STAT 302 – Statistical Computing

Course Syllabus Winter 2024

Course Personnel:

- Instructor: Yikun Zhang (`yikun@uw.edu`) (he/him).
- Office hours:
  - Mondays 2:00 pm – 3:00 pm (Padelford Hall B-226 or Zoom);
  - Tuesdays and Thursdays 5:50 pm – 6:30 pm (Mary Gates Hall 234).
- Grader: Hyojin Lim (`hyojinl@uw.edu`)
  - All questions regarding the lab scores must be **initially** emailed to the grader within **72 hours** after the scores are released. No further exceptions will be made past the 72-hours mark.

Email Policy: During workdays, we will try our best to respond to emails within 24 hours. The replying time could be slower during weekends. When emailing either the instructor or grader, please include "[STAT 302]" at the beginning of the email header. It will help us respond to the email faster.

Lecture Schedule:

- Time(s): Tuesdays and Thursdays 4:30 pm – 5:50 pm;
- Location: Mary Gates Hall 234 (MGH 234) (Zoom option is also provided but the lecture recording *will not* be provided; see Canvas for the Zoom links).
- **Students' participation in lectures will be highly appreciated.**

Course Website:

- **Course Materials:** https://zhangyk8.github.io/teaching/stat302_uw.
- **Canvas:** https://canvas.uw.edu/courses/1698283.
  - The Zoom links for lectures and office hours will be provided on Canvas.
  - Students are also encouraged to use the discussion board on Canvas to ask questions and/or answer questions from their classmates.
- **Gradescope:** https://www.gradescope.com/courses/684433.
  - Students are expected to submit all of their lab assignments through Gradescope.
  - Students can also submit the regrade request on Gradescope within **72 hours** after the grades are released.

Prerequisites: The prerequisite is any of the following courses:

- STAT 290: Advanced Placement (AP) Statistics;
- STAT 311: Elements of Statistical Methods;
- STAT 390: Statistical Methods in Engineering and Science;
- STAT 391: Quantitative Introductory Statistics for Data Science;

- QSCI 381: Introduction to Probability and Statistics.

Course Objectives:

- **Overview:** Driven by the era of big data, computational data analysis has become an indispensable cornerstone of modern statistics. Competitive statisticians must not only be capable of providing theoretical assurances for their proposed methodologies but also proficient in conducting robust simulation studies and meaningful data applications to support their proposals. These simulation studies and data applications demand strong statistical computing and programming skills. This proficiency entails the ability to read, modify, and create computer programs tailored to the specific requirements of their data analysis tasks.

  STAT 302 stands as an introductory statistical computing course designed to equip undergraduate students with essential programming skills, preparing them for more advanced data analysis and machine learning courses. Without assuming extensive programming background, students will learn the core of ideas of programming – data structures, functions, iteration, debugging, logical design, and abstraction – through writing code to tackle various statistical and numerical analysis problems. *The course will be taught in the R programming language.*

- **Outline:**
  - Basic R and data structures;
  - Programming fundamentals;
  - Functions, debugging, and testing;
  - Data manipulation (via `tidyverse`) and visualization (via basic R graphics and `ggplot2`);
  - Statistical simulation, sampling, and bootstrap;
  - Numerical analysis (root-finding methods, numerical differentiation, and numerical integration);
  - Basic linear regression and statistical prediction.

Logistic Requirements:

- **Laptop:** You are encouraged to bring your laptop to the lectures and install both R and R Studio on your personal device, both of which are free. Please consider contacting UW Student Technology Loan Program if you do not have a reliable personal laptop.
- **Install R and R studio:** Please make sure that you have both R and R studio. You may also find this R studio cheetsheet helpful.
- **R Markdown:** You are encouraged to install R Markdown for finishing the lab and homework problems. Here is a cheetsheet for R markdown.
- **Required textbook:** None.

Evaluation: Your final grade will be calculated as follows:

- 60% Weekly/Bi-weekly Lab Assignments (usually due on Mondays);
- 20% One In-class Midterm (Week 7);
- 20% One Final Project;
- 2% Extra Credit.

Notes:

- **Late submission of homework or lab assignments will be penalized; if your submission is late for x hours, then the score will be $\left(1 - \frac{x}{36}\right)^2 \times$ the original score. Namely, any late submission for 36 hours will receive 0 scores.**
- **Unless you have a proper reason (*e.g.*, emergency), there will be NO make-up exams for both the midterm exam and the final project.**

Academic Integrity: The University of Washington Student Conduct Code (WAC 478-121) defines prohibited academic and behavioral conduct and describes how the University holds students accountable as they pursue their academic goals. Allegations of misconduct by students may be referred to the appropriate campus office for investigation and resolution. More information can be found online at https://www.washington.edu/studentconduct/. The instructor reserves the right to assign a failing grade to the student for serious violations of student conduct.

- **Collaborations:** You are encouraged to discuss course materials (but not the exams) with your classmates. **All the solutions that you turn in, however, must be your own.** This includes but not limit to written explanations and code.
- **ChatGPT:** You are allowed to use ChatGPT or other AI tools to assist your learning and complete homework assignments. **However, if you have used an AI tool or technology in any other way in the process of completing any homework and lab questions, an acknowledgement of how you have used AI tools or technologies is required.**

Academic Accommodations: It is the policy and practice of the University of Washington to create inclusive and accessible learning environments consistent with federal and state law. If you have already established accommodations with Disability Resources for Students (DRS), please activate your accommodations via myDRS so we can discuss how they will be implemented in this course. If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations (conditions include but not limited to; mental health, attention-related, learning, vision, hearing, physical or health impacts), contact DRS directly to set up an Access Plan. DRS facilitates the interactive process that establishes reasonable accommodations. Contact DRS at disability.uw.edu.

Religious Accommodations: Washington state law requires that UW develop a policy for the accommodation of student absences or significant hardship due to reasons of faith or conscience or for organized religious activities. The UW's policy, including more information about requesting an accommodation, is available at Religious Accommodations Policy. Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form.

Diversity and Inclusion: Diverse backgrounds, embodiments, and experiences are essential to the critical thinking endeavor at the heart of university education. Therefore, I expect you to follow the UW Student Conduct Code in your interactions with your colleagues and course staffs in this course by respecting the many social and cultural differences among us, which may include, but are not limited to: age, cultural background, disability, ethnicity, family status, gender identity and presentation, citizenship and immigration status, national origin, race, religious and political beliefs, sex, sexual orientation, socioeconomic status, and veteran status.

Useful References:

- Previous Editions of STAT 302 at UW by Bryan Martin, Peter Gao, and Andrea Boskovic.
- STAT 133 Concepts in Computing with Data (UC Berkeley) by *Gaston Sanchez*: `https://github.com/ucb-stat133`.
- Statistical Computing at CMU: `https://www.stat.cmu.edu/~ryantibs/statcomp/`.
- Statistical Methods for Data Science by *Elizabeth Purdom*: `https://epurdom.github.io/Stat131A/book/index.html`.

- Advanced Statistical Computing by *Roger D. Peng*: https://bookdown.org/rdpeng/advstatcomp/.
- R for Data Science (Second Edition) by *Hadley Wickham* (based on the `tidyverse` package).
- Data Visualization: A Practical Introduction by *Kieran Healy* (gentle intro to `ggplot2` package).

Disclaimer: The instructor reserves the right to make changes to the syllabus as necessitated by circumstances.