

STAT 447B Topics in Statistics: Statistical computing Winter term 2021 (January to April)

Lecture times: Tue Thu: 9:30-10:50
Instructor: Dr. H. Joe

Web site: <http://uglab.stat.ubc.ca/~hjoe/stat447/>

Pre-requisites: Stat 306 or comparable course with multiple and binary regression, Stat 302 or introductory probability, basic familiarity with R for plots and regression, familiarity with adding libraries in R.

Statistical computing topics covered include: R statistical software, reproducible code, pseudo-code, writing functions, documentation, validation, debugging, numerical methods, profiling. A team project is the major activity.

Prerequisites: Stat 306 (or course with multiple regression), If you have taken Stat 406, then you have more options for the team project.

In the academic calendar, Stat 305 is the pre-requisite for a generic Stat 447 Topics in Statistics course. For this specific topics course, Stat 306 is the most necessary pre-requisite, and Stat 305 is helpful but not necessary.

The goal of this course is for you to enjoy doing a project making use of theory and methods learned in previous statistics courses. Code used for the project should be written in a modular, reproducible style and use of functions for algorithms and repetitive steps.

Course learning objectives. The emphasis is on development of skills: report writing, presentation, code writing, data sense, data analysis, statistical inference. These skills will help in post-graduation job hunting in data science and other career paths.

Course material will be relevant to the data sets and projects that the class is working on. For a project, possibilities are something from www.kaggle.com (if interested in data scientist as a career) or something involving COVID19 data (if interested in biostatistics, career in health sciences or pharmaceutical industry). Preferrably, choose a topic that is of special interest (minor or theme area) and then find teammates via piazza with similar interests.

Note that you would have to register with kaggle for data sets from current or expired data competitions.

References

1. Data Wrangling with R, by B C Boehmke, Springer. Available electronically at www.library.ubc.ca
2. Reproducible Code, published by British Ecological Society. [BES-Guide-Reproducible-Code-2019.pdf](#)

Assessment: Team project, regular homework assignments (upload at canvas.ubc.ca), midterm exam, final exam. Quality of code, validity of statistical methods, and proper statistical interpretations are assessed in the homework, project and exams.

For non-WebWork-based homework, code must be in R.

The project can be done in R or python (or mixture).

Team project: Topic of project to be chosen by week 3 of the term. Find something of interest to yourself and find 2 teammates with similar interest. Monthly interim reports are to be presented and submitted, and feedback will be given.

- Decide on team and project as soon as possible by end of January. Post your team members and project to the Stat 447 canvas agenda page. (This is to avoid two teams with the same kaggle project).
- There are three presentations for each team of 3 members. When ready to present, list the team member next to a date in February to April in the Stat 447 canvas agenda page.
 1. Project presentation 1: project description and goal, summary of relevant variables, show plots of how individual explanatory variables are related to the variable to be predicted.
 2. Project presentation 2: summarize the methods that have been attempted; class can provide feedback.
 3. Project presentation 3: the final analyses and conclusions on how methods compare based on out-of-sample performance methods; interpretations should be given.

For teams and decide on project by end of January, initial presentation of project weeks 4 and 5 (one person on each team makes initial presentation, 15 teams, 4 presentations per lecture period, 10 minute presentations, 5 minutes for discussion). For weeks 1 and 2 (Feb 22 to Mar 4) after midterm break (another person on each team makes presentation 2). After this, the final presentation is made when ready in the remaining weeks.

Submit presentation as pdf to instructor to get feedback.

Final report should be submitted after presentation; then (if allowed by UBC) an oral exam of team members to follow in April.

Format of final report and its accompanying code will be posted later.

In class presentation: create using Powerpoint, LaTeX or Rmarkdown of something similar.

Course evaluation

Individual part of course assessment, in-class participation is expected. Everyone will be asked at times to speak up in class (via a random rotation) -- (a) respond to questions in the middle of lecture notes; (b) pose a question or give feedback in a team presentation.

Final exam options: Written exam (questions randomized via WebWork, upload document) or Oral exam for you and your teammates concerning your project (TA and instructor will ask question to team members in a rotation).

Tentatively, the weighting scheme is the following.

Project 50% (1/2 for report and 1/2 for code).

Homework (combination of webwork and written assignments) 20%

Midterm exam (questions randomized via WebWork, upload document) 20%

In-class presentation and oral final exam (based on project) 10%

Alternative written final exam 10%
Bonus marks for earlier project completion.