OCEANS 140: Statistical Modeling
Syllabus, Winter 2024

Course description

Intermediate course focused on statistical modeling in a Bayesian framework, with applications in the biological and environmental sciences. Topics will include probability, generalized linear models, causal inference, and model comparison. We will take a hands-on, computational approach (R, Stan) to gain intuition so that students can later design their own inferential models. Prerequisites: Introductory statistics course, such as STATS 60, STATS 110, STATS 141 and some experience with scientific computing in R, such as STATS 32.

We will meet for 1.5 hours, twice a week. Classroom activities will be a mix of lecture, discussion, and collaborative problem solving. Outside of class, students will read selected chapters and solve problem sets. All participants in the course will abide by the Code of Conduct, described below.

Learning objectives

This course is a practical introduction to statistical modeling in a Bayesian framework. By the end of the course, you will be able to:

- Understand the basics of probability necessary for statistical inference
- Describe the components of Bayes theorem and their relevance for model specification
- Formulate and apply a Bayesian statistical model to your data
- Confidently learn new approaches to data analysis as you progress in your research

Instructor

Dr. Robin Elahi (elahi@stanford.edu)

Course website

All materials will be accessed through: https://elahi.github.io/stats2_site/
Assignments will be submitted through: https://canvas.stanford.edu/

Times and location

10:30am-12:00pm; T, Th in Agassiz 12

Office hours

TBD; you can also email to schedule an appointment

Course structure

We will meet for 1.5 hours, twice a week. Prior to class, students will complete a reading or homework assignment. In class, we will review the reading or assignment together, and then launch into the day's lecture or activity. **Please bring your laptop to all classroom sessions.**

Course texts
We will use the following texts, available online:

Statistical Rethinking, 2nd edition (McElreath 2020):
http://xcelab.net/rm/statistical-rethinking/

Book can be accessed here via Stanford's library:
https://searchworks.stanford.edu/view/13631911

McElreath's lectures and slides from his most recent course here:
https://github.com/rmcelreath/stat_rethinking_2024

Bayesian Models: a Statistical Primer for Ecologists (Hobbs and Hooten 2015):
https://press.princeton.edu/books/hardcover/9780691159287/bayesian-models

Book can be accessed here via Stanford's library:
https://searchworks.stanford.edu/view/13753652

Evaluation and grading
Your grade is based entirely on the submission of homeworks and labs, which will be submitted by 8am PST on the due date. They will be graded for completion only – you do not have to get the correct answers, but you will demonstrate that you made an honest effort (e.g., with an explanation of your answer; code). If you are taking the course for a letter grade, you must complete the project. If you are taking the course pass/fail, you will submit homework but not project assignments. If you are auditing the course, you need not submit any assignments.

Regardless of your grading status, it is important that you complete the readings and homework in a timely manner because we are building a community of statistical learners. You are expected to be an active colleague during our class meetings! I anticipate, many questions, failures, and growth spurts as we move through this difficult but intellectually satisfying material together. You should expect to allocate about 3-5 hours per week outside of class for the reading and homework.

Before class
Before the first class please download and install the latest versions of R (https://www.r-project.org/) and RStudio (https://www.rstudio.com/products/rstudio/).

To fit our models, we will be using Stan from within R (`rstan`), as well as several accessory packages. For installation, please refer to the following websites:

- https://mc-stan.org/users/interfaces/rstan
- https://github.com/rmcelreath/rethinking

If you have never used R, it is imperative that you get up to speed with the basics of R programming before the first day of class. I recommend working through this set of Data Carpentry lessons. It will take about two full days to complete, if you are starting from scratch:

- https://datacarpentry.org/R-ecology-lesson/index.html

Resources for learning R abound on the web, but here are some other starting points:

- https://education.rstudio.com/learn/beginner/
- https://ourcodingclub.github.io/course.html

Tips for success

- Read the text and complete the readings *before* class
- Be an active participant in class
- Ask questions – of me and your classmates. We will all benefit!
- Do not procrastinate. Don't let a week go by with unanswered questions as it will make the following week's material more difficult to follow

Additional resources

Bolker, B.M., 2008. *Ecological Models and Data in R*. Princeton University Press.

