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 Normal distribution vs. student t-distribution?

❸ What is t-statistic?

❹ How to calcuate a t-test?

❺ What are the hypotheses and the assumptions?

❻ How to interpret results?

Huge real-world applications

Development

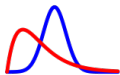The problems of **small sample sizes**

1876	1895	1908

t-distribution as
a posterior distribution

t-distribution as Pearson
type IV distribution

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

F. R. Helmert   **J. Lüroth**

K. Pearson

W. S. Gosset

(Source: Wiki)

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

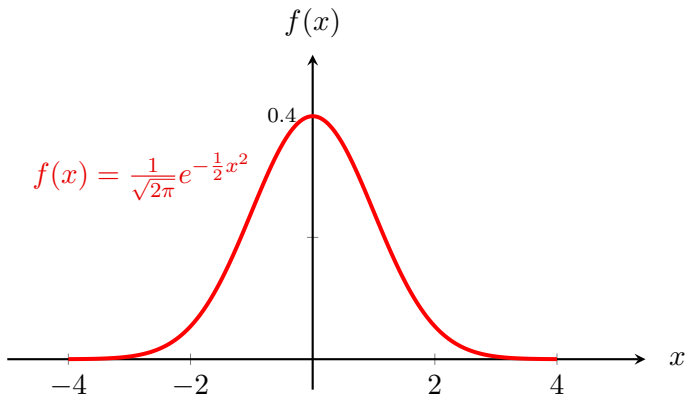(Source: Wiki)

Diagram of the Pearson system
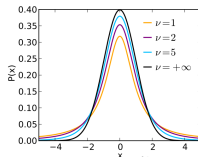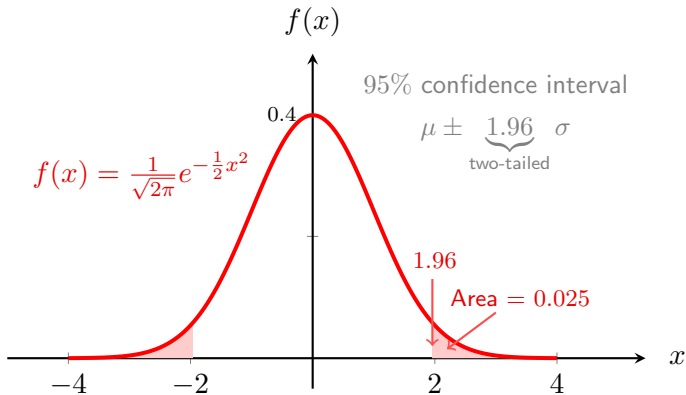
(Source: Wiki)

Probability density function

Revisiting Normal Distribution

Probability density function of the standard normal distribution

Revisiting Normal Distribution

Probability density function of the standard normal distribution

$$f(x)$$

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

95% confidence interval

$$\mu \pm \underbrace{1.96}_{\text{two-tailed}} \sigma$$

1.96

Area $= 0.025$

Implementing z-Test

Problem Statement

A company claims that the average daily energy consumption of households is 30 kWh with a population standard deviation of 5 kWh. A random sample of 40 households has an average daily energy consumption of 32 kWh. Conduct a two-tailed hypothesis test at a 95% confidence interval to determine if the sample provides sufficient evidence to reject the company's claim.

Implementing z-Test

Problem Statement

A company claims that the average daily energy consumption of households is 30 kWh
with a population standard deviation of 5 kWh. A random sample of 40 households
has an average daily energy consumption of 32 kWh. Conduct a two-tailed hypothesis
test at a 95% confidence interval to determine if the sample provides sufficient
evidence to reject the company's claim.

Steps:

❶ Formulate Hypotheses
  - Null Hypothesis (H_0): The population mean is $\mu = 30$ kWh.
  - Alternative Hypothesis (H_a): The population mean is not $\mu = 30$ kWh
    ($\mu \neq 30$).

❷ Use the z-test formula since the population standard deviation is known:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{32 - 30}{5/\sqrt{40}} = \frac{2}{5/6.32} = \frac{2}{0.79} \approx 2.53$$

  - $\bar{x} = 32$ (sample mean)   - $\mu = 30$ (population mean)

  - $n = 40$ (sample size)   - $\sigma = 5$ (population standard deviation)

Implementing z-Test

Problem Statement

A company claims that the average daily energy consumption of households is 30 kWh with a population standard deviation of 5 kWh. A random sample of 40 households has an average daily energy consumption of 32 kWh. Conduct a two-tailed hypothesis test at a 95% confidence interval to determine if the sample provides sufficient evidence to reject the company's claim.

Steps:

❷ Use the z-test formula since the population standard deviation is known:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{32 - 30}{5/\sqrt{40}} = \frac{2}{5/6.32} = \frac{2}{0.79} \approx 2.53$$

❸ Decision rule at a 95% confidence interval
  ○ Reject H_0 if $|z| > 1.96$.
 ○ Otherwise, fail to reject H_0.

