EEMB 174/274: Advanced Biostatistics and Bayesian Analysis

Winter 2022 Lecture: M-W 9:30-10:45 Girvetz 1115 Lab: Th 9-11:50 Phelps 1530; 2-4:50 Phelps 1526

Course Instructor:

Stephen Proulx sproul@ucsb.edu

Teaching Assistant:

Taom Sakal tsakal@ucsb.edu

Required Text

Richard MaElreath Statistical Rethinking. The book can be purchased through the bookstore or online in various formats. It is also available through the UCSB library: library e-copy.

Using the virtual machine to run R and Stan

All of the course work can be completed using a "virtual machine" that is provided for you to use. This virtual machine has the latest version of R and RStudio installed, as well as specific packages that we will use throughout the course. This allows each of you to access the precise same computational environment and will help to eliminate difficulties with installing software on each of your individual computers. An advantage of using the VM is that your work is saved remotely and will be preserved even if you leave the session and login later from a different computer or location. However, you will need internet access the entire time you are using the VM.

Setting up your own local environment

You may also choose to install the software tools on your own computer. Because each installation is unique I cannot provide detailed instructions on how to do this. However, the tools that you would need are

- R
- RStudio
- RStan
- Tidyverse package for R
- Rethinking package for R

Accessing the Virtual Machine

The VM can be accessed from a web browser by using the address: https://eemb-174-274.lsit.ucsb.edu/ You will be prompted to login with your UCSB credentials. This will launch RStudio which will have 4 panels visible. These panels are

- Top left: R code panel
- Top right: R environment panel
- Bottom left: Console and Terminal panel
- Bottom right: File system panel

Setting up your own directory

The bottom right panel will show the file system for your personal copy of the VM. This is where the course materials will show up when you add them using `git` and this is also where you will save your own work.

It is a good idea to create a directory to save your own work in. In the bottom right panel, click "New Folder" (top left of the panel, under "Files") and create a new folder. You will be prompted to enter a name for the folder, use your own name as part of this name.

Getting the course materials

Course materials, including homework and lab assignments, will be made available using github. This will allow us to update the materials throughout the quarter and for you to add them to your VM directory. We will also include various course resources, like tutorials and cheatsheets.

To get the course materials for the first time you will need to use the VM "Terminal". Access this in the lower left panel by clicking the tab "Terminal". This provides a command line prompt. You will use this tab to directly typ command line instructions.

To "clone" the git repository for the first time do:
`git clone https://github.com/srproulx/EEMB-174-274-2023`
This creates a directory named `EEMB-174-274-2023` with subdirectories containing the course materials.

To refresh with new material from the terminal you need to first "change directory":
`cd ~/EEMB-174-274-2023/`
and then "pull" the updated material:
`git pull https://github.com/srproulx/EEMB-174-274-2023`

HIGHLY RECOMENDED TO BOOKMARK THIS LNK: You will also be able to refresh the course material in one step using this link: https://bit.ly/EEMB174.

Coursework and grading

Coursework will include homework, lab reports, quizzes, and two assessment exercises (in place of midterm and final exams). Roughly, the breakdown is:

Category	Percentage
Homework	25%
Lab Reports	45%
Quizzes	15%
Midterm	5%
Final	10%

Homework and labs are numbered by the week in which they are assigned. Homework will not be assigned

in every week.

Gauchospace quizzes

These quizzes will be designed for you to self-check your progress and for us to get feedback on which concepts are more difficult and need more explanation. Some of the quizzes will be multiple choice and allow you to re-take the quiz until you get the answers right.

You should treat the quiz taking exercise as a learning opportunity. They are designed to probe your detailed understanding of the material. Make sure that you really understand why the wrong answers are wrong and why the correct answers are correct. Come to class after the quizzes with clarifying questions to ask.

Homework

Most weeks will have a homework assignment using problems from the textbook. These homeworks will be graded based on attempted completion, not correctness of the answers. To get credit you will have to explain your answer or describe, as best you can, where you got lost and were unable to finish the assignment. (But if you do get stuck we encourage you to work with your classmates to get unstuck.) Assignments will be due on Monday mornings at class time. If there is some reason that you are unable to complete the assignment by the due date, please contact us and let us know so that we can work with you to ensure that you can stay up to speed in the course.

