PLS 120: Applied Statistics in Agricultural Sciences

Confidence Intervals and T-Tests

Week 6 Tutorial Guide

Mohammadreza Narimani
Department of Biological and Agricultural Engineering
University of California, Davis

mnarimani@ucdavis.edu

October 2025

Contents

Important Links

Essential Course Resources

Course Website

All course materials available at:

Course Website Link

Interactive Binder Environment

Access Week 6 lab materials:

Week 6 Binder Link

Welcome to Week 6: Confidence Intervals and T-Tests

This week, we dive into **t-distribution and hypothesis testing** - essential tools for agricultural research when population parameters are unknown. You'll learn when to use t-tests instead of z-tests, construct t-based confidence intervals, and analyze treatment effectiveness!

T-Distribution vs Z-Distribution

When to Use T-Distribution

The t-distribution is used when the population standard deviation is unknown and must be estimated from sample data.

3.1.1 Key Differences

Use T-Distribution When:
- Population standard deviation (σ) is unknown
- Using sample standard deviation (s) as estimate
- Small to moderate sample sizes (especially n $<$ 30)
- More conservative than z-distribution

Use Z-Distribution When:
- Population standard deviation (σ) is known
- Large sample sizes (n $\geq$ 30) with CLT
- Population is normally distributed

Key Properties:
- T-distribution has heavier tails than normal
- As df increases, t approaches standard normal
- More uncertainty = wider confidence intervals

Degrees of Freedom

Understanding Degrees of Freedom

Degrees of freedom (df) represent the number of independent pieces of information available to estimate a parameter.

4.1.1 Degrees of Freedom Formula

One-Sample T-Test:

$$df = n - 1$$

Where:
`n` = sample size
`df` = degrees of freedom

Why n-1?

- We lose one degree of freedom when estimating the mean
- With n observations and known mean, only (n-1) are free to vary
- Smaller df = wider t-distribution = more conservative results

R Implementation:

```
df <- length(data) - 1
t_critical <- qt(0.975, df) # For 95% CI
```

T-Based Confidence Intervals

Confidence Interval Formula

T-based confidence intervals account for the additional uncertainty when estimating population standard deviation.

5.1.1 T-Confidence Interval Formula

T-Based Confidence Interval:

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

Where:
$\overline{x}$ = sample mean
$t_{\alpha/2,df}$ = critical t-value
s = sample standard deviation
n = sample size
df = n - 1

R Implementation:

```
sample_mean <- mean(data)
sample_sd <- sd(data)
n <- length(data)
df <- n - 1
t_critical <- qt(0.975, df) # 95% CI
margin_error <- t_critical * (sample_sd / sqrt(n))
ci_lower <- sample_mean - margin_error
ci_upper <- sample_mean + margin_error
```

Comparison: T vs Z Intervals

5.2.1 Width Comparison

Interval Width Comparison:

- T-based intervals are WIDER than z-based intervals
- Difference is larger for smaller sample sizes
- As n increases, t-intervals approach z-intervals

Example (n=10, 95% confidence):
Z-critical value: 1.96
T-critical value: 2.262
T-interval is about 15% wider

Example (n=30, 95% confidence):
Z-critical value: 1.96
T-critical value: 2.045
T-interval is about 4% wider

One-Sample T-Test

Hypothesis Testing with T-Tests

T-tests allow us to test hypotheses about population means when the population standard deviation is unknown.

6.1.1 One-Sample T-Test Formula

T-Test Statistic:

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Where:
$\overline{x}$ = sample mean
μ_0 = hypothesized population mean
`s` = sample standard deviation
`n` = sample size

Decision Rule:
If $|t| > t_{\alpha/2,df}$, reject H_0
If p-value $< \alpha$, reject H_0

R Implementation:
```
t.test(data, mu = hypothesized_mean)
t.test(data, mu = 60, conf.level = 0.95)
```

Interpreting T-Test Results

6.2.1 Understanding Output

T-Test Output Components:
- **t-statistic:** How many standard errors the sample mean is from μ_0
- **degrees of freedom:** n - 1
- **p-value:** Probability of observing this result if H_0 is true
- **confidence interval:** Range of plausible values for μ

Agricultural Example:
Testing if new fertilizer increases yield above 60 bu/acre
$H_0 : \mu = 60$ vs $H_a : \mu > 60$
If t = 2.34, df = 24, p = 0.028
Conclusion: Reject H_0, fertilizer significantly increases yield

Assignment 6 Overview

Assignment Structure (20 points total)

