TRƯỜNG ĐẠI HỌC TÀI CHÍNH-MARKETING
KHOA CƠ BẢN
BỘ MÔN TOÁN – THỐNG KÊ

BÀI GIẢNG

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN

Giảng viên
ThS. Lê Trường Giang

LÝ THUYẾT MẪU

Bài 1. TỔNG THỂ VÀ MẪU

1. Khái niệm về tổng thể và mẫu
2. Mẫu ngẫu nhiên và mẫu cụ thể
3. Hàm phân phối thực nghiệm

Bài 2. CÁC ĐẶC TRƯNG CỦA MẪU VÀ PHÂN PHỐI XÁC SUẤT.

1. Thống kê
2. Trung bình mẫu ngẫu nhiên
3. Tỉ lệ mẫu ngẫu nhiên
4. Phương sai mẫu ngẫu nhiên
5. Phương sai mẫu có điều chỉnh

Chương 4. DỮ LIỆU THỐNG KÊ

Bài 1. TỔNG THỂ VÀ MẪU

1. Khái niệm về tổng thể và mẫu
2. Mẫu ngẫu nhiên và mẫu cụ thể
3. Hàm phân phối thực nghiệm

Bài 1. TỔNG THỂ VÀ MẪU

Lấy mẫu ngẫu nhiên

Tổng thể

X: Biến ngẫu nhiên tổng thể

N: Kích thước tổng thể.

μ: Trung bình tổng thể.

σ: Độ lệch chuẩn tổng thể.

p: tỷ lệ tổng thể.

Mẫu

n: Kích thước mẫu.

$\overline{X}$: Trung bình mẫu.

S : Độ lệch chuẩn mẫu.

F_n : tỷ lệ mẫu.

★Ước lượng tham số

★Kiểm định giả thuyết

2. Mẫu nghẫu nhiên và mẫu cụ thể

a. Mẫu ngẫu nhiên

Mẫu ngẫu nhiên kích thước n được lập từ tổng thể X là một bộ gồm n biến ngẫu nhiên $X_i, i = 1,2,...,n$ độc lập và cùng phân phối với biến ngẫu nhiên X, kí hiệu là $W_n = (X_1, X_2, ..., X_n)$.

b. Mẫu cụ thể

Mẫu ngẫu nhiên này nhận n giá trị cụ thể $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$. Khi đó một bộ gồm n giá trị $w_n = (x_1, x_2, ..., x_n)$ được gọi là một mẫu cụ thể có kích thước n.

2. Mẫu nghẫu nhiên và mẫu cụ thể

c. Ví dụ

Thu nhập hàng tháng của mỗi gia đình tỉnh A (đơn vị triệu đồng).

{100,121, 230, 89,…197,… }.

Tập giá trị của biến ngẫu nhiên tổng thể X chỉ thu nhập của mỗi gia đình tỉnh A.

Một mẫu ngẫu nhiên gồm 50 hộ gia đình trong tỉnh A.

$$\{X_1, X_2, \ldots X_{50}\}.$$

Một mẫu cụ thể

{121, 203, 92,…120}

gồm 50 giá trị thu nhập của 50 hộ gia đình.

2. Mẫu nghẫu nhiên và mẫu cụ thể

Bảng phân phối thực nghiệm của mẫu cụ thể

Mẫu cụ thể $w_n = (x_1, x_2, ..., x_n)$, $x_1 < x_2 < ... < x_k$ và $n_1 + n_2 + ... + n_k = n$.

Bảng phân phối tần số thực nghiệm

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Bảng phân phối tần suất thực nghiệm

x_i	x_1	x_2	…	x_k
f_i	f_1	f_2	…	f_k

Trong đó, $f_i = \dfrac{n_i}{n}$.

2. Mẫu nghẫu nhiên và mẫu cụ thể

Bảng phân phối thực nghiệm của mẫu cụ thể

VD 1. Kiểm tra ngẫu nhiên 50 sinh viên. Ta sắp xếp điểm số X thu được theo thứ tự tăng dần và số sinh viên n có điểm tương ứng vào bảng như sau:

X (điểm)	2	4	5	6	7	8	9	10
n (số SV)	4	6	20	10	5	2	2	1

2. Mẫu nghẫu nhiên và mẫu cụ thể

Bảng phân phối thực nghiệm của mẫu cụ thể

Giá trị của mẫu cụ thể dạng ghép lớp

$[x_i; x_i + h]$	$[x_1; x_1 + h]$	$[x_2; x_2 + h]$	$\cdots$	$[x_k; x_k + h]$
n_i	n_1	n_2	$\ldots$	n_k

Trong trường hợp này ta sử dụng giá trị trung bình trên từng khoảng

x'_i	x'_1	x'_2	$\cdots$	x'_k
n_i	n_1	n_2	$\ldots$	n_k

với $x'_i = \dfrac{x_i + x_i + h}{2}$

2. Mẫu nghẫu nhiên và mẫu cụ thể

VD 2. Đo chiều cao X (cm) của $n = 100$ thanh niên. Vì chiều cao khác nhau nên để tiện việc sắp xếp, người ta chia chiều cao thành nhiều khoảng.

