STAT 139: Introduction to Linear Models

Fall 2025

Meeting Time: Tue/Thu 10:30-11:45 AM (Geological Museum 100)
Course Site: https://canvas.harvard.edu/courses/154816

Instructor: James Xenakis jxenakis@g.harvard.edu
Head TA: Zimeng Li zimengli@g.harvard.edu
Head TF: Jonas Freeman jonasfreeman@college.harvard.edu

1. Course Overview

Course Description

STAT 139 is an in-depth statistical introduction to linear models and related methods. Topics covered will include group comparisons (e.g., z- and t-based methods), with a focus on framing them as linear models. We will cover ANOVA and linear regression as well as their non-parametric extensions, permutation, and bootstrapping. The course will begin with a rigorous treatment of ordinary least squares (OLS) regression, including diagnostics and model selection/comparison. We will then move on to extensions including ridge regression and the LASSO, weighted least squares, and introductions to longitudinal and mixed models.

Learning Objectives

By the end of the course, students should be able to evaluate the strengths and weaknesses of a variety of statistical techniques. Given a dataset, students should be able to:

- state scientific hypotheses,
- explore the data using statistical software,
- determine which statistical model(s) may be appropriate,
- apply corresponding inferences,
- check the assumptions behind these tests and models,
- interpret the results of the analysis to draw conclusions about the hypotheses.

This course equips students with the foundational skills needed for advanced studies in Statistics - such as Stat 131, 140, 149, 160, 183, 186, and beyond - or for making data-driven conclusions across any discipline.

Prerequisites

Mathematics 21a (Multivariable Calculus) or equivalent, Mathematics 21b (Linear Algebra) or equivalent, and Statistics 110 (Theoretical Probability) are required. Statistics 111 (Theoretical Inference) and familiarity with R are *highly* recommended, though Statistics 104 or 109 is sufficient (students unsure about equivalents should email the instructor before enrolling). Concurrently taking Math 21b is allowed.

Textbooks

The recommended text for this class is:

– *Linear Models with R, 3rd Edition* by Julian J. Faraway, CRC press, 2014.

Purchasing or renting a textbook is not required, the notes will cover all course material. The second edition of the text is also acceptable, as is the (free) first edition. Recommended readings will be assigned semi-regularly, and sometimes supplemented with free materials from online sources, or of the teaching staff's creation.

Computing

This course will heavily use the R language and environment, as well as RStudio and RMarkdown. General familiarity with R is required for this course, but no other coding experience or knowledge is necessary (although it will be very helpful). Having taken Stat 111 will provide you with enough background in R. Further, these basic tutorials can be helpful in getting you started or as a basic review:

– DataCamp: https://www.datacamp.com/courses/free-introduction-to-r

– Rstudio.com: https://education.rstudio.com/

If you would like to install the software on your own machine (I do recommend doing this at some point, though it is not required for the class), R and RStudio are freely available for all common operating systems and can be downloaded at https://posit.co/download/rstudio-desktop/.

If you would prefer to use Posit Cloud (which allows you to do all your computing online through a browser), you can joing our Posit space here.

2. Learning Environment

Lectures

Lectures will be held on Tuesdays and Thursdays from 10:30-11:45 AM in Geological Museum 100. Lecture attendance is recommended, but not mandatory.

Sections

Weekly TF-led sections will be held throughout the course. The section schedule will be announced on the course website and there will be a range of day/time options to accommodate students' schedules. Sections will be devoted to review of concepts, practice problems, and problem set preparation/questions. Attendance will be taken in each section. More on this in the Bonus section of the syllabus.

Discussion Forum

Discussions through Ed (https://edstem.org/) are highly encouraged, but posting solutions of any kind is strictly prohibited. You will be able to access through the course website once the semester has begun.

Office Hours

Regular weekly office hours (time and location posted on Canvas) will take place throughout the course, during which you are free to ask questions about the course material or problem sets. More general statistics questions are also encouraged. In addition, you will be able to schedule one-on-one office hour appointments with Professor Xenakis through *Calendly* (link posted on Canvas).

3. Problem Sets (50 %)

Structure

There will be (tentatively) 9 problem sets throughout the semester, generally posted on Friday evenings, to be due the following Friday at 11:59 pm. All assignments should be submitted through Gradescope (both your PDF and your .Rmd file).

Most homeworks for this course will include two parts. **Part I** will contain typical homework questions covering the week's material. **Part II** will consist of questions that should be treated as milestones which build towards the completion of Homework 9.

