Bài giảng Xác suất Thống kê và ứng dụng

Nguyễn Thị Nhung

Bộ môn Toán - Đại học THĂNG LONG

Ngày 21 tháng 8 năm 2013

Chương VII

Ước lượng các tham số tổng thể

Chương VII

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu
2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Chương VII

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu

2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Nội dung chính trong chương

- Giới thiệu khái niệm ước lượng điểm và ước lượng khoảng.
- Giới thiệu về hàm ước lượng điểm và ước lượng điểm của những tham số tổng thể quan trọng như: trung bình, tỉ lệ, phương sai và độ lệch chuẩn.
- Giới thiệu về công thức tìm khoảng tin cậy đối xứng cho trung bình tổng thể trong các trường hợp:
  - Tổng thể tuân theo phân phối chuẩn, phương sai đã biết;
  - Tổng thể tuân theo phân phối chuẩn, phương sai chưa biết;
  - Tổng thể không biết phân phối, cỡ mẫu lớn.
- Giới thiệu về công thức khoảng tin cậy cho tỉ lệ tổng thể, cỡ mẫu lớn.

Những kiến thức sinh viên phải hiểu được trong chương

- Phân biệt ước lượng điểm và ước lượng khoảng.
- Nêu hàm ước lượng điểm và ước lượng điểm của những tham số tổng thể quan trọng như: trung bình, tỉ lệ, phương sai và độ lệch chuẩn.
- Hiểu ý nghĩa của khoảng tin cậy 90%, 95% hay $100(1-\alpha)\%$.
- Đưa ra được công thức khoảng tin cậy đối xứng cho trung bình tổng thể trong các trường hợp:
  - Tổng thể tuân theo phân phối chuẩn, phương sai đã biết;
  - Tổng thể tuân theo phân phối chuẩn, phương sai chưa biết;
  - Tổng thể không biết phân phối, cỡ mẫu lớn.
- Đưa ra công thức khoảng tin cậy cho tỉ lệ tổng thể, cỡ mẫu lớn.

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Bạn đang muốn ước tính điểm thi môn XSTK của sinh viên Thăng Long trong vài năm gần đây. Qua tiến hành điều tra bạn chọn được mẫu gồm điểm thi của 228 sinh viên. Với tập điểm thi này, bạn làm thế nào để trả lời được những câu hỏi sau về điểm của toàn bộ sinh viên Thăng Long:

- *Điểm trung bình khoảng bao nhiêu?*
- *Điểm trung bình tối thiểu là bao nhiêu? Tối đa là bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ khoảng bao nhiêu?*
- *Tỉ lệ số sinh viên thi đỗ tối thiểu là bao nhiêu và tối đa là bao nhiêu?*

Bài toán tình huống

Dữ liệu điểm thi

5.20	1.20	2.20	1.20	4.20	1.20	6.20	1.20	1.20	7.20	1.20	1.32
5.50	2.55	1.55	6.55	1.55	3.55	1.55	6.55	1.55	3.55	1.55	2.55
1.65	5.65	1.70	1.70	8.80	1.89	1.90	7.90	1.90	4.90	1.90	2.90
2.90	1.90	5.90	1.90	5.95	1.95	3.95	6.95	2.00	2.00	3.10	2.14
2.14	2.15	2.23	5.25	2.25	2.25	7.25	7.30	7.30	2.30	2.34	2.35
2.35	2.45	5.45	2.45	2.49	2.50	2.50	2.50	2.50	2.50	2.55	2.55
2.58	2.60	6.60	2.60	2.65	7.67	2.70	2.70	2.79	2.80	8.80	2.80
2.80	8.82	2.84	2.84	2.84	2.90	2.90	2.95	2.95	2.95	2.95	3.00
3.00	3.00	3.04	3.04	8.10	3.13	9.14	3.15	3.15	7.20	5.20	3.20
4.29	4.29	4.29	4.29	7.35	4.37	4.38	4.38	4.40	4.45	4.45	4.45
5.50	5.50	5.50	5.50	5.52	6.55	7.55	6.55	8.58	3.60	4.65	5.70
5.74	5.75	5.85	5.85	5.85	5.85	5.90	5.90	3.95	5.95	5.97	5.05
5.05	5.05	5.10	5.10	5.15	5.15	5.20	5.20	5.20	5.20	5.25	5.30
6.10	6.20	6.30	6.24	6.34	6.45	6.49	6.59	6.75	6.75	6.75	6.80
6.88	6.89	6.90	6.90	7.00	7.00	7.05	7.10	7.10	7.10	7.10	7.14
7.25	7.25	2.25	2.25	2.25	7.30	7.35	7.39	7.40	7.40	7.55	7.59
7.60	7.60	7.60	7.70	7.90	7.90	7.95	8.20	8.20	8.20	8.25	8.25
8.30	8.40	8.50	8.50	8.54	8.60	8.65	8.75	8.95	8.95	8.20	8.22
8.30	8.45	8.45	8.50	8.50	9.00	9.40	9.50	10.00	10.00	10.00	10.00

Bài toán tình huống

Câu hỏi tình huống

Giả sử bạn là nhân viên phòng nghiên cứu của một công ty chuyên về thực phẩm đóng hộp. Bạn được công ty giao cho nghiên cứu chiều dài của một loại dưa chuột để sản xuất hộp đóng cho phù hợp. Bạn chọn ngẫu nhiên ra 100 *quả dưa chuột đo đạc thấy chiều dài trung bình là* 9.3 *và độ lệch chuẩn là* 0.5*, với số liệu này thì bạn ước tính chiều dài trung bình của một quả dưa chuột khoảng là bao nhiêu?*

Bài toán tình huống

Câu hỏi tình huống

Gia đình bạn sở hữu một hồ cá. Để đánh giá trữ lượng cá trong hồ, bạn cho bắt 2000 *con cá, đánh dấu rồi thả xuống hồ. Sau đó bắt lại* 400 *con thì thấy có* 80 *con có dấu. Với số liệu này bạn ước tính trữ lượng cá trong hồ khoảng bao nhiêu?*

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu
2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Hàm ước lượng và ước lượng

Khái niệm

- *Giả sử cần ước lượng tham số θ của tổng thể, ta chọn mẫu ngẫu nhiên gồm n phần tử $X_1, X_2, \ldots, X_n$ và chọn hàm $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$ là một hàm biểu diễn tham số mẫu tương ứng với tham số của tổng thể cần ước lượng. Khi đó hàm $\hat{\theta}$ được gọi là hàm ước lượng của θ.*
- *Nếu lập một mẫu cụ thể và tính các giá trị $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ của hàm tham số mẫu $\hat{\theta}$ thì một giá trị tính trên mẫu cụ thể này được gọi là một ước lượng của θ.*
- *Nếu với mỗi một mẫu cụ thể hàm ước lượng $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ cho một giá trị duy nhất thì $\hat{\theta}$ được gọi là hàm ước lượng điểm của tham số tổng thể θ và giá trị tìm được này gọi là ước lượng điểm của tham số tổng thể θ.*

