90-711 Statistical Reasoning with R  
Fall 2017  
Professor Amelia M. Haviland  
TAs: Jackie Mauro, Maria Caballero Sanchez  
James Duguid, Shinai He, Kyle Furlong  
Colin Vasick, Brenda Mittelbuscher  

Statistical reasoning is essential to learning from data and understanding the strengths and limitations of data analyses. This course is grounded in questions of importance in public policy and focuses on how data and statistical reasoning can inform those questions. We will use a hands-on approach to develop skills and critical thinking in the fundamentals of causal inference, univariate and bivariate descriptive statistics, prediction, linear regression, uncertainty, and statistical inference. The hands-on approach involves learning the basics of how to use the R statistical language and weekly labs in which students will use R to carry out and interpret data analysis on real-world policy issues in a supervised setting. While useful, no R or computer programming experience is required for the course, and this course does not replace the stand-alone R course.

This is a rigorous graduate school introductory statistics and data analysis course, and I encourage students who take this course to continue using statistical reasoning and methods in your other courses, your summer internship, your Systems Synthesis Project, and in your career. In today’s information world, data are available everywhere and the role of statistics is rapidly increasing in public policy, health care, the arts, the entertainment industry, business, academia and many other parts of society. We echo the message of The New York Times which published an article entitled “For Today’s Graduate, Just One Word: Statistics.”

Required Text:


It is available in hardcover, paperback, and e-book

Required Software:

In this course we will use the open-source statistical software R (http://www.r-project.org). R can be more powerful than other statistical software such as STATA or SAS, but it can also be more difficult to learn. A variety of resources will be made available for 90-711 students in order to learn R as efficiently as possible. To help make using R easier, we’ll be using RStudio (http://www.rstudio.com/)—a user interface that simplifies many common operations.
Learning Objectives: Learn from data using statistical reasoning

- Use R and R Studio to explore, summarize, and visualize data.
- Apply the concept of potential outcomes to evaluate estimates of causal effects.
- Summarize and interpret univariate and bivariate distributions using histograms, bar plots, box plots, scatter plots, and quantile-quantile plots.
- Perform linear regression with single or multiple predictors and assess model fit.
- Interpret the results of linear regression models.
- Use probability to quantify uncertainty in estimators of parameters of interest.
- Make accurate statistical inferences using confidence intervals and standard errors.
- Appropriately interpret results of data analyses and statistical inferences.
- Create reports of data analyses and interpretations of results using R Markdown.

Grade Components:

5% Class Preparation Programming Exercises – prior to most class meetings (MW) you will be asked to answer some R programming exercises related to the R topics covered in the textbook reading required for that day’s class. Each student must complete the exercises themselves (this will be tracked automatically in R) but you are welcome to work together on them.

15% Weekly Labs – All students are required to attend the weekly labs (F). In these labs you will work with one or two other students to complete - in a supervised setting - a data analysis that uses the concepts covered in class that week and the R skills covered in the textbook up to that point. In general, the labs will be more difficult than but otherwise somewhat similar to the HW assigned the same week. Your lab with the lowest grade each mini will be dropped (not included in the calculation of your course grade) unless the lowest lab grade occurs the same week as the lowest homework grade, in which case we will drop the second lowest lab grade.

30% Weekly Homework - For the HW and labs, students are required to create a report of their work, code, results, and discussion, using Rmarkdown. An Rmarkdown file in turn can be used to create the report – a final HTML document. Students must upload an electronic copy of the Rmarkdown file to Canvas using xxx.Rmd as a file name where xxx is your Andrew ID. Students should attempt to create a final HTML report from this Rmarkdown file. At the start of the semester we will give up to full credit for an assignment based only on the Rmarkdown file. Later in the semester as students’ R skills develop, we will require the final HTML report in order to earn full credit on a homework assignment. Students may discuss the assignment in general, but all code, results, and interpretation of results are to be done individually. The homework assignment with the lowest grade each mini will be dropped (not included in the calculation of your course grade). No late homework assignments will be accepted.

20% Mid-term Exam – There will be one 1.5-hour mid-term exam held during class time the last week of the mini term. Please plan ahead to give yourself time at the end of the mini term to prepare for this exam.

30% Final Exam – There will be a 3-hour cumulative final exam held during the exam period at the end of the semester.
Active learning responsibility:

We are partners in this learning experience. I EXPECT YOU TO:

- Conduct your learning with academic integrity, see below for specifics and definitions
- Attend class and constructively participate in it, participate in Canvas Discussions of content
- Read assigned textbook sections before class and complete pre-class Swirl exercises
- Do individual homework assignments
- Prepare for exams
- Be aware of and proactive about your own learning style and time management
- Pursue your own understanding: What is your understanding? Does it fit with what else you know? What is solid, missing, or vague? What you can do to make what is missing or vague more solid: attend office hours, participate in Canvas Discussions, create a study group, review reading, review homework solutions, or make an appointment with the instructor or a TA, among other things.

