Common Statistical Tests

Bill Perry

# Common Statistical Tests Reference Guide

**Author:** Bill Perry

This is a comprehensive list of the common statistical tests we have used in class. I hope this serves as a resource you can use in the future when you are doing data analysis.

I have provided the common tests we have done with detailed information that will help you remember how to approach these questions and how to run the tests.

## Comparing Two Samples

### Parametric Two Sample T-Test

* **Webpage:** <01_two_sample_ttest.qmd>
* [📥 Download R Script](01_two_sample_ttest.r)
* [📥 Download sculpin data](data/t_test_sculpin_s07_ne14.csv)

**Data Types:**

* **X Variable:** Categorical (2 groups/levels)
* **Y Variable:** Continuous

**Assumptions:**

1. Independence of observations
2. Normal distribution of the dependent variable within each group
3. Homogeneity of variance (equal variances between groups)
4. Data should be approximately normally distributed
5. No extreme outliers

### Nonparametric Welch’s Two Sample Test

* **Webpage:** <02_two_sample_welches_ttest.qmd>
* [📥 Download R Script](02_welches_ttest_script.r)
* [📥 Download sculpin data](data/t_test_sculpin_s07_ne14.csv)

**Data Types:**

* **X Variable:** Categorical (2 groups/levels)
* **Y Variable:** Continuous

**Assumptions:**

1. Independence of observations
2. Normal distribution of the dependent variable within each group
3. Does NOT assume equal variances (this is the key difference from standard t-test)
4. Data should be approximately normally distributed
5. No extreme outliers

### Nonparametric Mann-Whitney U Two Sample Test

* **Webpage:** <03_two_sample_mann_whitney_u_test.qmd>
* [📥 Download R Script](03_mann_whitney_script.r)
* [📥 Download sculpin data](data/t_test_sculpin_s07_ne14.csv)

**Data Types:**

* **X Variable:** Categorical (2 groups/levels)
* **Y Variable:** Ordinal or continuous (can handle non-normal distributions)

**Assumptions:**

1. Independence of observations
2. Ordinal or continuous dependent variable
3. Does NOT require normal distribution
4. Does NOT require equal variances
5. The two groups should have similar distribution shapes for meaningful comparison of medians
6. No tied ranks (or minimal ties)

## Relationships Between Two Samples

### Correlation

* **Webpage:** [04\_correlation.qmd](04_correlation_analysis.qmd)
* [📥 Download R Script](04_correlation_analysis_script.r)
* [📥 Download m&m peanut data](data/data/mms_peanut.csv)

**Data Types:**

* **X Variable:** Continuous
* **Y Variable:** Continuous

**Assumptions of Pearson Correlation:**

1. Linearity: The relationship between variables is linear
2. Bivariate normality: Both variables are normally distributed
3. Independence: Observations are independent of each other

### Alternative: Non-parametric Correlation

**Data Types:**

* **X Variable:** Continuous or ordinal
* **Y Variable:** Continuous

#### When to Use Spearman’s Correlation

Use Spearman’s rank correlation when:

1. Data are not normally distributed
2. The relationship is monotonic but not necessarily linear
3. Data contain outliers that affect Pearson’s correlation
4. Working with ordinal data

### Linear Regression

* **Webpage:** <05_linear_regression.qmd>
* [📥 Download R Script](05_linear_regression_script.r)
* [📥 Download mozzie data](data/chap17q30DEETMosquiteBites.csv)

**Data Types:**

* **X Variable:** Continuous or ordinal
* **Y Variable:** Continuous

**Assumptions of Linear Regression:**

1. **Linearity**: The relationship between DEET concentration and mosquito bites is linear
2. **Independence**: Observations are independent of each other
3. **Homoscedasticity**: Constant variance of residuals across all DEET concentrations
4. **Normality**: Residuals are normally distributed

### Linear Regression with standard curve

* **Webpage:** <06_linear_regression_as_std_curve.qmd>
* [📥 Download paper weights](data/chap17paperweights.xlsx)
* [📥 Download leaf traces](data/leaf_trace_masses.xlsx)

**Data Types:**

* **X Variable:** Continuous or ordinal
* **Y Variable:** Continuous

**Assumptions of Linear Regression:**

1. **Linearity**: The relationship between DEET concentration and mosquito bites is linear
2. **Independence**: Observations are independent of each other
3. **Homoscedasticity**: Constant variance of residuals across all DEET concentrations
4. **Normality**: Residuals are normally distributed

### Multiple Regression

* **Webpage:** <07_multiple_linear_regression.qmd>
* [📥 Download R Script](07_multiple_regression_script.r)
* [📥 Download ants.csv](data/AntSpeciesDensity.csv)

**Data Types:**

* **X1 Variable:** Continuous
* **X2 Variable: Continuous**
* **Y Variable:** Continuous

**Assumptions of Multiple Linear Regression**

1. Linearity: Linear relationships between Y and each X, and Y and the combination of X’s
2. Independence: Observations are independent
3. Homoscedasticity: Constant variance of residuals
4. Normality: Residuals are normally distributed
5. No multicollinearity: Predictor variables are not highly correlated with each other
6. Sufficient sample size: Generally need at least 10-20 observations per predictor

### One Way ANOVA

* **Webpage:** <08_one_way_ANOVA.qmd>[.qmd](05_linear_regression.qmd)
* [📥 Download R Script](08_one_way_anova_script.r)
* [📥 Download mms.csv](data/mms.csv)

**Data Types:**

* **X Variable:** Categorical
* **Y Variable:** Continuous

**Assumptions:**

1. Independence: Observations within and between groups are independent
2. Normality: The residuals are normally distributed
3. Homogeneity of variances: The variances are equal across all groups

*This guide will be expanded as we cover additional statistical tests throughout the course.*