Ước lượng

Khoa Công nghệ Thông tin và Truyền thông

CMC UNIVERSITY

Nội dung

Ước lượng Điểm

Ước lượng Điểm

Định nghĩa 1

Cho mẫu ngẫu nhiên $\mathfrak{X} = (X_1, \ldots, X_n)$. *Hàm số*

$$\varphi(\mathfrak{X}) = \varphi(X_1, \ldots, X_n)$$

của mẫu ngẫu nhiên được gọi là một thống kê.

Ước lượng Điểm

Định nghĩa 1

Cho mẫu ngẫu nhiên $\mathfrak{X} = (X_1, \ldots, X_n)$. Hàm số

$$\varphi(\mathfrak{X}) = \varphi(X_1, \ldots, X_n)$$

của mẫu ngẫu nhiên được gọi là một thống kê.

Chú ý

Các vấn đề của thống kê chủ yếu được giải bằng cách xây dựng các hàm thống kê chỉ phụ thuộc vào mẫu ngẫu nhiên, không phụ thuộc vào các tham số chưa biết.

Ví dụ về hàm Thống kê

Một số hàm thống kê thường gặp là

$$\varphi(\mathfrak{X}) = \overline{X} = \frac{X_1 + \cdots + X_n}{n},$$

$$\varphi(\mathfrak{X}) = s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Vấn đề

- Để xác định một phân bố nhị thức, ta cần biết n và p; để xác định một phân bố Poisson, ta cần biết λ; để xác định một phân bố chuẩn, ta cần biết μ và σ^2. Các tham số đặc trưng của một phân bố được gọi chung là tham số quần thể.

Vấn đề

- Để xác định một phân bố nhị thức, ta cần biết n và p; để xác định một phân bố Poisson, ta cần biết λ; để xác định một phân bố chuẩn, ta cần biết μ và σ^2. Các tham số đặc trưng của một phân bố được gọi chung là tham số quần thể.
- Giả sử ta chưa biết giá trị chính xác của một tham số quần thể nào đó, ta lấy một mẫu ngẫu nhiên từ quần thể và bằng các phương pháp thống kê, ta sẽ tìm một ước lượng cho giá trị của tham số chưa biết.

Ước lượng Điểm

Cho mẫu ngẫu nhiên $(x_1, \ldots, x_n)$ từ phân bố $F(x, \theta)$, trong đó tham số θ chưa biết. Chúng ta muốn biết giá trị (gần đúng) của θ.

Ước lượng Điểm

Cho mẫu ngẫu nhiên $(x_1, \ldots, x_n)$ từ phân bố $F(x, \theta)$, trong đó tham số θ chưa biết. Chúng ta muốn biết giá trị (gần đúng) của θ. Hàm $\varphi(\mathfrak{X})$ không phụ thuộc vào θ nhưng có thể mang thông tin về θ. Trong trường hợp đó, $\varphi(\mathfrak{X})$ được gọi là một ước lượng điểm của θ.

Ước lượng Điểm

Cho mẫu ngẫu nhiên $(x_1, \ldots, x_n)$ từ phân bố $F(x, \theta)$, trong đó tham số θ chưa biết. Chúng ta muốn biết giá trị (gần đúng) của θ. Hàm $\varphi(\mathfrak{X})$ không phụ thuộc vào θ nhưng có thể mang thông tin về θ. Trong trường hợp đó, $\varphi(\mathfrak{X})$ được gọi là một ước lượng điểm của θ.

Chú ý

$(x_1, \ldots, x_n)$ là giá trị quan sát được của véc-tơ ngẫu nhiên $(X_1, \ldots, X_n)$, trong đó $X_1, \ldots, X_n$ là các biến ngẫu nhiên có cùng phân bố với X,i.e., phân bố của quần thể. Các biến $X_1, \ldots, X_n$ độc lập với nhau.

