha_2 = 0$):

$$\boxed{(-\infty; \overline{x} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}})}, \text{hàm laplace: } \phi(u_{1-\alpha}) = 0,5 - \alpha$$

- Khoảng ước lượng một phía ($\alpha_1 = 0; \alpha_2 = \alpha$):

$$\boxed{(\overline{x} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}; +\infty)}, \text{hàm laplace: } \phi(u_{1-\alpha}) = 0,5 - \alpha$$

Ước lượng khoảng cho kỳ vọng

Ví dụ

Doanh thu của một cửa hàng là biến ngẫu nhiên X(triệu/tháng) có độ lệch chuẩn 2 triệu/tháng. Điều tra ngẫu nhiên doanh thu của 500 cửa hàng có qui mô tương tự nhau ta tính được doanh thu trung bình là 10 triệu/tháng. Với độ tin cậy 95% hãy ước lượng khoảng cho doanh thu trung bình của cửa hàng thuộc qui mô đó.

Bài làm

- X là doanh thu của cửa hàng loại đang xét, $EX = \mu$, $VX = \sigma^2$ với $\sigma = 2$

  Chọn thống kê: $Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0;1)$
- Khoảng ước lượng đối xứng cho doanh thu trung bình μ là:

  $(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$
- Với $\overline{x} = 10, \sigma = 2, n = 500$

  $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$
- Thay các số liệu vào khoảng trên ta có kết quả: (9,825 ; 10,175)

Ước lượng khoảng cho kỳ vọng

Trường hợp 2: σ chưa biết

Do σ chưa biết nên ta thay thế bằng s.

- Chọn thống kê: $Z = \frac{\overline{X} - \mu}{s}\sqrt{n} \sim t(n-1)$
- Làm tương tự như trường hợp 1, ta chỉ thay phân vị chuẩn bằng phân vị Student.
- Mẫu cụ thể $(x_1, x_2, .., x_n)$, ta có khoảng ước lượng cho μ với độ tin cậy $1-\alpha$ là:

$$(\overline{x} - t(n-1, 1-\alpha_2)\frac{s}{\sqrt{n}}; \overline{x} + t(n-1, 1-\alpha_1)\frac{s}{\sqrt{n}})$$

- **Chú ý**:
  $n > 30$ thì phân phối chuẩn và phân phối student bậc tự do $(n-1)$ có thể coi là một.
  Do đó nếu $n > 30$ ta có thể chọn thống kê: $Z = \frac{\overline{X} - \mu}{s}\sqrt{n} \sim N(0;1)$
  Khoảng ước lượng cho μ với độ tin cậy $1-\alpha$ là:

$$(\overline{x} - u_{1-\alpha_2}\frac{s}{\sqrt{n}}; \overline{x} + u_{1-\alpha_1}\frac{s}{\sqrt{n}})$$

Ước lượng khoảng cho kỳ vọng

Trường hợp 2: σ chưa biết

- Khoảng ước lượng đối xứng $(\alpha_1 = \alpha_2 = \alpha/2)$:

$$\boxed{(\overline{x} - t(n-1, 1-\frac{\alpha}{2})\frac{s}{\sqrt{n}}; \overline{x} + t(n-1, 1-\frac{\alpha}{2})\frac{s}{\sqrt{n}})}$$

- Khoảng ước lượng một phía $(\alpha_1 = \alpha; \alpha_2 = 0)$:

$$\boxed{(-\infty; \overline{x} - t(n-1, 1-\alpha)\frac{s}{\sqrt{n}})}$$

- Khoảng ước lượng một phía $(\alpha_1 = 0; \alpha_2 = \alpha)$:

$$\boxed{(\overline{x} - t(n-1, 1-\alpha)\frac{s}{\sqrt{n}}; +\infty)}$$

Ước lượng khoảng cho kỳ vọng

Ví dụ

Ví dụ trước sẽ hợp với thực tế hơn nếu ta sửa lại như sau:
Doanh thu của một cửa hàng là biến ngẫu nhiên X(triệu/tháng). Điều tra ngẫu nhiên doanh thu của 500 cửa hàng có qui mô tương tự nhau ta tính được doanh thu trung bình là 10 triệu/tháng và độ lệch chuẩn mẫu hiệu chỉnh là 2 triệu/tháng. Với độ tin cậy 95% hãy ước lượng khoảng cho doanh thu trung bình của cửa hàng thuộc qui mô đó.

