Lý thuyết thống kê

Một số khái niệm

Tổng thể và mẫu

Định nghĩa

1. Một **tổng thể** (population) là tập hợp tất cả các đối tượng có chung một tính chất mà ta quan tâm. Số phần tử của tổng thể được gọi là *kích thước tổng thể*.
2. Một **mẫu** (sample) là một tập con của tổng thể. Số phần tử của mẫu được gọi là *kích thước mẫu*.
3. Việc chọn từ tổng thể một tập con nào đó được gọi là **phép lấy mẫu**.

Ví dụMột công ty sản xuất chip máy tính đóng gói mỗi hộp gồm 100 chip. Người ta muốn khảo sát tỉ lệ chip bị lỗi trong một lô hàng gồm 1000 hộp của công ty. Chọn ngẫu nhiên 80 hộp chip để kiểm tra. Ta có **tổng thể** là 1000 hộp chip và **mẫu** là 80 hộp chip được kiểm tra.

Mẫu ngẫu nhiên

Một trong những nhiệm vụ quan trọng của thống kê là xây dựng các phương pháp cho phép rút ra các kết luận hoặc đưa ra dự báo về toàn bộ tổng thể dựa trên một mẫu. Do đó, vấn đề lấy mẫu là một việc vô cùng quan trọng.

Định nghĩa (Mẫu ngẫu nhiên)

Một mẫu là ngẫu nhiên (random sample) nếu trong phép lấy mẫu đó, mỗi phần tử được chọn một cách độc lập và có xác suất được chọn như nhau.

Yêu cầu khi lấy mẫu:

- Mẫu được chọn là mẫu ngẫu nhiên.
- Kích thước mẫu đủ lớn.

Nếu kích thước mẫu càng lớn thì thông tin suy luận về tổng thể càng đáng tin cậy và có ý nghĩa.

Thống kê mô tả & Thống kê suy diễn

Định nghĩa

1. **Thống kê mô tả** (Descriptive Statistics) là các phương pháp sử dụng để tóm tắt hoặc mô tả một tập hợp dữ liệu, một mẫu nghiên cứu dưới dạng số hay biểu đồ trực quan.
2. **Thống kê suy diễn** (Inferential statistics) bao gồm các phương pháp được sử dụng để suy luận về các đặc điểm tổng thể từ thông tin có trong một mẫu được lấy từ tổng thể này.

Thống kê mô tả

Biểu diễn mẫu

Một mẫu có thể được biểu diễn theo các cách như sau:

- Mẫu dạng tần số (tần suất)
- Mẫu dạng khoảng
- Mẫu dạng biểu đồ

Biểu diễn mẫu

Giả sử ta quan tâm đến tính chất X (định lượng hoặc định tính) của các phần tử trong tổng thể nào đó. Tập giá trị của X là $\mathcal{S}$. Với một mẫu ngẫu nhiên với kích thước n được lấy từ tổng thể đó, tính chất X của các phần tử trong mẫu thuộc tập $\{x_1, x_2, \dots, x_m\} \subset \mathcal{S}$

- Bảng phân phối tần số:

X	x_1	x_2	$\cdots$	x_m
Tần số	k_1	k_2	$\cdots$	k_m

(k_i là số phần tử của mẫu có tính chất $X = x_i$)

- Bảng phân bố tần suất

X	x_1	x_2	$\cdots$	x_m
Tần suất	f_1	f_2	$\cdots$	f_m

($f_i = k_i/n$ là tần suất suất hiện của tính chất x_i)

Biểu diễn mẫu

Ví dụ

Kiểm tra 80 hộp (mỗi hộp chứa 100 chip bán dẫn) để tìm số lượng chip bị lỗi trong mỗi hộp.

