STAT 961: Statistical Methodology
Fall 2021

Course and Instructor Information

Course information

Classroom:	Steinberg-Dietrich Hall, Room 105
Class time:	Tue/Thu 10:15-11:45am
Canvas:	https://canvas.upenn.edu/courses/1597407
Github:	https://github.com/katsevich-teaching/stat-961-fall-2021
Piazza:	https://piazza.com/upenn/fall2021/stat961
Gradescope:	https://gradescope.com/courses/285243

Instructor: Eugene Katsevich

Office:	311 Academic Research Building
Email:	ekatsevi@wharton.upenn.edu
Office hour:	Thu 12:30-1:30pm

Teaching assistant: Hua Wang

Office:	TBA
Email:	wanghua@wharton.upenn.edu
Office hour:	TBA

Course Description

This goal of this course is to build a PhD-level foundation in frequentist statistical methodology, focusing on hypothesis testing and estimation in linear and generalized linear models. Important special cases will be considered, including one- and two-sample tests, analysis of variance, logistic regression, Poisson regression, and contingency table analysis. Most of the inferential tools covered will rely on parametric model assumptions, but non-parametric and robust alternatives will also be covered; these include the bootstrap, permutation tests, and rank-based methods. The additional topics of multiple testing, linear mixed models, and penalized regression will be covered to the extent time permits. Students will learn the theoretical basis of these methodologies as well as how to apply them in practice using the programming language R.

Prerequisites

STAT 961 is a fast-paced, PhD level course that requires significant prior preparation. Students are expected to have the following prerequisites:

- *Linear algebra* at the level of MATH 312 (including bases, vector spaces, inner products, orthogonal projections, and matrix decompositions)
- *Probability* at the level of STAT 430 (including random variables, probability distributions, multivariate normal random variables, and the central limit theorem)

- *Statistical inference* at the level of STAT 431 (including hypothesis testing, p-values, confidence intervals, and maximum likelihood estimation)
- *Programming* experience in R

List of topics
(tentative and subject to change)

Linear models
- Lease squares estimation: normal equations, geometric interpretation via projections, Gauss-Markov theorem, linear model examples, decomposition of variance, orthogonalization, partial correlation
- Inference in linear models: Normal random variables, hypothesis testing, confidence intervals, prediction intervals, power, collinearity
- Model-checking, model misspecification, and robust alternatives: Residual plots, leverage and influence, Huber-White estimator, pairs bootstrap, permutation tests, Wilcoxon test

Generalized linear models
- Exponential family distributions
- Maximum likelihood estimation for GLMs and iteratively reweighed least squares
- Testing and estimation in GLMs, deviances, goodness of fit
- Special cases of GLMs: Logistic regression, multinomial logistic regression, Poisson regression, negative binomial regression

Additional topics (if time permits)
- Linear mixed models
- Multiple testing
- Penalized regression and cross-validation

Logistics

- Course materials (lecture notes, homework, exams) will be distributed via Github. This private repository is accessible to enrolled students; see Canvas for instructions to join.
- Students will submit and receive feedback on homework and exams through Gradescope.
- The instructor and teaching assistant will hold office hours every week (times listed on the first page). Outside of office hours, students can ask questions about the course on Piazza. Students are encouraged to answer each others' questions, for which the instructor may award extra credit. Students should email the instructor with administrative questions.
- Students will use R, RStudio, knitr, LaTeX, Git, and Github to complete assignments and exams. Instructions to set up these tools are available on the course Github page.

Assignments and Exams

Assessment is based on homework, a midterm exam, and a final exam.

Homework (30%)
There will be five homework assignments, distributed roughly once every two weeks. These assignments will involve mathematical, programming, and conceptual questions. Students are permitted to work together on homework assignments, but solutions must be written up and submitted individually. Students must disclose any sources of assistance they received;

furthermore, they are prohibited from verbatim copying from any source and from consulting solutions to problems that may be available online and/or from past iterations of the course.

