Statistical Methods in Ecology and Evolution Part I IBIO/PLB/ENT 830

Fall 2024 T/Th 10:20-11:40 in the EEB HUB (first floor BPS)

Instructors

Dr. Joe Riedy
he/him/his
Office: STEM 1110M
Zoom Room: https://msu.zoom.us/j/8441613381 (Password: Sturgeon)
Email: riedyjos@msu.edu
Office hours: Thursday 2:00-4:00 in STEM 1110M and on Zoom

Dr. Emily Josephs
she/her/hers
Office: Plant Biology Lab 266
Zoom Room: https://msu.zoom.us/my/emjos (Password: plantsrule)
Email: josep993@msu.edu
Office hours: noon-1pm Tuesday + Thursday in PLB 266 (starting Oct 17)

TA: Sophie Buysse
she/her/hers
Office: Plant Biology Lab 266A
Zoom Room: https://msu.zoom.us/j/91936856477?pwd=aaiE720BlzJgWIcbeCmZdTdexfj33b.1 (Password: statshelp)
Email: buysseso@msu.edu
Office hours: 11am-noon Monday/Wednesday in the EEB Hub and on Zoom

Description and Objectives

This graduate level survey course is the first semester in a two-semester sequence, focused on the fundamental elements of data analysis in the fields of ecology and evolution. Students will learn how to interpret and model biological data with modern methods for estimation and inference using the R computing language.

Topics for this semester include: Quantitative toolbox, introduction to R, reproducibility and responsible coding practices, Rmarkdown, plotting, navigating errors/getting help, study design, introduction to probability theory, demystifying probability distributions, introduction to deterministic relationships, likelihood-based inference, modes of inference.

This will prepare students for topics that will be included next semester, including: Frequentist inference, linear models, generalized linear models, mixed and random effects, hierarchical models, zero-inflation models, model comparison and evaluation, power simulations, numerical

simulations, temporal and spatial correlations, and a holistic view of quantitative tools throughout the scientific process.

Recommended Background

There are no required prerequisites for 830. This course is designed to take students in the biological sciences of a variety of skill levels and backgrounds through a thorough quantitative grounding in statistical analysis in their field. However, this class (830) is a mandatory prerequisite for 831 and no students will be allowed into 831 without first taking 830. We strongly encourage all students to sequentially take both semesters of the sequence.

Course Goals

1. Provide the necessary background and quantitative foundation to learn how to analyze data for your thesis and for the rest of your career.
2. Introduce R, a powerful programming environment for data analysis and presentation. Develop skills at writing R functions, with the goal of being able to perform advanced, computationally intensive analyses.
3. Provide a conceptual introduction to model-based analysis and the translation from statistical to biological inference.
4. Lay the foundation for developing a sufficient and appropriate background to teach yourself new methods for data analysis as needed.
5. Build a network of quantitatively minded peers and forge a cohesive dynamic among incoming members of the grad cohort.

Learning Outcomes and Assessments

1. Students will develop a proficiency with programming in the R computing language. This will be assessed by lab assignments.
2. Students will develop a comprehensive understanding of probability and likelihood-based inference. This will be assessed by lab assignments and in-class presentations.
3. Students will develop a fluency with continuous and discrete probability distributions and data types, as well as deterministic functions for describing variable relationships. This will be assessed by lab assignments and in-class presentations.
4. Students will be able to teach themselves new statistical techniques as needed. This will be assessed by in-class presentations.

Course website: d2l.msu.edu

All class materials will be posted to the course's D2L website including lab materials and homework. All assignments will be submitted via the D2L site.

Course materials and books

For this course students will need a computer capable of running R and R Studio. Most decent laptops and desktops with >8GB RAM can run most R programs. Please talk to the instructors if you are having problems with computer access. This course does not have an official textbook. There is a list of useful references included at the end of this syllabus.

