Phép kiểm Z & t

Nguyễn Quang Vinh – Nguyễn Thị Từ Vân

Phân loại dữ liệu

(1) Định tính → thuộc tính (tần xuất)

Định lượng → số lượng

(2) Rời

Liên tục

(3) Định danh

Thứ tự

Khoảng

Tỷ số

SỐ ĐO VỊ TRÍ DỮ LIỆU (1)

Chuẩn hóa dữ liệu từ mẫu

z-score của mẫu

$$z = \frac{x - \bar{x}}{s}$$

$$z\text{-}scores = \text{standardized values}$$

SỐ ĐO VỊ TRÍ DỮ LIỆU (2)

Bách phân vị (p^{th}):

$$p\% \text{ or less} < p^{th} < (100-p)\% \text{ or less}$$

$$c = \frac{n.p}{100}$$

Where: n is the size of sample

Note: c is the position of data, not value

Tứ phân vị (Q):

$$Q_1 = 25^{th} \text{ percentile } = \frac{(n+1)^{th}}{4} \text{ ordered observation}$$

$$Q_2 = 50^{th} \text{ percentile } = \frac{2(n+1)^{th}}{4} \text{ ordered observation}$$

$$Q_3 = 75^{th} \text{ percentile } = \frac{3(n+1)^{th}}{4} \text{ ordered observation}$$

- Khoảng cách phân vị Q_{3-1}

$$IQR = Q_3 - Q_1$$

Diễn giải trung bình và độ lệch chuẩn

Quy tắc chung** áp dụng cho phân phối *bình thường* (*không chứng minh được*)

68% giá trị của dữ liệu nằm trong khoảng trung bình ± 1 S.D.

95% giá trị của dữ liệu nằm trong khoảng trung bình ± 2 S.D.

99.7% hoặc *hầu như tất cả* giá trị của dữ liệu nằm trong khoảng trung bình ± 3 S.D.

***Nói một cách chặt chẽ, quy tắc này chỉ áp dụng cho các giá trị của <u>tổng thể</u>. Tuy nhiên, có thể ứng dụng quy tắc trong trường hợp <u>mẫu lớn</u> có hình biểu diễn histogram xấp xỉ hình chuông.*

Phân phối Normal

- Phân phối ***quan trọng nhất***
- Có vai trò chính yếu trong ứng dụng ***nhiều*** kỹ thuật thống kê
- Do Abraham De Moivre* (1667 - 1754) đề xuất vào 12 tháng 11 năm 1733
- Carl Friedrich Gauss** (1777 - 1855) $\rightarrow$ phân phối Gaussian