Hàm ước lượng và ước lượng

Khái niệm

- *Giả sử cần ước lượng tham số θ của tổng thể, ta chọn mẫu ngẫu nhiên gồm n phần tử $X_1, X_2, \ldots, X_n$ và chọn hàm $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$ là một hàm biểu diễn tham số mẫu tương ứng với tham số của tổng thể cần ước lượng. Khi đó hàm $\hat{\theta}$ được gọi là hàm ước lượng của θ.*
- *Nếu lập một mẫu cụ thể và tính các giá trị $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ của hàm tham số mẫu $\hat{\theta}$ thì một giá trị tính trên mẫu cụ thể này được gọi là một ước lượng của θ.*
- *Nếu với mỗi một mẫu cụ thể hàm ước lượng $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ cho một giá trị duy nhất thì $\hat{\theta}$ được gọi là hàm ước lượng điểm của tham số tổng thể θ và giá trị tìm được này gọi là ước lượng điểm của tham số tổng thể θ.*

Hàm ước lượng và ước lượng

Khái niệm

- *Giả sử cần ước lượng tham số θ của tổng thể, ta chọn mẫu ngẫu nhiên gồm n phần tử $X_1, X_2, \ldots, X_n$ và chọn hàm $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$ là một hàm biểu diễn tham số mẫu tương ứng với tham số của tổng thể cần ước lượng. Khi đó hàm $\hat{\theta}$ được gọi là hàm ước lượng của θ.*
- *Nếu lập một mẫu cụ thể và tính các giá trị $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ của hàm tham số mẫu $\hat{\theta}$ thì một giá trị tính trên mẫu cụ thể này được gọi là một ước lượng của θ.*
- *Nếu với mỗi một mẫu cụ thể hàm ước lượng $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ cho một giá trị duy nhất thì $\hat{\theta}$ được gọi là hàm ước lượng điểm của tham số tổng thể θ và giá trị tìm được này gọi là ước lượng điểm của tham số tổng thể θ.*

Ví dụ

Ta muốn ước lượng điểm trung bình môn XSTK của toàn bộ sinh viên Thăng Long. Ta chọn một mẫu ngẫu nhiên gồm n phần tử $X_1, X_2, \ldots, X_n$ thì biến ngẫu nhiên chỉ trung bình mẫu

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

là một hàm ước lượng điểm của trung bình tổng thể sinh viên. Nếu ta chọn được một mẫu cụ thể với các giá trị $x_1, x_2, \ldots, x_n$ của $X_1, X_2, \ldots, X_n$ thì giá trị trung bình của mẫu đã chọn

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

là một ước lượng điểm của điểm trung bình của tổng thể sinh viên.

Một số hàm ước lượng điểm thông dụng

Tham số tổng thể	Hàm ước lượng	Ước lượng
Trung bình (μ_X)	$\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$	$\overline{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$
Phương sai (σ_X^2)	$S_X^2 = \dfrac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$	$s_x^2 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1}$
Độ lệch chuẩn (σ_X)	$S_X = \sqrt{\dfrac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}}$	$s_x = \sqrt{\dfrac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1}}$
Tỉ lệ (p)	$P_X = \dfrac{X}{n}$	$p_x = \dfrac{m}{n}$

Ví dụ

Bài toán

Tìm một ước lượng điểm cho điểm trung bình, phương sai, độ lệch chuẩn của điểm thi môn XSTK của toàn bộ sinh viên Thăng Long và tỉ lệ các sinh viên thi qua môn XSTK của toàn bộ sinh viên Thăng Long.

Ví dụ

Lời giải:

- Một ước lượng điểm của trung bình của tổng thể điểm thi môn XSTK là:

$$\overline{x} = \frac{5.20 + 1.20 + 2.20 + \ldots + 10.00 + 10.00 + 10.0}{228} = 5.15.$$

- Một ước lượng điểm của phương sai điểm thi của tổng thể điểm thi môn XSTK là:

$$s_x^2 = \frac{(1.20 - 3.73)^2 + \ldots + (10.0 - 3.73)^2}{228 - 1} = 6.03.$$

Ví dụ

Lời giải:

- Một ước lượng điểm của trung bình của tổng thể điểm thi môn XSTK là:

$$\overline{x} = \frac{5.20 + 1.20 + 2.20 + \ldots + 10.00 + 10.00 + 10.0}{228} = 5.15.$$

- Một ước lượng điểm của phương sai điểm thi của tổng thể điểm thi môn XSTK là:

$$s_x^2 = \frac{(1.20 - 3.73)^2 + \ldots + (10.0 - 3.73)^2}{228 - 1} = 6.03.$$

Ví dụ

Lời giải:

- Một ước lượng điểm của độ lệch chuẩn điểm thi của tổng thể điểm thi môn XSTK là:

$$s_x = \sqrt{s_x^2} = 2.46.$$

- Một ước lượng điểm của tỉ lệ sinh viên thi đỗ môn XSTK của tổng thể sinh viên Thăng Long

$$p_x = \frac{m}{n} = \frac{128}{228} = 0.56.$$

Ví dụ

Lời giải:

- Một ước lượng điểm của độ lệch chuẩn điểm thi của tổng thể điểm thi môn XSTK là:

$$s_x = \sqrt{s_x^2} = 2.46.$$

- Một ước lượng điểm của tỉ lệ sinh viên thi đỗ môn XSTK của tổng thể sinh viên Thăng Long

$$p_x = \frac{m}{n} = \frac{128}{228} = 0.56.$$

Tính toán trong R

```
> DiemThiXSTK = scan()
1:
Read 228 items

> #Mot uoc luong diem cho diem trung binh cua tong the la:
> mean(DiemThiXSTK)
[1] 5.151711

> #Mot uoc luong diem cho phuong sai diem cua tong the la:
> var(DiemThiXSTK)
[1] 6.031687

> #Mot uoc luong diem cho do lech chuan diem cua tong the la:
> sd(DiemThiXSTK)
[1] 2.455949

> #Mot uoc luong diem cho ti le sinh vien thi do cua tong the la:
> length(DiemThiXSTK[DiemThiXSTK >= 5.0])/length(DiemThiXSTK)
[1] 0.5614035
```

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu
2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Hàm ước lượng không chệch

Định nghĩa

Hàm ước lượng điểm $\hat{\theta}$ của tham số tổng thể θ được gọi là hàm ước lượng không chệch của θ nếu trung bình của $\hat{\theta}$ là θ, tức là $E(\hat{\theta}) = \theta$.

