The Relevance of t-Statistics for Small Sample Sizes

An Introductory Class to Higher Statistics

Xinyu Chen

December 6, 2024

Outline

Answering a lot questions, e.g.,

❶ How was t-statistic developed?

❷ Standard normal distribution vs. student t-distribution?

❸ What is t-statistic?

❹ What types of t-tests there are?

❺ What the hypotheses and the assumptions are?

❻ How a t-test is calculated?

❼ How you interpret results?

❽ Huge real-world applications

Development

The problems of small samples

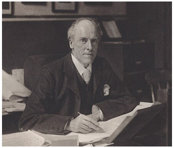
1876 — *t-distribution as a posterior distribution*

1895 — *t-distribution as Pearson type IV distribution*

1908 — *Student t-distribution* "The Probable Error of a Mean" (Biometrika)

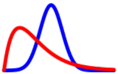
F. R. Helmert

J. Lüroth



K. Pearson

W. S. Gosset

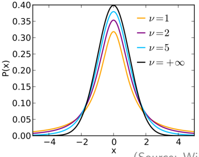


(Source: Wiki)

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

(Source: Wiki)

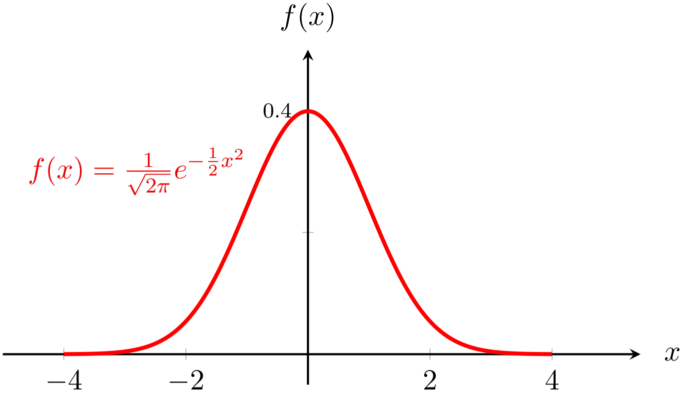
Diagram of the Pearson system



(Source: Wiki)

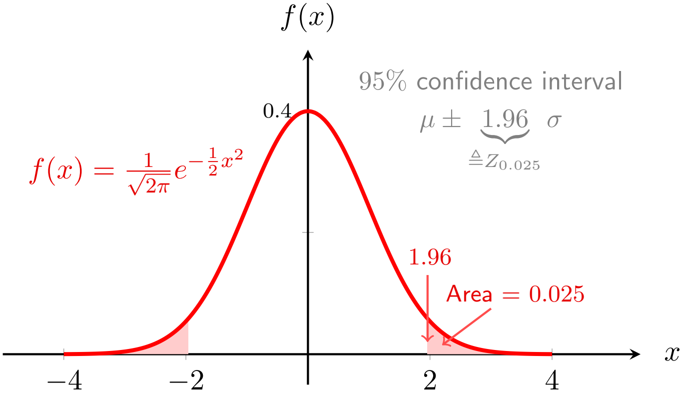
Probability density function

Revisiting Normal Distribution



Probability density function of standard normal distribution

Revisiting Normal Distribution



Probability density function of standard normal distribution

Z-Test

“Statistical tests are used to test the fit between a hypothesis and the data.”
Example

Student t-Distribution

- Probability density function (w/ random variable x):

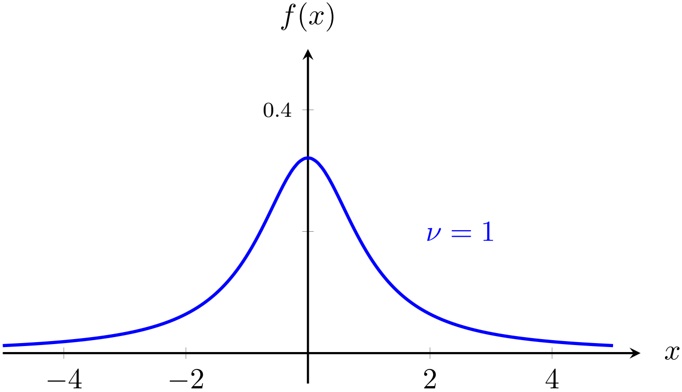
$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\big(1 + \frac{x^2}{\nu}\big)^{-\frac{\nu+1}{2}}$$

  - $\nu \in \mathbb{Z}^+$: Degrees of freedom
  - $\Gamma(\cdot)$: The Gamma function

- Example of $\nu = 5$ degrees of freedom:

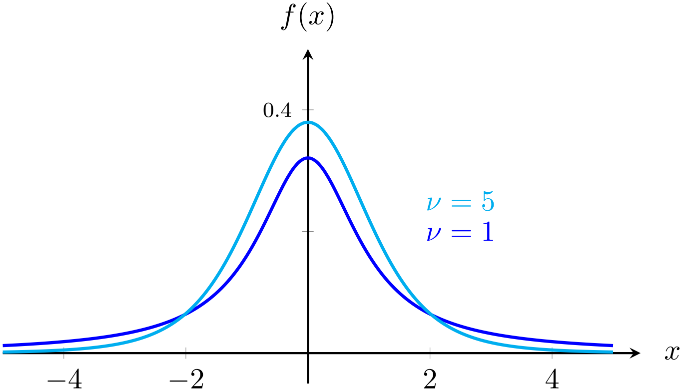
$$f(x) = \frac{\Gamma(3)}{\sqrt{5\pi}\,\Gamma(2.5)}\big(1 + \frac{x^2}{5}\big)^{-3}$$