Ước lượng không chệch

Định nghĩa 2

Thống kê $\hat{\theta} = \varphi(\mathfrak{X})$ được gọi là

- *một ước lượng không chệch (unbias) cho θ nếu $E[\hat{\theta}] = \theta$.*
- *ước lượng không chệch tốt nhất cho θ nếu*
  - *$\hat{\theta}$ là một ước lượng không chệch cho θ;*
  - *$\mathrm{Var}(\hat{\theta})$ là nhỏ nhất trong tập hợp tất cả các giá trị không chệch.*

Ước lượng cho Trung bình

Mệnh đề 1

Giả sử một quần thể có phân bố với giá trị kỳ vọng μ và phương sai σ^2, và $\mathfrak{X} = (X_1, \ldots, X_n)$ là một mẫu ngẫu nhiên lấy từ quần thể. Khi đó,

$$\hat{\mu} = \overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

là một ước lượng không chệch cho μ.

Ví dụ

Ví dụ 1

Bảng tần số sau đây thể hiện kết quả cân của 100 sản phẩm của một phân xưởng

Trọng lượng (g)	498	502	506	510
Tần số	40	20	20	20

Ví dụ

Ví dụ 1

Bảng tần số sau đây thể hiện kết quả cân của 100 sản phẩm của một phân xưởng

Trọng lượng (g)	498	502	506	510
Tần số	40	20	20	20

$$\bar{x} = \frac{498 \cdot 40 + 502 \cdot 20 + 506 \cdot 20 + 510 \cdot 20}{100} = 502.8\ (g).$$

Ví dụ

Ví dụ 1

Bảng tần số sau đây thể hiện kết quả cân của 100 sản phẩm của một phân xưởng

Trọng lượng (g)	498	502	506	510
Tần số	40	20	20	20

$$\bar{x} = \frac{498 \cdot 40 + 502 \cdot 20 + 506 \cdot 20 + 510 \cdot 20}{100} = 502.8\ (g).$$

Ước lượng: trọng lượng trung bình của tất cả sản phẩm trong phân xưởng là 502.8 g.

Ước lượng cho Phương sai

Mệnh đề 2

Giả sử một quần thể có phân bố với giá trị kỳ vọng μ *và phương sai* σ^2, *và* $\mathfrak{X} = (X_1, \ldots, X_n)$ *là một mẫu ngẫu nhiên lấy từ quần thể. Khi đó,*

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

là một ước lượng không chệch cho σ.

Ước lượng cho Tỉ lệ

Mệnh đề 3

Giả sử quần thể X có phân bố Bernoulli với tham số p chưa biết. $(X_1, \ldots, X_n)$ là một mẫu ngẫu nhiên quan sát từ quần thể. Khi đó,

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

là một ước lượng không chệch cho p.

Ước lượng cho Tỉ lệ

Mệnh đề 3

Giả sử quần thể X có phân bố Bernoulli với tham số p chưa biết. $(X_1, \ldots, X_n)$ là một mẫu ngẫu nhiên quan sát từ quần thể. Khi đó,

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

là một ước lượng không chệch cho p.

Nhận xét

Các giá trị quan sát được X_i's nhận một trong hai giá trị 1, 0 biểu thị đối tượng quan sát có tính chất A (mà ta đang quan tâm) hay không.

Khoảng tin cậy

Ước lượng Khoảng

Một cách khác để thu được thông tin về tham số quần thể chưa biết từ mẫu ngẫu nhiên là xây dựng một khoảng có chứa giá trị đúng của tham số.

Ước lượng Khoảng

Một cách khác để thu được thông tin về tham số quần thể chưa biết từ mẫu ngẫu nhiên là xây dựng một khoảng có chứa giá trị đúng của tham số.

Định nghĩa 3

Cho $(X_1, \ldots, X_n)$ là mẫu ngẫu nhiên từ phân bố $F(x, \theta)$. Khoảng $(\hat{\theta}_1, \hat{\theta}_2)$ được gọi là khoảng tin cậy cho tham số θ với xác suất $1 - \alpha$ nếu

$$\mathrm{P}(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha.$$

Xác suất $1 - \alpha$ được gọi là độ tin cậy của ước lượng, $2\varepsilon = \hat{\theta}_2 - \hat{\theta}_1$ được gọi là độ dài của khoảng tin cậy, ε được gọi là sai số của ước lượng.

