PLS 120: Applied Statistics in Agricultural Sciences

Functions and T-Tests

Week 7 Tutorial Guide

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

mnarimani@ucdavis.edu

November 2024

Contents

Important Links

Essential Course Resources

Course Website

All course materials available at:

Course Website Link

Interactive Binder Environment

Access Week 7 lab materials:

Week 7 Binder Link

Welcome to Week 7: Functions and T-Tests

This week, we explore **custom functions in R and two-sample t-tests** - essential tools for creating reusable code and comparing groups in agricultural research. You'll learn to write functions, perform statistical comparisons, and interpret results for evidence-based farming decisions!

Writing Functions in R

Why Use Functions?

Functions in R are powerful tools that simplify repetitive tasks and improve code efficiency. They make your analysis more organized, reduce errors, and allow for easy reuse across projects.

3.1.1 Function Syntax

Basic Function Structure:

```
myFunction <- function(arg1, arg2, ...) {
  # Code to execute
  result <- some_calculation
  return(result)
}
```

Components:
- `myFunction` = Function name
- `arg1, arg2` = Parameters (inputs)
- `return()` = Specifies output

Example - Square Function:

```
square <- function(x) {
  result <- x^2
  return(result)
}

# Usage
square(5)  # Returns 25
```

Sample Size Function

A practical example of function creation for statistical calculations.

3.2.1 Sample Size Calculation Function

Sample Size Formula:

$$n = \frac{z^2 \cdot p \cdot (1 - p)}{E^2}$$

Where:
`z` = critical value from normal distribution
`p` = expected prevalence
`E` = desired margin of error

R Function Implementation:

```
calculate_sample_size <- function(prev, alpha, margin_error) {
  z <- qnorm(1 - alpha / 2)
  n <- z^2 * prev * (1 - prev) / (margin_error^2)
  required_sample_size <- ceiling(n)
  return(required_sample_size)
}

# Usage
calculate_sample_size(0.1, 0.05, 0.05)  # Returns 139
```

Two-Sample T-Tests

When to Use Two-Sample T-Tests

Two-sample t-tests compare means between two independent groups, such as treatment vs control in agricultural experiments.

4.1.1 Welch's T-Test Formula

Two-Sample T-Test (Unequal Variances):

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:
$\overline{x_1}, \overline{x_2}$ = sample means
s_1, s_2 = sample standard deviations
n_1, n_2 = sample sizes

Degrees of Freedom (Welch's approximation):

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

R Implementation:

```
t.test(group1, group2, var.equal = FALSE)
```

Hypothesis Formation

Proper hypothesis setup is crucial for meaningful statistical testing.

4.2.1 Hypothesis Types

Two-Sided Test:
$H_0 : \mu_1 = \mu_2$ (means are equal)
$H_1 : \mu_1 \neq \mu_2$ (means are different)

One-Sided Tests:
Greater than: $H_1 : \mu_1 > \mu_2$
Less than: $H_1 : \mu_1 < \mu_2$

Agricultural Examples:

- Fertilizer vs Control: Does fertilizer increase yield?
- Variety A vs B: Which variety performs better?
- Irrigation methods: Is drip more efficient than flood?

R Implementation:

```
t.test(treatment, control, alternative = "two.sided")
t.test(treatment, control, alternative = "greater")
t.test(treatment, control, alternative = "less")
```

Interpreting T-Test Results

Understanding T-Test Output

T-test results provide multiple pieces of information for decision-making.

5.1.1 Key Output Components

T-Test Output Elements:

- **t-statistic:** Magnitude and direction of difference
- **degrees of freedom:** Based on sample sizes and variances
- **p-value:** Probability of observing result if H_0 true
- **confidence interval:** Range for true mean difference
- **sample estimates:** Group means

Decision Rules:
If p-value $< \alpha$ (0.05) $\rightarrow$ Reject H_0
If confidence interval excludes 0 $\rightarrow$ Significant difference

Agricultural Interpretation:
"The fertilizer treatment shows a statistically significant
increase in yield compared to control ($p = 0.023$).
The mean difference is 6.5 bu/acre with 95% CI [1.2, 11.8]."

One-Sided vs Two-Sided Tests

Understanding when to use directional hypotheses.

5.2.1 Test Selection Guidelines

Two-Sided Test Use When:

- Testing for any difference (better or worse)
- Exploratory research
- No prior expectation of direction

One-Sided Test Use When:

- Strong theoretical reason for direction
- Testing if treatment is better (not just different)
- More statistical power for detecting effects in predicted direction

Agricultural Examples:
Two-sided: "Does variety A differ from variety B?"
One-sided: "Does nitrogen fertilizer increase yield above control?"

Caution:
One-sided tests should be planned before data collection
Don't switch after seeing results!