❹ Interpretation
  ○ The test statistic $|z| = 2.53 > 1.96$ (exceeding the critical value).
  ○ Thus, we reject the null hypothesis.
  ○ The sample provides sufficient evidence to conclude that the average daily energy consumption is not 30 kWh.

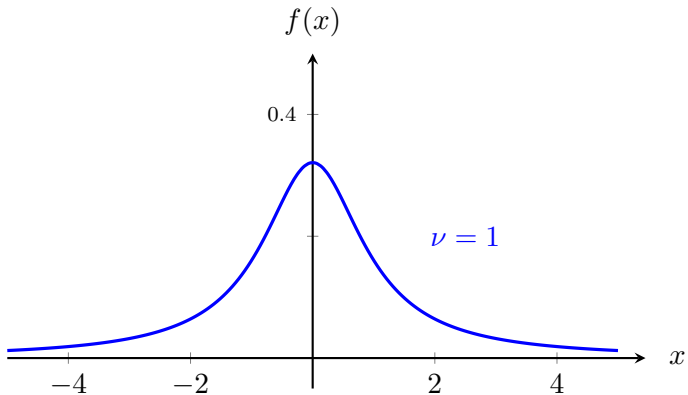# Student t-Distribution

- Probability density function:

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

  - $x \in \mathbb{R}$: The random variable
  - $\nu \in \mathbb{Z}^+$: Degrees of freedom
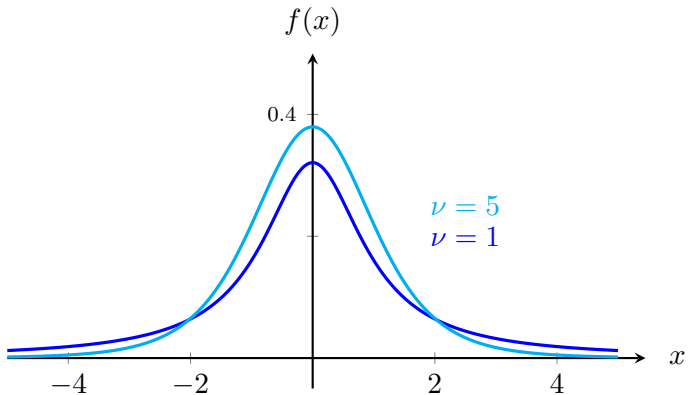 - $\Gamma(\cdot)$: The Gamma function
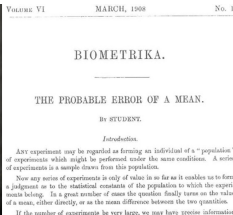
Gossset'1908 (known as "student")
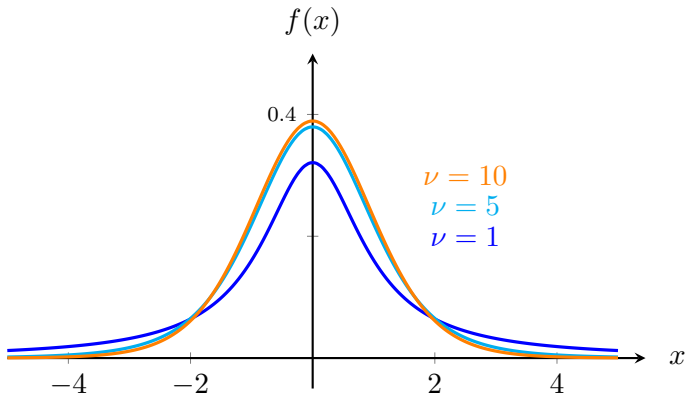
Student t-Distribution