Khoảng tin cậy cho μ khi biết σ^2

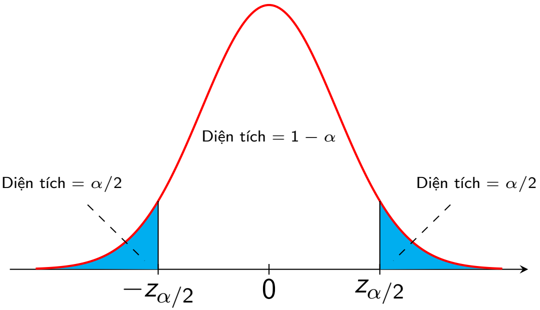
Mệnh đề 4

Giả sử quần thể X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$ với phương sai σ^2 đã biết. Khoảng ước lượng $(\hat{\mu}_1, \hat{\mu}_2)$ cho μ với độ tin cậy $1-\alpha$ được cho bởi

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

trong đó $z_{\alpha/2}$ thỏa mãn $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$.

Khoảng tin cậy cho μ khi biết σ^2



Hình 1: Các phân vị $\pm z_{\alpha/2}$ của phân bố chuẩn tắc.

Ví dụ

IQ

Giả sử chỉ số IQ là biến ngẫu nhiên có phân bố chuẩn với độ lệch chuẩn $\sigma = 15$. Để ước lượng chỉ số IQ của một cộng đồng, người ta chọn ngẫu nhiên 200 người và ước tính được chỉ số IQ trung bình của những người này là 105. Hãy ước lượng khoảng chỉ số IQ trung bình của cộng đồng đó với

(a) độ tin cậy 95%;

(b) độ tin cậy 99%.

Ví dụ (tiếp)

Theo bài ta có $n = 200, \quad \bar{x} = 105, \quad \sigma = 15.$

(a) Với tin cậy 95%, ta có $1 - \alpha = 0.95, \quad \alpha = 0.05.$
Để $\Phi(z_{0.025}) = 1 - \alpha/2 = 0.975$ thì $z_{0.025} = 1.96$. Khoảng ước lượng cho chỉ số IQ trung bình μ với độ tin cậy 95% là

$$(102.921, 107.079).$$

(b) Với độ tin cậy 99%, ta có $1 - \alpha = 0.99, \quad \alpha = 0.01.$
Để $\Phi(z_{0.005}) = 1 - \alpha/2 = 0.995$ thì $z_{0.005} = 2.58$. Khoảng ước lượng cho chỉ số IQ trung bình μ với độ tin cậy 99% là

$$(102.263, 107.737).$$

Áp dụng

Bài tập

Tìm các khoảng tin cậy 95% và 99% cho các mẫu sau đây

(a) $n = 100, \bar{x} = 250, \sigma = 80.$

(b) $n = 64, \bar{x} = 250, \sigma = 50.$

Áp dụng

Bài tập

Tìm các khoảng tin cậy 95% và 99% cho các mẫu sau đây

(a) $n = 100, \bar{x} = 250, \sigma = 80.$

(b) $n = 64, \bar{x} = 250, \sigma = 50.$

Bài tập

Khảo sát 18 giám đốc các công ty ở Mỹ cho thấy lương trung bình hàng năm của họ là 275000\$ với độ lệch chuẩn 62000\$. Tìm khoảng tin cậy 90% cho mức lương trung bình hàng năm của giám đốc các công ty ở Mỹ.

Khoảng ước lượng cho μ khi cỡ mẫu lớn

Mệnh đề 5

Giả sử mẫu ngẫu nhiên có cỡ mẫu $n \geq 30$. Khoảng ước lượng $(\hat{\mu}_1, \hat{\mu}_2)$ cho μ với độ tin cậy $1-\alpha$ được cho bởi

$$\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right),$$

trong đó $z_{\alpha/2}$ thỏa mãn $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$.

Ví dụ

Thu nhập trung bình

Để ước lượng thu nhập trung bình trong một năm của nhân viên công ty X, người ta chọn ngẫu nhiên 150 nhân viên và tính được lương trung bình một năm là 280 triệu đồng với độ lệch mẫu $s = 27$ triệu đồng. Xác định khoảng ước lượng cho thu nhập trung bình một năm của nhân viên công ty đó với độ tin cậy 95%.

Ví dụ

Thu nhập trung bình

Để ước lượng thu nhập trung bình trong một năm của nhân viên công ty X, người ta chọn ngẫu nhiên 150 nhân viên và tính được lương trung bình một năm là 280 triệu đồng với độ lệch mẫu $s = 27$ triệu đồng. Xác định khoảng ước lượng cho thu nhập trung bình một năm của nhân viên công ty đó với độ tin cậy 95%.