1	3	4	7	2	7	5	5	2	2	4	2	4	3	2
2	7	1	3	3	2	5	0	0	1	2	5	5	4	1
3	2	6	3	8	2	2	3	1	6	3	4	1	2	5
1	3	3	3	2	1	2	5	5	4	1	4	3	1	0
2	1	2	4	4	5	3	3	4	0	5	2	5	6	2
5	3	3	3	1										

Biểu diễn mẫu

Ta có bảng phân bố tần số và tần suất như sau

Số chip bị lỗi	Tần số	Tần suất
0	4	0,05
1	12	0,15
2	18	0,225
3	17	0,2125
4	10	0,125
5	12	0,15
6	3	0,0375
7	3	0,0375
8	1	0,0125
≥ 9	0	0
Tổng	80	1

Biểu diễn mẫu

Trong trường hợp tính chất X là định lượng với giá trị liên tục, ta có thể cho biểu diễn mẫu dạng khoảng như ví dụ sau:

Ví dụ

Chiều cao của 40 sinh viên trong một mẫu được biểu diễn như sau sau:

Khoảng	Tần số	Tần suất
$(146; 151]$	4	0,1
$(151; 156]$	2	0,05
$(156; 161]$	6	0,15
$(161; 166]$	10	0,25
$(166; 171]$	12	0,3
$(171; 176]$	6	0,15

Các khoảng như thế được gọi là **các lớp ghép** của X.

Biểu diễn mẫu

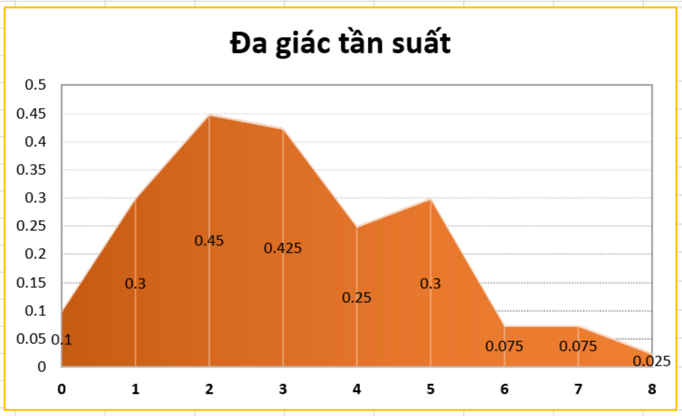
Giả sử ta có một mẫu với bảng phân bố tần số và tần suất như sau

x_i	x_1	x_2	$\cdots$	x_m
Tần số	k_1	k_2	$\cdots$	k_m
Tần suất	f_1	f_2	$\cdots$	f_m

Khi đó ta có thể biểu diễn mẫu dưới dạng biểu đồ thanh như sau

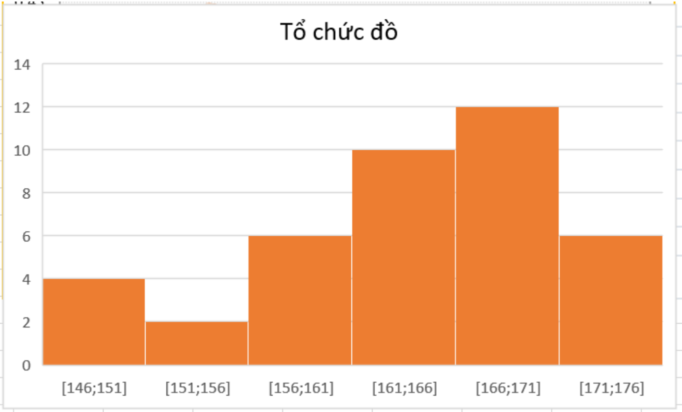
Biểu diễn mẫu

hoặc nối các điểm $(x_i; f_i)$ với $(x_{i+1}; f_{i+1})$ bởi đoạn thẳng ta được các đường gấp khúc được gọi là **biểu đồ đa giác tần suất**



Biểu diễn mẫu

Đối với bản phân bố lớp ghép, ta thường dùng đồ thị dạng **histogram** để biểu diễn.



(Biểu đồ histogram của mẫu trong Ví dụ 2.1.2)

Người ta có thể dùng các biểu đồ khác nhau để biểu diễn dữ liệu thu được.

Nguồn: Internet.

Các đặc trưng của mẫu

Cho một mẫu ngẫu nhiên có kích thước n được lấy từ một tổng thể. Các quan sát từ mẫu là các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$.

