

# **PHẦN 2.**

# **THỐNG KÊ**

**Từ chương này ta bắt đầu nghiên cứu về thống kê. Khoa học thống kê ra đời nhằm mục đích nghiên cứu các phương pháp thu thập, tổ chức và phân tích các dữ liệu nhằm thu nhận thông tin chân thực về đối tượng nghiên cứu một cách khách quan, đáng tin cậy và rút ra những kết luận hợp lý.**

**Thống kê được ứng dụng rộng rãi trong hầu hết các lĩnh vực và có vai trò cực kì quan trọng trong nhiều ngành khoa học, nhất là trong các ngành khoa học thực nghiệm như y khoa, sinh học, nông nghiệp, kinh tế,...**

**Đặc biệt thống kê rất cần cho các cấp lãnh đạo, các nhà quản lý, các nhà hoạch định chính sách. Khoa học thống kê cung cấp cho họ các phương pháp thu thập, xử lý và diễn giải các phân tích về dân số, kinh tế, giáo dục, ... để từ đó có thể vạch chính sách và ra các quyết định đúng đắn.**

**Thống kê** là bộ môn toán học nghiên cứu quy luật của các hiện tượng ngẫu nhiên có tính chất số lớn trên cơ sở thu thập và xử lý các số liệu thống kê – các kết quả quan sát. Như vậy, nội dung chủ yếu của thống kê toán là xây dựng các phương pháp thu thập và xử lý các số liệu thống kê nhằm rút ra các kết luận khoa học và thực tiễn.

**Các phương pháp thống kê toán là công cụ để giải quyết nhiều vấn đề khoa học và thực tiễn nảy sinh trong các lĩnh vực khác nhau của tự nhiên và kinh tế - xã hội.**

# **CHƯƠNG 3.**

## **MẪU THỐNG KÊ & ƯỚC LƯỢNG THAM SỐ**

**Chương 3 gồm hai phần:**

**Phần I nói về mẫu và thống kê mô tả, mẫu ngẫu nhiên và các đặc trưng mẫu.**

**Phần II nghiên cứu về các bài toán ước lượng tham số như ước lượng điểm, ước lượng khoảng cho tham số, và một số bài toán cần quan tâm khi xét đến bài toán ước lượng.**



## **TỔNG THỂ**

**Tập hợp các phần tử mà ta cần khảo sát về một hay một số dấu hiệu nào đó gọi là một tổng thể.**

**VÍ DỤ.** Ta cần nghiên cứu chiều cao của các cây bạch đàn trên một vùng đất. Dấu hiệu nghiên cứu là chiều cao. Tổng thể nghiên cứu là tập hợp các cây bạch đàn trên vùng đất đó.

Gọi  $X$  là chiều cao cây bạch đàn (chọn ngẫu nhiên) trong vùng. Khi đó  $X$  là một đại lượng ngẫu nhiên trên tổng thể và ta có thể coi việc nghiên cứu chiều cao của các cây này là nghiên cứu đại lượng ngẫu nhiên  $X$  trên tổng thể.

**Thông thường, tổng thể cần nghiên cứu là rất lớn, ta không thể chọn hết được tất cả các phần tử của tổng thể để nghiên cứu bởi vì:**

**Tốn kém nhiều (về thời gian, nhân lực, kinh tế, ...).**

**Việc quan sát, kiểm tra có thể làm hư hại các phần tử của tổng thể (chẳng hạn kiểm tra hàm lượng các chất có trong các hộp sữa, viên thuốc, ...).**

**Có nhiều trường hợp không thể xác định toàn bộ các phần tử của tổng thể (chẳng hạn kiểm tra các bệnh nhân HIV, chiều cao của một loài cây, ...).**

Do đó, ta chỉ thực hiện nghiên cứu trên các mẫu ( $n$  phần tử) là tập con được chọn ra từ tổng thể ( $N$  phần tử).