Lời giải

- Ta có $n = 150, \quad \bar{x} = 280, \quad s = 27.$
- Độ tin cậy là 95% nên $z_{0.025} = 1.96$.
- Khoảng ước lượng với độ tin cậy 95% cần tìm là

$$(275.679, 284.321).$$

Áp dụng

Bài tập

Đường kính của một loại piston sử dụng cho động cơ ô-tô được cho bởi bảng sau.

x (mm)	*Tần số*
23.94 – 23.96	1
23.96 – 23.98	19
23.98 – 24	28
24 – 24.02	32
24.02 – 24.04	17
24.04 – 24.06	3

Biết rằng đường kính piston có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. *Tìm khoảng ước lượng cho đường kính trung bình của piston với độ tin cậy* 95%.

Khoảng ước lượng cho μ khi cỡ mẫu nhỏ

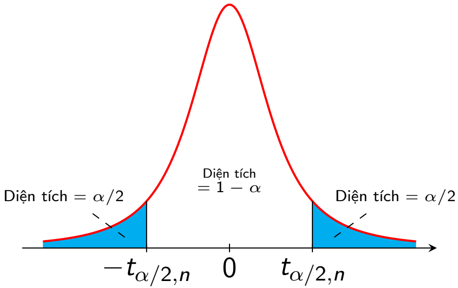
Mệnh đề 6

Giả sử quần thể có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$ *với phương sai* σ^2 *chưa biết và cỡ mẫu* $n < 30$. *Khi đó, khoảng ước lượng* $(\hat{\mu}_1, \hat{\mu}_2)$ *cho* μ *với độ tin cậy* $1 - \alpha$ *được cho bởi*

$$\left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right),$$

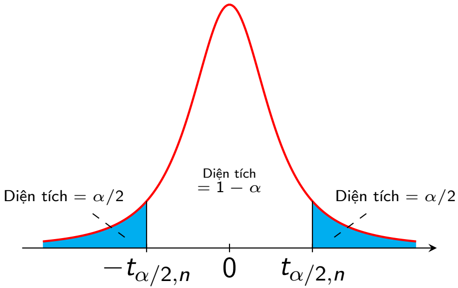
trong đó $t_{\alpha/2,n-1}$ *là số thực thỏa mãn* $\mathrm{P}(T_{n-1} > t_{\alpha/2,n-1}) = \alpha/2$, *với* T_{n-1} *có phân bố Student với* $n-1$ *bậc tự do.*

Đồ thị



Hình 2: Các phân vị $\pm t_{\alpha/2,n}$ của phân bố T_n.

Đồ thị



Hình 2: Các phân vị $\pm t_{\alpha/2,n}$ của phân bố T_n.

$t_{\alpha/2,n}$ thỏa mãn

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

Bảng giá trị

n/α	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240
4	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103
5	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178
13	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	0.6912	0.8662	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728
16	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150
17	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651

Bảng 1: Bảng giá trị của $t_{\alpha/2,n}$.

Ví dụ

Nếu một người không phải chủ sở hữu của một máy tính biết tên đăng nhập + mật khẩu và đăng nhập vào máy, liệu điều này có thể được phát hiện? Có một phương pháp phát hiện dựa vào thời gian gõ và nhả phím, tần số các phím được gõ. Tất cả những thông tin này được ghi lại và so sánh với chủ của máy. Nếu có sự khác biệt lớn, ta kết luận có sự xuất hiện của người lạ.
Bảng số liệu sau đây ghi lại thời gian gõ phím (đơn vị: giây) của một người dùng khi đăng nhập vào một máy tính.

0.24	0.22	0.26	0.34	0.35	0.32
0.33	0.29	0.19	0.36	0.30	0.15
0.17	0.28	0.38	0.40	0.37	0.27

Giả sử thời gian gõ phím có phân bố chuẩn, chúng ta xây dựng một khoảng với độ tin cậy 99% cho thời gian gõ phím trung bình.

