STAT 230A: Linear Models

Lectures: Mon, Wed, Fri 9:00am - 10:00am
101 Weill

Lab: Mon 12:00pm – 2:00pm, 330 Evans
Mon 2:00pm – 4:00pm, 330 Evans

Topics:

- Ordinary least squares (OLS)
- Gauss-Markov theorem
- Finite-sample exact and asymptotic inference
- Heteroskedasticity
- Cluster-robust standard errors
- Model checking
- Model selection and shrinkage
- Generalized linear models
- Generalized estimating equations

Instructor: Sam Pimentel
Office: 429 Evans
Email: spi@berkeley.edu
Preferred pronouns: he/him
Office Hours:
Wed 3:00-4:00 PM
Friday 10:30-11:30 AM
Location: 429 Evans

GSI: Xinyi Wang
Email: wang.xinyi@berkeley.edu
Preferred pronouns: she/her
Office Hours:
Tue 3:30 - 5:30 PM
Location: 444 Evans

Text: *Linear Model and Extensions* by Peng Ding. Available on Arxiv.

Equipment: Access to a computer with access to the Internet and R/Python installation will be necessary. If you do not have access to a computer, you can borrow one from the University library. See https://studenttech.berkeley.edu/hardware-lending for more details. The Student Technology Equity Program is another good resource. Feel free to contact the instructor if you have concerns about your access to needed technology.

This is a working draft of the syllabus and is subject to change.

Learning goals

By the end of the semester you should be able to:

1. Understand and interpret ordinary least squares regression models from a mathematical perspective.
2. Adapt the basic regression model to common practical complications including violations of standard assumptions, high-dimensional regimes, and non-continuous outcomes.
3. Evaluate the quality of a regression analysis in context and suggest improvements.
4. Use R or Python to fit, report, and clearly communicate the process and results of a regression analysis.

Prerequisites:

- Master's level probability at the level of STAT 201A. See Dr. Aditya Guntuboyina's STAT 201A course notes for specifics.
- Linear algebra, at the level of the appendix in the course notes. See also Gilbert Strang's

recorded lectures from the MIT course 18.06 Linear Algebra, Spring 2005 (available on YouTube).

- R or Python programming. R will be used for all examples, but assignments involving computing may be completed either in R or Python.

Attendance:

Regular attendance and participation in lecture is an important part of the learning experience in this course. As such, in-person attendance at lectures is expected and lecture recordings will not generally be released during the semester, although notes that I write on the display during lecture will be posted after class. Attendance is also encouraged at lab, and labs will not be recorded. Note that **attendance in your assigned lab is required on days when quizzes are given**.

Assessment:

Homework

We anticipate giving 7 homework assignments during the semester. Homework will be posted to bCourses, and will generally be due 2 weeks later at 5:00 PM via Gradescope (unless otherwise noted). All homework assignments must be submitted as **a single PDF file.** Homework will be a combination of analytical and computational exercises done "by hand" and data analysis using the computer. Handwritten submissions must be properly organized and easily readable to the grader. Computational problems should include all code with neatly presented tables and figures if necessary. Any submissions that are difficult for the grader to read are subject to a penalty for style/readability. The GSI will provide further instructions for homework submissions in lab sections.

Quizzes:

We will give 4 in-person quizzes **in lab** during the semester, on **February 9**, **February 23**, **April 13**, and **April 27**. These will test concepts from the homework due the week before; they will be closed-book, multiple choice, and last less than an hour. You **must attend your assigned section** on quiz days to ensure there are enough seats for everyone. I do not plan to offer makeup quizzes but your lowest quiz score will be dropped so missing one will not impact your grade.

Exams

An in-person cumulative midterm exam will be given from **7:00 to 9:00 PM on Thursday March 12th** and an in-person cumulative final exam will be given from **7:00 to 10:00 PM** on **Monday May 11th**. I do not plan to offer makeup times so **please confirm now** that you can attend both exams. You will be allowed to bring one (double-sided) page of notes with you into each exam; no other materials will be allowed.