Bài làm

- X(triệu/tháng) là doanh thu của cửa hàng loại đang xét, $EX = \mu$, $VX = \sigma^2$
  Chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{s}\sqrt{n} \sim t(n-1)$
- Khoảng ước lượng đối xứng cho doanh thu trung bình μ là:
  $(\overline{x} - t(n-1, 1-\frac{\alpha}{2})\dfrac{s}{\sqrt{n}}; \overline{x} + t(n-1, 1-\frac{\alpha}{2})\dfrac{s}{\sqrt{n}})$
- Với $\overline{x} = 10, s = 2, n = 500$
  $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow t(n-1, 1-\frac{\alpha}{2}) = t(499; 0,975) = 1,96$
- Thay các số liệu vào khoảng trên ta có kết quả: (9,825 ; 10,175)

Ước lượng khoảng cho tỷ lệ

Bài toán

Xác suất xảy ra sự kiện A là p.
Do không biết p nên người ta thực hiện n phép thử độc lập, cùng điều kiện.
Trong đó có m phép thử xảy ra A.
$f = {}^{m}/_{n}$ là ước lượng điểm không chệch cho p.
Câu hỏi: Với độ tin cậy $(1-\alpha)$ hãy ước lượng khoảng cho p.

Cách giải quyết: tương tự cách làm cho kỳ vọng

- Chọn thống kê: $Z = \dfrac{f - p}{\sqrt{p(1-p)}}\sqrt{n} \sim N(0;1)$
- Tuy nhiên do khó giải quyết nên người ta thay p dưới mẫu bởi f cho dễ tính.
  Thống kê trở thành: $Z = \dfrac{f - p}{\sqrt{f(1-f)}}\sqrt{n} \sim N(0;1)$
- Mẫu cụ thể $(x_1, x_2, .., x_n)$, ta có khoảng ước lượng cho p với độ tin cậy $1-\alpha$ là:

$$(f - u_{1-\alpha_2}\sqrt{\frac{f(1-f)}{n}}, f + u_{1-\alpha_1}\sqrt{\frac{f(1-f)}{n}})$$

Ước lượng khoảng cho tỷ lệ

Các trường hợp ước lượng hay dùng

- Khoảng ước lượng đối xứng ($\alpha_1 = \alpha_2 = \alpha/2$):

$$\boxed{(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}, f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}})}$$

- Khoảng ước lượng một phía ($\alpha_1 = \alpha; \alpha_2 = 0$):

$$\boxed{(-\infty; f + u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}})}$$

- Khoảng ước lượng một phía ($\alpha_1 = 0; \alpha_2 = \alpha$):

$$\boxed{(f - u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}}; +\infty)}$$

- **Chú ý**: Do tỷ lệ chỉ nhận giá trị từ 0 đến 1 nên ta có thể thay giá trị $-\infty$ bằng 0 và $+\infty$ bằng 1 trong khoảng ước lượng một phía.

Ước lượng khoảng cho tỷ lệ

Ví dụ

Tại một bến xe, kiểm tra ngẫu nhiên 100 xe thấy có 30 xe xuất phát đúng giờ. Với độ tin cậy 95% hãy ước lượng khoảng cho tỷ lệ xe xuất phát đúng giờ.

Bài làm

- Gọi p là tỷ lệ xe xuất phát đúng giờ.
  Chọn thống kê: $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n} \sim N(0;1)$
- Khoảng ước lượng đối xứng cho tỷ lệ xe xuất phát đúng giờ là:
  $(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}, f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}})$
- Với $n = 100, m = 30 \Rightarrow f = {}^{m}/_{n} = 0,3$
  $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$
- Thay các số liệu vào khoảng trên ta có kết quả: (0,21 ; 0,39)