

STAT 431: STATISTICAL INFERENCE

2019 Fall

Classes.

- Section 1: M/W 1:30–2:50 pm in 365 JMHH
- Section 2: M/W 3:00–4:20 pm in 365 JMHH

Instructor. Anderson Ye Zhang (ayz@wharton.upenn.edu)

Office: 444 JMHH

Office Hours: 4:30–5:30 pm on Monday

Teaching Assistant. Rachel Ruijia Wu (ruijiawu@wharton.upenn.edu)

Regular Office Hours: 1:30–2:30pm on Thursday, in F92 JMHH.

Extra Office Hours: 5:00–6:00pm on Thursday if homework is due that week, in F92 JMHH.

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 Websites. The course will use both the Canvas and Piazza platforms. Please check the Canvas site for announcements, handouts, sample codes, assignments, solutions, and other course materials. In addition, this course will also use Piazza, where students can ask and answer questions and interact with each other.

Homework. There will be 4 homework, assigned tri-weekly. Homework will be due 11:59pm Thursday. *No late homework will be accepted, but the lowest score will be dropped.*

- We only accept **online submission** of homework on Canvas. Please do not email your homework to the instructor or TAs.
 - You can **type** your homework (we accept pdf, doc, and docx documents) or
 - **scan** your **handwritten** homework and then upload it to *Canvas*.
- if you choose to scan your handwritten homework, please **do not submit raw camera photos** (not processed through a “scanning” app). Raw photos are essentially unreadable. Raw photos are also extremely difficult to scale appropriately to view on the screen within the Canvas grading tool.
 - You can use free smartphone apps, such as *CamScanner*, *Tiny Scanner*, and many basic apps to scan their handwritten work. These apps are easy to use, and automatically adjust the raw smartphone camera photos for maximum black/white contrast. Students can photograph each page one after the other and CamScanner automatically incorporates them into a single document.
 - You may also use a computer scanner (some multifunction printers can scan too) to scan your document. You can use any campus copy machine to scan to a pdf file directly to a personal USB stick they plug into the copier. The quality of scanning is generally good so that the output pdf file can be submitted directly without further processing.
- For each assignment, please combine your work (text, codes, and graphs) to a **single file** for submission, so that the graders just need to look at one file, not multiple ones.
- You can submit a single homework multiple times, before the due time. So if you want to correct a mistake in your submitted homework, you can just correct it and submitted it again before the deadline. We will grade the last version that is submitted prior to the submission deadline.
- You are encouraged to check the accuracy of grading of your exams and assignments. This includes checking if correct answers were mistakenly marked wrong or if points were added incorrectly. You can talk with or email the instructor or TAs to report grading mistakes.

Exams. There will be two midterms and one final exam, scheduled on the following dates:

- Midterm I
 - Date: Wednesday, October 2
 - Time: usual class meeting time
 - Location: classroom (365 JMHH)
- Midterm II
 - Date: Wednesday, November 6
 - Time: usual class meeting time
 - Location: classroom (365 JMHH)
- Final
 - Date: Wednesday, December 18
 - Time: 6-8 pm
 - Location: TBD

The final exam will be cumulative, with an emphasis on the material covered since the 2nd midterm. All the exams will be closed book, but students can bring up to 10 cheat-sheets (each no larger than a letter paper). Each student should also bring a calculator.

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

- Homework: 10% (lowest score dropped)
- Midterm I: 25%
- Midterm II: 25%
- 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 midterms 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.