Introduction to Statistical Model Building in R

EEB 429: M/W 11:30am-1pm
4151 Undergraduate Sciences Building (USB)

Instructor:

Dr. Gideon Bradburd
Pronouns: *he/him/his*
Office: BSB 3042
Email: bradburd@umich.edu
Office hours: TBD
BSB 3042

GSI:

Ivana Barnes
Pronouns: *she/her/hers*
Office: BSB 2020D
Email: ivbarnes@umich.edu
Office hours: TBD
TBD

Description and Objectives

This course is focused on the fundamental elements of data analysis in the fields of ecology and evolutionary biology. Students will learn how to interpret and model biological data with modern methods for estimation and inference using the R computing language.

Topics for the semester include: introduction to R/Rmarkdown, data plotting, navigating errors/getting help, introduction to probability, demystifying probability distributions, introduction to deterministic relationships, likelihood-based inference, Bayesianism vs. Frequentism, and modes of inference.

Recommended Background

Enforced prerequisites are: BIOLOGY 171, [BIOLOGY 172 or BIOLOGY 174], and BIOLOGY 173; or BIOLOGY 192 and BIOLOGY 173; or BIOLOGY 195 and BIOLOGY 173. STATS 250 and/or BIOLOGY 202 or BIOLOGY 131/BIOPHYS 117/COMPFOR 131 are recommended, but not required. This course counts toward the 60 credits of math/science required for a Bachelor of Science degree.

Course Goals

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

Learning Outcomes and Assessments

1. Students will demonstrate a proficiency with programming in the R computing language. Assessed via homework assignments, class engagement, and class projects.
2. Students will develop a comprehensive understanding of probability and likelihood- based inference. Assessed via homework assignments, class engagement, and class projects.

3. Students will develop a fluency with common continuous and discrete probability distributions and data types, as well as deterministic functions for describing variable relationships. Assessed via homework assignments, class projects, class engagement, and an in-class presentation.
4. Students will be able to teach themselves new statistical techniques as needed. Assessed via homework assignments, class projects, and class engagement.

Course website: https://umich.instructure.com/courses/737741

All class materials will be posted to the course's Canvas website, including lecture materials, lecture videos, and homework/project assignments. Assignments will be submitted via the Canvas site. Lecture slides will be posted before lectures so that students can use them as a study aide.

Grades

Course grades are based on the following components (see Letter Grades section for conversion):

- *Engagement (12.75%):* Students are required to contribute to discussion and actively participate in in-class group activities and coding sessions. To earn full credit for engagement, students must respond to the iClicker questions given during class to assess learning (questions are graded for correctness). You only need to answer 75% of the questions correctly to earn full credit for the course engagement portion of your grade.

- *Exercises (59.5%):* A homework exercise will be assigned after most classes. Assigned exercises will be due by the start of Monday's class the next week (unless otherwise specified). Students may work together on assignments, but each student must submit their own work. Wrestling with concepts outside of class is the best way to learn the material! Exercises will be assessed jointly for correctness and effort, and a key will be posted for your review. I strongly encourage you to compare your work with the posted key and schedule meetings with me or the GSI to clarify any concepts you're struggling with. Late homework will not be accepted, but there will be opportunities for students to earn full credit on late assignments (see Attendance & Late Assignment section below). You only need to answer 80% of the questions correctly to earn full credit for the homework portion of your grade.

- *Projects (12.75%):* The tools and skills I am providing are best absorbed when put into use. There will be three projects in this class, each of which will be worth 5% of your grade.

    - **R data exploration** (5%) *Students must use Rmarkdown to generate a report on a real-life dataset on a topic of their choice. This document should include clearly commented steps for at least four major components of data analysis in R: data exploration, cleaning, visualization, and interpretation. Each report must contain at least three figures. Note that these reports are not expected to include quantitative and/or statistical analysis; instead, interpretation of the data should be made on the basis of data visualization.*

    - **Probability Distribution Demos** (5%) *Each student will be assigned a probability distribution to introduce to the rest of the class. These introductions should be made in R markdown and feature 1) a formal description of the distribution (its parameters, its PMF/PDF, its mean/variance); 2) any relevant history or fun miscellanea; 3) a ShinyApp that allows students to explore the behavior of the distribution under different parameter values; 4) examples of biological data that are well-described by this distribution.*

Students will be split into small groups and present their demos to each other in class in order to learn about different distributions.