- Ta có $E(\overline{X}) = \mu_X$ nên $\overline{X}$ là một hàm ước lượng không chệch của trung bình tổng thể;
- Ta có $E(S_X^2) = \sigma_X^2$ nên S_X^2 là một hàm ước lượng không chệch của phương sai tổng thể;
- Ta có $E(P_X) = p$ nên P_X là một hàm ước lượng không chệch của tỉ lệ tổng thể;

Hàm ước lượng không chệch

Định nghĩa

Hàm ước lượng điểm $\hat{\theta}$ của tham số tổng thể θ được gọi là hàm ước lượng không chệch của θ nếu trung bình của $\hat{\theta}$ là θ, tức là $E(\hat{\theta}) = \theta$.

- Ta có $E(\overline{X}) = \mu_X$ nên $\overline{X}$ là một hàm ước lượng không chệch của trung bình tổng thể;
- Ta có $E(S_X^2) = \sigma_X^2$ nên S_X^2 là một hàm ước lượng không chệch của phương sai tổng thể;
- Ta có $E(P_X) = p$ nên P_X là một hàm ước lượng không chệch của tỉ lệ tổng thể;

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu
2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Hàm ước lượng không chệch

Định nghĩa

- *Giả sử $\hat{\theta}_1$ và $\hat{\theta}_2$ là hai hàm ước lượng không chệch của tham số tổng thể θ của cùng một mẫu ngẫu nhiên. Khi đó, $\hat{\theta}_1$ gọi là hữu hiệu hơn $\hat{\theta}_2$ nếu $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.*
- *Nếu $\hat{\theta}$ là một hàm ước lượng không chệch của θ mà không có một hàm ước lượng không chệch nào khác của θ hữu hiệu hơn thì $\hat{\theta}$ được gọi là hàm ước lượng không chệch hữu hiệu nhất của θ.*

Các hàm ước lượng sau đây

- Trung bình mẫu từ một phân phối chuẩn;
- Phương sai mẫu từ một phân phối chuẩn;
- Tỉ lệ mẫu từ một phân phối nhị thức

là các hàm ước lượng không chệch hữu hiệu nhất.

Hàm ước lượng không chệch

Định nghĩa

- *Giả sử $\hat{\theta}_1$ và $\hat{\theta}_2$ là hai hàm ước lượng không chệch của tham số tổng thể θ của cùng một mẫu ngẫu nhiên. Khi đó, $\hat{\theta}_1$ gọi là hữu hiệu hơn $\hat{\theta}_2$ nếu $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.*
- *Nếu $\hat{\theta}$ là một hàm ước lượng không chệch của θ mà không có một hàm ước lượng không chệch nào khác của θ hữu hiệu hơn thì $\hat{\theta}$ được gọi là hàm ước lượng không chệch hữu hiệu nhất của θ.*

Các hàm ước lượng sau đây

- Trung bình mẫu từ một phân phối chuẩn;
- Phương sai mẫu từ một phân phối chuẩn;
- Tỉ lệ mẫu từ một phân phối nhị thức

là các hàm ước lượng không chệch hữu hiệu nhất.

Hàm ước lượng không chệch

Định nghĩa

- *Giả sử* $\hat{\theta}_1$ *và* $\hat{\theta}_2$ *là hai hàm ước lượng không chệch của tham số tổng thể* θ *của cùng một mẫu ngẫu nhiên. Khi đó,* $\hat{\theta}_1$ *gọi là hữu hiệu hơn* $\hat{\theta}_2$ *nếu* $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.
- *Nếu* $\hat{\theta}$ *là một hàm ước lượng không chệch của* θ *mà không có một hàm ước lượng không chệch nào khác của* θ *hữu hiệu hơn thì* $\hat{\theta}$ *được gọi là hàm ước lượng không chệch hữu hiệu nhất của* θ.

Các hàm ước lượng sau đây

- Trung bình mẫu từ một phân phối chuẩn;
- Phương sai mẫu từ một phân phối chuẩn;
- Tỉ lệ mẫu từ một phân phối nhị thức

là các hàm ước lượng không chệch hữu hiệu nhất.

Nội dung trình bày

Định nghĩa khoảng tin cậy

Định nghĩa

Khoảng (L, U) *của tham số mẫu* $\hat{\theta}$ *được gọi là khoảng tin cậy của tham số tổng thể* θ *với xác suất bằng* $1-\alpha$ *nếu thỏa mãn điều kiện*

$$P(L < \theta < U) = 1 - \alpha.$$

Xác suất $1-\alpha$ *được gọi là độ tin cậy của ước lượng và thường được viết dưới dạng* $100(1-\alpha)\%$, $I = U - L$ *được gọi là độ rộng của khoảng tin cậy.*

Nhận xét: Khoảng ước lượng đã chỉ rõ mức độ chính xác của sự ước lượng nên được dùng nhiều hơn ước lượng điểm.

Định nghĩa khoảng tin cậy

Định nghĩa

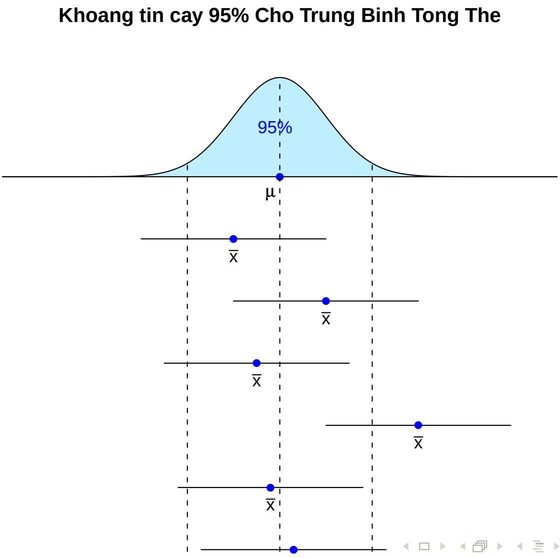
Khoảng (L, U) *của tham số mẫu* $\hat{\theta}$ *được gọi là khoảng tin cậy của tham số tổng thể* θ *với xác suất bằng* $1 - \alpha$ *nếu thỏa mãn điều kiện*

$$P(L < \theta < U) = 1 - \alpha.$$

Xác suất $1 - \alpha$ *được gọi là độ tin cậy của ước lượng và thường được viết dưới dạng* $100(1 - \alpha)\%$, $I = U - L$ *được gọi là độ rộng của khoảng tin cậy.*

Nhận xét: Khoảng ước lượng đã chỉ rõ mức độ chính xác của sự ước lượng nên được dùng nhiều hơn ước lượng điểm.

