STA2201H Methods of Applied Statistics II

Monica Alexander

Time and place:

- Wednesdays 2-5pm, starting 13 January
- Online synchronous via BBCollaborate

My contact details:

- monica.alexander@utoronto.ca. I do not check email after 5pm or on weekends.

Office hours:

- TBD, will be by appointment via Calendly

Course website:

https://github.com/MJAlexander/applied-stats-2021

Course description

This course covers a range of statistical methods, covering the theory, 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, identifying and dealing with data and modeling issues, and model interpretation and communication.

Textbooks

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

- Dobson, A. J., & Barnett, A. G. An introduction to generalized linear models. Chapman and Hall/CRC, 2008.
- 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.
- Hastie, T., Tibshirani, R., & Friedman, J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, 2009.
- Wickham, Hadley, and Garrett Grolemund. R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc., 2016.

Mode of Delivery

This course will be held entirely online. Lectures and office hours will occur synchronously through BBCollaborate. Given the current circumstances, I understand it may be difficult to attend class for a number of reasons, and as such all lectures will be recorded so that they can be accessed at a later time.

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 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://www.rstudio.com/products/rstudio/download/

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

Assessment

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

- Lab exercises (10 in total), 2% per week
- Two assignments worth 15% each
- Mid term exam (online), 15%
- Research project, 35%

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

Lab Exercises

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

Assignments

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

Mid-term

- A mix of multiple choice and short answer questions
- Covers all material in first 5 weeks
- Will need to be able to write R 'pseudo code' and interpret R code and output
- Held online via Quercus

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 2. 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 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, online via BBCollaborate.

Course outline

Planned content by week (note: subject to change)

- Week 1 (13/1/21): Introduction
- Week 2 (20/1/21): Generalized linear models recap
- Week 3 (27/1/21): Survival analysis
- Week 4 (3/2/21): Bayesian methods and inference (**Assignment 1 due**)
- Week 5 (10/2/21): Bayesian methods and inference II

READING WEEK (17/2/21)

- Week 6 (24/2/21): **Mid-term**
- Week 7 (3/3/21): Visualizing the Bayesian workflow and model checks
- Week 8 (10/3/21): Multilevel models I
- Week 9 (17/3/21): Multilevel models I
- Week 10 (24/3/21): Non-linear models I (**Assignment 2 and research proposal due**)
- Week 11 (31/3/21): Non-linear models II and extensions
- Week 12 (7/4/21): Recap and **Student presentations**
- **Research paper** due 16/4/21