Final project

Students will work in groups of one or two to carry out the final project, a regression analysis or related methodological investigation on a topic of your choice. A written project proposal will be due via Gradescope in April, and the final project report will be due via Gradescope during finals week.

Overall score

Your letter grade for the course will be based on the total points for all work in the semester, as follows:

- Homework (each assignment weighted equally): 14%
- Quizzes (each weighted equally, dropping lowest score): 18%
- Midterm: 15%
- Final: 30%
- Group project (including proposal): 23%

Your final exam grade will "clobber" your midterm score, in the sense that if you do better on the final than on the midterm the final exam score will replace the midterm score.

I do not plan to curve grades. Students scoring 90% or above overall will receive letter grades in the A-range, students scoring 75%-90% will receive letter grades in the B-range, and students scoring 60%-75% will receive letter grades in the C-range.

Online Resources

Course Website

https://stat230a.berkeley.edu/spring-2026. Contains information from this syllabus and links to the resources below.

bCourses

Homework assignments and notes from lecture and lab will be posted here. I will also make course announcements through bCourses.

Ed Discussion

This is an online forum (linked from the course website and from bCourses) to ask questions to fellow students and course staff and to answer other students' questions.

Gradescope

Homework assignments and regrade requests (see Policies section below) will be submitted through Gradescope, and graded exams will be posted here. This is also where you can view your grades on past assignments.

Policies

Possibility of revisions to course policies

All course policies, including assessment, are subject to change during the course of the semester in response to unforeseen events including but not limited to public health directives, natural disasters, and medical emergencies among members of the course staff.

Late Assignments

All students will have 5 late days that they may use for turning in homework after the due date. This will take the place of any extensions due to sickness or conflicts, unless there are extenuating circumstances, so use them wisely. Late days will automatically be deducted if an assignment is submitted past the deadline. **At most two late days can be used for a single assignment and submissions beyond this period will not be accepted.**

Submissions that lack the adequate number of late days will not be accepted. You will be notified when you are out of late days. Late days cannot be used for the group project, and they cannot be requested once homework solutions are posted. Late days are counted at the 24-hour-period level after the time the assignment was due. E.g., if the first assignment is due at 5pm, January 30 and you turn it in at 7pm, January 30, it counts as use of one late day; if you turn it in at 5:30pm, January 31, it counts as use of two late days.

Regrade requests

Regrade requests on an assignment are **due within one week of the release of the graded assignments and the solutions** (if applicable). Regrade requests should be submitted through Gradescope. Requests via email will not be answered. In writing a regrade request, please be specific about the nature and exact location of the error you feel the grader has made, with reference to the solutions if available.

Academic Honesty and AI Policy

The student community at UC Berkeley has adopted the following Honor Code: *"As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others."* My expectation is that you will adhere to this code. Beyond the importance of respecting your fellow students, acting with integrity in completing course assignments helps ensure that they achieve their purpose, which is to help you learn and develop valuable statistical understanding and skills.

- Homework must be done independently. If you get stuck or want to explore alternative approaches, feel free to discuss issues with students or course staff (including on the online forum); however, you may not do the homework jointly, nor may you ask for or share complete code or solutions. Sharing solutions or obtaining and/or using solutions from previous years or from the Internet, if such are available, is considered cheating.
- On all written assignments, including the homework, you should include a section listing all the sources you drew on in producing your answers; on the homework, you should also list the names of other students with whom you consulted.

Anyone caught cheating will be given a score of zero (0) on the assignment/exam and will be reported to the University's Office of Student Conduct.

