# PLS 120: Applied Statistics in Agricultural Sciences

Descriptive Statistics and Central Tendency


**Week 2 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 are available at:**

**Click Here to Access Course Website**

**Interactive Binder Environment**

**Access Week 2 lab materials directly:**

**Click Here to Launch Week 2 Binder**

## Welcome to Week 2: Descriptive Statistics

This week, we explore **descriptive statistics** and **measures of central tendency** - essential tools for summarizing and understanding data in agricultural research. You'll learn to calculate mean, median, mode, variance, and standard deviation using R!

## Key Statistical Concepts

### Measures of Central Tendency

Understanding where the "center" of your data lies is fundamental to statistical analysis.

#### 3.1.1 Mean ($\mu$ or $\bar{x}$)

**Definition:** The arithmetic average of all values

**Formula:** $\mu = \frac{\Sigma x}{n}$

**When to use:** Most common measure, but sensitive to outliers
**Example:** Average crop yield across all plots

#### 3.1.2 Median

**Definition:** The middle value when data is ordered

**Formula:** Middle value of sorted data

**When to use:** Better for skewed distributions or data with outliers
**Example:** Typical rainfall amount (not affected by extreme storms)

#### 3.1.3 Mode

**Definition:** The most frequently occurring value

**Formula:** Most common value in dataset

**When to use:** Useful for categorical data or finding the most common observation
**Example:** Most common pest species observed

### Measures of Variability

Understanding how spread out your data is from the center.

#### 3.2.1 Variance ($\sigma^2$)

**Definition:** Average squared deviation from the mean

**Formula:** $\sigma^2 = \frac{\Sigma(x-\mu)^2}{n}$

**Note:** Units are squared, making interpretation less intuitive

#### 3.2.2 Standard Deviation ($\sigma$)

**Definition:** Square root of variance - average distance from the mean

**Formula:** $\sigma = \sqrt{\sigma^2}$

**Advantage:** Same units as original data, easier to interpret
**Rule of thumb:** In normal distributions, 68% of data falls within $1\sigma$ of the mean

#### 3.2.3 Coefficient of Variation (CV)

**Definition:** Relative variability measure

**Formula:** $CV = \frac{\sigma}{\mu} \times 100\%$

**When to use:** Allows comparison of variability across different scales or units
**Example:** Compare consistency of wheat yield (tons/ha) vs. corn yield (bushels/acre)

## Key R Functions This Week

### Summary Statistics

`summary(data)` - Comprehensive summary
`mean(data$column)` - Calculate mean
`median(data$column)` - Calculate median
`var(data$column)` - Calculate variance
`sd(data$column)` - Calculate standard deviation
`quantile(data$column)` - Calculate quantiles

### Data Exploration

`head(data)` - First 6 rows
`str(data)` - Data structure
`nrow(data)` - Number of rows
`ncol(data)` - Number of columns

### Custom Mode Function

Since R doesn't have a built-in mode function, we create our own:

```
Mode <- function(x) {
ux <- unique(x)
ux[which.max(tabulate(match(x, ux)))]
}
```

## Assignment 2: Central Tendency Analysis

### Assignment Overview (20 points total)

1. **Part 1: Mean, Median, Mode (7 points)**
2. **Part 2: Variance & Standard Deviation (5 points)**
3. **Part 3: Quantiles (1 point)**
4. **Written Analysis (7 points)**

The assignment uses the LA crime dataset to compare statistics between all victims, male victims, and female victims.

## Why This Matters in Agriculture

**Agricultural Applications:**

- **Crop Yields** - Compare mean yields across varieties
- **Soil Properties** - Understand nutrient variability
- **Weather Patterns** - Analyze rainfall and temperature
- **Pest Populations** - Track abundance changes
- **Quality Control** - Monitor product consistency

## Getting Started

1. Launch Week 2 Binder environment
2. Navigate to `class_activity` folder
3. Open `Week2_Descriptive_Statistics.ipynb`
4. Work with the built-in iris dataset

## Learning Objectives

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

- Calculate and interpret mean, median, and mode
- Understand when to use each measure of central tendency
- Compute variance, standard deviation, and CV
- Use quantiles to understand data distribution
- Compare statistics across different subgroups

## Need Help?

**Mohammadreza Narimani**
Email: mnarimani@ucdavis.edu
Department of Biological and Agricultural Engineering, UC Davis