Ví dụ (tiếp)

Theo bài ta có $n = 18,\ \bar{x} = 0.29,\ s = 0.074$.
Giá trị $t_{\alpha/2,n-1}$ với $n = 18, \alpha = 0.01$ là $t_{0.005,17} = 2.898$. Do đó, khoảng với độ tin cậy 99% cho thời gian gõ phím trung bình là

$$0.29 \pm 2.898 \cdot \frac{0.074}{\sqrt{18}} = (0.24, 0.34).$$

Ví dụ (tiếp)

Trọng lượng trung bình

Để xác định trọng lượng trung bình của các bao bột mì được đóng gói tự động, người ta chọn ra 15 bao ngẫu nhiên và tính được $\bar{x} = 39.8$ kg và $s^2 = 0.144$. Xác định khoảng tin cậy với độ tin cậy 99% cho trọng lượng trung bình.

Ví dụ (tiếp)

Trọng lượng trung bình

Để xác định trọng lượng trung bình của các bao bột mì được đóng gói tự động, người ta chọn ra 15 bao ngẫu nhiên và tính được $\bar{x} = 39.8$ kg và $s^2 = 0.144$. Xác định khoảng tin cậy với độ tin cậy 99% cho trọng lượng trung bình.

- Từ giả thiết có $n = 15,\ \bar{x} = 39.8,\ s = \sqrt{0.144}$.
- Ta có $1 - \alpha = 0.99$ nên $\alpha/2 = 0.005$. Tra Bảng giá trị ta được $t_{0.005,14} = 2.9768$.
- Khoảng tin cậy 99% của trọng lượng trung bình là

$$39.8 \pm 2.9768 \cdot \frac{\sqrt{0.144}}{\sqrt{15}} = (39.51, 40.09).$$

Áp dụng

Bài tập

Để ước lượng chiều cao trung bình của thanh niên trong một quận, một mẫu ngẫu nhiên gồm 16 thanh niên được chọn, đo được chiều cao như sau (đơn vị cm)

172	173	173	174	174	175	175	176
166	166	167	165	173	171	170	171

Tìm khoảng ước lượng với độ tin cậy 95% cho chiều cao trung bình.

Khoảng ước lượng cho σ^2

Mệnh đề 7

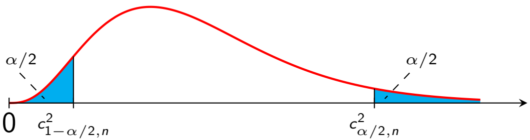
Cho $(X_1, \ldots, X_n)$ là một mẫu ngẫu nhiên từ một quần thể có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$, trong đó σ^2 chưa biết. Khoảng ước lượng cho σ^2 với độ tin cậy $1 - \alpha$ là

$$\left(\frac{(n-1)s^2}{c^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)s^2}{c^2_{1-\frac{\alpha}{2}, n-1}} \right),$$

với $c^2_{a,n-1}$ thỏa mãn

$$\mathrm{P}(\chi^2_{n-1} > c^2_{a,n-1}) = a.$$

Đồ thị



Hình 3: Các phân vị $c^2_{\alpha/2,n}$ và $c^2_{1-\alpha/2,n}$ của phân bố χ^2_n.

Bảng giá trị

n/a	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.0025
1	0.0000	0.0002	0.0010	0.0039	3.8415	5.0239	6.6349	7.8794	9.1406
2	0.0100	0.0201	0.0506	0.1026	5.9915	7.3778	9.2103	10.5966	11.9829
3	0.0717	0.1148	0.2158	0.3518	7.8147	9.3484	11.3449	12.8382	14.3203
4	0.2070	0.2971	0.4844	0.7107	9.4877	11.1433	13.2767	14.8603	16.4239
5	0.4117	0.5543	0.8312	1.1455	11.0705	12.8325	15.0863	16.7496	18.3856
6	0.6757	0.8721	1.2373	1.6354	12.5916	14.4494	16.8119	18.5476	20.2494
7	0.9893	1.2390	1.6899	2.1673	14.0671	16.0128	18.4753	20.2777	22.0404
8	1.3444	1.6465	2.1797	2.7326	15.5073	17.5345	20.0902	21.9550	23.7745
9	1.7349	2.0879	2.7004	3.3251	16.9190	19.0228	21.6660	23.5894	25.4625
10	2.1559	2.5582	3.2470	3.9403	18.3070	20.4832	23.2093	25.1882	27.1122
11	2.6032	3.0535	3.8157	4.5748	19.6751	21.9200	24.7250	26.7568	28.7293
12	3.0738	3.5706	4.4038	5.2260	21.0261	23.3367	26.2170	28.2995	30.3185
13	3.5650	4.1069	5.0088	5.8919	22.3620	24.7356	27.6882	29.8195	31.8831
14	4.0747	4.6604	5.6287	6.5706	23.6848	26.1189	29.1412	31.3193	33.4260
15	4.6009	5.2293	6.2621	7.2609	24.9958	27.4884	30.5779	32.8013	34.9496
16	5.1422	5.8122	6.9077	7.9616	26.2962	28.8454	31.9999	34.2672	36.4557
17	5.6972	6.4078	7.5642	8.6718	27.5871	30.1910	33.4087	35.7185	37.9461

