Spring 2021: Stat 230A Linear Models

Peng Ding
Email: pengdingpku@berkeley.edu

- Time: MW 5:00 pm–6:30 pm
- Office hours: MW 6:30 pm–7:30 pm after the lectures

  Zoom Meeting for both lectures and office hours (with your Berkeley zoom account):

  https://berkeley.zoom.us/j/93334380294?pwd=R3Q0MmNKd2wraGI3U2pJcmpvRzNEZz09

  Meeting ID: 933 3438 0294

  Passcode: gauss
- GSI: Sizhu Lu, Ph.D. student from Haas; Email: sizhu_lu@berkeley.edu
- Lab sessions: The GSI will give detailed instruction on the lab sessions, as well as her office hours, the homework submission, and the final project.

1. Course outline

The course covers the core materials of linear model including ordinary least squares (OLS), Gauss–Markov theorem, finite-sample exact and asymptotic inferences, heteroskedasticity and cluster robust standard errors, model checking, and model selection. It also covers important extensions of the classic linear model including ridge regression, lasso, generalized linear models (GLM), and generalized estimating equation (GEE). If time permits, it will also cover models beyond the conditional mean including quantile regression and the Cox proportional hazards model.

Textbook No textbook; lecture notes from the instructor, available on Bcourses.

You do not need to print the notes out in advance because they will be updated over the semester.

1.1. Theoretical foundations: linear algebra and some distribution theory

- Vector space and matrix algebra: rank, determinant, trace, and eigen-values/vectors of a matrix, eigen-decomposition, projection matrix, quadratic forms, positive (semi-)definite matrix, inverse of a block matrix, and Sherman–Morrison formula
- Multivariate normal and related distributions
- Limiting theorems: law of large numbers and central limit theorems

1.2. Classic theory of linear models

- Fixed design, sample least squares, Gauss–Markov Theorem
- Interpretation of least squares coefficients: Frisch–Waugh–Lovell Theorem, Cochran's formula
- Linear model with Gaussian errors, finite sample inference, t and F tests, analysis of variance, model checking
- Asymptotic inference in linear model with fixed design: heteroskedasticity-robust standard error
- Population and sample linear projections, Eicker–Huber–White standard error for misspecified linear models
- Transformations and dummy variables

1.3. Modern theory of linear models

- Linear predictor, model selection, cross-validation
- Shrinkage estimators: ridge, lasso, and elastic net

1.4. Generalized linear model and generalized estimating equation

- Binary data: logistic/Probit regression, case-control study
- Multinomial data: nominal and ordinal logistic regressions
- Count data: Poisson, Negative-Binomial, Zero-inflated Poisson, Zero-inflated Negative-Binomial regressions
- Models for repeated measurements/longitudinal outcomes: marginal model, generalized estimating equation, cluster-robust standard error

1.5. Models beyond conditional means—if time permits

- Quantile regression
- Cox proportional hazards model

2. Prerequisites

- Masters' level probability and statistics or beyond. For example, you should be familiar with the basics of probability theory in Aditya's notes:

  `https://www.stat.berkeley.edu/~aditya/resources/AllLectures2018Fall201A.pdf`
- Linear algebra. Please make sure you understand the appendix of linear algebra in my notes. You can also watch the youtube videos by Gilbert Strang: MIT 18.06 Linear Algebra, Spring 2005.
- R programming: The examples in class and homework problems are all in R.

3. Course requirements

The GSI will give guidance for the submission of homework, exam, and final project.

- Homework: 30% ($\approx$ 7 or 8 assignments, drop the one with the lowest score)
- Mid-term exam: 20%; take home, the week of March 15th–19th (before the spring break)
- Final project: 50% (group project with one or two group members; submission of the proposal during the week of March 29–April 2, after the spring beak; submission of the final project during the week of May 10–14)

Late Assignments: Due dates will be strictly enforced.
Academic Integrity: When solving the homework problems or writing the final project, you are welcome to discuss with each other or consult any materials online. But you should write up the solutions and final project by yourself. Importantly, you should acknowledge your collaborators and cite the references you used.