**nhà toán học Anh gốc Pháp*

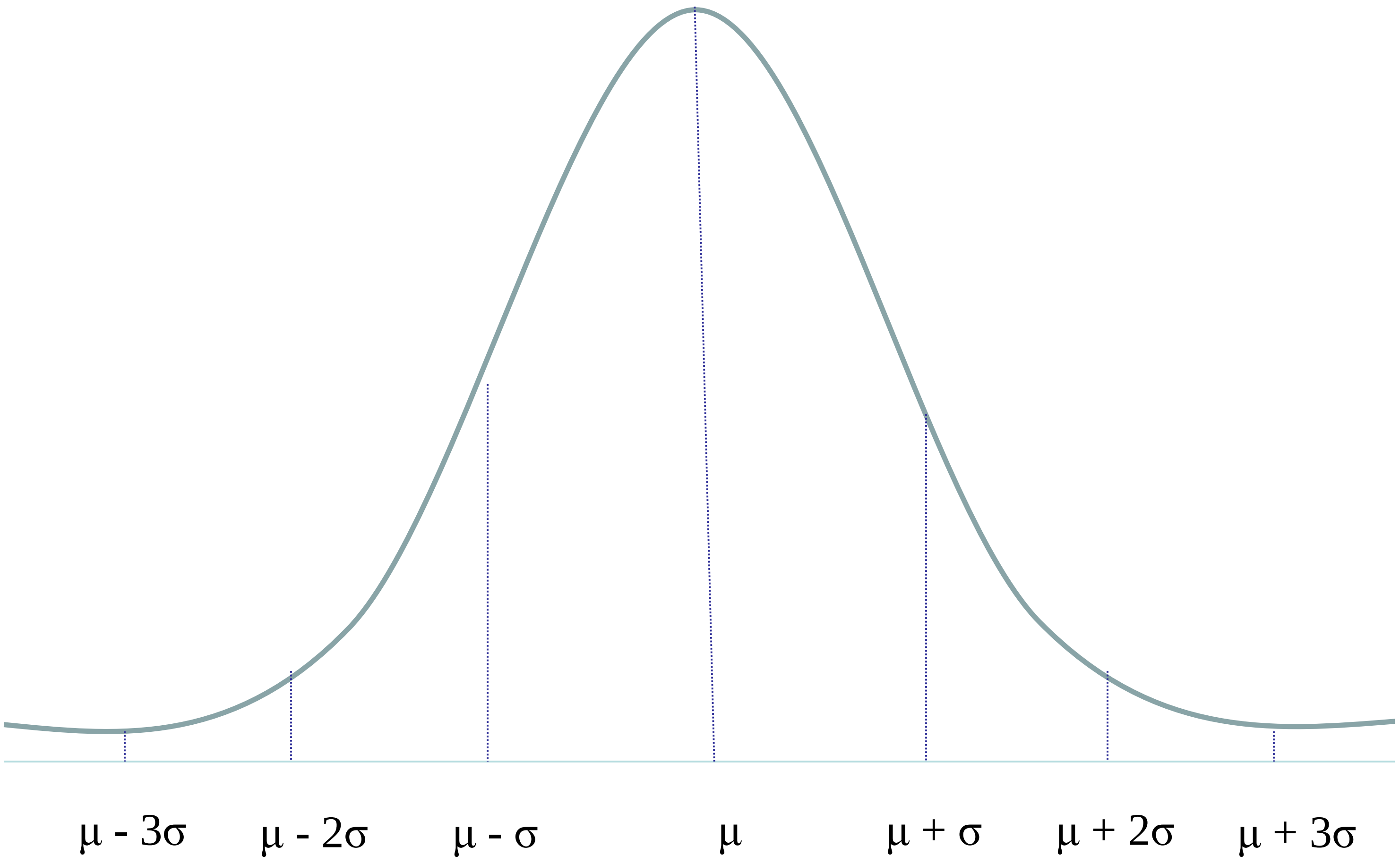
***nhà toán học Đức*

Phân phối Normal

Đặc điểm quan trọng

1. Đối xứng quanh trung bình μ
2. Trung bình = Trung vị = Mode
3. Diện tích giới hạn bởi đường cong và trục X = 1 đơn vị diện tích

Phân phối bình thường chuẩn



Phân phối Normal
Các tham số

- Bởi vì μ & σ biểu thị vị trí và độ phân tán của phân phối normal, chúng được gọi là *tham số*
- 2 tham số: μ & σ được dùng để xác định phân phối hoàn toàn

Phân phối Normal

Hàm mật độ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2}$$

$-\infty < x < \infty$

$\pi = 3.14159$

$e = 2.71828$

Phân phối bình thường chuẩn

Z-scores cho ***bất kỳ*** biến ngẫu nhiên X có phân phối normal =

Z chuẩn hóa:

$$Z = \frac{X - \mu}{\sigma}$$

Hàm mật độ xác suất bình thường chuẩn:

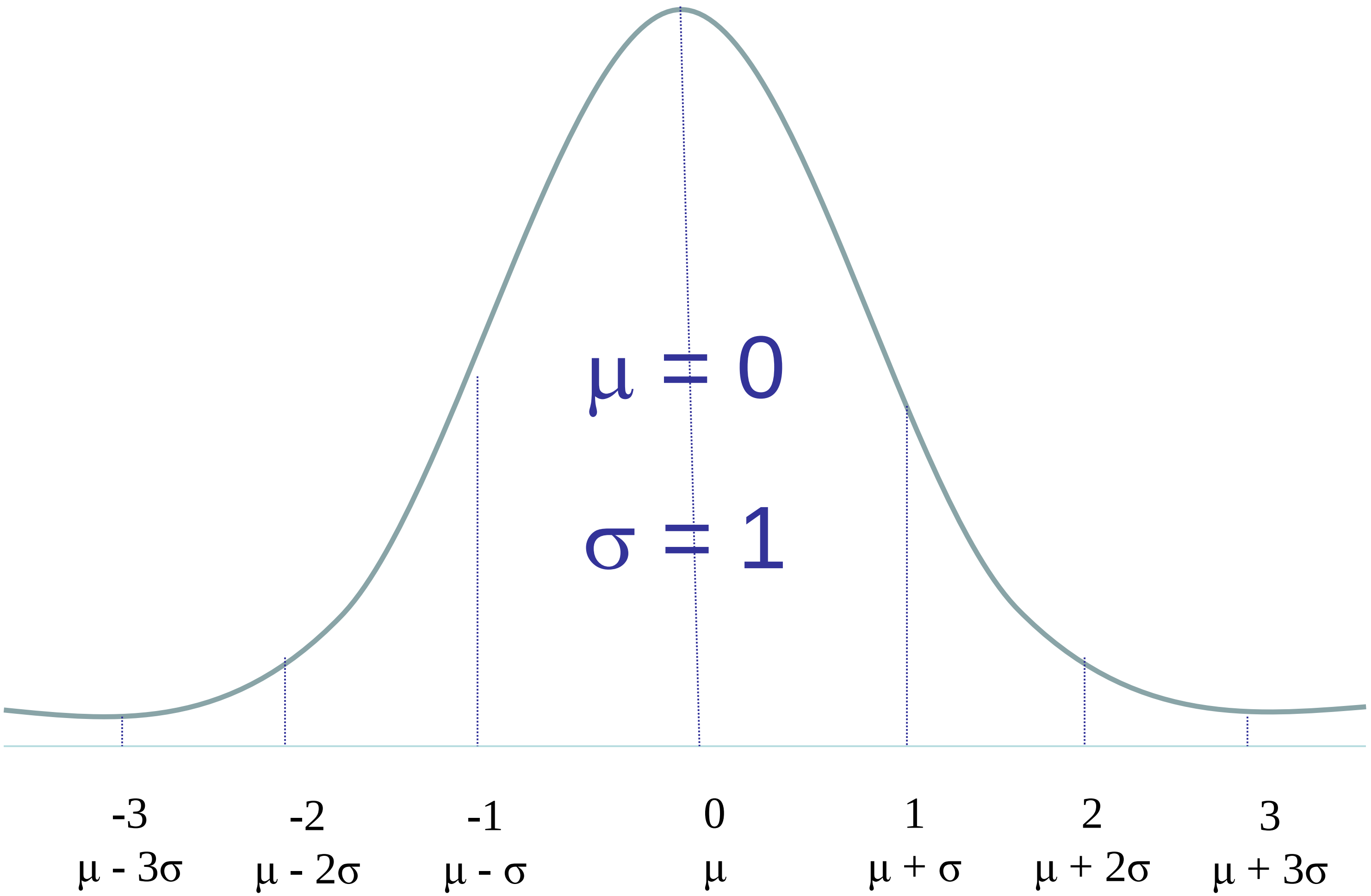
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-0.5z^2}$$