Định nghĩa (Trung bình mẫu)

1. Trung bình mẫu ngẫu nhiên $\{X_1, \ldots, X_n\}$ được xác định bởi

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

2. Nếu một mẫu có kích thước n nhận các giá trị $x_1, \ldots, x_n$ thì **trung bình mẫu cụ thể** là

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

Các đặc trưng của mẫu

Giả sử một mẫu có biểu diễn mẫu dạng khoảng với m lớp ghép $C_1, C_2, \ldots, C_m$ và tần số của lớp C_i là k_i. Khi đó trung bình mẫu $\overline{x}$ được xác định bởi

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

trong đó x_i là trung điểm (tâm) của khoảng C_i.

Các đặc trưng mẫu

Định nghĩa (Trung vị mẫu)

Trung vị mẫu, kí hiệu m, là một số mà số các giá trị của mẫu $\geq m$ bằng số các giá trị của mẫu $\leq m$.

Định nghĩa (Mode mẫu)

Mode của mẫu là giá trị của mẫu có tần số lớn nhất.

Các đặc trưng mẫu

Định nghĩa (Phương sai mẫu)

1. Một tổng thể có kích thước N và trung bình tổng thể là μ. Phương sai tổng thể (variance of a population), ký hiệu σ^2,

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}.$$

2. Cho một mẫu có kích thước n và trung bình mẫu là $\overline{x}$. Phương sai mẫu chưa hiệu chỉnh (variance of a sample), ký hiệu $\hat{s}^2$,

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

3. **Phương sai mẫu hiệu chỉnh**, ký hiệu s^2,

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

Phân phối mẫu

Ví dụ 3.1 Xét một tổng thể có kích thước $N = 5$ gồm các số $3, 6, 9, 12, 15$. Ta muốn tìm phân phối của trung bình mẫu đối với các mẫu ngẫu nhiên có kích thước $n = 3$.

Giải. Ta có 10 mẫu khác nhau có kích thước $n = 3$.

Mẫu	Các giá trị của mẫu	$\overline{x}$
1	3, 6, 9	6
2	3, 6, 12	7
3	3, 6, 15	8
4	3, 9, 12	8
5	3, 9, 15	9
6	3, 12, 15	10
7	6, 9, 12	9
8	6, 9, 15	10
9	6, 12, 15	11
10	9, 12, 15	12

Phân phối của trung bình và phương sai mẫu

Bảng phân phối của trung bình mẫu

$\overline{X}$	6	7	8	9	10	11	12
$P(\overline{X} = x_i)$	0,1	0,1	0,2	0,2	0,2	0,1	0,1

Nhận xét: Theo Định lý giới hạn trung tâm, ta có

1. Nếu tổng thể có phân phối chuẩn $N(\mu; \sigma^2)$ thì

$$\overline{X} \sim N(\mu; \sigma^2/n) \quad \text{và} \quad S \sim N(n\mu; n\sigma^2)$$

   với mọi mẫu có kích thước n.

2. Nếu tổng thể có phân phối không chuẩn thì

$$\overline{X} \approx N(\mu; \sigma^2/n) \quad \text{và} \quad S \approx N(n\mu; n\sigma^2)$$

   với mọi mẫu có kích thước $n \geq 30$.

Phân phối của trung bình và phương sai mẫu

Ví dụ

Giả sử chiều cao của sinh viên nam ở TPHCM có phân phối chuẩn với trung bình là 172cm và độ lệch chuẩn là 10cm. Chọn một mẫu ngẫu nhiên gồm 25 sinh viên.
a. Tìm quy luật phân phối của trung bình mẫu.
b. Tính xác suất mẫu đó có chiều cao trung bình lớn hơn 174cm.

Giải. Vì chiều cao X của tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ nên trung bình mẫu $\overline{X}$ có phân phối chuẩn $N(\mu,\sigma^2/n)$ với $\mu = 172; \sigma^2/n = 4$.

