OCN 683: Advanced Statistics in R

Sample Syllabus

Instructor: Kyle Edwards, office MSB 612, kfe@hawaii.edu, 956-6198

Course Goals

Biological data can be challenging to analyze, due to non-normal distributions, nonlinear relationships, spatial/temporal structure, and multivariate responses/predictors. The goal of this course is to introduce students to a variety of methods that can address these issues, while also developing the programming skills that allow one to better understand and troubleshoot statistical analyses. We will also look at the different approaches for testing hypotheses and drawing inferences from statistical analyses. Although the topics are "advanced", I will start from first principles and emphasize how most statistical analyses can be thought of as models that you build to describe/test for patterns in data. The class will be taught in R, which is freely available software that is ideal for flexible data analysis. The course works well as a sequel to OCN 682 - Introduction to Programming and Statistics in R. The methods and examples will focus on biological applications, especially ecology and oceanography, but the methods covered are broadly used across scientific fields.

Course Structure

3 credits. Two 1-hour lectures per week. MW 9:30-10:20, Watanabe 114.

Student Learning Outcomes

At the end of this course you will be able to:

- Use generalized linear models, generalized additive models, mixed models
- Analyze multivariate data with regression and ordination techniques
- Test hypotheses with frequentist, information criteria, and bayesian approaches
- Understand how maximum likelihood works
- Use simulation to understand how statistical models work
- Understand general principles of model/experimental design, and choose appropriate statistical methods for your analysis

Requirements/Prerequisites

There are no strict prerequisites for this course. In my lectures I will start from basic statistical principles, but we will move quickly into more advanced material. The lecture notes and homework assignments will be designed to help you learn programming skills as well as statistical methods, but I will not spend a lot of time on the basics of programming. I recommend that students have the equivalent of an undergrad stats class, or some experience with a programming language (ideally R, but Matlab, Python, etc. will suffice), or both.

Reading/Texts

The course will have no required reading, but book chapters/papers relevant to each lecture will be distributed for supplementary reading.

Software

R and other resources can be found here http://cran.us.r-project.org/. R Studio (http://www.rstudio.com/) is free software that makes the R GUI at little more user friendly, at least if you're using a PC (the basic Mac GUI is better than the basic PC GUI). R Studio also has integrated R Markdown features that make it useful for generating the homework assignments: https://rmarkdown.rstudio.com/.

Grading Scheme

The course can be taken CR/NC. Grades will be based entirely on weekly homework assignments that practice the methods described in lecture.

Homework Formatting

A single R Markdown document that alternates between your annotated R code and the corresponding requested output (figures, model output, etc). I.e., format it like the lecture notes. I recommend the html output format, it tends to work the best.

Office Hours

By request

Schedule

Lecture 1 — Refresher on classic statistical methods and linear models
- Linear regression, ANOVA are models of your data

Lecture 2 — Important probability distributions
- Normal, Poisson, binomial distributions

Lecture 3 — Important nonlinear functions
- Exponential, saturating, sigmoidal curves

Lecture 4	Programming in R Vectors, dataframes, loops and conditionals
Lecture 5-6	Maximum likelihood What is means, how it works
Lecture 7-8	Generalized linear models Binomial/logistic, Poisson models
Lecture 9	How to model and interpret multiple predictors
Lecture 10	Too many zeros, zero-inflated models
Lecture 11	Transformations and tricky data distributions
Lecture 12	Using simulation to understand statistical models
Lecture 13	Using models to make predictions, cross-validation
Lecture 14	Nonlinear least squares
Lecture 15-16	Model selection, AIC, multi-model inference
Lecture 17-18	Generalized additive models
Lecture 19-22	Mixed/hierarchical models
Lecture 23	Generalized additive mixed models
Lecture 24	Spatial and temporal autocorrelation, GLS
Lecture 25-27	Methods for multivariate data PCA, PCoA, NMDS, PERMANOVA
Lecture 28-29	Bayesian inference and MCMC