

STAT 213: Statistical Inference II, Spring 2024

Instructor: Pragya Sur (pragya@fas.harvard.edu)

Lectures: Mondays and Wednesdays 1:30 - 2:45 PM.

Course Webpage: <https://canvas.harvard.edu/courses/128099>

Discussion forum: We will use Ed for discussions.

Teaching Fellows: Souhardya Sengupta (ssengupta@g.harvard.edu)

Sections and Office Hours: TBD

Textbook: All course material is contained in class notes. Useful additional references are *Asymptotic Statistics* by Aad van der Vaart, *Testing Statistical Hypotheses* by Lehmann and Romano, *Theory of Point Estimation* by Lehmann and Casella.

Prerequisites: STAT 210, 211 (or equivalent).

Description: Stat 213 will build upon Stat 211, providing tools to study and validate statistical methods. A primary focus will be large-sample theory, specifically, inference for M- and Z-estimators under well-specified and mis-specified models, quadratic mean differentiability and its implications, local asymptotic analysis, contiguity, LeCam's lemmas, asymptotic analysis of tests including optimality, Bayesian asymptotics, e.g. consistency, Bernstein-von-Mises theorem; time and interest permitting, we will cover some special topics in high-dimensional statistics. This is a proof-based, theoretical class, and the focus will be on understanding the fundamentals of classical and modern statistical theory and methods.

Grading: HW (35%), 24-hr-take-home Midterm (20%), 24-hr-take-home Final (30%), Presentation (10%), Class participation (5%).

Homework Policies: Five homework sets will be assigned through the semester. Typically, assignments will be released on Wednesdays and be due on Canvas roughly 12-14 days later. Students may discuss the problems with colleagues, however, each student is expected to write their own solutions and report their collaborators when they turn in the assignment. Each student will have 48 cumulative hours of late time (measured on Canvas) forgiven, after which assignments turned in late will receive no credit, and each student's lowest homework score will be dropped. This is an advanced statistics class, and forms a foundational basis for starting to do research in modern statistics. As a result, the class is meant to deepen your own understanding of the subject. In the spirit of that, you are NOT allowed to use any generative AI for your homeworks.

Exams: The midterm will be a 24-hr-take-home exam, available under Quizzes starting Mar 1st, 9AM to Mar 4th, 3PM—choose any 24 hr block of your convenience within this time frame. *There will be no class on Mar 4th in lieu of the midterm.* The final will be also be a 24 hr take-home exam. The timing will be determined soon after the Registrar provides more information about exam scheduling. Both exams will be open book but NOT open internet, with no collaboration allowed. Also, please do not discuss the exams with anyone until after the window for taking the respective exam ends. You can cite any result from class or course handouts without rederiving it. From the Internet, you are allowed to use Wikipedia and Wolfram Alpha but can't cite such sources as an alternative to providing a mathematical derivation. You are NOT allowed to use any generative AI for the exams. You can write using a tablet, handwritten (and scan to a PDF),

or typeset your solutions. You should upload your exam solutions through Canvas—please do not email us the solutions.

Course Project/Presentation: This is a Ph.D. level class, which seeks to provide students the tools and background to make novel contributions in the fields covered. To achieve this, we will include a project component into the course. Modern day statistical research deals largely with high-dimensional data. I will release a list of recent papers in high-dimensional inference/prediction. Students will be requested to form groups of 2 (if necessary, 3), choose 1-2 papers from this list, and present its contents during the last 1-2 lectures of the semester. No new results are expected from the students. These presentations are crucial for students to get a glimpse of modern statistical research, and will contribute towards 10% of the course grade. Students will be required to declare their presentation partners and choice of papers by Feb 24, 11:59PM.

Generative AI policy Through the course, you are allowed to use Generative AI to deepen and broaden your understanding of the topics covered in class. In particular, if there are topics mentioned in passing in class on which modern research is performed, you may use Generative AI to find more sources to learn the corresponding material from. However, use of Generative AI for homeworks, exams, and in designing your course presentation, is strictly not allowed.

Class Participation: Lectures will be active: I will stop a few times during each lecture and have everyone work individually for roughly 2 minutes a question I pose. Because of the importance of the active learning component of the lectures, 5% of the course grade will come from class participation: after each such 2-minute working time, I will solicit thoughts on the proposed question from the class. As long as a student regularly attends lectures (at least $\sim 80\%$ of them) and participates in the discussion following the active learning exercises, they will receive full points for class participation.