Devlin, D., Guo, J., Kunin, D., and Xiang, D. 2018. *Seeing Theory*. https://seeing-theory.brown.edu/index.html

Edge, M.D., 2019. *Statistical Thinking From Scratch: a Primer for Scientists*. Oxford University Press.

Johnson, A.A., Ott, M.Q. and Dogucu, M., 2022. *Bayes Rules!: An Introduction to Applied Bayesian Modeling*. CRC Press. https://www.bayesrulesbook.com/index.html

Kruschke, J., 2014. *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan*.

Schedule

Week	Date	Topic	Reading
1	Tue., Jan. 7	Intro to course; Probability I	HH and SR prefaces
1	Thu., Jan. 9	Probability II	HH3.1-3.3
2	Tue., Jan. 14	Probability distributions	HH3.4
2	Thu., Jan. 16	Likelihood	HH4
3	Tue., Jan. 21	Intro to Bayesian statistical modeling	HH1, SR1
3	Thu., Jan. 23	Components of a Bayesian model	SR2
4	Tue., Jan. 28	Approximating the posterior distribution	SR3
4	Thu., Jan. 30	Linear models	SR4.1-4.4
5	Tue., Feb. 4	Polynomial regression and splines	SR4.5
5	Thu., Feb. 6	Multiple regression and DAGs	SR5.1
6	Tue., Feb. 11	Multiple regression and DAGs	SR5.2-5.4
6	Thu., Feb. 13	(More) causal inference	SR6
7	Tue., Feb. 18	Overfitting, information entropy	SR7.1-7.2
7	Thu., Feb. 20	Regularization, model comparison	SR7.3-7.6
8	Tue., Feb. 25	Building interactions	SR8.1-8.2
8	Thu., Feb. 27	Continuous interactions	SR8.3-8.4
9	Tue., Mar. 4	Markov chain Monte Carlo	SR9.1-9.2
9	Thu., Mar. 6	Hamiltonian MCMC	SR9.3-9.5
10	Tue., Mar. 11	Generalized linear models	SR10.2
10	Thu., Mar. 13	Generalized linear models	SR11.1-11.2

Stanford as an institution is committed to the highest quality education, and as your teaching team, our first priority is to uphold your educational experience. To that end we are committed to following the syllabus as written here, including through short or long-term disruptions, such as public health emergencies, natural disasters, or protests and demonstrations. However, there may be extenuating circumstances that necessitate some changes. Should adjustments be necessary we will communicate clearly and promptly to ensure you understand the expectations and are positioned for successful learning.

Plagiarism, dishonesty, and academic misconduct
It is expected that Stanford's Honor Code will be followed in all matters relating to this course. You are encouraged to meet and exchange ideas with your classmates while studying and working on homework assignments, but you are individually responsible for your own work and for understanding the material. Passing anyone else's scholarly work, which can include: written material, computer code, exam answers, graphics or other images, and even ideas as your own, without proper attribution, is considered academic misconduct. Plagiarism, cheating, and other misconduct, including bullying, discrimination, and harassment, are serious violations of the University's *Fundamental Standard* and *Honor Code*:
https://communitystandards.stanford.edu/policies-and-guidance

Affordability of course materials
Stanford University and its instructors are committed to ensuring that all courses are financially accessible to all students. If you are an undergraduate who needs assistance with the cost of course textbooks, supplies, materials and/or fees, you are welcome to approach us directly. If would prefer not to approach us directly, please note that you can ask the Diversity & First-Gen Office for assistance by completing their questionnaire on course textbooks & supplies: http://tinyurl.com/jpqbarn or by contacting Joseph Brown, the Associate Director of the Diversity and First-Gen Office (jlbrown@stanford.edu; Old Union Room 207). Dr. Brown is available to connect you with resources and support while ensuring your privacy.

Students with documented disabilities
Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty. Unless the student has a temporary disability, Accommodation letters are issued for the entire academic year. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 723-1066, URL: https://oae.stanford.edu/).