Ta có

$$\begin{aligned} P(\overline{X} > 174) &= 1 - P(\overline{X} \le 174) = 1 - \Phi\left(\frac{174-172}{2}\right) \\ &= 1 - \Phi(1) = 1 - 0,8413 = 0,1587. \end{aligned}$$

Phân phối của tỉ lệ mẫu

Giả sử tỉ lệ các phần tử có tính chất $\mathcal{P}$ trong một tổng thể là p. Xét một mẫu (ngẫu nhiên) với kích thước n và gọi f là tỉ lệ các phần tử có tính chất $\mathcal{P}$ trong mẫu. Khi đó, nếu $np > 5$ và $n(1-p) > 5$ thì

$$f \approx N(p; \frac{p(1-p)}{n}).$$

Ví dụ

Sản phẩm của một nhà máy sản xuất có 2% bị lỗi. Kiểm tra ngẫu nhiên 400 sản phẩm do nhà máy sản xuất. Tính xác suất trong 400 sản phẩm đó có không dưới 3% sản phẩm bị lỗi.

Giải. Gọi f là tỉ lệ sản phẩm bị lỗi của một mẫu ngẫu nhiên với kích thước $n = 400$. Khi đó $f \approx N(0,02; 0,007^2)$. Vậy nên

$$\begin{aligned} P(f \geq 0,03) &= 1 - P(F < 0,03) = 1 - \Phi_F(\frac{0,03 - 0,02}{0,007}) \\ &= 1 - \Phi(1,43) = 1 - 0,9236 = 0,0764. \end{aligned}$$

Lý thuyết ước lượng

Khái niệm

Ước lượng là một phỏng đoán về một giá trị chưa biết của tổng thể dựa vào quan sát trên mẫu lấy ra từ tổng thể.

Định nghĩa

1. Một ước lượng điểm là một giá trị dùng để ước lượng một tham số.
2. Một ước lượng khoảng là một khoảng giá trị dùng để ước lượng một tham số.

Ví dụ.

- Nếu nói chiều cao trung bình của sinh viên nam trường Đại học Công nghệ Thông tin là 174 cm thì đó là một **ước lượng điểm**.
- Nếu nói chiều cao trung bình đó nằm trong khoảng từ 159 cm đến 169 cm hay 164 ± 5 cm. Khi đó ta đã có một **ước lượng khoảng.**

Một số bài toán ước lượng

1. Ước lượng khoảng cho trung bình tổng thể khi biết độ lệch chuẩn tổng thể
2. Ước lượng khoảng cho trung bình tổng thể khi chưa biết độ lệch chuẩn tổng thể
3. Ước lượng tỉ lệ của tổng thể

Bài toán 1.

Giả sử rằng thời gian mua sắm của khách hàng tại một trung tâm thương mại có phân phối chuẩn với độ lệch chuẩn tổng thể là 20 phút. Chọn ngẫu nhiên 64 người đã mua sắm ở trung tâm đó. Người ta thấy rằng thời gian mua sắm trung bình của 64 người này là 75 phút. Tìm thời gian mua sắm trung bình của khách hàng tại trung tâm này với độ tin cậy 95%.

Định nghĩa

1. **Độ tin cậy** (confidence level), ký hiệu $1-\alpha$, của ước lượng khoảng của một tham số là xác suất khoảng ước lượng chứa tham số. Giả sử một số lượng lớn mẫu được chọn và quá trình ước lượng trên cùng một tham số được lặp lại.
2. **Khoảng tin cậy** (confidence interval) là một khoảng ước lượng cụ thể của một tham số tương ứng với độ tin cậy đã cho.

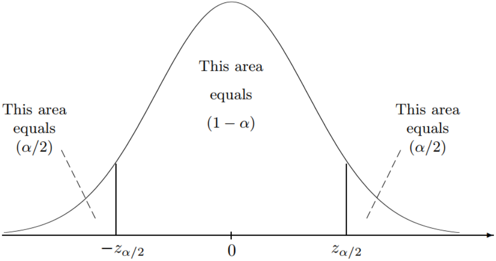
Ví dụ: Khi nói **khoảng tin cậy** của chiều cao trung bình của sinh viên các trường đại học tại TPHCM là $[155; 175]$ với **độ tin cậy** 95% có nghĩa là xác suất khoảng $[155; 175]$ chứa trung bình tổng thể là 95%.

Phương pháp Ước lượng trung bình tổng thể μ khi biết σ và $1-\alpha$.

