Stat 704: Data Analysis I

Fall 2022

- Lecture: 2:20–3:35 PM, Monday/Wednesday, LeConte 206

- Instructor: Dr. Yen-Yi Ho
  Office: LeConte 216A
  E-mail: hoyen@stat.sc.edu
  Office hours: Monday/Wednesday after class or by appointment (can be arranged virtually or in-person)
  Class website: http://people.stat.sc.edu/hoyen/Stat704/Stat704.html. Homework and announcements will be posted on the class website and it will be updated regularly.

- Prerequisite: STAT 704 has a co-requisite of STAT 712. If you register for STAT 704 for Fall 2021, please be sure that you are taking STAT 712 in Fall 2021 (or that you have previously taken STAT 712). If you wish to take an applied statistics sequence without taking STAT 712, please consider STAT 700-701 (or STAT J700-J701), which is designed for graduate students from departments other than statistics.

- Course description: Primarily for graduate students in statistics and the mathematical sciences. Informally, we will cover about a third of *Applied Linear Statistical Modelsa* (5th edition), by Kutner, Nachtsheim, Neter and Li as well as additional topics covered in notes.

- Learning Objective: Upon successfully completing this course, students will be able to:

  - Formulate a scientific question about the relationship of a continuous (or discrete) response variable Y and predictor variables X in terms of the appropriate linear (or logistic) regression model.

  - Interpret the meaning of regression coefficients in scientific terms as if for a substantive journal.

  - Develop graphical and/or tabular displays of the data to display the evidence relevant to model fitting.

  - Estimate the regression coefficients using statistical packages such as R or SAS or Stata.

- Check the major assumptions of the model including independence and model form (mean, variance and distribution of residuals) and make changes to the model appropriately to handle violations of the standard assumptions. Use regression diagnostics to prevent a small fraction of observations from having disproportionate influence on the results.

  - Write a methods and results section for a substantive journal, interpret the results and answer specific scientific questions with analysis results in a way that can be understood by substantive experts.

- Homework: Approximately 7-9 homework assignments will be announced in class and posted on the course web page. Use the homework template posted on the class website to write homework. There will be a take-home data analysis final project.

  Each student's homework must be done independently. You may discuss the homework, compare answers, et cetera, but *all submitted homework must be your own work.*

- Exams: There will be two in-class exams. The exams will help prepare students for the qualifying examination at the end of their first year. Exams may not normally be made up, except in extreme circumstances, for which written documentation of excuse (doctor's note, funeral notice, etc.) is required. If you suspect you may miss an exam day, it is important to contact me well in advance of the test date.

- **Students may not use previous student's homeworks or previous solutions.**

- Grading: The course grade will be based on homework (40%), the final take-home data analysis project (10%), and the two exams (25% each).

- Attendance: Attendance will not be taken, and it will not count towards final grades.

- If the instructor get sick or must miss a class for an academic/personal reason, the class will be pre-recorded or post-recorded.

- Computing: Some problems in this course involve significant computations, and for these, we will learn to use the software package R. R is

a free, open-source statistical programming language. All code used in class will be available on the course webpage; extensive coding examples are available on previous instructors' websites (Drs. Hitchcock and Grego) as well. R is an extremely useful statistical programming language that has become widely used in recent years. Student may also use other packages such as Stata, SAS.

- Disabilities: If you qualify for accommodations because of a disability, please submit a letter to me from the Office of Student Disability Services in a timely manner so that your needs can be addressed. The Office of Student Disability Services determines accommodations based on documented disabilities. Contact: 777-6142.
Close-Hipp, 1705 College St #102, Columbia, SC 29208

Some Important Questions

1. **Where can I find weekly schedule?** Please refer to the class website for important information such as lecture notes, weekly schedule, syllabus. The class website will be updated frequently.

2. **Can I hand in the homework late?** No. Late homework will lose 20% of total points per day, unless arrangements have been made with the instructor for an extension. Homework will not be accepted after the time at which graded homework are returned.

3. **Must I use R?** No. Use whatever you would like. You will be required to do simulation and programming, for which software like R are particularly well suited. The SAS software is also widely used by statisticians, so is STATA. Programs such as SPSS, Minitab, and Excel, which can be great for many analyses, are insufficient for the need of this class.

4. **What should I use to edit programs?** You should use an editor that has syntax highlighting, and automatic indenting. So, don't se Microsoft Word or Notepad to edit programs. In RStuido, you can open a new R script document. Or in R, under the File tab, open a new script by clicking on "New Document". Or if you're a hacker type, use emacs. SAS comes with its own editor that has auto-ident and syntax hightlighting.

5. **I have an event with dates that conflict with the final exam. Can I take the exam early?** No. I probably won't have the exam written to be taken early.