Minh họa hình học ý nghĩa khoảng tin cậy



Định nghĩa khoảng tin cậy

- Do tham số mẫu $\hat{\theta}$ là một biến ngẫu nhiên nên khoảng (L, U) cũng là một khoảng ngẫu nhiên, với một mẫu cụ thể ta sẽ có một khoảng số cụ thể (l, u).
- Với độ tin cậy $100(1-\alpha)\%$, nếu ta lấy tất cả các mẫu có thể cùng cỡ từ tổng thể, mỗi mẫu ta tính một khoảng ước lượng cụ thể kiểu (L, U) tương ứng với xác suất $(1-\alpha)$ thì ta sẽ có $100(1-\alpha)\%$ khoảng ước lượng tìm được chứa θ.

Định nghĩa khoảng tin cậy

- Do tham số mẫu $\hat{\theta}$ là một biến ngẫu nhiên nên khoảng (L, U) cũng là một khoảng ngẫu nhiên, với một mẫu cụ thể ta sẽ có một khoảng số cụ thể (l, u).
- Với độ tin cậy $100(1-\alpha)\%$, nếu ta lấy tất cả các mẫu có thể cùng cỡ từ tổng thể, mỗi mẫu ta tính một khoảng ước lượng cụ thể kiểu (L, U) tương ứng với xác suất $(1-\alpha)$ thì ta sẽ có $100(1-\alpha)\%$ khoảng ước lượng tìm được chứa θ.

Câu hỏi

Câu hỏi: Hãy nêu lại phân phối của trung bình mẫu và tỉ lệ mẫu.

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu
2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Bài toán

Cho tổng thể tuân theo phân phối chuẩn mà chưa biết trung bình μ *nhưng phương sai* σ^2 *đã biết. Tìm khoảng tin cậy* $100(1-\alpha)\%$ *cho trung bình* μ *của tổng thể dựa trên các thông tin của mẫu ngẫu nhiên chọn ra từ tổng thể.*

Phương sai đã biết

- Chọn một mẫu ngẫu nhiên gồm n phần tử $X_1, X_2, \ldots, X_n$ từ tổng thể tuân theo phân phối chuẩn với trung bình μ chưa biết và phương sai σ^2 đã biết. Khi đó trung bình mẫu $\overline{X}$ tuân theo phân phối chuẩn $N(\mu, \sigma^2/n)$ và ta có biến ngẫu nhiên $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ tuân theo phân phối chuẩn hóa $N(0,1)$.
- Do phân phối chuẩn hóa đối xứng qua trung bình $\mu = 0$ nên tồn tại số $z_{\alpha/2}$ sao cho:

$$P(Z > z_{\alpha/2}) = P(Z < -z_{\alpha/2}) = \frac{\alpha}{2}.$$

Do đó

$$P(z_{-\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Phương sai đã biết

- Chọn một mẫu ngẫu nhiên gồm n phần tử $X_1, X_2, \ldots, X_n$ từ tổng thể tuân theo phân phối chuẩn với trung bình μ chưa biết và phương sai σ^2 đã biết. Khi đó trung bình mẫu $\overline{X}$ tuân theo phân phối chuẩn $N(\mu, \sigma^2/n)$ và ta có biến ngẫu nhiên $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ tuân theo phân phối chuẩn hóa $N(0,1)$.
- Do phân phối chuẩn hóa đối xứng qua trung bình $\mu = 0$ nên tồn tại số $z_{\alpha/2}$ sao cho:

$$P(Z > z_{\alpha/2}) = P(Z < -z_{\alpha/2}) = \frac{\alpha}{2}.$$

Do đó

$$P(z_{-\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Phương sai đã biết

- Từ đó

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\
&= P(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) \\
&= P(\frac{-z_{\alpha/2}\sigma}{\sqrt{n}} < \overline{X}-\mu < \frac{z_{\alpha/2}\sigma}{\sqrt{n}}) \\
&= P(\overline{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} < \mu < \overline{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}).
\end{aligned}
$$

- Với một mẫu cụ thể được chọn ra thì $\overline{X}$ có giá trị $\overline{x}$, đẳng thức trên cho ta khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể là

$$\overline{x} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} < \mu < \overline{x} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}.$$

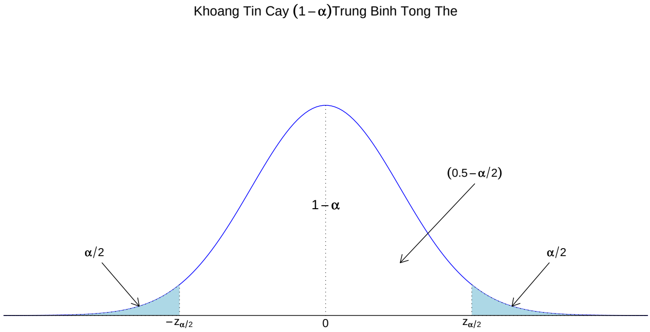
- Từ đó

$$\begin{aligned}1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2})\\ &= P(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2})\\ &= P(\frac{-z_{\alpha/2}\sigma}{\sqrt{n}} < \overline{X}-\mu < \frac{z_{\alpha/2}\sigma}{\sqrt{n}})\\ &= P(\overline{X}-\frac{z_{\alpha/2}\sigma}{\sqrt{n}} < \mu < \overline{X}+\frac{z_{\alpha/2}\sigma}{\sqrt{n}}).\end{aligned}$$

- Với một mẫu cụ thể được chọn ra thì $\overline{X}$ có giá trị $\overline{x}$, đẳng thức trên cho ta khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể là

$$\overline{x}-\frac{z_{\alpha/2}\sigma}{\sqrt{n}} < \mu < \overline{x}+\frac{z_{\alpha/2}\sigma}{\sqrt{n}}.$$

Minh họa hình học



Bảng một số giá trị tới hạn thường gặp

Theo định nghĩa z_α là số được xác định như sau: $P(Z > z_\alpha) = \alpha$, với Z là biến ngẫu nhiên tuân theo phân phối chuẩn hóa $N(0,1)$. Trong R, z_α được tính bởi lệnh

$$qnorm(\alpha, lower.tail = F) \quad \text{hay} \quad qnorm(1-\alpha)$$

$(1-\alpha)100\%$	α	z_α	$z_{\alpha/2}$
80%	0.2	$z_{0.2} = 0.84$	$z_{0.1} = 1.28$
85%	0.15	$z_{0.15} = 1.04$	$z_{0.075} = 1.44$
90%	0.1	$z_{0.1} = 1.28$	$z_{0.05} = 1.645$
95%	0.05	$z_{0.05} = 1.645$	$z_{0.025} = 1.96$
99%	0.01	$z_{0.01} = 2.33$	$z_{0.005} = 2.58$

Ví dụ

Bài toán

Một nhà máy sản xuất giấy theo dây chuyền tự động, giấy được sản xuất có chiều dài trung bình 29.7 cm và độ lệch chuẩn của chiều dài là 0.05 cm. Để kiểm soát tiêu chuẩn giấy thì định kì người ta sẽ chọn mẫu gồm 100 tờ giấy để tiến hành kiểm tra xem chiều dài của các tờ giấy sản xuất còn đạt tiêu chuẩn 29.7 cm hay không, nếu không cần phải kiểm tra xem có vấn đề gì xảy ra với dây chuyền sản xuất đã gây ảnh hưởng đến tiêu chuẩn giấy. Trong lần kiểm tra gần đây nhất chiều dài tờ giấy trung bình tính được từ mẫu là 29.698 cm. Hãy xác định khoảng ước lượng với độ tin cậy 95% và 99% cho chiều dài giấy trung bình của tổng thể các tờ giấy và đưa ra nhận xét.