- **Simulation/Inference Battles** (5%) *Students will be split into teams. 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. Teams will then exchange simulated datasets (without revealing the functions, distributions, or parameters chosen), and will work to infer the parameter values used for simulating the received data. In class, each group will give a brief presentation of the parameterization of the simulated/inferred models. The team that carried out the best inference will have their names inscribed eternally on the Wall of Glory.*

- *Lab Activities (15%):* Each week, there will be a lab activity designed to reinforce and complement the concepts covered in the lecture. As with the homework exercises and projects, the goal of these activities is to give students more opportunities for practicing new skills and cementing new knowledge. The exercises are designed to be completed and turned in during the lab (i.e., they will not add to the homework burden). All lab activities will cumulatively be worth 15% of the total grade for the course.

Attendance and Late Assignments Policy:

You are expected to attend all lectures and complete all homework assignments. Because we'll often be going over homework answers at the beginning of class, and because lecture capture videos will be posted to Canvas, late homework assignments will not be accepted. However, I understand that you may not be able to complete assignments or attend lecture due to illness or personal circumstances. Therefore, you can get full credit on the Lecture Engagement portion of your grade by answering just 75% correctly, and you can get full credit on the Homework portion of your grade by earning just 80% of total homework points. This policy is intended to also incorporate missed homeworks and lecture engagement questions that arise due to illness.

Accommodation

If you need an accommodation based on the impact of a disability, please let Dr. Bradburd know at your earliest convenience. We are happy to work with you and the Office of Services for Students with Disabilities (https://ssd.umich.edu; 734-763-3000) to help us implement personalized academic accommodations. Any information that you provide is private and confidential and will be treated as such.

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.

Student Well-being

Students may experience stressors that can impact both their academic experience and their personal well-being. If you are experiencing concerns, seeking help is a courageous thing to do for yourself and those who care about you. If the source of your stressors is academic, please contact the instructor via email so that we can find solutions together. For personal concerns, U-M offers the following resources:

- Counseling and Psychological Services (CAPS) (https://caps.umich.edu/): confidential; 734-764-8312; for after-hours urgent support, call and press 0; counseling, workshops, groups and more, counselors are embedded in some schools
- Dean of Students Office (https://deanofstudents.umich.edu/): 734-764-7420; provides support services to students and manages critical incidents impacting students and the campus community
- Ginsberg Center for Community Service Learning (https://ginsberg.umich.edu/): 734-763-3548; opportunities to engage as learners and leaders to create a better community and world
- Multi-ethnic Student Affairs (MESA) (https://mesa.umich.edu/): 734-763-9044; diversity and social justice through the lens of race and ethnicity
- Office of Student Conflict Resolution (https://oscr.umich.edu/): 734-936-6308; offers multiple pathways for resolving conflict
- Office of the Ombuds (https://ombuds.umich.edu/): 734-763-3545; students can raise questions and concerns about the functioning of the university.
- Services for Students with Disabilities (SSD) (https://ssd.umich.edu/): 734-763-3000; accommodations and access to students with disabilities
- Sexual Assault Prevention and Awareness Center (SAPAC) (https://sapac.umich.edu/) confidential; 734-764-7771 or 24-hour crisis line 734-936-3333; addresses sexual assault, intimate partner violence, sexual harassment, and stalking
- Spectrum Center (https://spectrumcenter.umich.edu/): 734-763-4186; support services for LGBTQ+ students
- Trotter Multicultural Center (https://trotter.umich.edu/): 734-763-3670; intercultural engagement and inclusive leadership education initiatives
- University Health Service (UHS) (https://uhs.umich.edu/): 734-764-8320; clinical services include nurse advice by phone, day or night
- Well-being for U-M Students website (https://wellbeing.studentlife.umich.edu/) - searchable list of many more campus resources
- Wolverine Wellness (https://uhs.umich.edu/wolverine-wellness): confidential; 734-763-1320; provides Wellness Coaching and much more

AI Policy

We are in the early days of a truly disruptive technology. Large Language Models (LLMs, e.g., ChatGPT, Claude, etc.) are transforming how we work and learn. While the impact of these tools on future employment, expertise, and daily life is not year clear, it seems likely that no one will hire you to copy and paste AI-generated output. At the same time, no one will hire you to ignore this technology. Success lies in learning how to critically evaluate and work with LLMs—to validate their output, improve your own understanding, and create high-quality results. Subject-level expertise, in conjunction with strong skills in working with AI, will be essential for the foreseeable future.

