Biology 133: Ecological Statistics and Data

Spring 2020

Instructor: Eric Scott
Office hours: Wednesday 1:00–2:00, Robinson 350 (across from the Biology office)
TA: Avalon Owens
Office hours: **TBD**

Lecture 3:00–4:15 Tuesday & Thursday, 226 Braker Hall
Lab: 1:30–4:00 Monday, Lincoln Filene 201 (former Rabb Room)

Readings:

- Bolker, B. 2008 *Ecological models and Data in R.* https://ms.mcmaster.ca/~bolker/emdbook/ (lecture)
- Grolemund & Wickham 2017. *R for Data Science.* https://r4ds.had.co.nz/ (lab)
- Other online readings TBA

Overview:

Emerging issues in health and environmental sciences require an understanding of how to make inferences and decisions in the face of limited data and uncertainty. In the past few decades, advances in computational sciences have opened new ways of using and interpreting data. These advances are fundamentally changing how biologists approach statistics. This course introduces students to the principles of probability and likelihood that underpin these growing areas of statistical computing. Examples come from ecological models and data, but reflect general principles of working with discrete data, maximum likelihood and Bayesian estimation techniques, and using fundamental principles of probability to build and interpret mixed and mixture models. Work will be done using the open-source statistics program, R, and exercises will emphasize building skills by working with peers, and by trial-and-error, both of which are key features of the open-source computing community.

Outline of Course Content

Schedule is subject to change.

Day	Date	Lecture	Lab
Thu	Jan-16	Probability	
Wed	Jan-22		Count Something (Tuft's Monday)
Tue	Jan-21	Likelihood	
Thu	Jan-23	Bayes' Theorem	
Mon	Jan-27		Working With Data
Tue	Jan-28	Distributions	
Thu	Jan-30	Binomial Distributions	
Mon	Feb-03		Functions & Loops
Tue	Feb-04	Model Competition	
Thu	Feb-06	Confidence	
Mon	Feb-10		Bayesian Data Analysis
Tue	Feb-11	Link Functions	
Thu	Feb-13	Population Dynamics	

Day	Date	Lecture	Lab
Tue	Feb-18	Poisson Distribution	
Thu	Feb-20		Binomial Models (Tuft's Monday)
Mon	Feb-24		Data Visualization
Tue	Feb-25	Deviance	
Thu	Feb-27	Exam 1	
Mon	Mar-02		Probability Distributions
Tue	Mar-03	Generalized Linear Models	
Thu	Mar-05	Distribution and Abundance	
Mon	Mar-09		Exam 1 Revisions
Tue	Mar-10	Zero-Inflated Poisson	
Thu	Mar-12	Why Be Normal?	
Mon	Mar-16	Spring Break	
Tue	Mar-17	Spring Break	
Thu	Mar-19	Spring Break	
Mon	Mar-23		Linear Models
Tue	Mar-24	Variance is Not Just Noise	
Thu	Mar-26	Space and Time	
Mon	Mar-30		Offsets and Zero-Inflated Models
Tue	Mar-31	Model Selection	
Thu	Apr-02	Design Matrix	
Mon	Apr-06		Mixed Models
Tue	Apr-07	Another Factorial Experiment	
Thu	Apr-09	Exam 2	
Mon	Apr-13		Exam 2 Revisions
Tue	Apr-14	Sources of Variance, Revisited	
Thu	Apr-16	Hierarchical Models	
Mon	Apr-20		Patriot's Day (no lab)
Tue	Apr-21	Multivariate Statistics	
Thu	Apr-23	TBD	

Grading:

Grades will be based on attendance/participation, approximately weekly homework assignments, which will be given a qualitative score (exceeds expectations, meets expectations, needs improvement, unacceptable), participation in lab exercises, and three graded tests. The basic calculation will be:

- Attendance/participation — 5%
- Lab exercises — 30%
- Homework — 25%
- In-class tests — 12.5% each (25% total)
- Final exam — 15%

Late homework assignments will be lowered by one qualitative category, per week (rounded up) they are late. Skills in data analysis and programming are cumulative, so it is especially important to stay on top of the assignments.

Homework should be printed out and turned in during class. I do not accept electronic submission, including email attachments. Homework questions should be answered in words or with calculations shown on the page. Unlike some Computer Science classes, I do not want you to turn in code. In the lab, you will learn to use R Markdown which allows you to mix written answers, code, and output from code, which you may use for completing your homework assignments if you like.

Lab

There are four main **objectives** for the lab portion of the class:

- Build/solidify programming skills in R
- Learn tools and workflows for reproducible analyses
- Review and revisit very basic statistical concepts (average/mean, variance, correlation, normal distribution), as needed
- Review material from the class

A reading will be assigned for each lab from texts available online. For each lab you should:

1. Read the assigned reading **before class** and attempt to implement any code it contains
2. Prepare two short questions or comments about the readings and turn these in on Canvas **before the lab** begins, that is, by 1:30 pm on Monday. These may be simple comments such as "I couldn't get X part of the code to work" or "what is a tibble?" or "what does 'regression' mean in this context?"
3. Spend the lab period completing exercises collaboratively with other students with help from your TA, Avalon Owens
4. Turn in final code **by the end of the lab period**, that is, by 4:00 pm on Monday

After each exam, all students in the class will have the opportunity to re-do problems they missed. Your final grade on each exam will be the average of the original grade and the corrected grade. I will NOT answer any questions about the exam during the regrading period. Two lab sessions will be devoted to time for exam revisions. TA Avalon Owens will be available in the lab sessions as well as during office hours to provide some guidance, but the primary onus is on you to work through the problems using past notes.

Lab grades will be based on the percentage of class periods attended, and assignments completed. Although the assignments will not be given a letter grade, partly completed exercises with little evidence of effort may not count as complete.