STAT 3011 Discussion 015

Week 10

Talha Hamza
University of Minnesota
College of Science and Engineering

Fall 2025

Two Types of Confidence Intervals

We will learn two main types of confidence intervals:

1. Population Proportion (p)

Used when dealing with categorical data (yes/no, success/failure)

2. Population Mean (μ)

Used when dealing with quantitative data (heights, weights, scores)

Both follow the general form:

$$\text{Point Estimate} \pm \text{Margin of Error}$$

Confidence Interval for Population Proportion (p)

When to use: Estimating proportion of successes in population
Formula:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Steps:

1. Find $\alpha = 1 -$ confidence level (e.g., for 95% CI: $\alpha = 0.05$)
2. Calculate sample proportion: $\hat{p} = \frac{\text{number of successes}}{n}$
3. Check conditions: $n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$
4. Find $z_{\alpha/2}$ using `qnorm(1-`α`/2)`
5. Calculate standard error: $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
6. Calculate margin of error: $moe = z_{\alpha/2} \times SE$
7. Construct CI: $\hat{p} \pm moe$

Confidence Interval for Population Mean (μ)

When to use: Estimating average value in population

Formula:

$$\overline{x} \pm t_{\alpha/2,df} \frac{s}{\sqrt{n}}$$

Steps:

1. Find $\alpha = 1 -$ confidence level (e.g., for 95% CI: $\alpha = 0.05$)
2. Calculate sample mean: $\overline{x} = \frac{\sum x_i}{n}$
3. Calculate sample standard deviation: $s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$
4. Find degrees of freedom: $df = n - 1$
5. Find $t_{\alpha/2,df}$ using `qt(1-`α`/2, df=`n`-1)`
6. Calculate standard error: $SE = \frac{s}{\sqrt{n}}$
7. Calculate margin of error: $moe = t_{\alpha/2,df} \times SE$
8. Construct CI: $\overline{x} \pm moe$

Proportion vs Mean

	Proportion	**Mean**
Parameter	p	μ
Point Estimate	$\hat{p}$	$\overline{x}$
Distribution	Normal	t-distribution
Multiplier	$z_{\alpha/2}$	$t_{\alpha/2,df}$
Standard Error	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\frac{s}{\sqrt{n}}$

Important Note: The t-distribution is robust to non-normality, in the absence of outliers.

Normal vs Student's t-Distribution

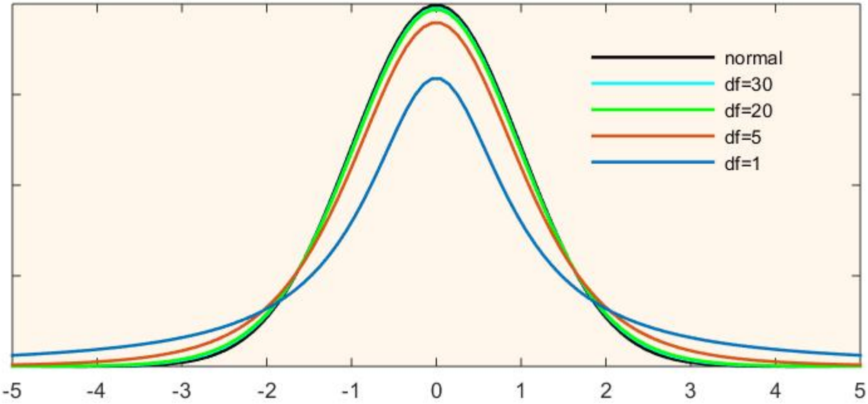

Image credit: jmp.com

R commands:

To Find Multipliers:

- $z_{\alpha/2}$: `qnorm(1-`α`/2)` or `qnorm(`α`/2, lower = FALSE)`
- $t_{\alpha/2,df}$: `qt(1-`α`/2, df=`n`-1)` or `qt(`α`/2, df=`n`-1, lower = FALSE)`

To Calculate Entire CI for Proportion:

- `prop.test(x=58, n=389, conf.level=0.95, alternative="two.sided", correct=FALSE)`

To Calculate Entire CI for Mean:

- `t.test(x = data, conf.level=0.95, alternative="two.sided", correct=FALSE)`

Interpreting 95% Confidence Intervals

Three Valid Interpretations

1. **Long-run frequency**:
   "If we repeated this procedure many times, 95% of calculated CIs would contain the true parameter."
2. **Single sample probability**:
   "There's a 95% probability that this procedure produces an interval containing the true parameter."
3. **Statistical significance**:
   "Values within the 95% CI are not significantly different from our estimate at α=0.05."

Note: The probability statement is about the **method***, not any particular interval.*

What Confidence Intervals Don't Mean

Frequent Misconceptions

- **Not**: "There's a 95% chance the true value is in this interval"
  (The interval either contains it or doesn't - no probability after calculation)
- **Not**: "95% of our data points are in this interval"
  (It's about the parameter estimate, not individual observations)
- **Not**: "If we repeat the study, there's 95% chance new estimate will be in this interval"
  (New intervals will center on new estimates)

Writing Hypotheses: H_0 and H_a

Null Hypothesis (H_0)

- Status quo claim
- Always contains "="
- Example: H_0: $\mu = \mu_0$

Alternative (H_a)

The claim we hope to find evidence for or against.

Three possible forms:

- $\mu < \mu_0$ (left-tailed)
- $\mu > \mu_0$ (right-tailed)
- $\mu \neq \mu_0$ (two-tailed)