

BIOS 60576 --- Advanced Biostatistics with R

Spring 2017

2:00-3:15 MW; 283 Galvin Hall

Instructor

Stuart Jones, Office: 264 Galvin, email: sjones20@nd.edu, Phone: 631-5703

Overview

This course will be taught in three modules:

M1 - Introduction to the R Statistics Environment

M2 - Maximum Likelihood to evaluate competing hypotheses

M3 - Application of Multivariate Statistics

Each module will use approximately nine class periods. These class periods will be a mixture of interactive lecture, group discussion of worked problems or literature, and real-time implementation of code. The overall goal of the course is to introduce the R Environment as a sophisticated, scientific programming platform. My hope for you is that at the end of the semester you will be able to use R as a vehicle to efficiently progress in your research. This may occur via direct application of approaches covered in the class, but more likely as a result of being equipped to self-teach the diverse domain-specific applications available in the R Environment.

Learning Goals

M1 – Complete basic research tasks, including data access, transformation, and manipulation, in an automated and reproducible manner using the R Environment.

M2 – Translate biological data to biological inference using multiple model comparison with maximum likelihood approaches.

M3 – Understand the strengths and weaknesses of a variety of multivariate statistical approaches and be able to implement these approaches in the R Environment.

Evaluation

As this is a graduate course, your grade is not of primary concern. Rather, my focus is on your progress towards the goals listed above. Having said that, it is useful for you and I to be able to evaluate whether we have achieved the learning goal of each module. Therefore, we will have a short take-home exam at the end of each module. Your course grade will be partially based upon these take-home exams (20%), but mainly based on in-class participation (80%).

Office hours

I won't have regular office hours for this course, but will be happy to meet with students by appointment.

Course text

There will be no required text for this course. I will provide you with weekly handouts. Here is a list of books that I would recommend as resources in biostatistics. I used many of these books in preparing this course.

-*A Primer of Ecological Statistics*, Gotelli & Ellison

-*The Art of R Programming*, Matloff

-*Data Analysis Using Regression and Multilevel/Hierarchical Models*, Gelman & Hill

-*Ecological Models and Data in R*, Bolker

-*The Ecological Detective*, Hilborn & Mangel

-*Bayesian Computation with R*, Albert

-*The R Book*, Crawley

-*Numerical Ecology with R*, Borcard, Gillet, Legendre

Course schedule

Week	Monday	Wednesday
January 16	-	M1 - R basics and data structures
January 23	M1 - Exercise 1	M1 - Programming and flow control
January 30	M1 - Exercise 2	M1 - Custom functions
February 6	M1 - Exercise 3	M1 - Text manipulation
February 13	M1 - Exercise 4	M1 - Plotting
February 20	M2- Likelihoods: linear models and the normal distribution	M2 - Exercise 1
February 27	M2 - Non-normal distributions and more likelihoods	M2 - Exercise 2
March 6	M2 - Investigate a field-specific example	M2 - Implement an example of personal interest
March 13	NO CLASS - Spring Break	NO CLASS - Spring Break
March 20	M2 - Generalized linear models	M2 - Exercise 3
March 27	M2 - Model comparison	M2 - Exercise 4
April 3	M3 - Why multivariate statistics and Association measures	M3 - Clustering
April 10	M3 - Indirect ordination	M3 - Exercise 1
April 17	NO CLASS - Easter Break	M3 - Classification and hypothesis testing
April 24	M3 - Exercise 2	M3 - Direct ordination
May 1	M3 - Exercise 3	M3 - Data analysis derby