Lab Assignments

Labs are graded for completion and accuracy. You may work with a partner, but the text/writing answers you give should be your own. You will write each lab assignment using R Markdown and you will submit both the R Markdown file and generated PDF as a single zipped file.

Schedule of the quarter

A graphical chart of the schedule for the quarter can be found in the github directory and on Gauchospace. This chart shows when you should be doing readings, watching videos, working on homework, doing labs, and when assessment exercises will be happening.

Exams

Date	Exam
2/9/23	Midterm Exam
3/16/23	Final Exam

Weekly topics/reading

Date	Subject	Reading
1/9/23	Intro to course, probabilitstic thinking, RStudio, Rmarkdown	Guide to RMarkdown
1/11/23	Bayesian reasoning	Chapter 1

Date	Subject	Reading
1/16/23	No Class, MLK Day	
1/18/23	Writing data stories and Bayesian updating	Chapter 2
1/19/23	Simulating data and describing distributions of outcomes	Chapter 3
1/23/23	in class lab work- characterizing distributions	
1/25/23	Fitting means	Chapter 4: 71-114
1/30/23	Linear regression	Chapter 4: 71-114
2/1/23	Multiple regression, spurious associations	Chapter 5: pages 123-150
2/6/23	Dealing with confounds	Chapter 6: 161-183
2/8/23	Post-treatment bias	
2/9/23	Midterm assessment exercise	
2/13/23	Information criteria for model comparison	Chapter 7: pages 191-214
2/15/23	Information criteria for model comparison	Chapter 7: pages 214-235
2/20/23	No Class	
2/22/23	Models with interactions	Chapter 8
2/27/23	MCMC methods and diagnostics	Chapter 9
3/1/23	Generalized Linear Models	Chapter 10
3/6/23	Binomial Regression	Chapter 11 (11.3 on is optional)
3/8/23	Multilevel models	Chapter 13 (to 13.4)
3/13/23	More Multilevel models	Chapter 13 (to 13.4)
3/15/23	Pooling models with covariance	Chapter 14: pages 435-446

Weekly video lectures

You should watch the recorded lectures by Richard McElreath before the date listed in the table. For some lectures, particular portions of the video are more relevant and this is printed in the "Notes" column.

Date	Title	Notes
1/5/22	The Golem of Prague (Lec 1)	Through 26:00
1/11/23	Bayesian Inference (Lec 2)	Through 21:00
1/18/23	Bayesian Inference (Lec 2)	21:00 - 57:00
1/12/22	Bayesian Inference (Lec 2)	57:00 - 1:12:00
1/25/23	Geocentric Models (Lec 3)	Through 37:30
2/1/23	Spurious Waffles (2019)	Through 38
2/6/23	Haunted DAG (2019)	Through 17:30
2/13/23	Overfitting (Lec 7)	Through 47:30
2/22/23	Conditional Manattees	Up to 36:00
2/23/23	MCMC Methods (lec 8)	Entire lecture, with a soft gaze
3/1/23	General Linear Models (Lec 9)	Through 30:30
3/6/23	Binomial Regression (Lec 9)	47:00-57:00
3/8/23	Multilevel Models (lec 12)	Through 57:00
3/8/23	Multi-Multilevel Models (lec 13)	Through 38:00

Maximizing your success in this course under our difficult circumstances

We are all experiencing difficult and unusual times right now. I understand that each of you has your own unique circumstance and that this will affect how you participate in the course this quarter. There will be due dates for homework and lab assignments and sticking to the due dates will help you to keep up to speed in the course and not fall behind. That said, we will not be taking off points for late homework or lab assignments. The quizzes will have cutoff times, again designed to ensure that you are staying up to date with the course.

Please reach out to either Professor Proulx (sproul@ucsb.edu) or Taom Sakal (tsakal@ucsb.edu) if you need to let us know about changes in your circumstances, have concerns about your ability to stay up to speed in the course, or have any other issues we can help with.