SDS 383D: Statistical Modeling II
MW 3:00PM — 4:30PM, Room: Zoom

Instructor: Dr. Antonio Linero, WEL5.244

Office Hours: 1:00pm — 2:00PM TTh, or by appointment.

Contact: antonio.linero@austin.utexas.edu

Prerequisites: This course primarily concerns probabilistic modeling using generalized linear models and hierarchical Bayesian modeling, although there will be other topics covered. This course is intended for graduate students in the Statistics and Data Sciences Department or any students from other departments who are interested in developing/using probabilistic models. In terms of substantive prerequisites, I assume that you are comfortable with the following topics:

- linear algebra;
- how to program in a language like `R` or `Python`;
- multivariate calculus;
- probability (although measure theory isn't necessary);
- basic inferential statistics at the level of Casella & Berger;
- linear regression.

If you have any doubt about your preparation for this course, or have any questions about whether this course is right for you, feel free to chat with me.

Course Website: Notes for this course are available at:

`https://github.com/theodds/2023-Statistical-Modelling-II`

Syllabus and zoom meeting link for office hours are available at the course site on `canvas.utexas.edu`.

Course Structure: This course is a blend between a traditional lecture-based course and a flipped classroom. Some of the time is spent on lectures in class. But a lot of the other class time will be student led. You will work on the exercises assigned on the course website and when you come to class you will share what you have done, and benefit from understanding what others have done. We will end up covering less than in an exclusively lecture-based course. But what you learn, you will learn deeply.

Textbook: Relevant readings will be given throughout the course. There are no formally required textbooks, but here are three recommended references that should be easily founder online:

- Data Analysis Using Regression and Multilevel/Hierarchical Models by Gelman and Hill. An e-book version is available through the UT Library website.
- Generalized Linear Models by McCullagh and Nelder.
- All of Statistics by Larry Wasserman.

Software: The examples in this course will use the `R` programming language, available at `www.r-project.org`. See also `www.rstudio.com` for a nice development environment. You will be asked to submit `R`-code as part of your homework solutions.

Course Objectives: The purpose of this course is to give students the mathematical tools and intuition required to develop and deploy sophisticated probabilistic models. Major topics will include:

- classical Frequentist and Bayesian approaches to inference;
- fundamentals of generalized linear models;
- multivariate distributions, such as the multivariate Gaussian and Dirichlet;
- nonparametric regression techniques (smoothing, additive models, and Gaussian processes);
- hierarchical linear/generalized linear models.

Grading: Your grade consists of three pieces: 40% exercises, 40% final project, and 20% participation.

Homework will be due on the first class day of each month. To lighten the burden, we will divide into groups of three each month (groups will not be allowed to overlap in subsequent months). **Groups will be randomly selected to present exercises for various problems, with supporting code included, throughout the semester.** At the end of each lecture, I will give a list of homework problems that I may ask you to present a solution to during the next lecture, but it is expected that you and your group will do all of the exercises in the notes.

I will grade homework submissions in the following way: each month I will select ten problems at random to grade. All parts of all problems will be weighted equally so that, for example, if I graded Problem 1, Problem 2a, and Problem 2b, then each of these would account for 1/3rd of the grade for the month. Each part will be graded either 0/2, 1/2, or 2/2.

As far as how submissions should look:

1. For mathematics questions, I'm fine with you turning in hand-written solutions, provided that you answer the questions using complete sentences. Answers should not consist of just long strings of calculations with no comments; you need to explain what you are doing.
2. For data analysis or coding questions, I expect results to be typed up with code given in-line and explained. It should be structured more-or-less like how I have things written up in the notes (i.e., code and discussion given together). The easiest way to do this would be to submit an R Markdown or Quarto document, or create a Jupyter notebook. If using one of

these frameworks, you should submit both the compiled document as a pdf as well as the raw notebook/.rmd/.qmd document.

For each homework submission, you will give a grade to each of your group members (including yourself) that reflects the grade that you would assign that group member based on their contribution to the homework assignments and in-class presentations, as well as any notes that you have about how the month went. **This, in addition to participation in class, will be used to assign participation credit.** This, in combination with the homework turned in, will be used to assign a grade for the homework each month. *Note:* I will still have final say on the grade each student gets on homework, and may overrule any good/bad evaluations if I feel (and have evidence supporting) that the evaluations are not representative of a student's contributions.

Final Project: Pick some relevant topic that interests you. Clear it with me ahead of time, and then do it. Basically, I trust you to choose something that will optimize your own learning experience, and that will dovetail with your research and educational goals. It certainly can overlap with your own research. Examples:

1. Analyze a data set from your own research, using techniques from class or closely related techniques.
2. Invent a new technique and show how awesome it is.
3. Prove something interesting about a procedure or algorithm related to what we're studying (admittedly unlikely, but certainly possible!).
4. Read a paper, or a group of related papers, that expands on some topic we've covered in class. Implement the method(s) and benchmark it (them) against something else.

Final projects are due on the day of the university-scheduled exam: Friday, May 4th, 2024. **Note: you should feel free to work either solo or in pairs for the projects.**

You should prepare a report up to 8 pages in length using the NeurIPS style files and guidelines, available on the course website. Reports are due on May 4th.

Students with Disabilities: Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259, `http://www.utexas.edu/diversity/ddce/ssd/`.

Religious Holy Days: By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

Scholastic Honesty: We expect students to behave with integrity. Students found cheating on exams or homework will receive a score of zero for that exam or assignment, and may be subject to additional disciplinary action. For more information on the University of Texas scholastic dishonesty policy, see the 2006-2007 General Information Catalog, Appendix C.

Campus Safety: Please note the following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, `http://www.utexas.edu/safety`:

- Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.
- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation should inform the instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors.
- Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- Behavior Concerns Advice Line (BCAL): 512-232-5050
- Further information regarding emergency evacuation routes and emergency procedures can be found at: `http://www.utexas.edu/emergency`.

Texas Senate Bill 17 and DEI: Texas Senate Bill 17, the recent law that outlaws diversity, equity, and inclusion programs at public colleges and universities in Texas, does not in any way affect content, instruction or discussion in a course at public colleges and universities in Texas. Expectations and academic freedom for teaching and class discussion have not been altered post-SB 17, and students should not feel the need to censor their speech pertaining to topics including race and racism, structural inequality, LGBTQ+ issues, or diversity, equity, and inclusion.