**TỔNG THỂ**

**MẪU**

## **PHƯƠNG PHÁP CHỌN MẪU**

**Để việc nghiên cứu trên mẫu cho ta kết quả gần với kết quả mà ta mong muốn khi nghiên cứu trên tổng thể thì mẫu được chọn phải có tính ngẫu nhiên, mang tính đại diện cho tổng thể và các số liệu phải đạt độ chính xác nào đó. Việc thu hẹp phạm vi nghiên cứu giúp việc xử lý cho ta kết quả vừa nhanh vừa đỡ tốn kém mà vẫn đạt được độ chính xác và tin cậy cần thiết.**

**Cụ thể, một mẫu ngẫu nhiên gồm  $n$  phần tử được chọn ra từ một tổng thể phải thỏa các điều kiện sau:**

- Mỗi phần tử trong tổng thể phải được chọn ngẫu nhiên và độc lập.**
- Mỗi phần tử trong tổng thể có khả năng được chọn như nhau (xác suất được chọn bằng nhau).**
- Mọi mẫu cỡ  $n$  đều có cùng khả năng khi được chọn từ tổng thể.**

**Thông thường, với kích thước mẫu đạt mức đủ lớn, việc coi mẫu thỏa yêu cầu bảo đảm những điều kiện trên nói chung là có thể chấp nhận được.**



**Tuy nhiên, nhiều khi do thực tế không thể làm như trên hay do tổng thể không xác định, nhiều trường hợp khó lấy được mẫu bảo đảm tính ngẫu nhiên và mang tính đại diện cho tổng thể.**

**Thường người ta yêu cầu:**

**Người lấy mẫu là các chuyên gia được đào tạo bài bản và có kinh nghiệm với các dạng tổng thể cần lấy mẫu.**

**Các dụng cụ, máy móc để cân đo và ghi dữ liệu phải được chọn lựa và tinh chỉnh để có độ chính xác đạt mức chấp nhận được.**

**Ngoài ra số phần tử  $n$  trong mẫu không được quá bé (chẳng hạn, tác giả Calvin Dytham đề nghị mức tối thiểu kích thước mẫu  $n \geq 20$  trong các thống kê sinh học thường gặp).**

**Phương pháp lấy mẫu sẽ được các chuyên gia lựa chọn phù hợp với các dạng tổng thể phải lấy mẫu và phù hợp với thực tế. Điều này sẽ được trình bày cụ thể hơn trong các giáo trình của thống kê ứng dụng.**

**Hiện nay, có nhiều phương pháp khác nhau để chọn mẫu ngẫu nhiên, nhưng khó có thể nói rằng phương pháp nào là tốt nhất. Việc chọn phương pháp phù hợp phụ thuộc vào đối tượng cụ thể và thói quen hay sở trường của nhà nghiên cứu.**

**Trong khuôn khổ bài giảng này, những ví dụ và các bài tập chỉ đơn giản mang tính minh họa cho lý thuyết thống kê nên ta chỉ xét các tổng thể là xác định và cho phép giả định mẫu thỏa yêu cầu bảo đảm tính ngẫu nhiên và tính đại diện cho tổng thể.**

Giả sử xét đại lượng ngẫu nhiên  $X$  trên một tổng thể. Ta sẽ  $n$  lần thực hiện loạt các hành động sau: lấy ngẫu nhiên một phần tử của tổng thể, “đo” giá trị  $X$ . Giá trị  $X$  “đo” được trên phần tử thứ  $i$  được gán là trị  $X_i$  ( $i = 1, \dots, n$ ). Các đại lượng  $X_1, \dots, X_n$  là các đại lượng ngẫu nhiên độc lập, chúng là các bản sao của  $X$  và có cùng qui luật phân phối xác suất với  $X$ . Ta gọi bộ thứ tự  $n$  đại lượng ngẫu nhiên  $(X_1, \dots, X_n)$  là mẫu ngẫu nhiên (kích thước  $n$ ) của đại lượng ngẫu nhiên  $X$ .

**VÍ DỤ.** Ta cần nghiên cứu chiều cao của các cây bạch đàn trên một vùng đất. Gọi  $X$  là chiều cao cây bạch đàn trong vùng. Ta dự định sẽ lấy mẫu ngẫu nhiên là 100 cây bạch đàn ở trong vùng rồi đo chiều cao.