Bảng 2: Bảng giá trị cho $c^2_{a,n}$.

Ví dụ

Số liệu đo đạc

Giả sử ta có một mẫu số liệu sau lấy từ một quần thể có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$

2.5	7.4	8.0	4.5	7.4	9.2

Hãy xây dựng một khoảng ước lượng với độ tin cậy 90% cho σ^2.

- Trung bình mẫu $\bar{x} = 6.5$.
- Phương sai mẫu $s^2 = 6.232$.
- Với độ tin cậy 90%, ta có $\alpha = 0.1$. Tra Bảng giá trị, ta có $c^2_{0.05,5} = 11.0705$, $c^2_{0.95,5} = 1.1455$.
- Khoảng tin cậy 90% cho phương sai là

$$\left(\frac{(n-1)s^2}{c^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{c^2_{1-\alpha/2,n-1}} \right) = (2.815, 27.2).$$

Áp dụng

Bài tập

Lương khởi điểm của kỹ sư máy tính có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$ *với các tham số* μ *và* σ *chưa biết. Ba kỹ sư được chọn ngẫu nhiên và có lương như sau (đơn vị: nghìn USD)*

30 50 70

Hãy xây dựng một khoảng ước lượng với độ tin cậy 95% cho σ.

Khoảng tin cậy cho Tỉ lệ

Vấn đề

Tần suất của các phần tử có cùng tính chất A trong mẫu là $f_n = \frac{n_A}{n}$, với n_A là số phần tử có tính chất A trong mẫu đã chọn. Ước lượng cho tham số quần thể p là tỉ lệ các cá thể có cùng tính chất A.

Khoảng tin cậy cho Tỉ lệ

Vấn đề

Tần suất của các phần tử có cùng tính chất A trong mẫu là $f_n = \frac{n_A}{n}$, với n_A là số phần tử có tính chất A trong mẫu đã chọn. Ước lượng cho tham số quần thể p là tỉ lệ các cá thể có cùng tính chất A.

Mệnh đề 8

Khoảng ước lượng $(\hat{p}_1, \hat{p}_2)$ *cho* p *với độ tin cậy* $1-\alpha$ *được cho bởi*

$$\left(f_n - z_{\alpha/2}\sqrt{\frac{f_n(1-f_n)}{n}}, f_n + z_{\alpha/2}\sqrt{\frac{f_n(1-f_n)}{n}}\right),$$

trong đó $z_{\alpha/2}$ *thỏa mãn* $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$.

Ví dụ

Tỉ lệ nảy mầm

Một viện nghiên cứu vừa phát triển một loại hạt giống. Để ước lượng tỉ lệ nảy mầm, họ gieo thử nghiệm 100 hạt được chọn ngẫu nhiên. Kết quả có 90 hạt nảy mầm. Hãy tìm khoảng ước lượng cho tỉ lệ nảy mầm của loại hạt giống đó với độ tin cậy 95%.

Ví dụ

Tỉ lệ nảy mầm

Một viện nghiên cứu vừa phát triển một loại hạt giống. Để ước lượng tỉ lệ nảy mầm, họ gieo thử nghiệm 100 hạt được chọn ngẫu nhiên. Kết quả có 90 hạt nảy mầm. Hãy tìm khoảng ước lượng cho tỉ lệ nảy mầm của loại hạt giống đó với độ tin cậy 95%.

