STAT 151A: Linear Modeling: Theory and Applications

Lectures: Mon, Wed, Fri 11:00am-12:00pm
9 Lewis

Lab: Wed 9am –11am, 344 Evans
Wed 2 pm – 4 pm, 344 Evans

Bring your laptop. If you do not have access to a laptop, 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.

Instructor: Sam Pimentel
Office: 429 Evans Hall
Email: spi@berkeley.edu
Preferred pronouns: he/him
Office Hours:
Thursdays 4-5 pm, 429 Evans
Fridays 12-1 pm, 429 Evans

GSI: Florica Constantine
Email: florica@berkeley.edu
Preferred pronouns: she/her
Office Hours:
Tuesdays 9 am - 12 pm, 428 Evans
Wednesdays 12-1 pm, 428 Evans

Topics:

- Regression modeling
- Visualization and diagnostics
- Confidence Intervals
- Hypothesis Testing
- Generalized Linear Models
- Model selection and shrinkage
- Generalized linear models
- Nonlinear approaches

Textbook:

Applied Regression Analysis and Generalized Linear Models, J. Fox, 3rd Edition. Electronic course reserve available for 2-hour checkout; see user guide for electronic course reserves.

Selected readings from:

Statistical Models: Theory and Practice, D. Freedman. Available electronically via Berkeley Library.

An Introduction to Statistical Learning with Applications in R, G. James et al. Available online.

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 the purposes and benefits of linear modeling in common applied contexts.
2. Design and conduct regression analyses in R for common data settings.
3. Interpret statistical models, estimates of model parameters, and inferences correctly.
4. Evaluate the quality of a regression analysis and suggest improvements.
5. Communicate the process and results of a data analysis simply and clearly for a broad audience, using well-organized prose and code and effective data visualizations.

Prerequisites:

STAT 135. STAT 133 is strongly recommended. STAT 135 implies other prerequisite courses (**STAT 134 and its prerequisites**). In particular, you must have had linear algebra, so you should be familiar with basic matrix operations, vector subspaces and projections, rank and invertibility of matrices, and quadratic forms. This will form a core component of the course and is a real requirement. We also assume **familiarity with R**, and unless otherwise noted assignments involving computing must be completed in that language.

Lecture:

Lectures will cover core theory and concepts, with supporting data analysis examples. To get the full benefit of lecture, it is best to read the supporting material ahead of time. I encourage active engagement and discussion during lectures, and I will frequently pose questions and call on students to answer.

During lectures I will primarily write notes on an iPad and project them on the screen; occasionally I will also use slides or other materials or do live coding demonstrations in R. Copies of the written lecture notes and any slides or R files used in lecture will be posted on bCourses following class.

While I will not take attendance, lecture recordings will not be posted so attending in-person is vital to your success in the course.

Lab:

Lab time will be spent working on practice problems and data analysis in R, and you should plan to bring your laptop. You may attend a lab for which you are not enrolled (physical space permitting), **except on days when a quiz is given**. Slides and other materials for lab will be posted on bCourses, but in-person attendance is strongly recommended.

Assessment:

Homework

We anticipate giving five homework assignments during the semester. Homework will be posted to bCourses, and will generally be due 2 weeks later. All homework is due **via Gradescope** (linked through bCourses) unless otherwise noted. Homework will be a combination of analytical and computational exercises done "by hand" and data analysis using the computer.

Exams

Three in-class quizzes and an in-person, proctored final exam will be given. The final exam time is **Monday December 15, 11:30 AM – 2:30 PM**. For each quiz or exam you will be permitted to use one double-sided page of notes but no other written materials or electronic devices.

Final project

Students will work in groups of three to carry out the final project, a regression analysis on a research question of your choice. A written project proposal will be due by **11:59 PM PST on Friday, November 14th**, and the final project report will be due in Gradescope by **11:59 AM PST on Friday, December 19** (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): 10%

- Quizzes (each quiz weighted equally): 36%
- Final exam: 30%
- Group project (including proposal): 24%

Your final exam grade will "clobber" your individual quiz scores, in the sense that if you do better on the final than on any individual quiz the final exam score will replace that quiz score. For this reason **no makeup quizzes will be given**.

Grades will not be curved. 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

bCourses

Homework assignments, grades, and material from lecture (where applicable) will be posted here. I will also make course announcements through bCourses.

Ed Discussion

I have created an Ed Discussion site for this course, which you can access through the link in bCourses. This is an online forum to ask questions to fellow students and course staff. Involvement in the discussion on Ed Discussion will factor into your course participation score.

Gradescope

Homework assignments, take-home exams, and regrade requests (see Policies section below) will be submitted through Gradescope, which you can also access through the link in bCourses.

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 emergencies, power outages, forest fires, 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. To use a late day, **you must submit a Google Form at this link requesting a late day before the homework is due**, or you risk receiving a large penalty or a zero. Late day requests by email will not be answered. Late days cannot be used for the group project, and a maximum of three late days can be used on any one assignment.

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. 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 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 Ed Discussion); 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.
- During the take-home midterm, you must not consult with any other person besides the course staff, although you will be allowed free use of books, class notes, and online resources.
- 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 approximately 2/3 of your grade comes from in-class exams during which you will have no access to GenAI, and that homework in particular is provided primarily as a way to help you prepare for those exams. Make sure that you are using homework as a way 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 "**151A**."
2) Neither I (nor the GSIs) explain course material over email and will not respond to emails with such requests. Please use Ed Discussion, office hours, discussion section, or GSI's office hours (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 as a result of 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 Profs. Deborah Nolan and Aditya Guntuboyina 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 Fox unless otherwise noted)
Aug 27	Course intro, transformation		Ch 3, 4.1-4.3, 4.5
Sep 1	(no lecture Monday) Transformation and simple regression		Ch 5.1 Freedman Ch. 1
Sep 8	Multiple regression, geometric perspective		5.2, 10.1-10.2 Freedman 2.3-2.4
Sep 15	Probability model for multiple regression, collinearity	HW #1 due Friday	10.3
Sep 22	Quiz review Statistical inference	Quiz #1 Wednesday	6.1-6.2
Sep 29	Categorical variables as predictors	HW #2 due Friday	9.2 (skip 9.2.1), 9.3.1-2, 9.4.1-3
Oct 6	ANOVA	Form project groups by Friday	Ch. 7 (skip 7.2.1), 9.1, 9.2.1, 10.4
Oct 13	Bootstrap	HW #3 due Friday	21.1- 21.4 (skip 21.2.3)
Oct 20	Quiz review Influential Observations, Diagnostics	Quiz #2 Wednesday	11.1—11.5 (skip 11.3.2), 11.7-11.8.2, 12.1-12.2 (skip 12.1.1, 12.2.2)
Oct 27	Interpreting Models Model Selection		13.2.2, 22.1 (skip "Closer look at AIC," "Closer Look at BIC")
Nov 3	Shrinkage methods	HW#4 due Friday	13.2.3 James et al. 6.2 (skip "Bayesian interpretation")
Nov 10	Logistic Regression	Project proposal due Friday	14.1
Nov 17	Binomial logistic model, polytomous outcomes, GLMs	Quiz #3 Wednesday	14.2,1 4.3 (skip p. 396)
Nov 24	Thanksgiving Break (no lecture on Wed or Fri)		
Dec 1	Nonlinear regression, regression trees	HW #5 due Friday	James et al. 7.1-7.4, 7.7, 8.1
Dec 8	RRR week		
Dec 15		Final exam: Monday Dec 15, 11:30am-2:30pm PST. Final projects due: 11:59 AM, Friday Dec 19	