ESM 206 (Fall 2017)
Statistics and Data Analysis in Environmental Science and Management

Bren School of Environmental Science and Management
UC Santa Barbara

Instructor: Allison Horst
Email: ahorst@bren.ucsb.edu
Allison's Office: Bren 4406
Office Hours: Thursday 12 pm – 2 pm (Sycamore Room, BH 1510)

TA: Sean Fitzgerald
TA Email: spfitzgerald@ucsb.edu
TA Office: Bren Hall 1306
Office Hours: Tuesday 11am – 12pm (Sean's office, BH 1306)

Course Description: Develop critical thinking, technical and communication skills to successfully approach and answer environmental questions using quantitative and qualitative data. Topics include: descriptive statistics, experimental design, uncertainty, exploratory data analysis, basic probability theory, hypothesis testing, data visualization and communication. Skills will be developed through analysis of real-world datasets using R and RStudio software.

Grading: 55% Homework and Lab
20% Midterm (in class)
25% Final Exam (take-home)

Course Policies:

- Assignments submitted late will only be accepted within one week of the due date, and will be worth 50% of the original score. **Homework submitted more than one week after the original due date will not be accepted.**
- Lab attendance is **mandatory.**

Bring to each lecture: Course reader (required), pen/pencil, calculator

Tentative Course Schedule and Topics:

Week 1: Exploratory data analysis, intro to data visualization; accessing, loading and filtering datasets in R

Week 2: The normal distribution; Central Limit Theorem; exploring normality and basic data manipulation in R (sorting, subsetting, classes of data, histograms, QQ-plots, skewness and kurtosis); exploratory data analysis in R (tables, scatterplots, summary statistics continued)

Week 3: Considering, calculating and expressing uncertainty (data spread, SD, SE, etc.; residuals (and assumptions); probability continued (the z- and t-distribution); intro to hypothesis testing; communicating uncertainty (graphics/tables/text)

Week 4: Hypothesis testing continued, assumptions, null and alternative hypotheses, errors and considerations in hypothesis testing, limitations (why isn't a *p*-value enough?); effect size; power; communicating results of hypothesis tests (graphically, in tables and in text); hypothesis testing in R (performing, understanding outputs, communicating results); finding effect size and power (*a priori* power analysis for experimental design)

Week 5: Linear regression (intro to ordinary least squares); regression assumptions, considerations, concerns; regression diagnostics in R; coefficient of determination, correlation (Pearson's *r*)

Week 6: Analysis of Variance (ANOVAs), conceptually, mathematically and in RStudio; post-hoc testing (by Tukey's HSD); communicating results. Working with count/categorical data; chi-square analysis, interpretation and communication; tables and counts in R; updating classes of data (factors/characters/integers) in R and specifying levels.

Week 7: Multiple linear regression (conceptually, mathematically); interpreting multiple regression results; considerations in multivariate analyses; dummy variables and reference variables; pitfalls of the *p*-value; model diagnostics and concerns

Week 8: Multiple linear regression continued: exploring AIC, VIF, multicollinearity. What do we do when data doesn't satisfy assumptions of normality/parametric tests? Exploring alternatives: rank-based tests, non-parametric approaches, transformation; dealing with data limitations; types of bias in data collection, analysis and interpretation

Week 9: Rank-based tests continued (Wilcoxon Signed Rank, Mann-Whitney U, Kruskall-Wallis), pitfalls and considerations in rank-based tests

Week 10: Multi-factor ANOVA, non-parametric regression (binary logistic, ordinal logistic regression), course review