ĐẶC TRƯNG TỔNG THỂ	ĐẶC TRƯNG MẪU
Trung bình tổng thể: $E(X) = \mu$	Trung bình mẫu: $\bar{X}$
Phương sai tổng thể: $D(X) = \sigma^2$	Phương sai mẫu: $s^2$
Độ lệch tổng thể: $\sigma$	Độ lệch mẫu: $s$
Tỷ lệ tổng thể: $P = M / N$	Tỷ lệ mẫu: $F = m / n$

# TRUNG BÌNH MẪU:

$$\bar{X} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}$$

## PHƯƠNG SAI MẪU (hiệu chỉnh):

$$S^2 = \left[ \frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k}{n} - (\bar{X})^2 \right] \frac{n}{n-1}$$

**Khi kích thước mẫu  $n$  đủ lớn, có thể coi:**

**Trung bình mẫu  $\approx$  Trung bình tổng thể**

**Phương sai mẫu  $\approx$  Phương sai tổng thể**

**Độ lệch mẫu  $\approx$  Độ lệch tổng thể**

**Tỉ lệ mẫu  $\approx$  Tỉ lệ tổng thể**

**VÍ DỤ.** Điều tra năng suất lúa trong một vùng lớn. Kiểm tra 100 ha trồng lúa được chọn ngẫu nhiên, người ta thu được bảng số liệu:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích	10	20	30	15	10	10	5

- a) Tính trung bình mẫu, phương sai mẫu, độ lệch mẫu?
- b) Xấp xỉ năng suất trung bình của lúa trong vùng?
- c) Xấp xỉ độ phân tán năng suất của lúa trong vùng?

a) **TRUNG BÌNH MẪU:**

$$\bar{X} = \frac{41 \cdot 10 + 44 \cdot 20 + 45 \cdot 30 + 46 \cdot 15 + 48 \cdot 10 + 52 \cdot 10 + 54 \cdot 5}{100} = 46$$

**PHƯƠNG SAI MẪU:**

$$S^2 = \left( \frac{41^2 \cdot 10 + 44^2 \cdot 20 + 45^2 \cdot 30 + 46^2 \cdot 15 + 48^2 \cdot 10 + 52^2 \cdot 10 + 54^2 \cdot 5}{100} - 46^2 \right) \frac{100}{99}$$

$\approx 10.9091$

**ĐỘ LỆCH MẪU:**  $S \approx 3,3029$  (tạ/ha)

**b) Trung bình tổng thể  $\approx$  Trung bình mẫu.**

Năng suất trung bình của lúa trong vùng xấp xỉ 46 (tạ/ha).

**c) Độ lệch tổng thể  $\approx$  Độ lệch mẫu.**

Độ phân tán năng suất của lúa trong vùng xấp xỉ 3,3029 (tạ/ha).

Cách xấp xỉ như trên  
gọi là **ước lượng điểm**  
(dùng một giá trị để  
xấp xỉ).



Khi nghiên cứu đại lượng ngẫu nhiên  $X$  trên một tổng thể, ta thường quan tâm đến tham số đặc trưng của đại lượng ngẫu nhiên này (trung bình của tổng thể, tỉ lệ của tổng thể, phương sai của tổng thể, ...) và nếu tham số này vẫn chưa được xác định thì một trong các bài toán cơ bản của thống kê là **ước lượng** tham số bằng phương pháp mẫu, tức dựa vào mẫu ngẫu nhiên thu thập được, ta đưa ra một giá trị hoặc một khoảng giá trị để đánh giá tham số đó.

**Phương pháp ước lượng điểm nói trên có một nhược điểm cơ bản là khi kích thước mẫu nhỏ thì ước lượng điểm tìm được có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng, tức là sai số của ước lượng có thể rất lớn. Mặt khác, dùng ước lượng điểm không thể đánh giá được khả năng mắc sai lầm khi ước lượng bằng bao nhiêu.**

**Do đó, người ta thường  
sử dụng phương pháp  
ước lượng bằng khoảng  
tin cậy.**

## ƯỚC LƯỢNG ĐIỂM

**Trung bình mẫu  $\approx$  Trung bình tổng thể**

**Phương sai mẫu  $\approx$  Phương sai tổng thể**

**Độ lệch mẫu  $\approx$  Độ lệch tổng thể**

**Tỉ lệ mẫu  $\approx$  Tỉ lệ tổng thể**

## ƯỚC LƯỢNG KHOẢNG

Giả sử trong tổng thể có đặc trưng  $A$  nào đó chưa biết và cần ước lượng.