Ví dụ

Bài toán

Một nhà máy sản xuất giấy theo dây chuyền tự động, giấy được sản xuất có chiều dài trung bình 29.7 cm và độ lệch chuẩn của chiều dài là 0.05 cm. Để kiểm soát tiêu chuẩn giấy thì định kì người ta sẽ chọn mẫu gồm 100 tờ giấy để tiến hành kiểm tra xem chiều dài của các tờ giấy sản xuất còn đạt tiêu chuẩn 29.7 cm hay không, nếu không cần phải kiểm tra xem có vấn đề gì xảy ra với dây chuyền sản xuất đã gây ảnh hưởng đến tiêu chuẩn giấy. Trong lần kiểm tra gần đây nhất chiều dài tờ giấy trung bình tính được từ mẫu là 29.698 cm. Hãy xác định khoảng ước lượng với độ tin cậy 95% và 99% cho chiều dài giấy trung bình của tổng thể các tờ giấy và đưa ra nhận xét.

Ví dụ

Bài toán

Một nhà máy sản xuất giấy theo dây chuyền tự động, giấy được sản xuất có chiều dài trung bình 29.7 cm và độ lệch chuẩn của chiều dài là 0.05 cm. Để kiểm soát tiêu chuẩn giấy thì định kì người ta sẽ chọn mẫu gồm 100 tờ giấy để tiến hành kiểm tra xem chiều dài của các tờ giấy sản xuất còn đạt tiêu chuẩn 29.7 cm hay không, nếu không cần phải kiểm tra xem có vấn đề gì xảy ra với dây chuyền sản xuất đã gây ảnh hưởng đến tiêu chuẩn giấy. Trong lần kiểm tra gần đây nhất chiều dài tờ giấy trung bình tính được từ mẫu là 29.698 cm. Hãy xác định khoảng ước lượng với độ tin cậy 95% và 99% cho chiều dài giấy trung bình của tổng thể các tờ giấy và đưa ra nhận xét.

Ví dụ

Bài toán

Một nhà máy sản xuất giấy theo dây chuyền tự động, giấy được sản xuất có chiều dài trung bình 29.7 cm và độ lệch chuẩn của chiều dài là 0.05 cm. Để kiểm soát tiêu chuẩn giấy thì định kì người ta sẽ chọn mẫu gồm 100 tờ giấy để tiến hành kiểm tra xem chiều dài của các tờ giấy sản xuất còn đạt tiêu chuẩn 29.7 cm hay không, nếu không cần phải kiểm tra xem có vấn đề gì xảy ra với dây chuyền sản xuất đã gây ảnh hưởng đến tiêu chuẩn giấy. Trong lần kiểm tra gần đây nhất chiều dài tờ giấy trung bình tính được từ mẫu là 29.698 cm. Hãy xác định khoảng ước lượng với độ tin cậy 95% và 99% cho chiều dài giấy trung bình của tổng thể các tờ giấy và đưa ra nhận xét.

Lời giải

Ví dụ

Lời giải:

- Nếu độ tin cậy là 99% ta có $z_{\alpha/2} = z_{0.005} = 2.58$.
- Tương tự khoảng ước lượng cho μ với độ tin cậy 99% là:

$$29.6871 \leqslant \mu \leqslant 29.7129.$$

Với độ tin cậy 99%, trung bình tổng thể các tờ giấy được sản xuất được ước lượng trong khoảng từ 29.6871 cm đến 29.7129 cm chứa 29.7 cm nên ta có thể khẳng định tiến trình sản xuất vẫn bình thường.

Nhận xét: Khi độ tin cậy tăng thì khoảng ước lượng rộng hơn, khoảng ước lượng càng lớn thì khả năng để khoảng này chứa trung bình càng tăng. Tuy nhiên với khoảng ước lượng càng rộng thì độ chính xác của ước lượng càng thấp. Với một độ tin cậy đã xác định cách duy nhất để tăng độ chính xác của bài toán ước lượng là tăng cỡ mẫu vì tăng cỡ mẫu làm khoảng ước lượng hẹp lại.

Ví dụ

Lời giải:

- Nếu độ tin cậy là 99% ta có $z_{\alpha/2} = z_{0.005} = 2.58$.
- Tương tự khoảng ước lượng cho μ với độ tin cậy 99% là:

$$29.6871 \leqslant \mu \leqslant 29.7129.$$

Với độ tin cậy 99%, trung bình tổng thể các tờ giấy được sản xuất được ước lượng trong khoảng từ 29.6871 cm đến 29.7129 cm chứa 29.7 cm nên ta có thể khẳng định tiến trình sản xuất vẫn bình thường.

Nhận xét: Khi độ tin cậy tăng thì khoảng ước lượng rộng hơn, khoảng ước lượng càng lớn thì khả năng để khoảng này chứa trung bình càng tăng. Tuy nhiên với khoảng ước lượng càng rộng thì độ chính xác của ước lượng càng thấp. Với một độ tin cậy đã xác định cách duy nhất để tăng độ chính xác của bài toán ước lượng là tăng cỡ mẫu vì tăng cỡ mẫu làm khoảng ước lượng hẹp lại.

Độ rộng của khoảng tin cậy

Độ rộng của khoảng tin cậy $100(1-\alpha)\%$ là:

$$\omega = \overline{x} + z_{\alpha/2}\sigma/\sqrt{n} - (\overline{x} - z_{\alpha/2}\sigma/\sqrt{n}) = 2z_{\alpha/2}\sigma/\sqrt{n}$$

Nhận xét:

- Độ lệch chuẩn σ của tổng thể càng lớn thì độ rộng của khoảng tin cậy càng lớn.
- Độ tin cậy $(1-\alpha)$ càng lớn thì độ rộng của khoảng tin cậy càng lớn.
- Khi cỡ mẫu càng lớn thì độ rộng của khoảng tin cậy càng nhỏ.

