STA2201H Methods of Applied Statistics II

Monica Alexander

Time and place:

- Wednesdays 2-5pm, starting 8 January
- Bissell 325

My contact details:

- monica.alexander@utoronto.ca. I do not check email after 5pm or on weekends.
- Room 380, Sociology (725 Spadina, level 3)
- Room 6010, Statistics (100 St. George, level 6)

Course website:

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

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.

Computing

Throughout the course we will be using R in all examples, labs and homework assignments. Exams will also require interpretation of R output. Those not familiar with are encouraged to do so in the first few weeks of the course.

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

- Lab exercises, 1% per week
- Four assignments worth 12.5% each
- Final exam worth 40%

Short lab exercises will need to be submitted via GitHub. If possible, you will need to bring a laptop to class.

The assignments will require writing and submitted R code. It is expected that all assignments are completed in RMarkdown. 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.

Textbooks

There is no one 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.

Course outline

Planned content by week (note: subject to change based on time/priorities)

- Week 1 (8/1/20): Introduction
- Week 2 (15/1/20): Generalized linear models I
- Week 3 (22/1/20): Generalized linear models II
- Week 4 (29/1/20): Bayesian Inference (**Assignment 1 due**)
- Week 5 (5/2/20): Multilevel models I
- Week 6 (12/2/20): Multilevel models II (**Assignment 2 due**)
- Week 7 (26/2/20): Temporal models
- Week 8 (4/3/20): Non-linear and non-parametric models
- Week 9 (11/3/20): Sampling, measurement issues and other problems (**Assignment 3 due**)
- Week 10 (18/3/20): Sampling, measurement issues and other problems II
- Week 11 (25/3/20): Model diagnostics, checking and selection
- Week 12 (1/4/20): Recap (**Assignment 4 due**)