Use of generative AI tools

This course allows limited use of generative AI tools (GenAI) such as ChatGPT and Google Gemini on homework assignments and projects. For example, GenAI may be used to perform research in ways similar to search engines such as Google. It may be used to support coding tasks, including by generating code chunks to solve specific issues that arise in the data analysis process, and it may be used as a writing assistant in its capacity as a word processor, such as Word or Pages, i.e., for correcting grammar and spelling, and other functions like synonym suggestion. Please keep in mind the following principles when using GenAI:

- GenAI should never be employed for a use that would constitute plagiarism if the GenAI source were a human or organizational author. For example, you should not be using

GenAI to generate large chunks of text and copying them verbatim into your project report.

- If you use GenAI on a homework or project, you should include a statement acknowledging your use and share the specific ways in which it was used. The following template may be helpful:

  I acknowledge the use of [insert AI system(s) and link] to [specific use of GenAI]. The prompts used include [list of prompts]. The output from these prompts was used to [explain the use].

- Remember that over 60% of your grade comes from in-class quizzes and exams during which you will have no access to GenAI, and that homework is provided primarily to help you prepare for those assessments. Make sure that you are using homework to help yourself master the material.

Email

1) If you wish for your email to make it into my inbox, the subject of your email must contain the text "**230A**."
2) Neither I nor the GSI explains course material over email and we will not respond to emails with such requests. Please use the online forum, office hours, and lab (or schedule another time to meet if you have irreconcilable conflicts with the office hours).
3) I respond to email regarding the class roughly once a day, and rarely during the weekend.

Inclusivity and Accommodation

My hope is to establish a learning environment in this course that welcomes diversity of thought, perspective, and experience, and to be respectful of your individual identity as a student. I am happy to use your preferred name and/or personal pronoun. If you feel uncomfortable due to anything that is said in class, or if you feel that your performance in the course is being impacted by experiences outside of class, please do not hesitate to reach out to me about your concerns.

In addition, if you need accommodations for any physical, psychological, or learning disability, please speak to me after class or during office hours. Please note that **you must make arrangements in a timely manner through DSP** so that I can make the appropriate accommodations.

Acknowledgments

Most of the materials used in this course, including this syllabus, are close adaptations from materials originally created or compiled by Prof. Peng Ding and generously provided for the current semester. In writing this syllabus I also adapted content from Prof. Chris Paciorek, from Prof. Monica Linden of Brown University, and from the Berkeley Academic Senate AI Working Group.

Anticipated Course Schedule

Week	Topics	Assignments Due & Exams	Assigned Reading (from lecture notes)
Jan 19	(MLK Day) Motivation, linear algebra review		Ch. 1, Appendix A-B
Jan 26	Distribution theory review, multiple regression		Ch. 2-3, Appendix C
Feb 2	Gauss-Markov, normal linear model	HW#1 due Monday	Ch. 4-5
Feb 9	Asymptotic inference, Frisch-Waugh-Lovell	Quiz #1 Monday	Ch. 6-8
Feb 16	(President's Day) Cochran's Theorem, R^2	HW#2 due Monday	Ch. 9-10
Feb 23	Leverage, population least squares	Quiz #2 Monday	Ch. 11-12
Mar 2	Overfitting, model selection	HW#3 due Monday	Ch. 13
Mar 9	Ridge regression Midterm review	Midterm (Thursday)	Ch. 14
Mar 16	LASSO, transformations, interaction.	HW#4 due Friday	Ch. 15-17
Mar 23	Spring break		
Mar 30	Weighted/restricted OLS, logistic regression		Ch. 18-20
Apr 6	Categorical and count outcomes	HW#5 due Monday Project proposal due Friday	Ch. 21-22, Appendix D
Apr 13	Generalized linear models	Quiz #3 Monday	Ch. 23-24
Apr 20	Generalized estimating equations, Quantile regression	HW#6 due Monday	Ch. 25-26
Apr 27	Survival outcomes Guest speakers	Quiz #4 Monday HW#7 due Friday	Ch. 27
May 4	RRR week		
May 12	Finals week	Final (Monday) Project due Wednesday	