STA2201H Methods of Applied Statistics II

Winter 2024

Time and place:

- Wednesdays 10-1pm, starting 10 January
- Location: LA248

Instructor:

- Monica Alexander, monica.alexander@utoronto.ca

TA:

- Michael Chong, myc.chong@mail.utoronto.ca

Office hours:

- Wednesdays 2-3pm 700 University Avenue Room 9135

Course website:

Quercus, and https://github.com/MJAlexander/applied-stats-2024

1 Course description

This course covers a range of statistical methods, covering the application and interpretation of models on a range of different datasets. Topics will include generalized linear models, Bayesian inference, generalized linear mixed models, generalized additive models involving nonparametric smoothing, model evaluation and selection. We will also cover some core statistical computing techniques.

A large focus of the outcomes on this course will also be on reproducible research, dealing with 'real' data, identifying and dealing with data and modeling issues, and model interpretation and communication.

2 Textbooks

There is no required textbook for this course. The following texts may be useful as reference:

- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. Bayesian data analysis. Chapman and Hall/CRC, 2013.
- Gelman, Andrew, and Jennifer Hill. Data analysis using regression and multilevel/hierarchical models. Cambridge university press, 2006.
- McElreath, Richard. Statistical Rethinking. CRC Press, 2020.
- Wickham, Hadley, and Garrett Grolemund. R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc., 2016.

3 Computing

Throughout the course we will be using R in all examples, labs and homework assignments. Exams will also require interpretation of R output. In particular, you will be learning and expected to code in the `tidyverse` style, and all labs and assignments will be written using Quarto (or R Markdown).

You will need to have R and RStudio installed on your computer:

- Download R here: https://www.r-project.org/
- Download RStudio (free version) here: https://posit.co/download/rstudio-desktop/
- Make sure you can render/knit a Quarto or R Markdown document (this may involve installying `tinytex`)

You will also need to have a GitHub account: https://github.com/

4 Assessment

Broadly, there are four components to the assessment for this course:

- Lab exercises (8 in total), 2.5% each
- Three assignments worth 10% each
- Mid-term, 15%
- Research project, 35%

All labs and assignments, and the written component of the research project will be completed in Quarto or R Markdown. The '.qmd' or '.Rmd', resulting pdf and any relevant files should be submitted, such that the file is reproducible and compiles to a pdf with no errors.

4.1 Lab Exercises

- Short hands-on exercises in R
- To be handed in via GitHub
- Will be due by 9am on the following Monday

4.2 Assignments

- A mix of short answer questions and applying methods to a dataset
- To be handed in via Quercus

4.3 Mid-term

- Short answer questions
- Interpretation of R output
- In person

4.4 Research project

For the research project you will be required to investigate a research question of interest using a dataset of your choice and applying methods learned during the semester. The project consists of three main components:

- **Research proposal (7.5%)** describing your research question of interest, why it is interesting, any hypotheses you may have, the dataset and variables you are using, and what methods you intend to use. Also include some initial EDA. 1-2 pages excluding graphs and tables, due at the same time as Assignment 3. Submitted via Quercus.
- **Research paper (20%)** on your topic. It is expected that the report be written in the style of an academic paper, including the following sections: Abstract, Introduction, Data, Methods, Results, Discussion, and Conclusion. Must be written in Quarto or R Markdown, and all files to reproduce the paper and analysis must be submitted via Quercus. Due at the end of semester.
- **Presentation (7.5%)** on you research project, including question, data, methods and results. Around 5 minutes. It is expected you will present with the aid of slides. To take place in Week 12 in class.

5 Course policies

5.1 Communication

The best way to ask questions about course material or assignments is in person during the lecture, lab or your instructor's office hours. The following are guidelines for email communication: please make sure that you have a legitimate need before you write and that you cannot resolve your question during the lecture, lab or office hours; email messages should state the course number and the purpose of the email clearly in the subject line. Please be respectful and clear when emailing.