Midterm exam (30%)
The midterm exam will be a take-home exam structured like a homework assignment, consisting of mathematical, programming, and conceptual questions. This exam is open-book and open-notes but students must complete the exam individually. **The time limit for the midterm is 24 hours; it will be released on Friday, October 22 at 9am EST and will be due on Saturday, October 23 at 8:59am EST.**

Final exam (40%)
The final exam will be a take-home exam structured like a homework assignment, consisting of mathematical, programming, and conceptual questions. This exam is open-book and open-notes but students must complete the exam individually. **The time limit for the final is 24 hours; it will be released on Wednesday, December 15 at 9am EST and will be due on Thursday, December 16 at 8:59am EST.**

Course Grades

An overall numeric grade will be computed for each student at the end of the semester by weighting the homework, midterm, and final according to the above percentages. Letter grades will then be assigned based on numeric grade thresholds chosen at the discretion of the instructor.

Course Textbooks

Our primary textbook (required) is
  • Alan Agresti. *Foundations of Linear and Generalized Linear Models.* 2015.

This textbook is available for purchase at the Penn Bookstore and is on reserve at Lippincott Library.

A helpful reference for R programming (optional) is
  • Hadley Wickham and Garrett Grolemund. *R for Data Science.* 2016.

This textbook is freely available online.
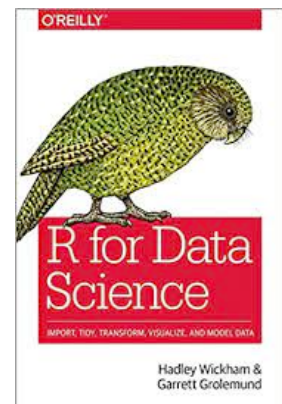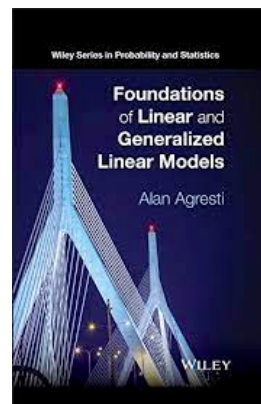
Course Policies

Late homework
To offset the effect of relatively common difficult circumstances (computer crash, job interview, PDF compilation problem), **each student will get three "free" late days for homework submission over the course of the semester. No late penalty will be assessed for these three late days, with no need to request or justify this accommodation.** After a student has used his or her late days, each additional late day will come with a penalty of 10 points (out of

100). No homework will be accepted more than three days after the deadline. Lateness will be determined by the Gradescope timestamp and measured in whole days. Exceptions to this policy will be provided to students encountering major unforeseen circumstances (e.g. family emergencies), provided they obtain a letter from their academic advisor or a departmental representative.

Exam makeups

Students unable to take the midterm or final exam at the scheduled time should notify the instructor as soon as possible, and makeups will be offered at the discretion of the instructor. A foreseen conflict (e.g. another class has an exam scheduled at the same time) must be corroborated with evidence of the conflict and an unforeseen conflict (e.g. family emergencies) must be corroborated with a letter from an academic advisors or departmental representative.

Regrades

All assignments will be graded through Gradescope, where points will be awarded or deducted based on clear rubrics. Regrade requests, which can also be submitted through Gradescope, will be considered only in cases when there is a clear discrepancy between the rubric and the grade. **A regrade request must be submitted within a week of the date the grade was posted.**

Academic integrity

In accordance with Penn's Code of Academic Integrity, students must comply with the course collaboration policies described in this syllabus and in the assignment instructions. **All academic integrity violations will be reported to the Office of Student Conduct and all assignments where violations occurred will receive grades of zero.** If you have any questions about collaboration policies, please do not hesitate to contact the instructor.

Accessibility for students with disabilities

The instructor is committed to creating a learning experience that is as accessible as possible. Students with disabilities should reach out to the Office of Student Disabilities Services (SDS) by calling 215-573-9235 (services are confidential) and email the instructor. The instructor will then work with the student and SDS to provide reasonable accommodations.