

# STAT 431: STATISTICAL INFERENCE

2021 Spring

## Classes.

- Section 1: T/R 10:30 am–11:50 am on Zoom
- Section 2: T/R 12:00 pm–1:20 pm on Zoom

Visit the course Canvas site for Zoom links.

**Instructor.** Anderson Ye Zhang ([ayz@wharton.upenn.edu](mailto:ayz@wharton.upenn.edu))

Office Hours: TBD

**Teaching Assistant.** Rachel Ruijia Wu ([ruijiawu@wharton.upenn.edu](mailto:ruijiawu@wharton.upenn.edu))

Office Hours: TBD

**Course Website.** <https://canvas.upenn.edu/courses/1568630>. Please check the Canvas site for announcements, slides, sample codes, assignments, solutions, and other course materials.

**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.

Online Access: [https://franklin.library.upenn.edu/catalog/FRANKLIN\\_9948495913503681](https://franklin.library.upenn.edu/catalog/FRANKLIN_9948495913503681). Then click the “Online access”, use Penn credential to log in, and then click “check out.” The number of copies is limited, so it would become unavailable when many students are accessing it simultaneously.

**Homework.** There will be 5 homework assignments, assigned every 2-3 weeks. Homework will be due 11:59pm Thursday (Philadelphia time).

- *No late homework will be accepted, but the lowest score will be dropped towards to the final grade.*
- Gradescope will be used for homework submission and grading.
- Students can help each other on solving the problems, but are expected to prepare the final writeup individually with acknowledgment of the help received.

**Exams.** There will be two midterms and one final exam. All the exams will NOT be cumulative. They are scheduled on the following dates:

- Midterm I
  - Date: Thursday, Feb 18
  - Time: 6-8pm
  - Format:
    1. It is an open-book exam. Students can access their own notes, course Canvas site, textbook(s), and references. Other usage of the Internet (especially communication with others) is strictly prohibited.
    2. The 2-hour exam is administered over Zoom. Students are required to be present on Zoom, and are encouraged to keep their cameras on.
    3. There will be multiple sittings of the exam to accommodate students in different time zones.
- Midterm II
  - Date: Thursday, Apr 1
  - Time: 6-8pm
  - Format: Same to Midterm I.
- Final
  - Date: TBD
  - Time: TBD
  - Format: TBD

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

- Components:
  - Homework: 30% (lowest score dropped)
  - Midterm I: 20%
  - Midterm II: 20%
  - Final: 30%
- Letter grade assignment:

- A+/A/A- if  $90 \leq \text{Grade} \leq 100$ ,
  - B+/B/B- if  $80 \leq \text{Grade} < 90$ ,
  - C+/C/C- if  $70 \leq \text{Grade} < 80$ .
- There will be no make-ups, extensions, or extra credit opportunities under any circumstances.
  - You will not be excused from an exam due to a conflict with a job interview or other personal events.

**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.