

# Statistics 139 Syllabus

## Spring 2017

- Instructor: Michael Parzen  
Email: [mparzen@fas.harvard.edu](mailto:mparzen@fas.harvard.edu) or [michaelparzen@gmail.com](mailto:michaelparzen@gmail.com)  
Office: SC-300B. Office Hours: Monday & Wednesday, 12-1pm.
- Lectures: Monday & Wednesday 9:30-11am in Geological Lecture Hall  
Due to the lecture location, lectures will **not** be videotaped.
- Web Site: <https://canvas.harvard.edu/courses/21564>
- Textbooks: No required textbook-several references will be used and electronically available on the course web site.
- Software: R, free for download:  
Download Link: <https://www.r-project.org/>

### **Course Objective:**

An in-depth introduction to statistical inference with linear models and related methods. Topics include *t*-tools and permutation-based alternatives including bootstrapping, multiple-group comparisons, analysis of variance, linear regression, model checking and refinement, and a little about generalized linear models (GLM), like logistic regression. Emphasis is made on thinking statistically, evaluating assumptions, and developing tools for real-life applications.

By the end of the course, students should be able to evaluate the strengths and weaknesses of a variety of statistical techniques appearing in the media, scientific literature, or students' own work. 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 100/102/104/110 or equivalent (Multivariable Calculus, Linear Algebra, and Intro Statistics...AP stats is fine). Concurrently taking Math 21b is fine.

### **Sections:**

Optional (but **strongly** suggested) TA-led sections will be held throughout the course. Sections will mostly meet on Wednesday and Thursdays. Sections will go over practice problems and review difficult material.

**Computing:**

We will be heavily using the statistical software package, *R*. *R* is available to download for free for both Macs and PCs (and on Unix) here: <http://cran.us.r-project.org/>

Some students prefer to use *R* through the interface *R Studio*. *R studio* is not required, but it does help organize your computing projects. It can be downloaded here: <http://www.rstudio.com/>

No previous knowledge of the computer programming or the software is required; though having some programming experience would be very helpful, like having taken CS 50 or Stat 102, 107, or 135.

**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 Mike by the end of the third week of the term. 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 instructor and TAs, of course), 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; your handed-in assignment must represent your own thoughts. All computer output you submit must come from work that you have done yourself. All exams (midterm and final) are individual work.

**Grading Guidelines:**

Your final score for the course will be computed using the following weights; 20% for homework, 10% for project, 30% for midterm and 40% for final exam.

**Homework:**

There will be approximately 8 homework assignments. No HW scores will be “dropped.” You are allowed **one** late homework submission, due 24 hours later, no questions asked (solutions will be posted then). Any other late homework submissions will not be accepted without a note from UHS or your resident dean’s office. Homework will be submitted via Canvas.

**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 4-6 page paper. More details to come in October.

**Exams:**

There will be one midterm (on Wed, March 8), and a take-home final exam. You will be allowed two reference sheets of notes, front-and-back, for the midterm. The final exam will be open notes.

## **Projected Course Schedule** (may change slightly as the semester goes on)

<b>Class Number</b>	<b>Date</b>	<b>Topic</b>
1	Monday, January 23, 2017	Introduction and R
2	Wednesday, January 25, 2017	Random Vars, Normality, Central Limit Theorem
3	Monday, January 30, 2017	Central Limit Theorem and Confidence Intervals
4	Wednesday, February 01, 2017	Hypothesis Testing
5	Monday, February 06, 2017	In depth, Testing a Population Proportion
6	Wednesday, February 08, 2017	In depth, the Two Sample t Test
7	Monday, February 13, 2017	The Bootstrap and Permutation Tests
8	Wednesday, February 15, 2017	Power and Sample Size Calculations
9	Monday, February 20, 2017	<b>No Class-President's Day</b>
10	Wednesday, February 22, 2017	Multiple Comparisons and ANOVA
11	Monday, February 27, 2017	Simple Linear Regression
12	Wednesday, March 01, 2017	Assumptions for Simple Linear Regression
13	Monday, March 06, 2017	Midterm Review
14	Wednesday, March 08, 2017	Exam 1
15	Monday, March 13, 2017	<b>No Class-Spring Break</b>
16	Wednesday, March 15, 2017	<b>No Class-Spring Break</b>
17	Monday, March 20, 2017	Diagnostics and Transformations for SLR, I
18	Wednesday, March 22, 2017	Diagnostics and Transformations for SLR, II
19	Monday, March 27, 2017	Multiple Regression with Matrices
20	Wednesday, March 29, 2017	Diagnostics and Transformations for MLR, I
21	Monday, April 03, 2017	Diagnostics and Transformations for MLR, II
22	Wednesday, April 05, 2017	Model Checking and Refinement
23	Monday, April 10, 2017	Strategies for Variable Selection
24	Wednesday, April 12, 2017	Extensions to Regression (Quantile, Robust)
25	Monday, April 17, 2017	Regression with Time Series Data
26	Wednesday, April 19, 2017	Introduction to Logistic Regression
27	Monday, April 24, 2017	Multivariate Basics-PCA
28	Wednesday, April 26, 2017	<b>Last Class (Course Review)</b>