Z test & t test

Nguyen Quang Vinh - Nguyen Thi Tu Van

Classification of data types

(1) Qualitative $\rightarrow$ attribute (frequencies)

Quantitative $\rightarrow$ amount

(2) Discrete

Continuous

(3) Nominal

Ordinal

Interval

Ratio

MEASURES OF POSITION (1)

Standardizing the sample data

Sample z-score:

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

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

MEASURES OF POSITION (2)

Percentiles (p^{th}):

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

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

Where: n is the size of sample

Note: c is the position of data, not value

Quartiles (Q):

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

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

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

Interquartile range:

$$IQR = Q_3 - Q_1$$

Interpreting mean and S.D.

Empirical rule** for *normal* distribution (*unproved*):

68% of our data values lie within the interval of mean ± 1 S.D.

95% of our data values lie within the interval of mean ± 2 S.D.

99.7% or *almost all* of our data values lie within the interval of mean ± 3 S.D.

***Strictly speaking, the empirical rule applies to <u>population</u> values. However, this rule works very well for <u>large</u> <u>samples</u> having an approximate bell-shaped histogram.*

Normal distribution

- ***The most important*** distribution
- A ***key role*** in the application of ***many*** statistical techniques
- Abraham De Moivre* (1667 - 1754) on November 12th, 1733
- Carl Friedrich Gauss** (1777 - 1855) $\rightarrow$ Gaussian distribution