Các thanh niên có chiều cao trong cùng 1 khoảng được xem là cao như nhau. Khi đó, ta có bảng số liệu ở dạng khoảng như sau:

X	148-152	152-156	156-160	160-164	164-168
n	5	20	35	25	15

Khi cần tính toán, người ta chọn ***số trung bình*** của mỗi khoảng để đưa số liệu trên về dạng bảng:

X	150	154	158	162	166
n	5	20	35	25	15

3. Hàm phân phối thực nghiệm

Giả sử $\left(X_1;X_2;...;X_n\right)$ là một mẫu ngẫu nhiên được xây dựng từ đại lượng ngẫu nhiên X với hàm phân phối xác suất $F_X\left(x\right)$.

Định nghĩa: Hàm phân phối thực nghiệm ngẫu nhiên tương ứng với mẫu $\left(X_1;X_2;...;X_n\right)$, kí hiệu là $F_n\left(x\right)$, xác định bởi công thức sau

$$F_n\left(x\right)=\begin{cases} 0 & \text{nếu } x \le \min(X_1,X_2,...,X_n), \\ \dfrac{k}{n} & \text{nếu có k phần tử trong mẫu} < \text{x}, \\ 1 & \text{nếu } x > \max(X_1,X_2,...,X_n). \end{cases}$$

3. Hàm phân phối thực nghiệm

Định lí Glivenko:

$$P\left(\lim_{n\to\infty}\sup_{-\infty<x<+\infty}\left|F_n(x)-F_X(x)\right|=0\right)=1$$

Ý nghĩa: Hàm phân phối thực nghiệm là một xấp xỉ của hàm phân phối lý thuyết. Xấp xỉ đó càng tốt khi cỡ mẫu *n* càng lớn.

Bài 2. CÁC ĐẶC TRƯNG CỦA MẪU VÀ PHÂN PHỐI XÁC SUẤT.

1. Thống kê
2. Trung bình mẫu ngẫu nhiên
3. Tỉ lệ mẫu ngẫu nhiên
4. Phương sai mẫu ngẫu nhiên
5. Phương sai mẫu có điều chỉnh

1. Thống kê

Thống kê là một hàm xác định trên các biến ngẫu nhiên của mẫu. Một thống kê của mẫu $W_n = (X_1, X_2, ..., X_n)$ được kí hiệu là $G = G(X_{1,} X_2, ..., X_n)$.

Chẳng hạn, $\bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$ là một thống kê trên mâu ngẫu nhiên $\mathrm{W}_n = (X_1, X_2, ..., X_n)$.

2. Trung bình mẫu ngẫu nhiên

Mẫu ngẫu nhiên $W_n = \left(X_1, X_2, ..., X_n\right)$, trung bình mẫu ngẫu nhiên là một thống kê $\bar{X}$ được xác định

$$\bar{X} = \frac{1}{n}\left(X_1 + X_2 + ... + X_n\right).$$

Mẫu cụ thể $\mathrm{w}_n = \left(x_1, x_2, ..., x_n\right)$, trung bình thực nghiệm $\bar{x}$ được cho bởi

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \bar{x} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i .$$

Một số đặc trưng của trung bình mẫu ngẫu nhiên

$i.\ E\left(\bar{X}\right) = \mu \quad ii.\ \mathrm{Var}\left(\bar{X}\right) = \dfrac{\sigma^2}{n}$

$iii.$ Nếu $X \sim N\left(\mu, \sigma^2\right)$ thì $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ và $\dfrac{\left(\bar{X} - \mu\right)\sqrt{n}}{\sigma} \sim N\left(0;1\right)$.

2. Trung bình mẫu ngẫu nhiên

Ví dụ 1. Chiều cao (cm) của một loại cây công nghiệp là BNN tuân theo luật phân phối chuẩn với trung bình là 75 và độ lệch chuẩn là 10. Người ta đo ngẫu nhiên 25 cây loại trên, tính xác suất để chiều cao trung bình của 25 cây đó nằm trong khoảng từ 71cm đến 79cm

ĐS: 0,9554

2. Trung bình mẫu ngẫu nhiên

Định lý giới hạn trung tâm. Cho mẫu ngẫu nhiên $W_n = (X_1, X_2, ..., X_n)$ được thành lập từ biến ngẫu nhiên X có kỳ vọng μ, phương sai σ^2.

Khi đó $\forall x \in \mathbb{R}$

$$\lim_{n \to +\infty} P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < x \right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = P(Z < x), \quad Z \sim N(0;1).$$

Nhận xét. Khi $n > 30$ ta có thể xem thống kê $\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$ có luật phân phối chuẩn tắc $N(0;1)$ cho dù biến ngẫu nhiên tổng thể X có bất kì phân phối nào.