Office Hours (aka Drop in help sessions)
We will be doing office hours both in person and on zoom (see schedule above for locations). This time is for you to come ask questions about course content, get help with homework questions you missed, work on homework questions for the next class, or chat about anything you'd like to. If you would like to meet over zoom during an in-person office hour, please let the instructor know and we will make sure to open up zoom and check for you. If you cannot make office hours, you can email the instructors to set up a separate time to meet.

Grades
Course grades are based on the following components:

Participation (5%): To earn full credit for participation, students are required to contribute to discussions and actively participate in in-class group activities and coding sessions.

Exercises (40% assignment, 10% self-assessment): Most weeks we will be doing exercises using R during class. Students may work together on assignments, but each student must submit their own work. Late homeworks will not be accepted so please turn in whatever you have at the deadline. We understand that life happens and everyone can skip one homework without a grade penalty. Please contact the TA if you will need to miss more than one homework.

After the due date, we will post a key to the homework and ask each of you to grade your own homework and fill out a short self-assessment on d2l within one week. This self-assessment is intended to help you figure out what parts of the material you may need to spend more time on, and help us see what we may need to review in class.

We also encourage you to come to office hours or schedule meetings with any of the instructors to clarify any concepts you're struggling with. Wrestling with concepts outside of class is the best way to learn the material, but we also want to help you succeed.

Group projects (40%): The tools and skills we are providing are best absorbed when put into use. There will be three rounds of group-presentations during this class:

1) R data analysis (15%)
Students will form groups of 3 based on shared research interests to analyze an uncleaned dataset. This dataset can be generated from available data (i.e. observations

on iNaturalist) or come from their PI, but should not have been previously cleaned or analyzed. Each group will create an Rmarkdown file that describes their data analysis pipeline for the dataset. This document should include clearly commented steps for the four major components of data analysis in R: data exploration, cleaning, analysis, and visualization. Each group will produce one Rmarkdown file with at least three figures created by the individual members of the group.

2) Distribution maps (10%)

In groups of 3-4, students will work together to build a map to help them pick the distributions that best fit different types of data.

3) Simulation/Inference Battles (15%)

Students will be split into teams of 3-4. On day 1, each team will pick a deterministic function and a probability distribution and combine them to simulate a dataset and create an associated biological story of the data. On day 2, teams will exchange simulated datasets (without revealing the functions/distributions/parameters chosen), and will then work together to infer the parameter values used for simulating the received data. In class on day 3, each group will give a presentation of the parameterization of the simulated/inferred models. The team that carried out the best inference will be awarded a prize.

End of semester self-assessment (5%): We will work with you to set your own goals at the beginning of the class, evaluate your progress in the middle of the semester, and self-evaluate your success in the course at the end of the semester.

Grading Scale

Grade in course	Percent earned
4.0	90-100%
3.5	85-89%
3.0	80-84%
2.5	75-80%
2.0	70-74%
1.5	65-69%
1.0	60-64%
0.0	<60%

Absences: Attendance will be essential to your success. If you have to miss class, you are still responsible for any assignments or presentations due that day. As assignments may not be posted in advance, please check D2L and ask classmates to see what you missed and turn in any assignments in a timely fashion. If you will be missing a class presentation (dates underlined on syllabus), please contact the professors well in advance of the date to make a plan for your participation that accommodates your absence. If you miss class and would like a recording of the previous year's lectures (when the course was taught online), please email the instructor.

Religious holidays: If you will miss an in-class presentation assignment to observe a religious holiday, please let us know within the first two weeks of class, so we can make arrangements.

Respect for Diversity
Our classroom will be a place where diversity is accepted and valued. The differences between class members will be embraced. Language that degrades an individual or group because of gender, ethnicity, nationality, socioeconomic status, religious preference, sexual orientation, or intellectual ability will not be tolerated.

