Chương 4: Mở đầu về Thống kê

Phan Quang Sáng

Học viện Nông nghiệp Việt Nam

Hà Nội, Ngày 24 tháng 10 năm 2018

http://fita.vnua.edu.vn/vi/pqsang

Nội dung chính

1. Tổng thể và mẫu
   - Khái niệm
   - Các phương pháp lấy mẫu
   - Bố trí mẫu và hàm phân phối mẫu
2. Mẫu ngẫu nhiên và các đặc trưng mẫu
   - Mẫu ngẫu nhiên và hàm mẫu
   - Các đặc trưng mẫu
3. Một số phân phối XS thường gặp trong thống kê
   - Các định lý về phân phối chuẩn
   - Phân phối khi bình phương
   - Phân phối Student
   - Phân phối Fisher-Snedecor
4. Phân vị mức $1 - \alpha$

Tổng thể hay đám đông là tập hợp tất cả các cá thể (phần tử) của một tập hợp cần quan sát nào đó.
Kích thước tổng thể là số phần tử của tổng thể, ký hiệu N (thường rất lớn).
Mẫu là tập hợp một số cá thể lấy ra từ tổng thể. Số lượng cá thể của một mẫu gọi là kích thước mẫu, giả sử n.
Giả sử muốn nghiên cứu đặc tính định lượng X. Với một mẫu cụ thể có kích thước n, cá thể thứ i trong mẫu có đặc tính $X = x_i$ thì bộ số $(x_1, x_2, \cdots, x_n)$ cũng được gọi là một mẫu.
Tập tất cả các bộ số như vậy tương ứng với tập tất cả các mẫu có cùng kích thước n được gọi là không gian mẫu.

Sắp xếp theo bảng tần số và tần suất

Bảng tần số

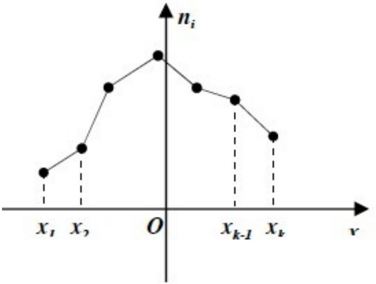
x_i	x_1	x_2	$\ldots$	x_i	$\ldots$	x_k
n_i	n_1	n_2	$\ldots$	n_i	$\ldots$	p_k

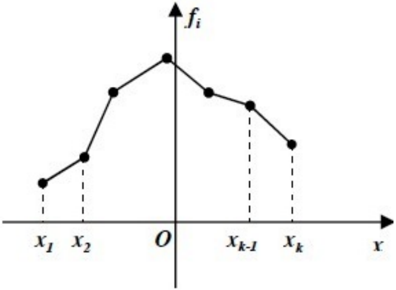
Giá trị $f_i = \frac{n_i}{n}$ gọi là tần suất của cá thể có $X = x_i$ trong mẫu.

Bảng tần suất

x_i	x_1	x_2	$\ldots$	x_i	$\ldots$	x_k
f_i	f_1	f_2	$\ldots$	f_i	$\ldots$	f_k

Biểu diễn đồ thị tần số và tần suất

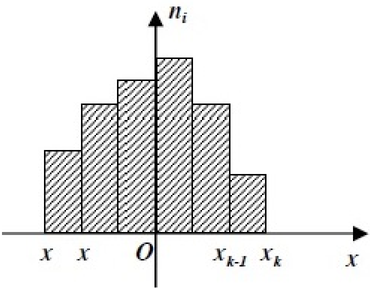


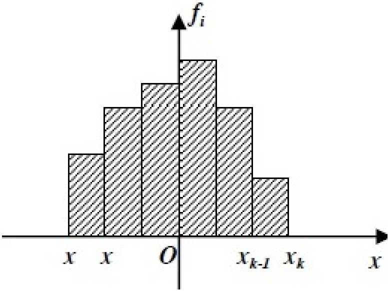


Sắp xếp số liệu theo lớp

Ví dụ:

Lớp	[155, 160)	[160, 165)	[165, 170)	[170, 175)
Tần số	15	13	12	5
Tần suất				





Hàm phân phối mẫu tương ứng với bảng tần suất là

$$F_n(x) = \begin{cases} 0 & \text{nếu} & x \leq x_1 \\ f_1 & \text{nếu} & x_1 \leq x \leq x_2 \\ \dots & \dots & \dots \\ f_1 + f_2 + \cdots + f_i & \text{if} & x_i \leq x \leq x_{i+1} \\ \dots & \dots & \dots \\ 1 & \text{nếu} & x \geq x_k \end{cases}$$

Khi kích thước mẫu n đủ lớn ta có thể lấy hàm phân phối mẫu thay cho hàm phân phối xác suất (chưa biết) của X ở tổng thể.