Cách ước lượng khoảng cho  $A$  là cách làm như sau: chỉ ra khoảng  $(a, b)$  nào đó sao cho  $A$  nằm trong khoảng đó với xác suất cao.

$$P[A \in (a, b)] = 1 - \alpha \quad (\alpha \text{ nhỏ})$$

**Mong muốn:  $A \in (a, b)$ .**

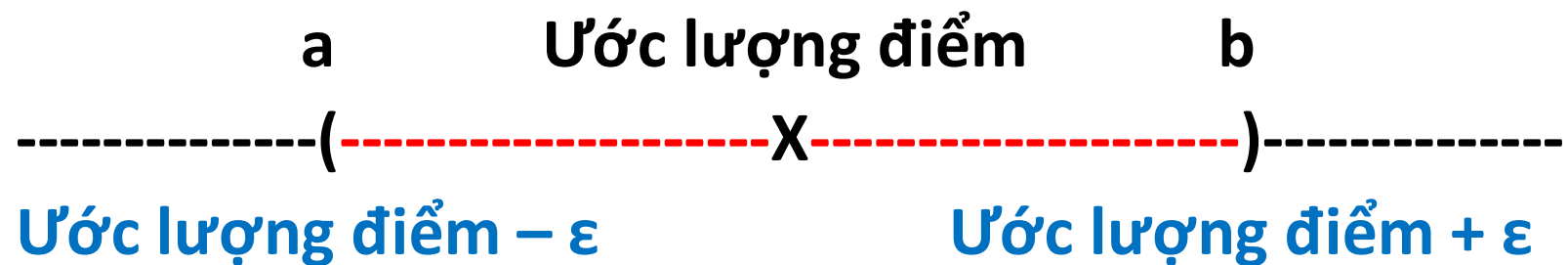
**Nếu  $A \in (a, b) \Rightarrow$  ước lượng đúng.**

**Nếu  $A \notin (a, b) \Rightarrow$  ước lượng sai.**

**Xác suất đúng =  $P[A \in (a, b)]$  (tăng lên)**

**Xác suất sai =  $P[A \notin (a, b)]$  (giảm xuống)**

**ƯỚC LƯỢNG ĐIỂM** nằm trong **ƯỚC LƯỢNG KHOẢNG**



**Bán kính =  $\epsilon$  (độ chính xác, sai số)**

**Đường kính =  $2\epsilon$**

**Mong muốn:**

- **Tăng xác suất đúng lên.**
- **Giảm đường kính xuống.**

**Giảm đường kính xuống**

**=> Tăng xác suất sai lên !**



**Xác suất đúng  $\leq 100\%$**

**$0 \leq \text{xác suất sai} \leq \alpha$**

**Xác suất đúng  $\geq 1 - \alpha$**

**Độ tin cậy**

**Tìm đường kính bé nhất.**

# **BÀI TOÁN TỐI ƯU:**

**Xác suất đúng  $\geq 1 - \alpha$**

**Đường kính bé nhất.**

**Phương pháp ước lượng bằng khoảng tin cậy khắc phục được các nhược điểm của phương pháp ước lượng điểm. Chẳng những nó làm tăng độ chính xác của ước lượng mà còn đánh giá được mức độ tin cậy của ước lượng đó nữa. Tuy nhiên nó cũng chứa đựng khả năng mắc sai lầm bằng  $\alpha$ .**

## ƯỚC LƯỢNG KHOẢNG CHO TRUNG BÌNH CỦA TỔNG THỂ $E(X)$

Xét biến ngẫu nhiên  $X$ .

Từ tổng thể lấy ra mẫu ngẫu nhiên.