3. Tỉ lệ mẫu ngẫu nhiên

Mẫu $W_n = (X_1, X_2, ..., X_n)$ được lập từ tổng thể $X \sim B(1; p)$, khi đó trung bình $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ được gọi là tỉ lệ mẫu ngẫu nhiên, kí hiệu là F_n.

Ta tính được các đặc trưng sau

$i,\ E(F_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p.$ $\qquad$ $ii.\ Var(F_n) = \frac{p(1-p)}{n}.$

Mẫu cụ thể $w_n = (x_1, x_2, ..., x_n)$, ta có tỉ lệ phần tử có tính chất A trong mẫu là

$$f = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{n_A}{n}$$

với n_A là số phần tử có tính chất A.

3. Tỉ lệ mẫu ngẫu nhiên

Định lý De Moivre – Laplace, từ mẫu ngẫu nhiên $W_n = (X_1, X_2, ..., X_n)$

được lập từ tổng thể $X \sim B(1; p)$, tỉ lệ mẫu là $F_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,

$\forall x \in \mathbb{R}$ ta có

$$\lim_{n\to\infty} P\left(\frac{F_n - p}{\sqrt{\frac{p(1-p)}{n}}} < x.\right) = P(Z < x), \qquad Z \sim N(0;1).$$

Cụ thể $n>30$; $n.p>5$ và $n(p\text{-}1) > 5$ ta có thể sử dụng xấp xỉ trên.

$$\frac{F_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0;1).$$

4. Phương sai mẫu ngẫu nhiên

Cho mẫu ngẫu nhiên $W_n = (X_1, X_2, ..., X_n)$ được lập từ tổng thể X có kỳ vọng μ và phương sai σ^2, thống kê S

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

được gọi là phương sai mẫu.

Độ lệch chuẩn mẫu được định nghĩa $S = \sqrt{S^2}$.

Chú ý. Thống kê S còn được viết dưới dạng sau

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

5. Phương sai mẫu có điều chỉnh

Cho mẫu ngẫu nhiên $W_n = (X_1, X_2, ..., X_n)$ được lập từ tổng thể X có kỳ vọng μ và phương sai σ^2, thống kê S

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

được gọi là phương sai mẫu điều chỉnh. Độ lệch chuẩn mẫu điều chỉnh được định nghĩa $S = \sqrt{S^2}$.

5. Phương sai mẫu có điều chỉnh

Chú ý. Ta có thể biểu diễn phương sai mẫu điều chỉnh

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

Mẫu cụ thể $w_n = (x_1, x_2, ..., x_n)$ kích thước *n* được cho theo bảng tần số sau

x_i	x_1	x_2	…	x_k	
n_i	n_1	n_2	…	n_k	$\sum_{i=1}^{k} n_i = n$

Khi đó, sai sai mẫu điều chỉnh được cho bởi

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - n\left(\overline{x}\right)^2\right].$$

Ví dụ 2. Thống kê lượng đường cát trắng bán ra mỗi ngày của của hàng A cho trong bảng sau

25	27,5	22	25	18	16	20	21,5	16	25
18	17,5	21,5	30	18	25	19,5	20	18,5	21

Tính trung bình và độ lệch chuẩn mẫu điều chỉnh?

Mô tả sự biến thiên của số trung bình: sai số chuẩn

(Trích bài giảng của GS. Nguyễn Văn Tuấn – Australia)
http://www.nguyenvantuan.com

- Nếu chúng ta chọn mẫu *N* lần (mỗi lần với *n* đối tượng), thì chúng ta sẽ có *N* số trung bình. **Độ lệch chuẩn của *N* số trung bình này chính là sai số chuẩn**. Do đó, sai số chuẩn phản ảnh độ dao động hay biến thiên của các số trung bình mẫu (*sample averages*).
- Công thức tính sai số chuẩn (SE – *standard error*):

$$\boxed{SE = \frac{s}{\sqrt{n}}.}$$

Ý nghĩa của độ lệch chuẩn và sai số chuẩn

- Gọi số trung bình của một quần thể là μ (nên nhớ rằng chúng ta không biết giá trị của μ). Gọi số trung bình tính từ mẫu là $\overline{x}$ và độ lệch chuẩn là *s*. Theo lý thuyết xác suất của phân phối chuẩn, chúng ta có thể nói rằng:
  - 95% *cá nhân* trong quần thể đó có giá trị từ $\overline{x} - 1,96 \times s$ đến $\overline{x} + 1,96 \times s$.
  - 95% *số trung bình* tính từ mẫu có giá trị từ $\overline{x} - 1,96 \times SE$ đến $\overline{x} + 1,96 \times SE$.

- Như vậy, *độ lệch chuẩn* phản ảnh *độ biến thiên của một số cá nhân* trong một quần thể. Còn *sai số chuẩn* phản ảnh *độ dao động của các số trung bình* chọn từ quần thể.