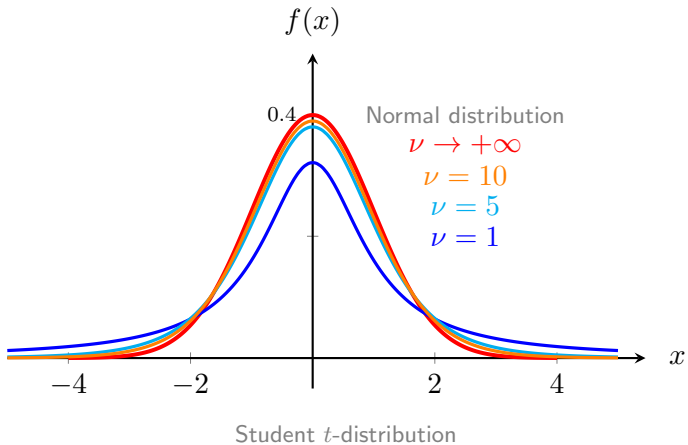Student t-distribution

Student t-Distribution

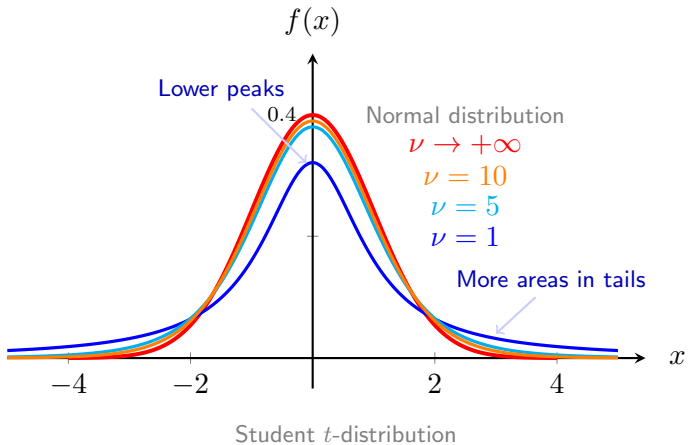Student t-distribution

Student t-Distribution

Student t-distribution

Student t-Distribution

Student t-distribution

Student t-Distribution

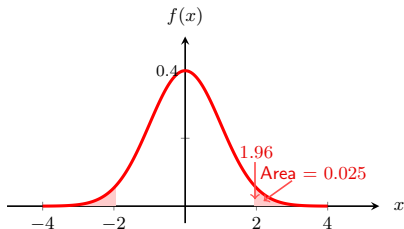Student t-distribution

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

- If population standard
  deviation σ is known

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

Standard normal distribution

95% **Confidence Interval**

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

- If population standard deviation σ is known

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

Standard normal distribution

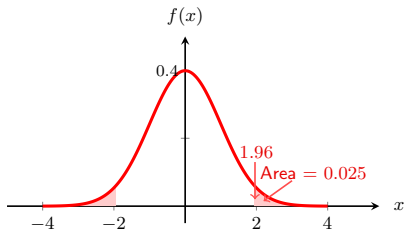- If σ is unknown, using sample standard deviation s instead

$$\bar{x} \pm \ ? \times \frac{s}{\sqrt{n}}$$
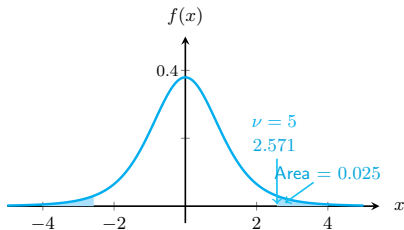
Student t-distribution

- **Heavier tail** in student t-distribution ($\nu = n - 1$ degrees of freedom) is important for small sample size n

- Formula of t-statistic:

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

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

- 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.

- The problem of small sample size!

Implementing t-Test

Problem Statement

A company claims that the average daily energy consumption of households is 30 kWh. A random sample of 6 **households** has an average daily energy consumption of 32 kWh, with a sample standard deviation of 6 kWh. Conduct a two-tailed hypothesis test at a 95% confidence interval to determine if the sample provides sufficient evidence to reject the company's claim.

Implementing t-Test

Problem Statement

A company claims that the average daily energy consumption of households is 30 kWh. A random sample of 6 **households** has an average daily energy consumption of 32 kWh, with a sample standard deviation of 6 kWh. Conduct a two-tailed hypothesis test at a 95% confidence interval to determine if the sample provides sufficient evidence to reject the company's claim.

Steps:

❶ Formulate Hypotheses
  ○ Null Hypothesis (H_0): The population mean is $\mu = 30$ kWh.
  ○ Alternative Hypothesis (H_a): The population mean is not $\mu = 30$ kWh ($\mu \neq 30$).

❷ Use the t-test formula since the population standard deviation is not known:
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{32 - 30}{6/\sqrt{6}} = \frac{2}{6/2.449} \approx 0.816$$

  ○ $\bar{x} = 32$ (sample mean)   ○ $s = 6$ (sample standard deviation)

  ○ $n = 6$ (sample size)   ○ $\sigma = 30$ (population mean)

Small sample sizes

- Degrees of freedom for a t-test:

$$\nu = \underbrace{n}_{\text{sample size}} -1 = 6 - 1 = 5$$

- t-distributions with ν degrees of freedom at a 95% confidence interval (two-tailed)

$\nu = 1$	$\nu = 5$	$\nu = 10$	$\nu \to +\infty$
12.706	2.571	2.228	1.960

- The critical t-value

$$t_{\nu,(1-0.95)/2} = t_{5,0.025} = 2.571$$

Implementing t-Test

Problem Statement

A company claims that the average daily energy consumption of households is 30 kWh. A random sample of 6 **households** has an average daily energy consumption of 32 kWh, with a sample standard deviation of 6 kWh. Conduct a two-tailed hypothesis test at a 95% confidence interval to determine if the sample provides sufficient evidence to reject the company's claim.

Steps:

❷ Use the t-test formula:
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{32 - 30}{6/\sqrt{6}} = \frac{2}{6/2.449} \approx 0.816$$

❸ Decision rule at a 95% confidence interval
  ○ Reject H_0 if $|t| > 2.571$.
  ○ Otherwise, fail to reject H_0.

❹ Interpretation
  ○ The test statistic $|t| = 0.816 < 2.571$.
  ○ Thus, we fail to reject the null hypothesis.
  ○ There is not enough evidence to conclude that the average daily energy consumption differs from the company's claim of 30 kWh.

Summary

To be determined...

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