STAT 431/511: STATISTICAL INFERENCE

Classes.
- Section 1: Tu/Th 10:30–11:50 am in F65 JMHH
- Section 2: Tu/Th 12–1:20 pm in F65 JMHH

Instructor. Weijie Su (`suw@wharton.upenn.edu`)

Office: 472 JMHH

Office Hours: 2–3 pm on Thursday

Teaching Assistant. Raiden Hasegawa (`raiden@wharton.upenn.edu`)

Office: F96 JMHH

Office Hours: 3–4:30 pm on Tuesday. Additional office hours will be held at 3–4 pm Monday in weeks that homework is due

Course Overview. The course aims to equip students with ideas and tools in statistics which range from the very beginning of the subject to an intermediate level. Together, we will examine a collection of basic concepts and commonly used methods, with an emphasis on the understanding of when and how to apply them, and why. Students will also experiment the ideas on data examples using the statistical software `R`. We will cover the following topics in class:

- Collecting, summarizing and visualizing data
- Distribution of sampling statistics
- Point estimation and confidence intervals
- Hypothesis testing
- Inference with two populations
- Goodness of fit
- Regression (simple linear regression, multiple regression, ANOVA, logistic regression)
- Maximum likelihood
- Nonparametric methods

Prerequisites. The official prerequisite of the course is STAT 430. The effective prerequisite is fluency with basic probabilistic reasoning and analysis (e.g., probability distributions and densities; joint distributions; conditional probability, independence, correlation, and covariance; moment generating functions; law of large numbers; central limit theorem; etc). It would be helpful to have previous exposure to linear algebra, but it is not required. Previous exposure to the statistical computing language `R` is not required, either.

Textbook. *Statistics and Data Analysis*, Ajit C. Tamhane & Dorothy D. Dunlop. We will cover selected topics from Chapters 3-15. We will not have time to cover all details and examples. The students should read along in the textbook to gain most from the class. Exams will only cover topics discussed in class; however, students are responsible for keeping track of which topics are covered and which are not.

Course Website. The course website uses the Canvas platform. Please check the course website for announcements, handouts, sample codes, assignments, solutions, and other course materials. Students are encouraged to use the discussion board on Canvas to interact with the TA.

Homework. There will be 6 homework, assigned approximately biweekly. Homework will be due on Thursday and must be handed in during class or in Instructor's mailbox (located on the 4th floor of JMHH) by 5 pm on the date the assignment is due. *No late homework will be accepted, but the lowest score will be dropped.*

Exams. There will be a midterm and a final exam, scheduled on the following dates:

- Midterm
    - Date: Tuesday, February 13
    - Time: usual class meeting time
    - Location: classroom (F65 JMHH)
- Final
    - Date: Tuesday, May 8
    - Time: 6–8 pm
    - Location: TBD

The final exam will be cumulative, with an emphasis on the material covered since the midterm. Both exams will be closed book, but students can bring two cheat-sheets that are no larger than an A4 paper. Each student should also bring a calculator.

Grading. The course grade will be based on the homework, the midterm, and the final.

- Homework: 30% (lowest score dropped)
- Midterm: 30%
- Final: 40%

Collaboration Policy. Working together on homework is allowed and encouraged. However, students must write up their homework solutions by themselves. Names of collaborating students should be provided on the front page of each homework write-up.

Statistical Computing. Few homework assignments will involve coding and statistical analysis on datasets provided. It is recommended that students download and use R for this purpose. Other software such as Excel or Matlab may also be used for data analysis on homework sets. Knowledge of R will not be tested in the midterm and the final. However, R will be the language referred to in class, so students who choose to use another statistical computing platform will need to figure out the equivalent commands on their own.