Student t-Distribution



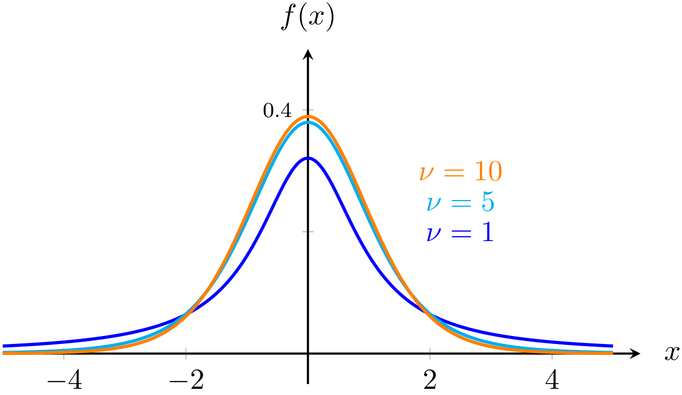
Student t-distribution

Student t-Distribution



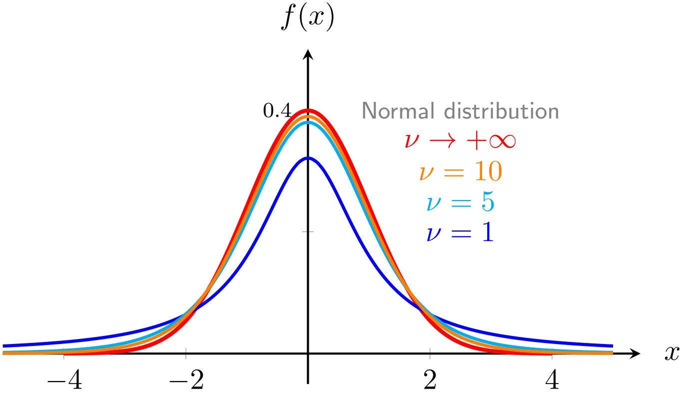
Student t-distribution

Student t-Distribution



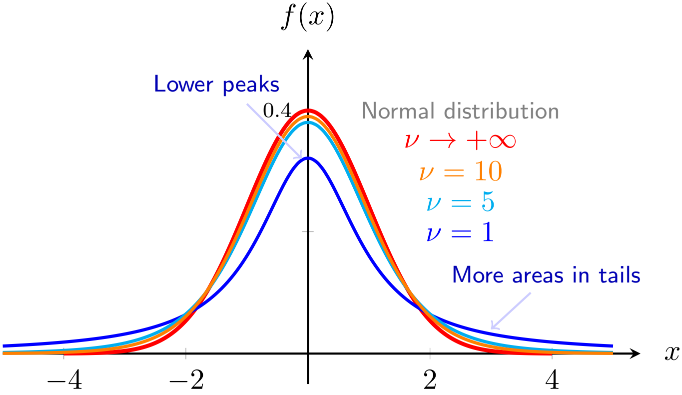
Student t-distribution

Student t-Distribution



Student t-distribution

Student t-Distribution



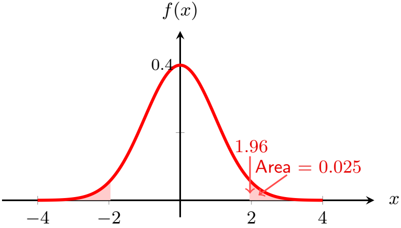
Student t-distribution

95% Confidence Interval

For the population mean μ (given) and standard deviation σ (given or not?)

- If σ is known

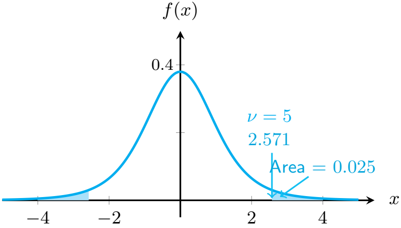
$$\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$



Standard normal distribution

- If σ is not known

$$\bar{x} \pm ? \times \frac{s}{\sqrt{n}}$$



Student t-distribution

Development

- The t-statistic depends on the type of test, but for a one-sample t-test:

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

  - $\bar{x}$ sample mean
  - μ population mean
  - s sample standard deviation
  - n sample size

- The t-statistic quantifies the difference relative to variability in the data.
- (Interpretation) A high absolute value of t (larger than the critical value from the t-table) suggests a statistically significant difference.

t-Table

Teaching Concept

Method

use math
use figures
use examples
use data
use codes
use latex to create all examples

Thanks for your attention!

Any Questions?

About me:

- Homepage: https://xinychen.github.io
- How to reach me: chenxy346@gmail.com