**French English mathematician*

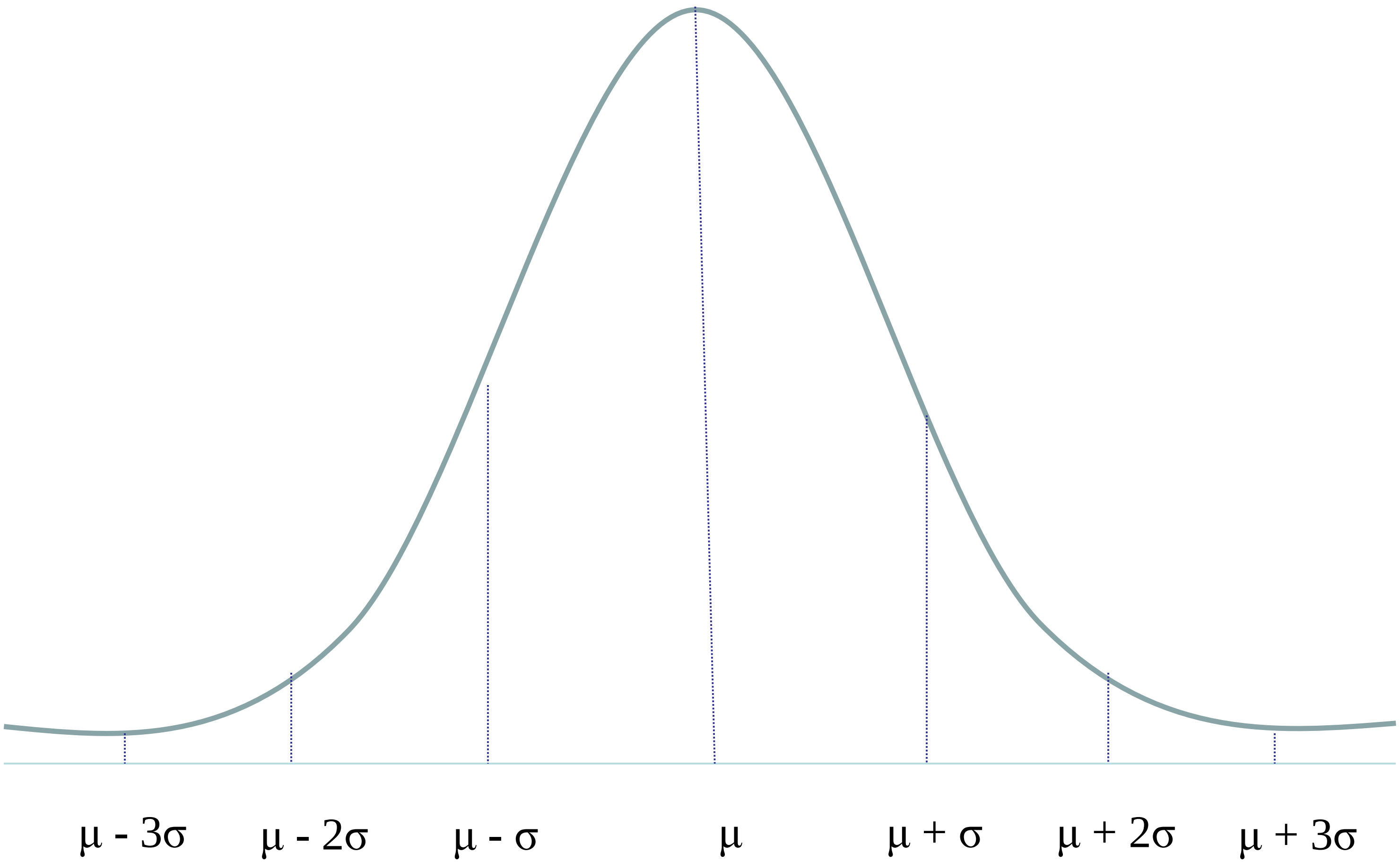
***German mathematician*

Normal distribution

Important characteristics

1. It is symmetrical about its mean μ
2. Mean = Median = Mode
3. The total area under the curve & above the X-axis = one square unit

Normal distribution



Normal distribution
Parameters

- Because μ & σ represent the location and spread of the normal distribution, they are called *parameters*
- The 2 parameters: μ & σ are used to define the distribution completely

Normal distribution

The normal density

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

$-\infty < x < \infty$

$\pi = 3.14159$

$e = 2.71828$

Standard normal distribution

Z-scores for ***any*** normal random variable X =

Z standardized values:

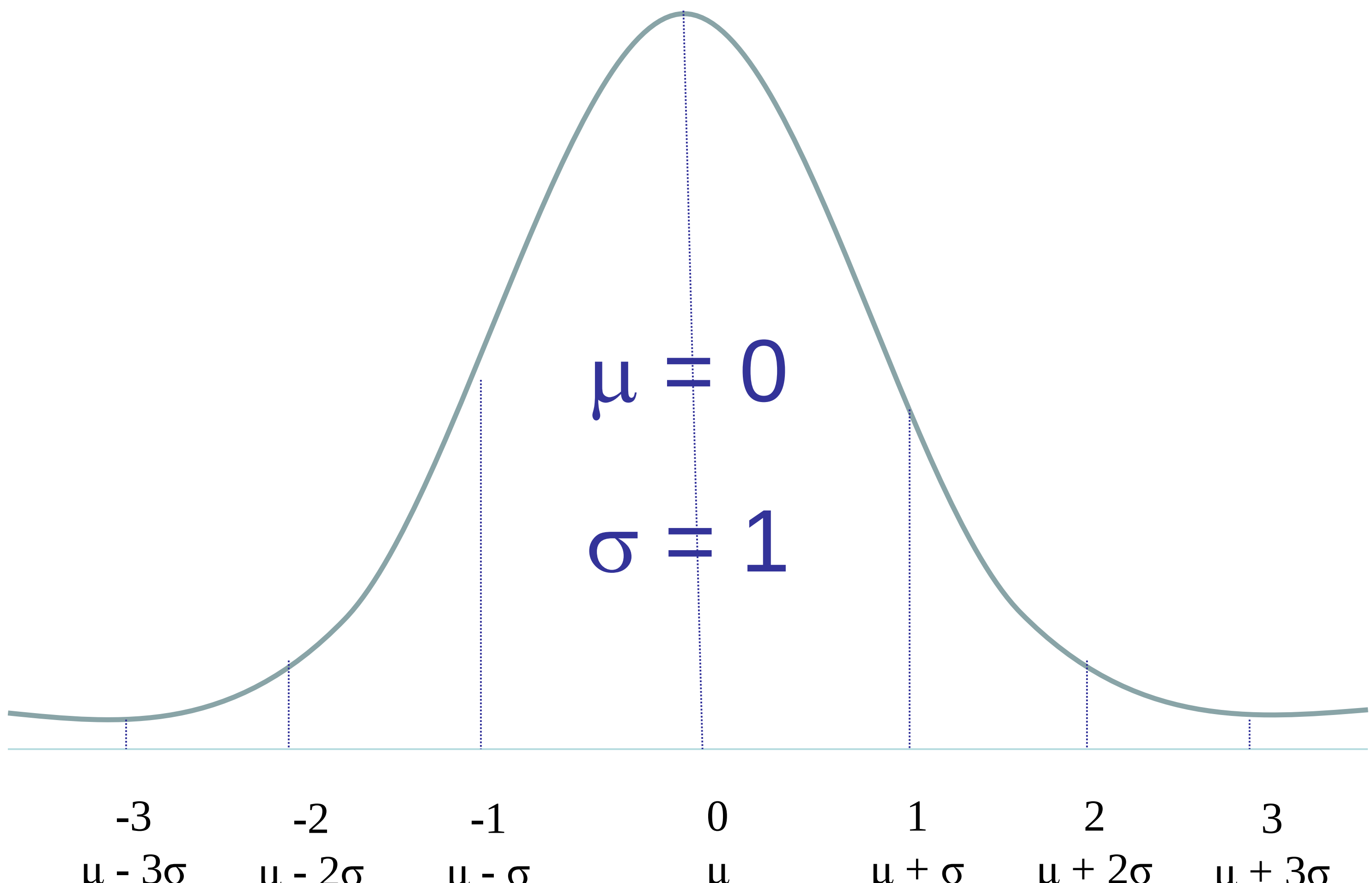
$$Z = \frac{X - \mu}{\sigma}$$

The standard normal density:

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

-∞ < z < ∞
π = 3.14159
e = 2.71828

Standard normal distribution



Areas Under the Standard Normal Curve

- The area between z_0 & z_1

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

THE *t* DISTRIBUTION

- PROBLEM

□ σ is known & not known μ *(!)*

□ *Indeed, it is the usual case,* σ *&* μ *is unknown*

- We cannot make use the statistic $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$\rightarrow$ because σ is *unknown, even when n is large,*

$\Rightarrow$ use $s = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$ to replace σ

THE *t* DISTRIBUTION

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

$\rightarrow$ *Student's t distribution*

= *t distribution*

- The quantity:

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

follows this distribution.

THE *t* DISTRIBUTION
Properties

1. It has a mean of 0.

2. It is symmetrical about the mean.

3. In general, it has a variance greater than 1, but the variance approaches 1 as the sample size becomes large.

   For $\nu > 2$, the variance of the t distribution is $\nu/(\nu - 2)$

$\Leftrightarrow$ For $n > 3$, the variance of the t distribution is $(n-1)/(n-3)$

THE *t* DISTRIBUTION Properties

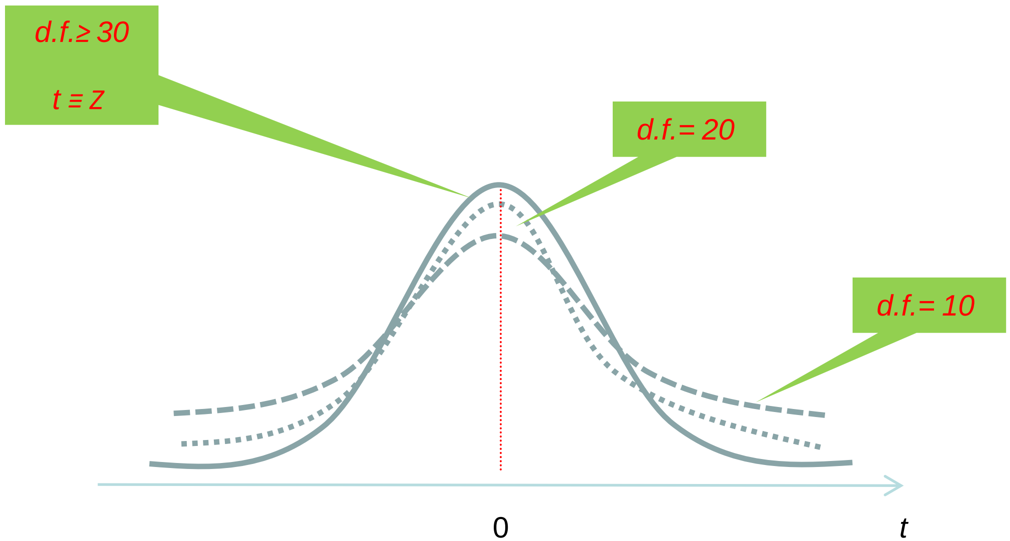
4. The variable t ranges from $-\infty$ to $+\infty$

5. The t distribution = a family of distributions, since there is a different distribution for each sample value of $\nu = n - 1$

6. Compared to the normal distribution the t distribution is less peaked in the center & has higher tails

7. The t distribution approaches the normal distribution as n - 1 approaches infinity.

The t distributions
(a family of distributions)



t distributions with degrees of freedom

Notice

A *requirement* for valid use of the t distribution:

- ☼ sample must be drawn from a normal distribution

- ☼ an assumption of, *at least*, a mound-shaped population distribution be tenable