Type	Question Breakdown	Grade Contribution
Homework 1–8	Part I: ∼ 90% Part II: ∼ 10%	40%
Homework 9	Part I: ∼ 20% Part II: ∼ 80%	10%

Late Work Flexibility

Students are permitted up to five late weekends. Each late weekend extends the deadline by up to 72 hours (until Monday 11:59 PM). Because solutions cannot be released until all work is submitted, no extensions beyond this window are allowed without penalty. Any submission within

the 72-hour window counts as one late weekend.

Extra flexibility can be granted in extenuating circumstances; email Professor Xenakis and your resident dean.

Homework Drops

Homework drops are **not permitted** in this class.

Collaboration

You are encouraged to discuss homework with other students, but you must write your final answers yourself in your own words and reference any human collaborators (you do not have to reference ChatGPT). Solutions that are copied from someone else's work are not acceptable and will be treated as Honor Code violations.

4. Exams (50 %)

General Information

Midterms will be taken **in-person** during regular class time (Midterm 2 may also include a take-home component). The Final exam will also be in-person, with the date set by the registrar. Students may bring **up to four sides of notes on A4 paper** (two double-sided sheets) into the exam.

Type	Question Breakdown	Length
Midterm 1 (Oct 2)	Homework & Section style problems: ∼75% Short answer: ∼25%	1h 15 min
Midterm 2 (Nov 6)	Homework & Section style problems: ∼50% Short answer: ∼25% Understanding R output: ∼25%	1h 15 min
Final (TBD)	*Designed to be completed in 2 hours*	3 hrs

5. Grading

As a teaching staff, we understand that every student has different approaches to learning, and unforeseen circumstances can affect exam performance. With this in mind, we have designed two possible grading paths. Your final grade will be determined by the following formula:

$$\text{Final Grade} = \max\{\text{Path 1}, \text{Path 2}\} + \text{Bonus}$$

Component	Path 1	Path 2
Problem Sets	50%	50%
Midterms	$20\% \cdot \max(M_1, M_2) + 5\% \cdot \min(M_1, M_2)$	$40\% \cdot (M_1 + M_2)$
Final	25%	10%

6. Bonus

Active participation and engagement in the course is optional but encouraged; it can boost your grade but never lowers it.

- **High participation:** Attending > 90% of all sections *may* earn an upward grade adjustment, depending on the overall class distribution.
- **Late weekend bonus:** Homework averages are multiplied by 1.03 with 1-2 late weekends used, and by 1.05 if no late weekends are used (capped at 100%).

Path 1 example scenario: Jonas struggled on Midterm 1 due to illness, but excelled in Midterm 2 and plans to focus heavily on the Final. He took *no late weekends*, so his homework average of 90% receives a 5% boost to 94.5%. Path 1 lets Jonas benefit from the higher midterm score (20%) while minimizing the impact of the lower midterm (5%). With a strong Final (25%), Jonas can still secure a high course grade despite one weak midterm. For example, suppose $M_1 = 50$, $M_2 = 95$, and $F = 85$. Then:

$$\text{Path 1} = 0.50(94.5) + 0.25(85) + 0.20(95) + 0.05(50) = 89.48$$

$$\text{Path 2} = 0.50(94.5) + 0.10(85) + 0.20(50) + 0.20(95) = 87.73$$

Path 2 example scenario: Sam performed steadily on both Midterms and wanted to prioritize other finals. He used *two late weekends*, so his homework average of 92% receives a 3% boost to 94.76%. With both Midterms equally weighted (20% each) and only 10% on the Final, Path 2 rewards Sam's consistent effort. Suppose $M_1 = 88$, $M_2 = 90$, and $F = 75$. Then:

$$\text{Path 1} = 0.50(94.76) + 0.25(75) + 0.20(90) + 0.05(88) = 87.59$$

$$\text{Path 2} = 0.50(94.76) + 0.10(75) + 0.20(88) + 0.20(90) = 89.16$$

Bonus: With **high participation**, both Jonas and Sam would receive a grade bump if they were on the wrong side of the A-/B+ boundary (exact cutoffs will vary by year).

7. Additional Information

Accessibility

Harvard College is committed to working with all students. Accommodation requests can be made with University Disability Resources. Advance notice and appropriate documentation (in the form of a letter from the Disability Access Office) are required for accommodations.

Regrade Requests

The process to submit regrade requests is as follows:

- Regrade requests must be submitted to Gradescope within a week of when the grade is released,
- In your request, clearly and succinctly state what error you believe occurred,
- Note that your grade may increase, decrease, or stay the same, and you will not be able to revert to your original grade if your grade does happen to decrease.
- Impolite requests will not be honored.

Generative AI

This course encourages students to explore the use of generative artificial intelligence tools such as ChatGPT to gain conceptual and theoretical insights, as well as assistance with coding.
We draw your attention to the fact that different classes at Harvard could implement different AI policies, and it is the student's responsibility to conform to expectations for each course.