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu

2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết

Bài toán

Cho tổng thể tuân theo phân phối chuẩn với trung bình μ chưa biết và phương sai σ^2 chưa biết. Tìm khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể.

Định lí

Lấy một mẫu ngẫu nhiên độc lập từ tổng thể tuân theo phân phối chuẩn. Gọi $\overline{X}$ là trung bình mẫu và S_X^2 là phương sai mẫu của mẫu ngẫu nhiên. Khi đó

$$t = \frac{\overline{X} - \mu}{S_X/\sqrt{n}}$$

là biến ngẫu nhiên tuân theo phân phối Student với $(n-1)$ bậc tự do.

Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết

Bài toán

Cho tổng thể tuân theo phân phối chuẩn với trung bình μ chưa biết và phương sai σ^2 chưa biết. Tìm khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể.

Định lí

Lấy một mẫu ngẫu nhiên độc lập từ tổng thể tuân theo phân phối chuẩn. Gọi $\overline{X}$ là trung bình mẫu và S_X^2 là phương sai mẫu của mẫu ngẫu nhiên. Khi đó

$$t = \frac{\overline{X} - \mu}{S_X/\sqrt{n}}$$

là biến ngẫu nhiên tuân theo phân phối Student với $(n-1)$ bậc tự do.

Phân phối Student

- Phân phối Student do William S.Goset tìm ra năm 1908 dưới bút danh Student nên ngày nay ta quen gọi là phân phối Student, kí hiệu là phân phối t.
- Phân phối t có dạng đối xứng với phần đuôi dài hơn và độ tập trung ở trung tâm kém hơn phân phối z;
- Phân phối t có một họ phân phối với các bậc tự do khác nhau; t_n là kí hiệu của phân phối t với n bậc tự do;
- Khi cỡ mẫu nhỏ (tức bậc tự do nhỏ), phân phối t khá khác phân phối z nhưng khi cỡ mẫu tăng lên phân phối t dần dần tiến đến hội tụ với phân phối z và khi cỡ mẫu từ 30 trở lên hai phân phối xấp xỉ nhau với độ chính xác cao.

Phân phối Student

- Phân phối Student do William S.Goset tìm ra năm 1908 dưới bút danh Student nên ngày nay ta quen gọi là phân phối Student, kí hiệu là phân phối t.
- Phân phối t có dạng đối xứng với phần đuôi dài hơn và độ tập trung ở trung tâm kém hơn phân phối z;
- Phân phối t có một họ phân phối với các bậc tự do khác nhau; t_n là kí hiệu của phân phối t với n bậc tự do;
- Khi cỡ mẫu nhỏ (tức bậc tự do nhỏ), phân phối t khá khác phân phối z nhưng khi cỡ mẫu tăng lên phân phối t dần dần tiến đến hội tụ với phân phối z và khi cỡ mẫu từ 30 trở lên hai phân phối xấp xỉ nhau với độ chính xác cao.

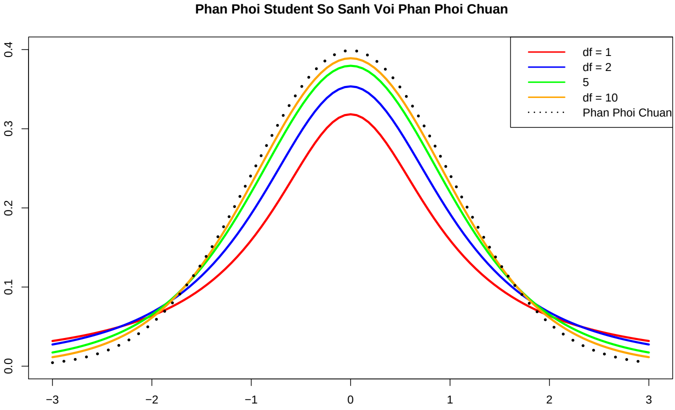
Phân phối Student

- Phân phối Student do William S.Goset tìm ra năm 1908 dưới bút danh Student nên ngày nay ta quen gọi là phân phối Student, kí hiệu là phân phối t.
- Phân phối t có dạng đối xứng với phần đuôi dài hơn và độ tập trung ở trung tâm kém hơn phân phối z;
- Phân phối t có một họ phân phối với các bậc tự do khác nhau; t_n là kí hiệu của phân phối t với n bậc tự do;
- Khi cỡ mẫu nhỏ (tức bậc tự do nhỏ), phân phối t khá khác phân phối z nhưng khi cỡ mẫu tăng lên phân phối t dần dần tiến đến hội tụ với phân phối z và khi cỡ mẫu từ 30 trở lên hai phân phối xấp xỉ nhau với độ chính xác cao.

Phân phối Student

- Phân phối Student do William S.Goset tìm ra năm 1908 dưới bút danh Student nên ngày nay ta quen gọi là phân phối Student, kí hiệu là phân phối t.
- Phân phối t có dạng đối xứng với phần đuôi dài hơn và độ tập trung ở trung tâm kém hơn phân phối z;
- Phân phối t có một họ phân phối với các bậc tự do khác nhau; t_n là kí hiệu của phân phối t với n bậc tự do;
- Khi cỡ mẫu nhỏ (tức bậc tự do nhỏ), phân phối t khá khác phân phối z nhưng khi cỡ mẫu tăng lên phân phối t dần dần tiến đến hội tụ với phân phối z và khi cỡ mẫu từ 30 trở lên hai phân phối xấp xỉ nhau với độ chính xác cao.

Minh họa hình học



Khoảng ước lượng theo phân phối student

- Ta có

$$t_{n-1} = \frac{\overline{X} - \mu}{S_X/\sqrt{n}}$$

tuân theo phân phối Student với (n - 1) bậc tự do.