Accommodations for Students with Disabilities:

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Statement of Support for Students’ Health & Well-being

Your graduate school experience might prove to be exciting, stimulating, and enjoyable, but it is likely to entail stress as well. The University Provost provides the following thoughts for students. I consider them important.

*Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.*

*All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is almost always helpful.*

*If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at [http://www.cmu.edu/counseling/](http://www.cmu.edu/counseling/). Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.*
Academic Integrity:
Students are expected to honor the letter and the spirit of the Carnegie Mellon University Policy on Cheating and Plagiarism. All activities cited in that policy will be treated as cheating in this course. Students are expected to familiarize themselves with this policy. Students are also encouraged to review the Carnegie Mellon University Academic Disciplinary Actions Overview for Graduate Students, which details penalties and sanctions, as well as students’ rights. I will take a zero-tolerance policy on cheating and plagiarism and will consult with Departmental leadership on appropriate action for all instances of cheating and plagiarism. As the aforementioned policies indicate, penalties can include course failure, suspension, and dismissal from the program.

Carnegie Mellon University Policy on Cheating and Plagiarism
http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html

Carnegie Mellon University Academic Disciplinary Actions Overview for Graduate Students
http://www.cmu.edu/academic-integrity/documents/academic-disciplinary-actions-overview-for-graduate-students.2013.pdf

What constitutes plagiarism in a coding class?¹
“The course collaboration policy allows you to discuss the problems with other students, but requires that you complete the work on your own. Every line of text and line of code that you submit must be written by you personally. You may not refer to another student's code, or a "common set of code" while writing your own code. You may, of course, copy/modify lines of code that you saw in lecture or lab.

The following discussion of code copying is taken from the Computer Science and Engineering Department at the University of Washington. You may find this discussion helpful in understanding the bounds of the collaboration policy.

‘[It is] important to make sure that the assistance you receive consists of general advice that does not cross the boundary into using code or answers written by someone else. It is fine to discuss ideas and strategies, but you should be careful to write your programs on your own. You must not share actual program code with other students. In particular, you should not ask anyone to give you a copy of their code or, conversely, give your code to another student who asks you for it; nor should you post your solutions on the web, in public repositories, or any other publicly accessible place. [You may not work out a full communal solution on a whiteboard/blackboard/paper and then transcribe the communal code for your submission.] Similarly, you should not discuss your algorithmic strategies to such an extent that you and your collaborators end up turning in [essentially] the same code. Discuss ideas together, but do the coding on your own.

¹ This section is taken from Alexandra Chouldechova’s class “Programming R for Analytics.”
Expected Semester Schedule:

- **Introduction & Causality – Week 1**
  - Topic: Overview of the course, Potential Outcomes • Textbook: 1, 2.3
  - HW #1 posted on Monday 8/28; Due by 8pm Tuesday, 9/5
- **Causality – Week 2**
  - Topic: Randomized Studies • Textbook: Chapter 2.4
  - HW #2 posted on Tuesday 9/5; Due by 8pm Tuesday, 9/12
- **Causality – Week 3**
  - Topic: Observational Studies & Describing Variability • Textbook: 2.5, 2.6
  - HW #3 posted on Monday 9/11; Due by 8pm Tuesday, 9/19
- **Measurement – Week 4**
  - Topic: Plotting to Learn about Univariate Distributions • Textbook: 3.3, 3.4
  - HW #4 posted on Monday 9/18; Due by 8pm Tuesday, 9/26
- **Measurement - Week 5**
  - Topic: Describing Bivariate Relationships • Textbook: 3.4, 3.6
  - HW #5 posted on Monday 9/25; Due by 8pm Tuesday, 10/3
- **Prediction - Week 6**
  - Topic: Linear Regression with 1 predictor • Textbook: 4.2
  - HW #6 posted on Monday 10/2; Due by 8pm Tuesday, 10/10
- **Prediction – Week 7**
  - Topic: Multiple Linear Regression • Textbook: 4.2, 4.3
  - HW #7 posted on Monday 10/9; Due by 8pm Monday, 10/16
- **Mid-term Exam – Week 8**
  - Topic: Review • Textbook: Chapters 1 - 4
  - Mid-term Exam on Wednesday 10/18
  - No Labs on Friday 10/20

- Details of the schedule for the second half of the semester will be released by the end of September. We will cover:
  - Probability – From Populations to Samples • Textbook: Chapter 6
  - Statistical Inference – Learning about a Population from a Sample • Textbook: Chapter 7