-∞ < z < ∞

π = 3.14159

e = 2.71828

Phân phối bình thường chuẩn



Diện tích dưới đường cong bình thường chuẩn

- Diện tích giữa z_0 & z_1

$$\int_{z_0}^{z_1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

PHÂN PHỐI *t*

- VẤN ĐỀ

□ σ biết & không biết μ *(!)*

□ *thực ra, thông thường cả* σ *&* μ *đều không biết*

- Không thể dùng số thống kê này $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$\rightarrow$ bởi vì không biết σ, *ngay cả khi cỡ mẫu n lớn,*

$\Rightarrow$ khi đó dùng $s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$ để thay thế σ

PHÂN PHỐI *t*

- William Sealy Gosset "*Student*" (1908)

→*Phân phối student's*

= *phân phối t*

- Trị số:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

theo phân phối này.

PHÂN PHỐI *t*
Các đặc tính

1. Trung bình là 0.

2. Hình dạng đối xứng quanh trung bình.

3. Thông thường, phương sai có giá trị lớn hơn 1, nhưng phương sai tiến gần 1 khi cỡ mẫu càng lớn.

Khi $\nu > 2$, phương sai của phân phối t là $\nu/(\nu -2)$

$\Leftrightarrow$ Khi n > 3, phương sai của phân phối t là (n-1)/(n-3)

PHÂN PHỐI *t*
Các đặc tính

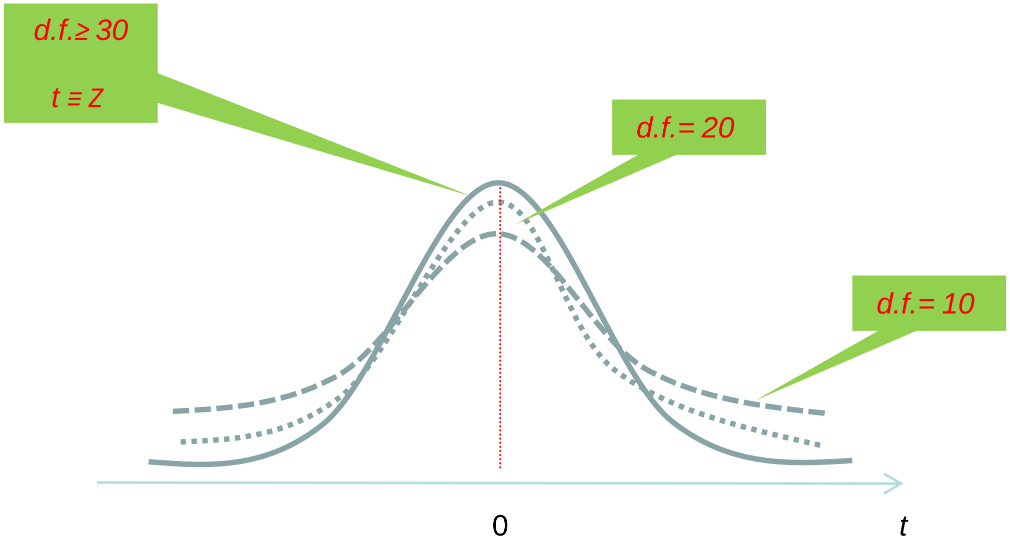
4. Giá trị t thay đổi từ -∞ đến +∞

5. Phân phối t = một họ phân phối, bởi vì có một phân phối khác nhau tương ứng với mỗi mẫu với ν = n - 1

6. So với phân phối normal, phân phối t có đỉnh thấp hơn & hai đuôi cao hơn

7. Phân phối t trở thành phân phối bình thường khi n - 1 tiến tới vô cực.

Các phân phối t
(một họ phân phối)



Các phân phối t tương ứng với độ tự do

Lưu ý

Một *điều kiện cần* để sử dụng phân phối t:

☼ mẫu phải được rút ra từ tổng thể có phân phối bình thường

☼ giả định phân phối của tổng thể *ít ra* cũng có hình mô đất