Ta biết rằng khi tổng thể có phân phối chuẩn hoặc không có phân phối chuẩn nhưng kích thước mẫu $n \geq 30$ thì ta đều có

$$\overline{X} \approx N(\mu; \sigma^2/n).$$

Suy ra $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0,1)$



- Tìm giá trị $z_{\alpha/2}$ sao cho

$$\boxed{\Phi(z_{\alpha/2}) = 1 - \alpha/2}.$$

- Ước lượng khoảng của trung bình tổng thể với độ tin cậy $1 - \alpha$ là

$$\boxed{\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Giá trị $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là **sai số của ước lượng** hoặc **độ chính xác** của ước lượng.

Giải Bài toán 1.

- Ta có $\overline{x} = 75$; $n = 64$ và $\sigma = 20$;
- Độ tin cậy $1 - \alpha = 95\%$. Suy ra $\alpha = 0,05$ và $z_{\alpha/2} = 1,96$
- Độ chính xác $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1,96\dfrac{20}{\sqrt{64}} = 4,9$
- Khoảng tin cậy của trung bình tổng thể với độ tin cậy 95% là

$$[70,1 ; 79,9]$$

Bài tập 1. Chọn ngẫu nhiên 30 người để kiểm tra thời gian sử dụng chiếc điện thoại di động đầu tiên. Người ta thấy rằng thời gian sử dụng trung bình của 30 người này là 5,6 năm. Giả sử thời gian sử dụng chiếc điện thoại di động đầu tiên có phân phối chuẩn với độ lệch chuẩn là 0,8 năm. Tính khoảng thời gian trung bình sử dụng chiếc điện thoại đầu tiên với độ tin cậy 99%.

xtb=5.6, n=30, dlc=0.8 a=0.01, z(a/2)=2.58
dcx= 0.377 ktc = [5.223; 5.797]

Bài tập 2. Thu nhập trung bình hàng tháng của 30 hộ dân trong một thành phố được cho như sau (đơn vị triệu đồng)

12.23	16.56	4.39	2.89	1.24	2.17	13.19	9.16	1.42	73.25
1.91	14.64	11.59	6.69	1.06	8.74	3.17	18.13	7.92	4.78
16.85	40.22	2.42	21.58	5.01	1.47	12.24	2.27	12.77	2.76

Tìm khoảng tin cậy 90% của thu nhập trung bình hàng tháng của toàn thành phố. Biết thu nhập trung bình hàng tháng có phân phối chuẩn và có độ lệch chuẩn 14.405.

xtb=11.09, n=30, dlc=14.405, a=0.1,
z(a/2)=1-a/2=1.65, dcx=4.34
ktc=[6.75, 15.43]

Ước lượng khoảng cho trung bình tổng thể

Khi chưa biết độ lệch chuẩn tổng thể

Bài toán 2. Thu nhập trung bình hàng tháng của **30** hộ dân trong một thành phố được cho như sau (đơn vị triệu đồng)

12.23	16.56	4.39	2.89	1.24	2.17	13.19	9.16	1.42	73.25
1.91	14.64	11.59	6.69	1.06	8.74	3.17	18.13	7.92	4.78
16.85	40.22	2.42	21.58	5.01	1.47	12.24	2.27	12.77	2.76

Tìm khoảng thu nhập trung bình hàng tháng của toàn thành phố với độ tin cậy 90%.

Bài toán 3. Theo một thống kê cho thấy số thu nhập của 7 công nhân trong năm 2021 của một công ty được cho như sau (đơn vị triệu đồng)

54,6 59 60,9 63,1 71,6 84,4 99,3

Giả sử thu nhập trong năm 2021 của công ty có phân phối chuẩn. Tìm khoảng thu nhập trung bình của công ty này với độ tin cậy 99%.

Phương pháp

Trường hợp 1: Kích thước mẫu $n \geq 30$.

