STAT 211: Statistical Inference I, Fall 2024

Instructor: Lucas Janson (`ljanson@fas.harvard.edu`)

Lectures: Mon/Weds 9:00–10:15 AM, Science Center Hall E

Course Webpage: `https://canvas.harvard.edu/courses/138002`

Course Staff: Yuanchuan Guo (`yguo1@g.harvard.edu`),
Kevin Du (`kevindu@college.harvard.edu`),
Kenny Gu (`kgu@college.harvard.edu`)

Office Hours:
Lucas Janson: Thursdays 3:00–5:00 PM (Science Center 710)
Yuanchuan Guo: TBD 2h (location TBD)
Kevin Du: TBD 1h (location TBD)
Kenny Gu: TBD 1h (location TBD)

Sections:
Wednesdays 4:30–5:30 PM (location TBD)
Thursdays 9:00–10:00 AM (location TBD)
Fridays 10:30–11:30 AM (location TBD)

Text: All course material is contained in detailed, textbook-style lecture notes (written by the instructor expressly for this course), which will be posted to Canvas after each lecture. Useful additional references are *Testing Statistical Hypotheses* by Lehmann and Romano, *Statistical Inference* by Casella and Berger, and *Theory of Point Estimation* by Lehmann and Casella.

Prerequisites: STAT 110 and 111 (or equivalents). This is a proof-based class, so comfort with proof-based math, especially real analysis, is recommended. For instance, Math 112 or Stat 210 would provide good exposure to the tools and mathematical approach in this class, though measure theory will not be used in Stat 211. The probability tools we'll use are: the central limit theorem, the law of iterated expectation, Jensen's inequality, Slutsky's theorem, the continuous mapping theorem. Other concepts that will arise in class that might be good to be familiar with are uniform functional convergence, dominated convergence, basic analysis about polynomials, and the difference between convergence in mean, convergence in probability, convergence in distribution, and almost sure convergence.

Grading: Homework (45%), 16h take-home midterm (25%), 16h take-home final (30%).

Policies: A total of five homework assignments will each be assigned on Wednesdays and be due on Canvas 12 days later on the next-to-next Monday (or the following Wednesday if the Monday is a holiday or the homework spans Thanksgiving) at 11:59 PM, covering material from three to four lectures. Assignments will be handled through Gradescope, which is accessible through Canvas as one of the tabs on the lefthand side. Collaboration is allowed but students must write up and fully understand their own solutions and report their collaborators and cite any online resources used when they turn in the assignment. Use of AI is allowed, but should be treated as you would a collaborator: you can use it for assistance (and if you do, you should indicate how in your submission) but you may not directly ask it for the answer. Each student will have 48 cumulative hours of late time (measured on Canvas) forgiven after which assignments turned in late will receive no credit. Both exams will be 24-hour take-home exams with no collaboration allowed. Lectures will be active: I will stop a few times each class and have everyone work individually for 1–2 minutes on a question I pose. Students will also be asked to fill out a short 1-minute online feedback form at the end of each lecture so I can gauge their learning and address common points of confusion at the beginning of the next lecture; the forms will ask for (but not require) students' names so I can follow up individually in certain cases if needed, but no part of the course grade will depend on these forms.

Description: Foundations of frequentist and Bayesian inference, and decision theory. Likelihood, sufficiency, and ancillarity. Point estimation, unbiasedness, maximum likelihood, method of moments, minimum-variance. Parametric and non-parametric hypothesis testing, confidence intervals. Selective inference: multiple testing, familywise error rate, false discovery rate. Bayesian inference, conjugate priors, credible intervals. Admissibility, Stein's phenomenon, empirical Bayes. Time permitting: post-selection inference and the bootstrap.

Goals: This is a proof-based, theoretical class, so the focus will be on understanding the fundamentals of classical and modern statistical methods. Since it is a PhD-level class, the main goal is to give students the tools and background to make novel contributions in the fields covered. To this end, every lecture or two I will mention a relatively recent paper that relies heavily on a concept from that lecture, so that students interested in how the tools we learn are useful in modern research can learn more (students will not be tested on this extra material). The goal of the active learning component of the class is to enhance understanding by having students engage with lecture material while it is being taught.

Tentative Schedule:

Date	Homework	Topics
9/4		Class schedule/policies, overview of topics that will be covered
9/9		Frequentist and Bayesian philosophies, sufficiency
9/11		Minimal sufficiency, completeness
9/16		Ancillarity, point estimation, unbiaseness, MSE
9/18	1 out	Rao–Blackwell, Lehmann–Scheffe (UMVUE), score function
9/23		Fisher information, Cramér–Rao lower bound
9/25		Method of moments, maximum likelihood estimation
9/30	1 due	Consistency of MLE, asymptotic normality of MLE
10/2	2 out	Delta method, hypothesis testing, Neyman–Pearson lemma
10/7		Monotone likelihood ratio, uniformly most powerful tests, likelihood ratio tests
10/9		Asymptotic tests (likelihood ratio, score, Wald)
10/14		Holiday (no class)
10/16	2 due, 3 out	Univariate nonparametric tests
10/21		Two-sample nonparametric tests, permutation tests
10/23		Selective inference, Bonferroni, Šidák, Holm
10/28	3 due	False discovery rate, Benjamini–Hochberg, Benjamini–Yekutieli
10/30		Confidence intervals, inverting hypothesis tests, pivotal quantities
10/30–11/4		Take-home midterm (excludes 10/28 and 10/30 material): choose any 16h window between 10:15am 10/30 and 9:00am 11/4
11/4		Test statistic distributions, shortest intervals, asymptotic intervals
11/6	4 out	Bayesian inference, conjugate priors
11/11		Jeffrey's prior, credible intervals
11/13		Prediction intervals, decision theory, admissibility, minimax
11/18	4 due	Bayes risk, admissibility of Bayes
11/20	5 out	Admissibility of univariate mean, James–Stein estimator
11/25		Stein's lemma, Stein's paradox
11/27		Holiday (no class)
12/2		Post-selection inference
12/4	5 due	The bootstrap
12/8–12/10		Take-home final: choose any 16h window between 7am 12/8 and 11:59 PM 12/10