- Do tính đối xứng của phân phối t nên tồn tại số $t_{n-1,\alpha/2}$ sao cho:

$$P(t_{n-1} > t_{n-1,\alpha/2}) = P(Z < -t_{n-1,\alpha/2}) = \frac{\alpha}{2}.$$

Do đó

$$P(-t_{n-1,\alpha/2} < t_{n-1} < t_{n-1,\alpha/2}) = 1 - \alpha.$$

Khoảng ước lượng theo phân phối student

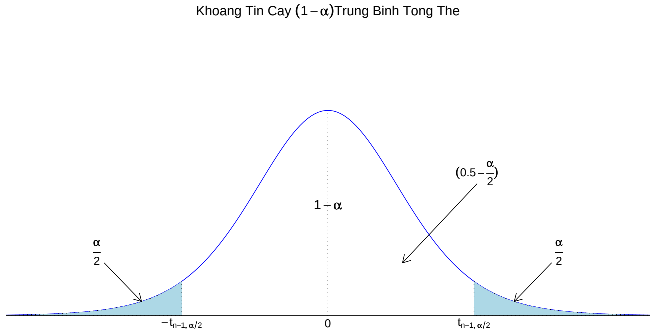
- Từ đó

$$\begin{aligned}1-\alpha &= P(-t_{n-1,\alpha/2} < t_{n-1} < t_{n-1,\alpha/2}) \\ &= P(-t_{n-1,\alpha/2} < \frac{\overline{X}-\mu}{s_X/\sqrt{n}} < t_{n-1,\alpha/2}) \\ &= P(\frac{-t_{n-1,\alpha/2}S_X}{\sqrt{n}} < \overline{X}-\mu < \frac{t_{n-1,\alpha/2}S_X}{\sqrt{n}}) \\ &= P(\overline{X} - \frac{t_{n-1,\alpha/2}s_X}{\sqrt{n}} < \mu < \overline{X} + \frac{t_{n-1,\alpha/2}s}{\sqrt{n}}).\end{aligned}$$

- Với một mẫu cụ thể được chọn ra thì $\overline{X}$ có giá trị $\overline{x}$, đẳng thức trên cho ta khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể là

$$\overline{x} - \frac{t_{n-1,\alpha/2}s_x}{\sqrt{n}} < \mu < \overline{x} + \frac{t_{n-1,\alpha/2}s_x}{\sqrt{n}}.$$

Minh họa hình học



Tính toán phân phối Student trong R

Cho biến ngẫu nhiên $X \backsim t_n$ có hàm mật độ xác suất là $f(x)$, khi đó

- `dt(x, n)` cho giá trị của hàm mật độ $f(x)$ tại x.
- `pt(x, n, lower.tail = TRUE(FALSE))` cho giá trị $P(X \leqslant x)$ $(P(X > x))$.
- `qt(p, n, lower.tail = TRUE(FALSE))` cho ta giá trị là x sao cho `pt(x, n, lower.tail = TRUE(FALSE)) = p`.
- `rt(p, n)` cho mẫu ngẫu nhiên cỡ n tuân theo phân phối Student t_n.

Tính giá trị tới hạn của phân phối Student trong R

Cho biến ngẫu nhiên $X \backsim t_n$, khi đó giá trị tới hạn $t_{n,\alpha}$ được xác định bởi công thức:

$$P(t_n > t_{n,\alpha}) = \alpha.$$

Do đó trong R, $t_{n,\alpha}$ được tính bởi công thức

$$qt(1-\alpha, n) \qquad \text{hoặc} \qquad qt(\alpha, n, lower.tail = F)$$

Ví dụ giá trị $t_{9,0.025}$ được tính trong R như sau:

```
> qt(1 - 0.025, 9)
[1] 2.262157
> #Hoặc
> qt(0.025, 9, lower.tail = F)
[1] 2.262157
```

Ví dụ

Bài toán

Ban giám đốc một nhà máy nhiệt điện muốn ước lượng số than dùng trung bình hàng tuần. Họ chọn ngẫu nhiên 10 tuần và xem xét lượng than dùng thì thấy trung bình $\overline{x} = 11400$ *tấn và độ lệch chuẩn của mẫu* $s_x = 700$ *tấn. Tìm khoảng tin cậy* 95% *cho lượng than trung bình hàng tuần. Giả sử rằng lượng than dùng hàng tuần tuân theo phân phối chuẩn.*

Lời giải

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu

2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Bài toán

Cho một tổng thể có trung bình μ chưa biết và phương sai σ^2 chưa biết. Tìm khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể dựa trên các thông tin của mẫu ngẫu nhiên chọn ra từ tổng thể có số phần tử $n \geqslant 30$.

Theo định lí giới hạn trung tâm khi n lớn ($n \geqslant 30$), trung bình mẫu có thể xấp xỉ phân phối chuẩn, phương sai σ^2 chưa biết được thay thế bằng phương sai mẫu S_X^2. Thiết lập tương tự như trên ta được khoảng ước lượng:

$$\overline{x} - z_{\alpha/2} s_x/\sqrt{n} \leqslant \mu \leqslant \overline{x} + z_{\alpha/2} s_x/\sqrt{n}.$$

Phương sai chưa biết

Bài toán

Cho một tổng thể có trung bình μ chưa biết và phương sai σ^2 chưa biết. Tìm khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của tổng thể dựa trên các thông tin của mẫu ngẫu nhiên chọn ra từ tổng thể có số phần tử $n \geqslant 30$.

Theo định lí giới hạn trung tâm khi n lớn ($n \geqslant 30$), trung bình mẫu có thể xấp xỉ phân phối chuẩn, phương sai σ^2 chưa biết được thay thế bằng phương sai mẫu S_X^2. Thiết lập tương tự như trên ta được khoảng ước lượng:

$$\overline{x} - z_{\alpha/2} s_x/\sqrt{n} \leqslant \mu \leqslant \overline{x} + z_{\alpha/2} s_x/\sqrt{n}.$$

Ví dụ

Bài toán

Một nhân viên phòng nghiên cứu của một công ty chuyên về thực phẩm đóng hộp đang nghiên cứu chiều dài của một loại dưa chuột để sản xuất hộp đóng cho phù hợp. Nhân viên này chọn ngẫu nhiên ra 100 *quả dưa chuột đo đạc thấy chiều dài trung bình là* 9.3 *và độ lệch chuẩn là* 0.5, *với số liệu này hãy tìm khoảng tin cậy* 99% *cho chiều dài trung bình của các quả dưa chuột.*

Lời giải

Tìm khoảng tin cậy cho trung bình tổng thể trong R

Trong R khi tìm khoảng tin cậy cho trung bình tổng thể, ta dùng các hàm sau:

- Khi phương sai của tổng thể đã biết và dữ liệu mẫu dạng sơ cấp ta sử dụng hàm `z.test`;
- Khi phương sai của tổng thể đã biết và dữ liệu mẫu dạng thứ cấp ta sử dụng hàm `zsum.test`;
- Khi phương sai của tổng thể chưa biết và dữ liệu sơ cấp ta sử dụng hàm `t.test`;
- Khi phương sai của tổng thể chưa biết và dữ liệu mẫu dạng thứ cấp ta sử dụng hàm `tsum.test`.

Khoảng tin cậy cho trung bình tổng thể, phương sai đã biết, dữ liệu sơ cấp trong R

- `z.test(x, sigma.x =, conf.level = 0.95)`
- trong đó,
  - `x` là véc tơ số chỉ dữ liệu mẫu;
  - `sigma.x` là độ lệch chuẩn của tổng thể;
  - `conf.level` là số thuộc $[0, 1]$ chỉ độ tin cậy, mặc định độ tin cậy là 0.95.