1. $\overline{x}, s$ là trung bình và độ lệch chuẩn hiệu chỉnh của một mẫu cụ thể
2. Đổi biến $Z = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$, khi đó $Z \sim N(0;1)$.
3. Tra bảng $A4$, tìm $z_{\alpha/2}$.
4. Khoảng tin cậy của μ với độ tin cậy $1-\alpha$ là

$$[\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}].$$

Trường hợp 2: Kích thước mẫu $n < 30$ và tổng thể có phân phối chuẩn

1. Đổi biến $T = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$, khi đó $T \sim St(n)$.
2. Tra bảng $A5$ dòng $n-1$, tìm $t_{\alpha/2}$ thỏa mãn $P(T > t_{\alpha/2}) = \dfrac{\alpha}{2}$.
3. Khoảng tin cậy của μ với độ tin cậy $1-\alpha$ là

$$[\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}].$$

Cách tìm t_β trong bảng A5: bậc tự do của phân phối Student là $n-1$.

Ví dụ. Tìm giá trị $t_{0,005}$ với bậc tự do 17.

ν	α, the right-tail probability									
(d.f.)	.10	.05	.025	.02	.01	.005	.0025	.001	.0005	.0001
1	3.078	6.314	12.706	15.89	31.82	63.66	127.3	318.3	636.6	3185
2	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60	70.71
3	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92	22.20
4	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610	13.04
5	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.894	6.869	9.676
6	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959	8.023
7	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408	7.064
8	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041	6.442
9	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781	6.009
10	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587	5.694
11	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437	5.453
12	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318	5.263
13	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221	5.111
14	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140	4.985
15	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073	4.880
16	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015	4.790
17	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965	4.715
18	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922	4.648
19	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883	4.590
20	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850	4.539

Theo bảng A5, ta có $t_{0,005} = 2,898$.

Bài toán 3. Theo một thống kê cho thấy số thu nhập của 7 công nhân trong năm 2021 của một công ty được cho như sau (đơn vị triệu đồng)

54,6   59   60,9   63,1   71,6   84,4   99,3

Giả sử thu nhập trong năm 2021 của công ty có phân phối chuẩn. Tính khoảng thu nhập trung bình của công ty này với độ tin cậy 99%.

- Trung bình mẫu: $\overline{x} =$ ……… 70,414
- Độ lệch chuẩn mẫu hiệu chỉnh: $s =$ ……… 16,1
- Tìm $t_{\alpha/2}$ với độ tin cậy $1 - \alpha = 0,99$ và bậc tự do 6: $t_{\alpha/2} =$ ……… 3,707
- Khoảng tin cậy cần tìm

$$\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$47,852 \leq \mu \leq 92,976$$

Ví dụ

Kiểm tra tuổi thọ (tính bằng giờ) của 50 bóng đèn do nhà máy A sản xuất, người ta được bảng số liệu sau

Tuổi thọ	3300	3500	3600	4000
Số bóng đèn	10	20	12	8

a. Ước tính tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất với độ tin cậy 97%.
b. Dựa vào mẫu trên để ước lượng tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất có độ chính xác 59,02 giờ thì phải đảm bảo độ tin cậy là bao nhiêu?
c. Dựa vào mẫu trên, nếu muốn ước lượng tuổi thọ trung bình của các bóng đèn do nhà máy A sản suất có độ chính xác nhỏ hơn 40 giờ với độ tin cậy 98% thì cần phải kiểm tra tối thiểu bao nhiêu bóng đèn?

Giải a. (Kích thước mẫu là 50 và chưa biết độ lệch chuẩn tổng thể)

- Trung bình mẫu: $\overline{x} =$ 3564
- Độ lệch chuẩn mẫu hiệu chỉnh: $s =$ 217,3683
- Độ tin cậy $1 - \alpha = 0,97$. Suy ra $z_{\alpha/2} =$ 2,17
- Độ chính xác:

$$z_{\alpha/2}\frac{s}{\sqrt{n}} = 2,17\frac{217,3683}{\sqrt{50}} = 66,707$$

- Khoảng tin cậy của tuổi thọ trung của bóng đèn với độ tin cậy 97% là

[3497,293; 3630,707]