I want you to think about this class like learning to play an instrument. You're here to practice, make mistakes, and build mastery over time. Sure, you could ask an LLM to "play the piece" for you by doing your homework. But if you do that, you're not the one learning. Instead, use the LLM as a tutor and "practice partner"—a tool to get feedback, refine your technique, and expand what you're capable of doing. Doing so will allow you to get truly good not just at stats but at using the best tools available to do even better stats.

To aid you in achieving this goal, there will be plenty of opportunities for in-class, computer-free efforts to show your mastery of the subject material. I will also provide guidance the appropriate use of AI to help maximize its impact on your learning.

Class schedule (class dates in **bold** indicate an in-class presentation)

Date	Day	Title	Description
1/8	Wed	Course Intro	*Introductions, course expectations, pre-quiz*
1/13	Mon	Intro to R	*What is it and why use it?*
1/15	Wed	Data in R	*Data types/structures, R studio, getting help*
1/20	Mon	Martin Luther King Jr. Day	No class
1/22	Wed	Data Manipulation	*Indexing & subsetting*
1/27	Mon	Data Visualization	*All the plots!*
1/29	Wed	DataViz II & DEEP BREATH	*Multi-panel plots; review & solidify concepts*
2/3	Mon	Programming I	*Functional functions*
2/5	Wed	Programming II	*Lexical scoping: is it logical?*
2/10	Mon	Programming III	*Style and Speed*
2/12	Wed	Probability I	*Putting the fun in fundamentals*
2/17	Mon	Probability II	*Taking probability out for a spin*
2/19	Wed	Probability Distributions	*Math is all around us*
2/24	Mon	DEEP BREATH	*Review & solidify concepts*
2/26	**Wed**	**Get to know a distribution**	***Interactive R demos***
3/3,3/5		Spring Break	No class
3/10	Mon	DEEP BREATH	*Review & solidify concepts*
3/12	Wed	Intro to likelihood	*More coin flipping than a bored gumshoe in a noir film*
3/17	Mon	Intro to Bayesianism	*Back to Bayesics*
3/19	Wed	Intro to inference	*Markov chain Monte Carlo*
3/24	Mon	Inference II	*MCMC as a team! By our powers combined!*
3/26	Wed	Intro to model building	*Deterministic functions*
3/31	Mon	Model Building II	*Nonlinearity, link functions, GLMs*
4/2	Wed	DEEP BREATH	*Review & solidify concepts*
4/7	Mon	Inference III	*Putting all the pieces together*
4/9	Wed	Simulation/Inference Battles	*In-class group project work time*
4/14	Mon	Simulation/Inference Battles	*In-class group project work time*
4/16	Wed	Simulation/Inference Battles	*In-class group project work time*
4/21	Mon	Simulation/Inference Battles	*In-class group project work time*
TBD		**Battle Presentations**	***ETERNAL GLORY TO THE VICTORS***

Letter Grades

Letter grades will be determined based on the following grading scale according to the percentage of total points earned:

Letter Grade	Percentage total points earned
A+	97 - 100
A	93 - 96.99
A-	90 - 92.99
B+	87 - 89.99
B	83 - 86.99
B-	80 - 82.99
C+	77 - 79.99
C	73 - 76.99
C-	70 - 72.99
D+	67 - 69.99
D	63 - 66.99
D-	60 - 62.99
E	< 60

Course materials and books

For this course, students will need a laptop computer capable of running R and R Studio and accessing the internet. Most decent laptops with >8GB RAM will be sufficient. Macs and Linux machines will be easier than PCs, but both are supported in this course. For any student who needs it, the university has a "laptop loaner" program, but requests are on a first-come, first-serve basis, so students interested in this service should contact the program early. Please talk to the instructor 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.

Helpful References

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.