Khoảng tin cậy cho trung bình tổng thể, phương sai đã biết, dữ liệu thứ cấp trong R

- **`zsum.test(mean.x, sigma.x =, n.x =, conf.level = 0.95)`**
- trong đó,
  - `mean.x` là trung bình mẫu;
  - `sigma.x` là độ lệch chuẩn của tổng thể;
  - `n.x` là cỡ mẫu;
  - `conf.level` xem hàm `z.test`.

Ví dụ trong R

```
> zsum.test(mean.x=29.698,sigma.x=0.05,n.x=100,conf.level=0.95)

        One-sample z-Test

data:  Summarized x
z = 5939.6, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 29.6882 29.7078
sample estimates:
mean of x
   29.698
```

Khoảng tin cậy cho trung bình tổng thể, phương sai chưa biết, dữ liệu sơ cấp trong R

- `t.test(x, conf.level = 0.95)`
- trong đó,
  - `x` là dữ liệu mẫu;
  - `conf.level` xem hàm `z.test`.

Khoảng tin cậy cho trung bình tổng thể, phương sai chưa biết, dữ liệu thứ cấp trong R

- `tsum.test(mean.x, s.x =, n.x =, conf.level = 0.95)`
- trong đó,
  - `mean.x` là trung bình mẫu;
  - `s.x` là độ lệch chuẩn của mẫu;
  - `n.x` là cỡ mẫu;
  - `conf.level` xem hàm `z.test`.

Ví dụ trong R

```
> tsum.test(mean.x=9.3, s.x=0.5, n.x=100, conf.level=0.99)

        One-sample t-Test

data:  Summarized x
t = 186, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 9.16868 9.43132
sample estimates:
mean of x
      9.3
```

Ví dụ trong R

```
> tsum.test(mean.x=11400, s.x=700, n.x=10, conf.level=0.95)

        One-sample t-Test

data:  Summarized x
t = 51.5, df = 9, p-value = 1.971e-12
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 10899.25 11900.75
sample estimates:
mean of x
    11400
```

Nội dung trình bày

1. Ước lượng điểm
   - Hàm ước lượng và ước lượng
   - Hàm ước lượng không chệch
   - Hàm ước lượng hữu hiệu

2. Ước lượng khoảng
   - Định nghĩa khoảng tin cậy
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai đã biết
   - Khoảng tin cậy cho trung bình của phân phối chuẩn, phương sai chưa biết
   - Khoảng tin cậy cho trung bình tổng thể với cỡ mẫu lớn
   - Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

Bài toán

Ước lượng khoảng cho tỉ lệ số phần tử của một tổng thể có một đặc tính nào đó từ tỉ lệ của mẫu ngẫu nhiên.

Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

- Gọi p_x là ước lượng điểm của tỉ lệ tổng thể p. Theo định lí giới hạn trung tâm ta có khi cỡ mẫu n lớn ($np \geqslant 5, n(1-p) \geqslant 5$), biến ngẫu nhiên
$$Z = \frac{P_X - p}{\sqrt{p(1-p)/n}}$$
có phân phối xấp xỉ phân phối chuẩn.
- Khi cỡ mẫu n lớn thì tỉ lệ tổng thể có thể xấp xỉ bởi tỉ lệ mẫu p_x, nghĩa là
$$\sqrt{p(1-p)/n} \approx \sqrt{p_x(1-p_x)/n}.$$

Khoảng ước lượng cho tỉ lệ tổng thể với cỡ mẫu lớn

- Vậy với cỡ mẫu lớn thì biến ngẫu nhiên

$$Z = \frac{P_X - p}{\sqrt{p_x(1-p_x)/n}}$$

có phân phối xấp xỉ phân phối chuẩn.

- Lập luận như phần tìm khoảng ước lượng cho trung bình tổng thể ta có khoảng tin cậy $100(1-\alpha)\%$ cho tỉ lệ tổng thể là:

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

Ví dụ

Bài toán

Để đánh giá trữ lượng cá trong hồ, người ta bắt 2000 con cá, đánh dấu rồi thả xuống hồ. Sau đó bắt lại 400 con thì thấy có 80 con có dấu.

a. *Với độ tin cậy 95%, hãy tìm khoảng tin cậy 95% cho tỉ lệ cá được đánh dấu trong hồ.*
b. *Với độ tin cậy 95%, hãy ước lượng trữ lượng cá hiện có trong hồ.*

Lời giải

Khoảng tin cậy cho tỉ lệ một tổng thể trong R

- `prop.test(x, n, conf.level = 0.95, correct = TRUE)`
- trong đó,
  - `x` là véc tơ đếm số lần "thành công" trong mẫu, `n` là số phần tử của mẫu;
  - `conf.level` xem hàm `z.test`.
  - `correct` là tham số chỉ có hay không có sự điều chỉnh liên tục Yate, mặc định là `correct=TRUE`.

Ví dụ trong R

```
> prop.test(80,400,conf.level=0.95,correct=F)

        1-sample proportions test without continuity correction

data:  80 out of 400, null probability 0.5
X-squared = 144, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.1637371 0.2419703
sample estimates:
  p
0.2
```

Bài tập tự học

Bài toán

Một mẫu ngẫu nhiên gồm 8 nhà phân tích tài chính đưa ra dự báo về tiền lãi (triệu VND) trên mỗi phần vốn trong một công ty vào năm tới là $450, 530, 460, 480, 560, 500, 430, 515$. *Tìm một ước lượng điểm cho trung bình, phương sai, độ lệch chuẩn và tỉ lệ những dự báo về tiền lãi cao hơn 500 của công ty đó.*

Bài tập tự học

Bài toán

Một nhân viên phòng nghiên cứu của một công ty chuyên về thực phẩm đóng hộp đang nghiên cứu chiều dài của một loại dưa chuột để sản xuất hộp đóng cho phù hợp. Nhân viên này chọn ngẫu nhiên ra 100 *quả dưa chuột đo đạc thấy chiều dài trung bình là* 9.3 *và độ lệch chuẩn là* 0.5*, với số liệu này hãy tìm khoảng tin cậy* 99% *cho chiều dài trung bình của các quả dưa chuột.*

Bài tập tự học

Bài toán

Sở vệ sinh muốn thăm dò ý kiến về giờ lấy rác từ các hộ gia đình ở thành phố Hồ Chí Minh. Một mẫu ngẫu nhiên khác gồm 174 hộ được chọn ra thì có 62 số hộ cho rằng tốt nhất là lấy rác vào lúc 6h30 sáng. Tìm khoảng tin cậy 90% cho tỉ lệ của toàn thể số hộ đồng ý với quan điểm này.