PLS 120: Applied Statistics in Agricultural Sciences

ANOVA and Linear Regression Analysis

Week 10 Tutorial Guide

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

mnarimani@ucdavis.edu

December 2025

Contents

Important Links

Course Website
All course materials available at: Course Website Link
Interactive Binder Environment
Access Week 10 lab materials: Week 10 Binder Link

Welcome to Week 10: ANOVA and Linear Regression

This week, we explore **Analysis of Variance (ANOVA)** and **Linear Regression** - powerful statistical methods for comparing multiple groups and modeling relationships between variables in agricultural research.

Analysis of Variance (ANOVA)

When to Use ANOVA

ANOVA compares means across multiple groups simultaneously, such as testing different diet treatments on pig growth or comparing crop yields across varieties.

ANOVA Components and Formulas

ANOVA Table Structure:

Source	SS	df	MS	F
Between Groups	SSB	$k-1$	MSB	MSB/MSE
Within Groups	SSE	$N-k$	MSE	
Total	SST	$N-1$		

Sum of Squares Formulas:
$SSB = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$ (Between groups)
$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$ (Within groups)
$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2$ (Total)
Mean Squares:
$MSB = \frac{SSB}{k-1}$ (Between groups mean square)
$MSE = \frac{SSE}{N-k}$ (Error mean square)
F-Statistic:
$F = \frac{MSB}{MSE}$ (Ratio of between to within group variation)

ANOVA Hypotheses

ANOVA Hypotheses:
Null Hypothesis (H_0): All group means are equal
$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
Alternative Hypothesis (H_1): At least one group mean differs
H_1 : Not all μ_i are equal
Decision Rule:
If $F > F_{critical}$ or $p < \alpha \rightarrow$ Reject H_0
If $F \leq F_{critical}$ or $p \geq \alpha \rightarrow$ Fail to reject H_0
R Implementation:

```
model <- lm(response ~ factor, data)
anova(model)
summary(model)
```

Manual ANOVA Calculations

Step-by-Step ANOVA Construction

Building ANOVA tables from scratch helps understand the underlying statistical concepts.

Manual ANOVA Steps:
1. Calculate Group Means:
$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$ for each group i
2. Calculate Overall Mean:
$\bar{x} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}}{N}$ where $N = \sum n_i$
3. Calculate Sum of Squares:
SSB: Weighted squared deviations of group means from overall mean
SST: Squared deviations of all observations from overall mean
SSE: SST - SSB (or sum of within-group squared deviations)
4. Calculate Degrees of Freedom:
Between: $df_1 = k - 1$ (number of groups - 1)
Within: $df_2 = N - k$ (total observations - number of groups)
Total: $df_3 = N - 1$ (total observations - 1)
5. Calculate Mean Squares and F-statistic:
$MSB = SSB/df_1$, $MSE = SSE/df_2$
$F = MSB/MSE$

Linear Regression Analysis

Simple Linear Regression

Linear regression models the relationship between a continuous predictor and response variable.

Simple Linear Regression Model:
$y = \beta_0 + \beta_1 x + \epsilon$
Where:
y = response variable (dependent)
x = predictor variable (independent)
β_0 = intercept (y-value when x = 0)
β_1 = slope (change in y per unit change in x)
ϵ = random error term
Least Squares Estimates:
$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
Correlation Coefficient:
$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$
R-squared (Coefficient of Determination):
$R^2 = r^2$ = Proportion of variance explained by the model
R Implementation:

```
model <- lm(y ~ x, data)
summary(model)
cor(x, y)
```

Multiple Linear Regression

Multiple Linear Regression Model:
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon$
Categorical Predictors:
When using factors (like Species), R creates dummy variables:
Reference category (baseline)
Coefficients represent differences from reference
Model Interpretation:
Each β_i represents change in y per unit change in x_i
Holding all other variables constant
Intercept is y-value when all predictors = 0
R Implementation:

```
model <- lm(y ~ x1 + x2 + factor, data)
model <- lm(Petal.Width ~ Species, data)
model <- lm(Petal.Width ~ Petal.Length, data)
```

Assignment 10 Overview

Assignment Structure (20 points total)

1. **Part 1: Data Loading and Study Design (6 points)**
   - Identify treatment, response variable, and experimental unit (3 points)
   - Formulate null and alternative hypotheses (2 points)
   - Load and examine pig weight data (1 point)
2. **Part 2: ANOVA Table Calculations (7 points)**

- Calculate sum of squares (SSB, SSE, SST) (2 points)
- Calculate degrees of freedom (1 point)
- Calculate mean squares and F-statistic (2 points)
- Complete ANOVA table (2 points)

3. **Part 3: Statistical Interpretation and Conclusions (7 points)**
   - Interpret F-statistic and p-value (3 points)
   - Draw conclusions about diet effects (2 points)
   - Discuss practical implications (2 points)

Agricultural Applications

Real-World ANOVA and Regression Applications:

- **Variety Trials** - Compare crop performance across multiple varieties
- **Treatment Comparisons** - Test fertilizer, pesticide, or management effects
- **Feed Efficiency Studies** - Analyze animal growth under different diets
- **Soil Management** - Compare tillage or amendment effects on soil health
- **Environmental Factors** - Model relationships between climate and yield
- **Quality Assessment** - Analyze factors affecting product quality
- **Breeding Programs** - Compare genetic lines or breeding methods
- **Economic Analysis** - Model cost-benefit relationships in agriculture

Getting Started

1. Launch Week 10 Binder environment
2. Navigate to `class_activity` folder
3. Open `Week10_ANOVA_Regression.ipynb`
4. Work through ANOVA and regression examples
5. Complete Assignment 10 in `assignment` folder

Learning Objectives

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

- Build ANOVA tables from scratch with manual calculations
- Understand sum of squares, degrees of freedom, and F-statistics
- Perform one-way ANOVA using R functions
- Fit and interpret simple and multiple linear regression models

- Check assumptions for ANOVA and regression analyses
- Choose appropriate statistical methods for different research questions
- Interpret statistical output in agricultural research context
- Understand the relationship between ANOVA and regression

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