18 April 2021

COURSE SYLLABUS

Advanced Ecological Data Analysis 16:215:599 Spring 2021
Thursdays 12:35 – 3:35

Course materials on Canvas and Github at https://github.com/aeda2021/

Instructors

Rachael Winfree (128 ENR building) – rachael.winfree@rutgers.edu
Malin Pinsky (130 ENR building) – malin.pinsky@rutgers.edu

Pre-requisites: Knowledge of basic statistics consistent with an introductory statistics course (i.e., probability distributions, t-tests, ANOVA, simple linear regression) and basic knowledge of R or other scientific programming language (e.g., Matlab). The course will be conducted in R. For those with limited knowledge of R, we suggest spending substantial time becoming familiar with the language before the semester.

Summary: This course provides an overview of advanced statistical methods commonly used to model ecological data. Such data often violates assumptions of simpler techniques, including: non-normal distribution of response variables, non-linear and non-monotonic relationships between predictors and response variables, and the presence of spatial and temporal autocorrelation. We will focus on application of these methods with only minimal discussion of their theoretical basis. Students will be encouraged to work with their own datasets.

Course Objectives

- Introduce students to best practices for data organization and coding.
- Develop skills in the application of statistical methods to complex ecological data.

Learning Outcomes

At the end of this class, students will be able to:

- Organize their data and code for transparent, efficient, and repeatable analysis
- Determine the appropriate statistical method for answering ecological questions given a set of characteristics of the data
- Apply and interpret generalized linear models
- Understand when and how to use fixed and random effects in linear models
- Develop and fit non-linear models to data using maximum likelihood

Format: One 3 hour meeting per week. Class time will be largely devoted to discussion and problem-solving exercises in R. Lectures that students must absorb prior to each class will be provided on Canvas as pre-recorded videos, i.e., a “flipped classroom” design.

Evaluation: mid-term exam (30%), final project (30%), in class work and participation (40%)

Readings: All required readings to be posted on Canvas.

Topics: Suggested topics as follows, with flexibility based on student interests to be determined prior to the start of the course.

- Week 1 (Jan 21; Winfree) Course intro; data manipulation and visualization
- Week 2 (Jan 28; guest instructor Alexa Fredston) Using Git and Github to stay organized
- Week 3 (Feb 4; Winfree) Review of introductory statistics
- Week 4 (Feb 11; guest instructor Dylan Simpson) Maximum likelihood
- Week 5 (Feb 18; Winfree) Generalized linear models (GLMs)
- Week 6 (Feb 25; Pinsky) Random effects and mixed effects models
- Week 7 (March 4) Mid-term
- Week 8 (March 11; guest instructor Dylan Simpson) Bayesian methods
- Week 9 (March 25; Pinsky) Generalized Additive Models (GAMs)
- Week 10 (April 1; Pinsky) Generalized Linear Mixed Effect Models (GLMMs)
- Week 11 (April 8; guest instructor Brendan Reid) Machine learning
- Week 12 (April 15; Winfree) Randomization, permutation, and null models
- Week 13 (April 22) Guided in-class work on final projects
- Week 14 (April 29) Final project presentations
- Final paper due Monday, May 3 by 5:00pm

Final projects

- Develop a solid, well-thought-out statistical analysis of a dataset. Use what we have done in class, or go beyond if you want. We encourage you to use one of your own datasets related to your thesis research project so that the class can help move you forwards in your PhD.
- Present this in class (10 min with 5 min questions)
    - No need to show us your code, unless it is helpful to show a small part to explain what you did
- Write this up in a scientific paper format, e.g., Intro, Methods, Results, Discussion. Explain the question, study system, etc. Walk through your logic and analytical process. Interpret what you found.
- Submit your paper as a commit to your Git repo as either of
    - RMarkdown document with text and code
    - Regular document (Word, Google Doc, etc.) with associated R script to do the analysis, make the figures, etc.