Assignment 7 Overview

Assignment Structure (20 points total)

1. **Part 1: Load and Visualize Data (4 points)**
   - Load wheat yield dataset and convert factors (2 points)
   - Create boxplot visualization (2 points)
2. **Part 2: Form Hypotheses (3 points)**
   - State null and alternative hypotheses clearly (3 points)
3. **Part 3: Perform T-Test (5 points)**
   - Separate data by treatment groups (2 points)
   - Execute two-sample t-test with proper parameters (3 points)
4. **Part 4: Interpret Results (4 points)**
   - Interpret p-value correctly (1 point)
   - Draw conclusions about fertilizer treatments (2 points)
   - Make decision about null hypothesis (1 point)
5. **Part 5: One-Sided Test (4 points)**
   - Perform directional t-test (2 points)
   - Interpret one-sided test results (2 points)

Agricultural Applications

Real-World Function and T-Test Applications:

- **Treatment Efficacy** - Compare fertilizer, pesticide, or irrigation treatments
- **Variety Trials** - Test new crop varieties against established standards
- **Quality Control** - Automated functions for routine quality assessments
- **Experimental Design** - Sample size calculations for field trials
- **Breeding Programs** - Compare genetic lines for yield and quality traits
- **Environmental Studies** - Test effects of farming practices on soil health
- **Economic Analysis** - Compare profitability of different management strategies
- **Precision Agriculture** - Analyze zone-specific treatment effects

Key Concepts Summary

Function Development

Function Best Practices:

- Use descriptive function names
- Include parameter validation
- Add comments explaining purpose
- Test with simple examples first
- Return meaningful results

Common Function Types in Agriculture:

- Statistical calculations (mean, variance, sample size)
- Unit conversions (acres to hectares, etc.)
- Economic calculations (profit, cost per unit)
- Data transformations and cleaning

Function Benefits:

- Reduces code duplication
- Easier debugging and maintenance
- Consistent calculations across analyses
- Shareable with colleagues

Statistical Testing Fundamentals

T-Test Assumptions:

- Independent observations
- Approximately normal distributions
- Random sampling from populations
- For Welch's t-test: unequal variances OK

Effect Size Considerations:

- Statistical significance $\neq$ practical importance
- Consider magnitude of difference
- Economic significance in agricultural context
- Confidence intervals show range of plausible effects

Multiple Comparisons:

- Be cautious with many t-tests
- Consider family-wise error rate
- Use appropriate corrections when needed

Data Analysis Workflow

Step-by-Step Analysis

1. **Data Preparation**
   - Load and examine data structure
   - Convert categorical variables to factors

- Check for missing values and outliers
- Create exploratory visualizations

2. **Hypothesis Formation**
   - Define research question clearly
   - State null and alternative hypotheses
   - Choose appropriate test type (one/two-sided)
   - Set significance level
3. **Statistical Testing**
   - Separate data by groups
   - Check test assumptions
   - Perform appropriate t-test
   - Calculate effect sizes
4. **Interpretation and Communication**
   - Interpret results in agricultural context
   - Consider practical significance
   - Make evidence-based recommendations
   - Communicate uncertainty appropriately

Getting Started

1. Launch Week 7 Binder environment
2. Navigate to `class_activity` folder
3. Open `Week7_Regression_Analysis.ipynb`
4. Work through function examples and t-test exercises
5. Complete Assignment 7 in `assignment` folder

Learning Objectives

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

- Create custom functions with parameters and return values
- Build reusable functions for statistical calculations
- Form appropriate null and alternative hypotheses
- Perform two-sample t-tests using Welch's method
- Interpret t-statistics, p-values, and confidence intervals
- Distinguish between one-sided and two-sided tests
- Apply statistical testing to agricultural research problems
- Make data-driven decisions about treatment effectiveness

Tips for Success

Best Practices:

- Test functions with simple examples before complex data
- Always check t-test assumptions (normality, independence)
- Use `var.equal = FALSE` for Welch's t-test (safer default)
- Interpret p-values correctly (probability under null hypothesis)
- Consider both statistical and practical significance
- Visualize data before testing (boxplots, histograms)
- Choose test type (one/two-sided) before seeing results
- Always interpret results in agricultural context

Common Mistakes to Avoid

Avoid These Errors:

- Writing functions without testing them first
- Forgetting to use `return()` in functions
- Using one-sided tests without strong justification
- Misinterpreting p-values as probability hypotheses are true
- Ignoring practical significance when p-value is small
- Not checking t-test assumptions
- Switching between one/two-sided tests after seeing data
- Over-interpreting non-significant results

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

Last updated: November 2024 | PLS 120 - Applied Statistics in Agriculture | UC Davis
Week 7: Functions and T-Tests