b. Ta có độ chính xác bằng 59,02 giờ, tức là

$$z_{\alpha/2}\frac{s}{\sqrt{n}} = 59,02$$

Suy ra

$$z_{\alpha/2} = 59,02\frac{\sqrt{n}}{s} = 1,92$$

Do đó

$$\Phi(1,92) = \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}.$$

Trang bảng A4, ta có $\alpha =$ 0,05486

Như vậy, độ tin cậy là 94,514%

c. Ta có độ chính xác nhỏ hơn 40 giờ với độ tin cậy 98%, tức là

$$z_{\alpha/2}\frac{s}{\sqrt{n}} < 40.$$

Suy ra

$$\sqrt{n} > z_{\alpha/2}\frac{s}{40}$$

Vì $1-\alpha = 0,98$ nên suy ra $z_{\alpha/2} =$ ...2,33.....Và do đó

$$\sqrt{n} > z_{\alpha/2}\frac{s}{40} = 2,33\frac{217,3683}{40} = \text{...16,6612...}$$

Kết luận: $n >$ ...160.........

Bài tập 3. Một thống kê cho thấy chi phí (tính bằng triệu) của các mẫu quảng cáo 30-giây trên một số đài truyền hình được cho như sau

14	55	165	9	15	66	23	30	150	22	12	13	54	73	55	41	78

Giả sử chi phí cho một video quảng cáo 30-giây có phân phối chuẩn. Ước tính chi phí trung bình cho một quảng cáo 30-giây trên truyền hình với độ tin cậy 90%.

Ước lượng khoảng cho tỉ lệ của tổng thể

- p : tỉ lệ tổng thể (tỉ lệ phần tử có tính chất $\mathcal{P}$ trong tổng thể)
- f : tỉ lệ mẫu cụ thể (tỉ lệ phần tử có tính chất $\mathcal{P}$ trong mẫu)
- Khi $\begin{cases} nf \geq 5 \\ n(1-f) \geq 5 \end{cases}$ thì tỉ lệ mẫu ngẫu nhiên xấp xỉ phân phối

$$N(f; \frac{f(1-f)}{n}).$$

- Với độ tin cậy $1-\alpha$, khoảng tin cậy chứa tỉ lệ tổng thể là

$$\left[f - z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}; f + z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\right]$$

- Độ chính xác (sai số) của ước lượng là $\epsilon = z_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}}$.

  Theo bất đẳng thức Cauchy, ta có

$$\epsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \leq z_{\alpha/2}\frac{1}{2\sqrt{n}}.$$

  Do đó, sai số tối đa trong ước lượng tỉ lệ tổng thể là $\dfrac{z_{\alpha/2}}{2\sqrt{n}}$.

Bài toán 4. Thăm dò ý kiến của 100 cử tri được chọn ngẫu nhiên tại một thành phố cho thấy có 80% trong số cử tri này ủng hộ ứng viên A. Với độ tin cậy 98%, hãy ước lượng tỉ lệ của tất cả các cử tri ủng hộ ứng viên A tại thành phố này.

Giải Bài toán 4.

- Tỉ lệ mẫu cụ thể $f = 0,8$
- Kích thước mẫu $n = 100$
- Độ tin cậy $1 - \alpha = 0,98$ suy ra $z_{\alpha/2} = 2,33$
- Độ chính xác (sai số) là

$$\epsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} = 2,33\frac{0,4}{10} = 0,0932.$$

- Khoảng tin cậy $[0,7068; 0,8932]$.

Như vậy có từ 70,68% đến 89,32% cử tri ủng hộ ứng viên A.

Bài tập 4. Lấy ngẫu nhiên 200 sản phẩm trong một kho hàng để kiểm tra thì thấy có 21 sản phẩm có lỗi.

a. Với độ tin cậy 95%, hãy ước lượng tỉ lệ sản phẩm lỗi của cả kho hàng.

b. Dựa vào mẫu trên, để ước tính tỉ lệ sản phẩm bị lỗi có độ chính xác là 0,035 thì độ tin cậy bằng bao nhiêu?

c. Dựa vào mẫu trên, nếu muốn ước lượng tỉ lệ sản phẩm bị lỗi với độ chính xác nhỏ hơn 0,01 với độ tin cậy 93% thì cần kiểm tra ít nhất bao nhiêu sản phẩm.