- Theo bài ta có $n = 100, f_n = 90/100 = 0.9$.
- Độ tin cậy là 95% nên $1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05$.
  Để $\Phi(z_{0.025}) = 0.975$ thì $z_{0.025} = 1.96$.
- Khoảng ước lượng cho p với độ tin cậy 95% là

$$(0.8412, 0.9588).$$

Áp dụng

Bài tập

Một cuộc thăm dò trước bầu cử tổng thống Mỹ được tiến hành. Người ta chọn ngẫu nhiên 100 người để hỏi ý kiến thì có 60 người nói rằng họ sẽ bỏ phiếu cho Donald Trump. Tìm khoảng ước lượng cho tỉ lệ cử tri bỏ phiếu cho Trump với độ tin cậy 90%.

Tính toán với R

Các hàm thông dụng với biến Gauß

Code	Tên
`dnorm(x, mean, sd)`	hàm mật độ (pdf)
`pnorm(x, mean, sd)`	hàm xác suất (cdf)
`qnorm(q, mean, sd)`	hàm ngược (quantile function)
`rnorm(n, mean, sd)`	sinh n giá trị ngẫu nhiên

Bảng 3: Một số lệnh thông dụng cho biến Gauß.

Ví dụ

E.g., với $\mu = 1, \sigma = 2$,

$$\texttt{dnorm(4,1,2)} = \frac{1}{\sqrt{8\pi}} e^{-\frac{(4-1)^2}{8}} = 0.0647588,$$

$$\texttt{pnorm(4,1,2)} = \frac{1}{\sqrt{8\pi}} \int_{-\infty}^{4} e^{-\frac{(x-1)}{2\sigma^2}} dx = 0.9331928,$$

$$\texttt{qnorm(0.75,1,2)} = x_0 = 2.34898 \text{ do } \frac{1}{\sqrt{8\pi}} \int_{-\infty}^{x_0} e^{-\frac{(x-1)}{2\sigma^2}} dx = 0.75.$$

Câu lệnh với Biến chuẩn tắc

Trong trường hợp tính toán với biến chuẩn tắc, các câu lệnh có thể được đơn giản hóa, *e.g.*, `dnorm(x)`, `pnorm(x)`, `qnorm(q)`, `rnorm(n)`.

Tính toán với Phân bố Chuẩn tắc

Để tính giá trị

$$P(Z < t_0) = \Phi(t_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_0} e^{-\frac{x^2}{2}} dx,$$

ta dùng câu lệnh

```
> pnorm(t_0)
```

Tính toán với Phân bố Chuẩn tắc

Để tính giá trị

$$P(Z < t_0) = \Phi(t_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_0} e^{-\frac{x^2}{2}} dx,$$

ta dùng câu lệnh

```
> pnorm(t_0)
```

E.g., các giá trị $\Phi(0), \Phi(1.5), \Phi(-1.5)$ có thể được tính như sau

```
> pnorm(0)
[1] 0.5
> pnorm(1.5)
[1] 0.9331928
> pnorm(-1.5)
[1] 0.0668072
```

Phân bố Chuẩn tắc (tiếp)

Để tính giá trị z_α biết rằng

$$\Phi(z_\alpha) = \alpha,$$

ta dùng câu lệnh

```
> qnorm(alpha)
```

Phân bố Chuẩn tắc (tiếp)

Để tính giá trị z_α biết rằng

$$\Phi(z_\alpha) = \alpha,$$

ta dùng câu lệnh

```
> qnorm(alpha)
```

E.g., với $\alpha = 0.9, 0.95, 0.99$ ta có

```
> qnorm(0.9)
[1] 1.28152
> qnorm(0.95)
[1] 1.644854
> qnorm(0.99)
[1] 2.326348
```

Tính toán với Phân bố χ^2

Code	Tên
`dchisq(x, df)`	hàm mật độ (pdf)
`pchisq(x, df)`	hàm xác suất (cdf)
`qchisq(q, df)`	hàm ngược (quantile function)
`rchisq(n, df)`	sinh n giá trị ngẫu nhiên

Bảng 4: Một số lệnh thông dụng cho biến χ^2.

Tính toán với Phân bố Student-t

Code	Tên
`dt(x, df)`	hàm mật độ (pdf)
`pt(x, df)`	hàm xác suất (cdf)
`qt(q, df)`	hàm ngược (quantile function)
`rt(n, df)`	sinh *n* giá trị ngẫu nhiên

Bảng 5: Một số lệnh thông dụng cho biến Student.

End of Chapter 4