Giả sử đặc trưng X ở mỗi cá thể ở tổng thể có hàm phân phối xác suất là $F(x)$.
Tiến hành lấy mẫu ngẫu nhiên có kích thước n:
$(X_1, X_2, \cdots, X_n)$.
Khi đó các X_i là các Bnn độc lập (do việc lấy mẫu là ngẫu nhiên) và có cùng phân phối với X.
Ngược lại một véc tơ $(X_1, X_2, \cdots, X_n)$ như thế cũng được gọi là một mẫu ngẫu nhiên lấy ra từ tổng thể đang xét.
Với mỗi mẫu cụ thể ta thu được một thể hiện tương ứng của mẫu ngẫu nhiên là bộ $(x_1, x_2, \cdots, x_n)$.
Hàm mẫu: là một hàm của mẫu ngẫu nhiên dạng
$g(X_1, X_2, \cdots, X_n)$

Trung bình mẫu

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Phương sai mẫu chưa hiệu chỉnh

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

Phương sai mẫu đã hiệu chỉnh

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

Trung bình mẫu

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Phương sai mẫu chưa hiệu chỉnh

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

Phương sai mẫu đã hiệu chỉnh

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

Ví dụ...

Định lý

Nếu $X \sim N(\mu, \sigma^2)$ thì $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.

Định lý

Nếu $(X_1, X_2, \cdots, X_n)$ là mẫu ngẫu nhiên của $X \sim N(\mu, \sigma^2)$ thì

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

Định nghĩa

Bnn $Z \sim \chi^2_n$ (khi bình phương bậc n) nếu

$$Z = \sum_{i=1}^{n} X_i^2,$$

trong đó các $X_i \sim N(0,1)$ và độc lập.

Định nghĩa

Bnn $Z \sim T_n$ (hoặc $Z \sim S_n$) nếu

$$Z = \frac{X}{\sqrt{Y}}\sqrt{n},$$

trong đó X, Y là các Bnn độc lập, $X \sim N(0,1)$, $Y \sim \chi^2_n$.

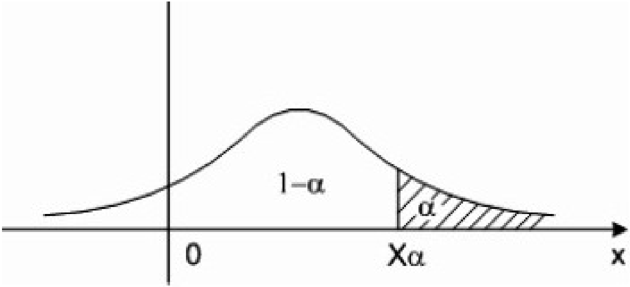
Định lý

Nếu $(X_1, X_2, \cdots, X_n)$ là mẫu ngẫu nhiên của $X \sim N(\mu, \sigma^2)$ thì

$$Z = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim T_{n-1}.$$

Định nghĩa

Cho X là Bnn liên tục có hàm pp là $F(x)$. Với $\alpha \in (0; 1)$, phân vị mức $1 - \alpha$ là giá trị X_α sao cho $F(X_\alpha) = 1 - \alpha$.



Như vậy $P(X > X_\alpha) = \alpha$.

Ký hiệu của phân vị trong một số T/h đặc biệt

Khi $X \sim N(0,1)$ thì ký hiêu $X_\alpha \to U_\alpha$.

Khi $X \sim \chi^2_n$ thì ký hiêu $X_\alpha \to \chi^2_{n;\alpha}$, hoặc $\chi^2(n;\alpha)$.

Khi $X \sim T_n$ thì ký hiêu $X_\alpha \to t_{n;\alpha}$, hoặc $t(n;\alpha)$.

Cách tra bảng:
+ U_α thì tra $\Phi^{-1}(1-\alpha)$
+ $\chi^2_{n;\alpha}$ thì tra hàng n cột α
+ $t_{n;\alpha}$ thì tra hàng n cột 2α

Ví dụ: $U_{0,025} = 1,96$; $\chi^2_{10;\ 0,05} = 18,307$; $t_{0,025;\ 12} = 2,179$.