Statistics 139 Syllabus
Fall 2021

Instructor: Kevin Rader
Email: krader@fas.harvard.edu
Office Hours: Mon, 12-1pm, Wed 3-4pm (SC-300D)

Lectures: Mon & Wed 1:30-2:45pm. Location: SC-Hall E
Lectures (and labs) will be recorded and posted online within 24 hours.

Labs: Fri 1:30-2:45pm. Location: SC-Hall E

Web Site: https://canvas.harvard.edu/courses/89311

Textbook: *Linear Models with R*, Faraway, 2nd edition.
(recommended) (Amazon Link: https://www.amazon.com/Linear-Models-Chapman-Statistical-Science/dp/1439887330)
Other Useful Texts:
Regression and Multilevel/Hierarchical Models, Gelman and Hill.
The Statistical Sleuth, Ramsey and Schafer.

Software: R and RStudio, both free for download:
R: http://lib.stat.cmu.edu/R/CRAN/
RStudio: https://www.rstudio.com/products/rstudio/download/
We will also offer a course-specific workspace on https://rstudio.cloud/

Course Objective:
An in-depth introduction to statistical methods with linear models and related methods. Topics include group comparisons (t-based methods, non-parametric methods, bootstrapping, analysis of variance), linear regression models and their extensions (ordinary least squares, ridge, LASSO, weighted least squares, multi-level models), regression trees and random forests, model checking and refinement, model selection/comparison, and cross-validation. The probabilistic basis of all methods will be emphasized.

By the end of the course, students should be able to evaluate the strengths and weaknesses of a variety of statistical technique. Given a data set, students should be able to

- state hypotheses,
- explore the data using statistical software,
- determine which statistical model may be appropriate,
- apply corresponding hypotheses tests,
- check the assumptions behind these tests and models,
- interpret the results of the analysis to draw conclusions about the hypotheses.

This course is designed to prepare students for further coursework in statistics (such as Stat 131, Stat 140, Stat 149, Stat 160, Stat 183, Stat 186, and others) or for drawing conclusions from data in any field.

Prerequisites:
Mathematics 21a and 21b or equivalent, and Statistics 110 (Multivariable Calculus, Linear Algebra, and Theoretical Probability). Statistics 111 (Theoretical Inference) is **highly** recommended; having taken Statistics 104 or 109 will suffice. Concurrently taking Math 21b is allowed.

Lecture Format:
Lectures will be a *half-flipped* classroom: new content will be presented for roughly half the time (~30-40 min) via annotated lecture slides, and the other half will be working through handout problems (a mix of conceptual, mathematical, and *R*-based problems) in small groups. Friday Labs will have even more emphasis on the practice problems. Active participation during these group work sessinos.

Sections:
Optional (but **strongly** suggested) TF-led sections will be held throughout the course. Sections schedule will be announced on Canvas and will be spread out across various times to accommodate students' schedules. Sections will go over practice problems and review difficult material. One section (or more) will be recorded.

Office Hours:
Office Hours will be held at various times of the day to accommodate students/ schedules. There will also be evening *study breaks* in which much of the teaching staff will be present to help you on the HW assignments.

Discussion Forum: Ed: https://us.edstem.org/. You will be able to link through the Canvas page.

Computing:
We will be heavily relying on the statistical software package, *R*, along withusing *R studio* and R Markdown. *R* is available to download for free for both Macs and PCs (and on Unix) here: http://cran.us.r-project.org/. *R* Studio can be downloaded here: http://www.rstudio.com/. General familiarity with *R* is required, but no other coding experience is necessary (though is VERY helpful). Stat 111 will be enough background in *R*. These basic tutorials can be helpful for getting you started or as basic review:
DataCamp: https://www.datacamp.com/courses/free-introduction-to-r
Rstudio.com: https://education.rstudio.com/

Accommodations for students with disabilities:
Students needing academic adjustments or accommodations because of a documented disability must present their Faculty Letter from the Accessible Education Office (AEO) and speak with Kevin by the end of the third week of the term: Friday, September 17. Failure to do so may result in us being unable to respond in a timely manner. All discussions will remain confidential.

Collaboration:
You are encouraged to discuss homework with other students (and with the teaching staff), but you must write your final answers yourself, in your own words. Solutions prepared "in committee" or by copying or paraphrasing someone else's work are not acceptable. All computer output you submit must come from work that you have done yourself. **Please indicate on your problem sets the names of the students with whom you worked.** All exams are individual work.

Homework:

There will be ~7 homework assignments and are due on a Wednesday or Friday at 11:59pm (due every week and a half, approximately). The assignments will be posted on the course website 7-10 days before they are due. Each HW can be done completely using R-markdown. No HW scores will be "dropped." You are allowed **2** late homework days throughout the semester, and at most 1 day (24 hours) can be used on any one assignment. Any other late HW submissions will not be accepted without a note from UHS or your resident dean's office.

Project:

A group project will be due during reading period. It will be based on a data analysis of your choice and will result in a 6-8 page paper. More details to come in October. The group project will be due during reading period (Dec 4-9).

Exams:

There will be one take home midterm (due Wed, Oct. 27), and one take-home final exam (tentatively due Thurs, Dec. 16). The midterm and final exam will be open-notes and open-book and will include coding in R. Both are roughly one-week exams.

Grading Guidelines:

Your final score for the course will be computed using the following weights. Letter grades will be determined based on the "standard" cut-offs: for example, 89.5 will be guaranteed at least an A-. Curving is up to the teaching staff's discretion.

Component	Weight
Problem Sets	35%
Midterm	20%
Final Exam	30%
Group Project	15%
Total	100%