Ước lượng điểm cho  $E(X)$  là:  $\bar{X}$

Ước lượng khoảng cho  $E(X)$  là:  $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$

**Biết  $\sigma$** 

$$\varepsilon = K(\alpha / 2) \frac{\sigma}{\sqrt{n}}$$

**Không biết  $\sigma$** 

$$\varepsilon = K(\alpha / 2) \frac{s}{\sqrt{n}}$$

Trường hợp 1	Trường hợp 2	Trường hợp 3
Biết $X$ phân phối chuẩn. <b>Chưa biết <math>\sigma</math>.</b>	$n$ đủ lớn $(n \geq 30)$ .	Biết $X$ phân phối chuẩn. <b>Biết <math>\sigma</math>.</b>
$K(\alpha/2)$ tra Bảng phân vị Student.	$K(\alpha/2)$ tra Bảng phân vị Chuẩn tắc.	$K(\alpha/2)$ tra Bảng phân vị Chuẩn tắc.

BẢNG 2. GIÁ TRỊ TỚI HẠN PHÂN PHỐI CHUẨN TẮC:  $u(\alpha)$ 

$u$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

BẢNG 3. GIÁ TRỊ TỚI HẠN PHÂN PHỐI STUDENT:  $t_{\alpha}$ 

$n \backslash \alpha$	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
1	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.3	318.309
2	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.327
3	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215
4	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893
6	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501
9	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297
10	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930
13	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733
16	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686
17	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646
18	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610
19	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579
20	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552
21	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527
22	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505
23	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467
25	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450
26	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435
27	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421
28	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408
29	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396
30	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385
40	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307
50	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261
60	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232
70	0.847	1.044	1.294	1.667	1.994	2.381	2.648	2.899	3.211
80	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195
90	0.846	1.042	1.291	1.662	1.987	2.368	2.632	2.878	3.183
100	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174
120	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160

**VÍ DỤ.** Lấy ngẫu nhiên 25 bao thuốc do một công ty được sản xuất ra. Tính được trung bình mẫu bằng 39,8 (g) và phương sai mẫu bằng 0,144. Giả thiết trọng lượng các bao thuốc bột là biến ngẫu nhiên có phân phối chuẩn.

a) Hãy ước lượng trọng lượng trung bình của các bao thuốc do công ty sản xuất với độ tin cậy 95% ?

b) Hãy ước lượng trọng lượng trung bình của các bao thuốc do công ty sản xuất với độ tin cậy 99%, biết thêm độ lệch chuẩn  $\sigma = 0,4$  (g) ?



a) Gọi  $X$  là trọng lượng của bao thuốc do công ty sản xuất.

Trọng lượng trung bình của các bao thuốc là:  $E(X) = ?$

Ước lượng điểm cho  $E(X)$  là:  $\bar{X}$

Ước lượng khoảng cho  $E(X)$  là:  $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$

$$\bar{X} \pm K(\alpha / 2) \frac{s}{\sqrt{n}} = 39,8 \pm 2,0639 \frac{\sqrt{0,144}}{\sqrt{25}} \approx 39,8 \pm 0,1566$$

Độ tin cậy 95% ( $1 - \alpha = 0,95$ )  $\Rightarrow \alpha = 0,05$

Tra Bảng phân vị Student ( $n - 1$  bậc tự do):

$$K(\alpha/2) = t_{24}(0,025) = 2,0639$$

Vậy trọng lượng trung bình của các bao thuốc do công ty sản xuất là vào khoảng  $(39,6434; 39,9566)$  (g).

$$\bar{X} \pm K(\alpha / 2) \frac{\sigma}{\sqrt{n}} = 39,8 \pm 2,5758 \frac{0,4}{\sqrt{25}} \\ \approx 39,8 \pm 0,2061$$

**b) Độ tin cậy 99% ( $1 - \alpha = 0,99$ )  $\Rightarrow \alpha = 0,01$**

**Tra Bảng phân vị Chuẩn tắc:**

$$K(\alpha/2) = u(0,005) = 2,5758$$

**Trọng lượng trung bình của các bao thuốc do công ty sản xuất là vào khoảng (39,5939; 40,0061) (g).**

## ƯỚC LƯỢNG KHOẢNG CHO TỶ LỆ CỦA TỔNG THỂ $P = M/N$

Từ tổng thể lấy ra mẫu ngẫu nhiên  
(cỡ mẫu  $n$  đủ lớn).