5.2 Accessibility Services

It is the University of Toronto's goal to create a community that is inclusive of all persons and treats all members of the community in an equitable manner. In creating such a community, the University aims to foster a climate of understanding and mutual respect for the dignity and worth of all persons. Please see the University of Toronto Governing Council "Statement of Commitment Regarding Persons with Disabilities" at http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppnov012004.pdf. In working toward this goal, the University will strive to provide support for, and facilitate the accommodation of individuals with disabilities so that all may share the same level of access to opportunities, participate in the full range of activities that the University offers, and achieve their full potential as members of the University community. We take seriously our obligation to make this course as welcoming and accessible as feasible for students with diverse needs. We also understand that disabilities can change over time and will do our best to accommodate you. Students seeking support must have an intake interview with a disability advisor to discuss their individual needs. In many instances it is easier to arrange certain accommodations with more advance notice, so we strongly encourage you to act as quickly as possible. To schedule a registration appointment with a disability advisor, please visit Accessibility Services at http://www.studentlife.utoronto.ca/as, call at 416-978-8060, or email at: accessibility.services@utoronto.ca. The office is located at 455 Spadina Avenue, 4th Floor, Suite 400. Additional student resources for distressed or emergency situations can be located at distressedstudent.utoronto.ca; Health & Wellness Centre, 416-978-8030, http://www.studentlife.utoronto.ca/hwc, or Student Crisis Response, 416-946-7111.

5.3 Academic Integrity Clause

Copying, plagiarizing, falsifying medical certificates, or other forms of academic misconduct will not be tolerated. Any student caught engaging in such activities will be referred to the Dean's office for adjudication. Any student abetting or otherwise assisting in such misconduct will also be subject to academic penalties. Students are expected to cite sources in all written work and presentations. See this link for tips for how to use sources well: (http://www.writing.utoronto.ca/advice/using-sources/how-not-to-plagiarize). According to Section B.I.1.(e) of the Code of Behaviour on Academic Matters it is an offence "to submit, without the knowledge and approval of the instructor to whom it is submitted, any academic work for which credit has previously been obtained or is being sought in another course or program of study in the University or elsewhere." By enrolling in this course, you agree to abide by the university's rules regarding academic conduct, as outlined in the Calendar. You are expected to be familiar with the Code of Behaviour on Academic Matters (http://www.artsci.utoronto.ca/osai/The-rules/code/the-code-of-behaviour-on-academic-matters) and Code of Student Conduct (http://www.viceprovoststudents.utoronto.ca/publicationsandpolicies/codeofstudentconduct.htm) which spell out your rights, your duties and provide all the details on grading regulations and academic offences at the University of Toronto.

5.4 Equity and Diversity

The University of Toronto is committed to equity and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect. As a course instructor, I will neither condone nor tolerate behaviour that undermines the dignity or self-esteem of any individual in this course and wish to be alerted to any attempt to create an intimidating or hostile environment. It is our collective responsibility to create a space that is inclusive and welcomes discussion. Discrimination, harassment and hate speech will not be tolerated. Additional information and reports on Equity and Diversity at the University of Toronto is available at http://equity.hrandequity.utoronto.ca.

5.5 Use of Generative AI in Assignments

In general, students are discouraged from using generative AI tools, such as ChatGPT4 in this course. Specific course policies around the use of generative AI tools are as follows:

- R code: Students may wish to use generative AI tools to aid in initial development and writing of R code to answer assignment questions and carry out the research project analysis. If this is the case, the use of such tools should be explicitly acknowledged in the submitted work, and the relevant prompts should be included as an appendix to the assignment. In addition, all code must be thoroughly commented or described in the submitted work. Failure to do so may result in penalties.

- Written work: Using generative AI tools to generate written answers to assignment questions of text contained in the final research project is prohibited in this course. Representing as one's own an idea, or expression of an idea, that was AI-generated may be considered an academic offense in this course.

This course policy is designed to promote your learning and intellectual development and to help you reach course learning outcomes.

6 Course outline

Planned content by week (note: subject to change based on learning conditions)

- Week 1 (10/1/24): Introduction

- Week 2 (17/1/24): Generalized linear models recap
- Week 3 (24/1/24): Bayesian methods and inference
- Week 4 (31/1/24): Bayesian methods and inference II (**Assignment 1 due**)
- Week 5 (7/2/24): Estimating Bayesian models
- Week 6 (14/2/24): Visualizing the Bayesian workflow and model checks

READING WEEK (21/2/24)

- Week 7 (28/2/24): MID-TERM (**Assignment 2 due**)
- Week 8 (6/3/24): Multilevel models
- Week 9 (13/3/24): Non-linear models I
- Week 10 (20/3/24): Non-linear models II (**Assignment 3 and research proposal due**)
- Week 11 (27/3/24): Dealing with correlation and other extensions
- Week 12 (3/4/24): **Student presentations**
- **Research paper** due 12/4/24 (tentative: I'll try and make this as late as possible)