1. **Part 1: Overall Wheat Yield Analysis (9 points)**
   - Load and explore wheat yield dataset (1 point)
   - Calculate sample size and basic statistics (1 point)
   - Compute mean and standard deviation (2 points)
   - Calculate z-score for specific value (1 point)
   - Construct 95% confidence interval (3 points)
   - Interpret results in agricultural context (1 point)
2. **Part 2: Treatment Comparison (11 points)**

- Separate control and fertilizer groups (1 point)
- Calculate treatment statistics (2 points)
- Compute standard errors (2 points)
- Calculate margins of error (2 points)
- Construct confidence intervals for each treatment (3 points)
- Compare treatments and interpret results (1 point)

Agricultural Applications

Real-World T-Test Applications:

- **Fertilizer Effectiveness** - Test if new fertilizer significantly increases crop yield
- **Variety Trials** - Compare new crop varieties against established standards
- **Treatment Efficacy** - Evaluate pesticide or herbicide effectiveness
- **Quality Control** - Test if product meets quality standards
- **Environmental Impact** - Assess effects of farming practices on soil health
- **Breeding Programs** - Compare performance of new genetic lines
- **Irrigation Studies** - Test optimal water application rates
- **Harvest Timing** - Determine optimal harvest dates for maximum yield

Key Concepts Summary

T-Test Fundamentals

When to Use T-Tests:
- Population σ is unknown (most real situations)
- Sample size is small to moderate
- Data is approximately normally distributed
- Testing hypotheses about means

T-Test Assumptions:
- Random sampling from population
- Observations are independent
- Data is approximately normally distributed
- For small samples, normality is more critical

Confidence Interval Interpretation:
- 95% CI: We're 95% confident the true mean lies in this range
- Wider intervals = more uncertainty
- T-intervals are more conservative than z-intervals

Statistical Decision Making

Hypothesis Testing Steps:
1. State null and alternative hypotheses
2. Choose significance level (α)
3. Calculate test statistic
4. Find p-value or critical value
5. Make decision and interpret in context

Type I and Type II Errors:
Type I Error: Rejecting true H_0 (false positive)
Type II Error: Failing to reject false H_0 (false negative)

Practical vs Statistical Significance:
- Statistical significance: p $< \alpha$
- Practical significance: Effect size matters in real world
- Large samples can detect tiny, unimportant differences

Data Analysis Workflow

Step-by-Step Analysis

1. **Data Exploration**
   - Load data and examine structure
   - Check for missing values and outliers
   - Create summary statistics and visualizations
2. **Assumption Checking**
   - Assess normality (histograms, Q-Q plots)
   - Check for independence
   - Identify potential issues
3. **Statistical Analysis**
   - Calculate appropriate test statistics
   - Construct confidence intervals
   - Perform hypothesis tests
4. **Interpretation**
   - Interpret results in agricultural context
   - Consider practical significance
   - Make recommendations based on findings

Getting Started

1. Launch Week 6 Binder environment
2. Navigate to `class_activity` folder

3. Open `Week6_Confidence_Intervals.ipynb`
4. Work through interactive exercises
5. Complete Assignment 6 in `assignment` folder

Learning Objectives

By the end of this week, you will be able to:

- Distinguish between t-distribution and z-distribution applications
- Calculate degrees of freedom for one-sample t-tests
- Construct t-based confidence intervals
- Perform one-sample t-tests for hypothesis testing
- Compare treatment groups using confidence intervals
- Interpret t-test results in agricultural contexts
- Understand the relationship between sample size and interval width
- Make data-driven decisions about treatment effectiveness

Tips for Success

Best Practices:

- Always check if you should use t or z distribution
- Remember: df = n - 1 for one-sample t-tests
- T-intervals are wider than z-intervals (more conservative)
- Use `t.test()` function in R for complete analysis
- Check normality assumptions, especially for small samples
- Interpret confidence intervals correctly (about the parameter, not individual observations)
- Consider both statistical and practical significance
- Always interpret results in the context of the agricultural problem

Common Mistakes to Avoid

Avoid These Errors:

- Using z-distribution when σ is unknown
- Forgetting to subtract 1 for degrees of freedom
- Misinterpreting confidence intervals
- Ignoring assumptions (normality, independence)
- Confusing statistical significance with practical importance
- Not considering the agricultural context in interpretations
- Using wrong tail for one-sided vs two-sided tests

Need Help?

Mohammadreza Narimani
Email: mnarimani@ucdavis.edu
Department of Biological and Agricultural Engineering, UC Davis
Office Hours: Thursdays 10 AM - 12 PM (Zoom)
Zoom Link: Join Office Hours