Ước lượng điểm cho  $p$  là:  $f$

Ước lượng khoảng cho  $p$  là:  $(f - \varepsilon, f + \varepsilon)$

$$\varepsilon = K(\alpha / 2) \frac{\sqrt{f(1-f)}}{\sqrt{n}}$$

**$K(\alpha/2)$  tra Bảng phân vị Chuẩn tắc.**

**VÍ DỤ.** Nghiên cứu nhu cầu tiêu dùng một mặt hàng ở một thành phố, người ta điều tra 1000 người được chọn ngẫu nhiên của thành phố và thấy 600 người có nhu cầu. Hãy ước lượng tỷ lệ người có nhu cầu về mặt hàng đó trong toàn thành phố với độ tin cậy 95%?

Gọi tỷ lệ người có nhu cầu về mặt hàng đó trong toàn thành phố là  $p$ .

Ước lượng điểm cho  $p$  là:  $f = m/n = 600/1000 = 0,6$

Ước lượng khoảng cho  $p$  là:  $(f - \varepsilon, f + \varepsilon)$

$$f \pm K(\alpha / 2) \frac{\sqrt{f(1-f)}}{\sqrt{n}} = 0,6 \pm 1,96 \frac{\sqrt{0,6(1-0,6)}}{\sqrt{1000}} = 0,6 \pm 0,0304$$

Độ tin cậy 95% ( $1 - \alpha = 0,95$ )  $\Rightarrow \alpha = 0,05$

Tra Bảng phân vị Chuẩn tắc:  $K(\alpha/2) = K(0,025) = 1,96$

Vậy tỷ lệ người trong thành phố có nhu cầu về mặt hàng đó là khoảng (56,96%; 63,04%).

# **CÁC BÀI TOÁN LIÊN QUAN ĐẾN BÀI TOÁN ƯỚC LƯỢNG**

- Bài toán xác định cỡ mẫu.**
- Bài toán xác định độ tin cậy.**

# XÁC ĐỊNH KÍCH THƯỚC MẪU

$$\varepsilon = K(\alpha / 2) \frac{\sigma}{\sqrt{n}} \quad \Rightarrow n = \left( K(\alpha / 2) \frac{\sigma}{\varepsilon} \right)^2$$

$$\varepsilon = K(\alpha / 2) \frac{s}{\sqrt{n}} \quad \Rightarrow n = \left( K(\alpha / 2) \frac{s}{\varepsilon} \right)^2$$

$$\varepsilon = K(\alpha / 2) \frac{\sqrt{f(1-f)}}{\sqrt{n}} \quad \Rightarrow n = \left( K(\alpha / 2) \frac{\sqrt{f(1-f)}}{\varepsilon} \right)^2$$



**VÍ DỤ.** Giả sử cần ước lượng năng suất lúa trung bình cho một vùng lớn. Sau khi điều tra 100 ha lúa chọn ngẫu nhiên trong vùng, ta có ước lượng năng suất trung bình của vùng ước chừng 45,35 đến 46,65 (tạ/ha) với độ tin cậy 95%. Tìm số ha lúa phải điều tra thêm để khoảng ước lượng có độ dài như cũ và có độ tin cậy là 99%?

Gọi X là năng suất lúa trong vùng. Năng suất lúa trung bình trong vùng là E(X).

Khoảng ước lượng cũ:  $\bar{X} \pm \varepsilon = (45,35; 46,65) \Rightarrow \varepsilon = (46,65 - 45,35) / 2 = 0,65$

Độ tin cậy cũ 95% ( $1 - \alpha = 0,95$ )  $\Rightarrow K(\alpha/2) = K(0,025) = 1,96$

$$\varepsilon = K(\alpha/2) \frac{s}{\sqrt{n}} \Rightarrow s = \frac{\varepsilon \sqrt{n}}{K(\alpha/2)} = \frac{0,65 \cdot \sqrt{100}}{1,96} \approx 3,3163$$

Khoảng ước lượng mới có độ dài như cũ ( $\varepsilon' = \varepsilon$ ). Tìm cỡ mẫu mới  $n' = ?$

Độ tin cậy mới 99% ( $1 - \alpha' = 0,99$ )  $\Rightarrow \alpha' = 0,01$

Tra Bảng phân vị Chuẩn tắc:  $K(\alpha'/2) = K(0,005) = 2,5758$

$$\varepsilon' = K(\alpha'/2) \frac{s'}{\sqrt{n'}} \Rightarrow n' = \left( K(\alpha'/2) \frac{s'}{\varepsilon'} \right)^2 = \left( 2,5758 \times \frac{3,3163}{0,65} \right)^2 \approx 172,7$$

Vậy diện tích lúa phải điều tra thêm là:  $173 - 100 = 73$  (ha).

