STAT 151A: Linear Modeling: Theory and Applications

Lectures: Tuesday, Thursday 9:30-11am
213 Wheeler

Lab: Wed 9-11am
Wed 11-1pm
332 Evans

Instructor: Elizabeth Purdom
Office: 433 Evans Hall
Phone: 642-6154 (office)
Email: epurdom@stat.berkeley.edu
Office Hours: M 3-5pm (start 1/27)

GSI: Steven Pollack
Email: steven.pollack@berkeley.edu
Office Hours: W 2-4pm

Topics:
- Regression modeling
- Visualization and diagnostics
- Confidence Intervals and Hyp. Testing
- Generalized Linear Models
- Large numbers of predictors

Textbook:
Applied Regression Analysis and Generalized Linear Models, J. Fox

Additional readings will be posted on bspace

Prerequisites:

STAT 135. Statistics 133 is recommended. STAT 135 implies other prerequisite courses (134 and **its** prerequisites). In particular, you must have had linear algebra, so you should be are familiar with: basic matrix operations, vector subspaces and projections, rank and invertibility of matrices, quadratic forms. This will form a core component of the course and is a real requirement.

Lab:

Lab time will be spent working on practice problems, and how to analyze the data in R.

We assume familiarity with R. If you do not have experience with R, we will make online introductions to the programming language available (see resources on bspace).

Class Activities and Attendance:

Past student evaluations have said that attending lecture was critical for doing well in my class, and this is exactly what I expect. To get the full benefit of the class, you should read the supporting material -- ahead of time, if possible.

I encourage active engagement in the class, and frequently will pose questions and call on people during class. I will also occasionally ask for you to answer questions on an index card in class and turn in the card or otherwise use them during class.

Handouts given in class deal with in-class activities and discussions and will generally **not** be posted online, unless they cover material not in the book.

Course Work:

Homework

Homework will be posted to bspace, and will generally be due 1-2 weeks later. All homework is due **in lecture**. Homeworks will be a combination of computational exercises and data analysis using the computer.

The final homework score will be the sum of all homework grades, so 15 points in HW1 counts the same as 15 points in HW5, with points allocated to questions across the semester that aim to be compatible. This means the homework assignment as a block will not count equally, and it is always worth your while to turn in what you have done of your assignment, even if it is incomplete.

Exams

There will be two midterms, **tentatively** scheduled for the week of Mar 3, and April 14th, each of which will count equally.

Final Project

I am planning on creating a final project/take-home exam where you will have to analyze a dataset from scratch using all the tools of the class, as appropriate. This is still in flux, and I may convert it to an in-class final exam.

Overall score

Your letter grade for the course will be based on the total points for all work in the semester.
- Homework: 20%
- Final Project/Exam (1): 35%
- Midterms (2): 45%

Policies

Late Assignments

Late assignments will not be accepted except with my personal approval, which will general result in a penalty except in extraordinary circumstances. I expect any requests to be done as timely as possible.

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 GSI) explain course material over email and will not respond to emails with such requests. Please come to 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 almost never in the evening nor weekend.

Academic Honesty Policy
- Homework must be done independently. Obtaining and/or using solutions from previous years or from the internet, if such are available, is considered cheating. You may discuss issues about the homework with other students, but you must **not** sit down and do the assignment jointly.
  Please note that while the homework is time-consuming, they are *en masse* 20% of your grade; the cumulative weight is large, but becoming desperate over single questions is **not** worth the risk of cheating!

- For exams cheating includes, but is not limited to, bringing written or electronic materials into an exam or quiz, using written or electronic materials during an exam or quiz, copying off another person's exam or quiz, allowing someone to copy off of your exam or quiz, and having someone take an exam or quiz for you.

In fairness to students who put in an honest effort, cheaters will be harshly treated. Any evidence of cheating will result in a score of zero (0) on the **entire** assignment or examination. I will **always** report incidences of cheating to Student Judicial Affairs, which may administer additional punishment.

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

Syllabus

The following is only a guide, and there is likely to be slight variation as the semester progresses. The reading described below is a guide to where the relevant material can be found in the book for the subjects described under 'Topic.' It is not a prescription to when **you** should actually read the material.

Week	Topics		Assigned Reading
Jan 20	Course Logistics Exploratory Data Analysis Simple Regression		Ch 1-4, 5.1
Jan 27	Multiple Regression		Ch 5.2, 6
Feb 3	Inference: hypothesis testing		Ch 6
Feb 10	Boostrap Inference		Ch 21
Feb 17	Matrix representation		Ch 9.2-9.4
Feb 24	Dummy Variables and ANOVA		Ch 7, 8 9.1.1-9.1.2
Mar 3		Midterm	
Mar 10	Regression Diagnostics		Ch 11,12
Mar 17	Model Building & Interpreting Models		Ch 13.2, 22
Mar 24	SPRING BREAK		
Mar 31	Large number of Predictors		Ch 13.1
Apr 7	Logistic Regression		Ch 14.1
Apr 14	GLMs & Exponential families	Midterm	Ch 14.2, 15
Apr 21	GLM Diagnostics, Inference		
Apr 28	Non-parameteric Regression		
May 5 (RR)			
	Final Exam/Project: May 14, 11:30-2:30pm		