Mental Health Resources for Students
As your instructors, we care about your well-being and we encourage you, when needed, to access resources that are available to help you cope with life's struggles. These resources include:

MSU Counseling Center:
Web: counseling.msu.edu
Email: counseling@cc.msu.edu
Phone: 517-355-8270

MSU Olin Student Health Center
Web: olin.msu.edu 24/7
Phone nurse: 517-353-5557

Listening Ear Crisis Center – Lansing, MI
(517) 337-1717 (24-hour Crisis Line)
(517) 337-1728 (Business Line)
2504 E Michigan Ave., Lansing, MI

Learning Continuity Statement

Your health is essential to learning! If you fall ill or have other reasons that you are unable to attend this course for an extended period of time, please contact Sophie Buysse and the instructor (either Dr. Riedy or Dr. Josephs) as soon as possible. We will work with you to make sure to accommodate your individual situation.

Course Continuity Statement: If either Dr. Riedy or Dr. Josephs are unable to teach due to illness, the other will take over their lectures. In the unlikely event that neither professor can teach, the course will be put on hiatus. The course will resume where it left off on the same order of lectures. If possible, we will combine lectures to cover all material as originally intended and issue a revised schedule ASAP. If this hiatus is during the time scheduled for a group project, Sophie Buysse will continue to help students complete the project.

Class schedule IBIO/PLB/ENT 830 Fall 2024

Primary Instructor: Dr. Joe Riedy

Date		Day	Topic
8/27	Tuesday	1	Course Intro Introduction to instructors
8/29	Thursday	2	Intro to R What is it and why use it?
9/3	Tuesday	3	RStudio R studio data types/structures
9/5	Thursday	4	Data Manipulation indexing & subsetting in base & dplyr
9/10	Tuesday	5	Data Manipulation II Adding loops
9/12	Thursday	6	Data Visualization Base plotting
9/17	Tuesday	7	Data Visualization II Multi-panel plots
9/19	Thursday	8	Data Analysis Intro Intro to data analysis pipeline using lm
9/24	Tuesday	9	Programming I Functional functions
9/26	Thursday	10	Programming II Lexical scoping: is it logical?
10/1	Tuesday	11	Programming III Style and Speed
10/3	Thursday	12	RMarkdown Like Word
10/8	Tuesday	13	Data analysis projects Data exploration in your own project

10/10	Thursday	14	Wrap-up
10/15	Tuesday	14	Reflections

Primary Instructor: Dr. Emily Josephs

Date		Day	Topic
10/17	Thursday	15	Probability I – simulations
10/22	Tuesday		NO CLASS FALL BREAK
10/24	Thursday	16	Probability II – math
10/29	Tuesday	17	The Binomial Distribution
10/31	Thursday	18	Probability Distributions and Distribution maps
11/5	Tuesday	19	**Election day!! No class! Go vote and/or help your friends vote!**
11/7	Thursday	20	Distribution maps
11/12	Tuesday	21	Intro to likelihood
11/14	Thursday	22	Intro to model building: deterministic functions
11/19	Tuesday	23	Model Building II: nonlinearity
11/21	Thursday	24	Simulation/Inference Battles
11/26	Tuesday	25	Experimental design
11/28	Thursday		NO CLASS THANKSGIVING
12/3	Tuesday	26	More Simulation/Inference Battles
12/5	Thursday	27	Simulation/Inference Battle Presentations, Wrap-up

Helpful References

(Copies of these books are available in the Josephs lab bookshelf)

Bolker, B.M. (2008). Ecological models and data in R. Princeton University Press.

McElreath, R. (2015). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. Chapman & Hall/CRC Texts in Statistical Science.

Hobbs, N.T., & Hooten, M.B. (2015) Bayesian models: a statistical primer for ecologists. Princeton University Press.

Whitlock, M.C. & Schluter, D. (2008) Analysis of Biological Data. Roberts and Company Publishers.

Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., & Smith, G.M. (2009). Mixed effects models and extensions in ecology with R. Springer Science & Business Media.

Edge, M.D. (2019). Statistical Thinking from Scratch: A Primer for Scientists. Oxford University Press.