# XÁC ĐỊNH ĐỘ TIN CẬY

$$\varepsilon = K(\alpha / 2) \frac{\sigma}{\sqrt{n}} \Rightarrow K(\alpha / 2) = \frac{\varepsilon \sqrt{n}}{\sigma} \Rightarrow \alpha / 2 \Rightarrow \alpha \Rightarrow 1 - \alpha$$

$$\varepsilon = K(\alpha / 2) \frac{s}{\sqrt{n}} \Rightarrow K(\alpha / 2) = \frac{\varepsilon \sqrt{n}}{s} \Rightarrow \alpha / 2 \Rightarrow \alpha \Rightarrow 1 - \alpha$$

$$\varepsilon = K(\alpha / 2) \frac{\sqrt{f(1-f)}}{\sqrt{n}} \Rightarrow K(\alpha / 2) = \frac{\varepsilon \sqrt{n}}{\sqrt{f(1-f)}} \Rightarrow \alpha / 2 \Rightarrow \alpha \Rightarrow 1 - \alpha$$

**VÍ DỤ.** Nghiên cứu nhu cầu tiêu dùng một mặt hàng ở một thành phố, người ta điều tra 1000 người được chọn ngẫu nhiên và được tỉ lệ mẫu bằng 60%. Ước lượng tỉ lệ người trong thành phố có nhu cầu về mặt hàng đó khoảng 55% đến 65% thì độ tin cậy bao nhiêu ?

$$\varepsilon = K(\alpha / 2) \frac{\sqrt{f(1-f)}}{\sqrt{n}} \Rightarrow K(\alpha / 2) = \frac{\varepsilon \sqrt{n}}{\sqrt{f(1-f)}} = \frac{0,05 \sqrt{1000}}{\sqrt{0,6(1-0,6)}} \approx 3,2275$$

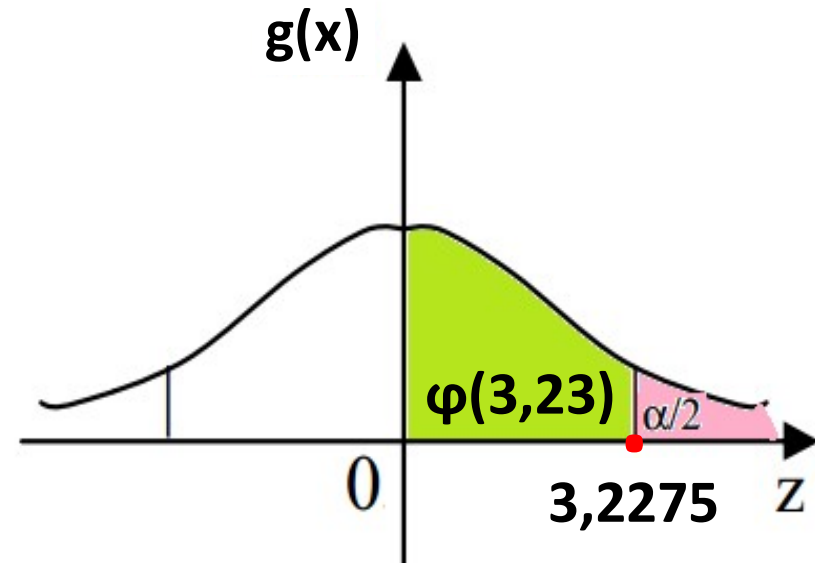
Khoảng ước lượng (55%, 65%)

$$\varepsilon = (0,65 - 0,55) / 2 = 0,05$$

$$\begin{aligned} \alpha/2 &= 0,5 - \varphi(3,23) \\ &= 0,5 - 0,49938 \\ &= 0,00062 \end{aligned}$$

$$\alpha = 0,00124 \Rightarrow 1 - \alpha = 0,99876$$

Độ tin cậy cần tìm là 99,876%.



Phân phối chuẩn tắc