

36-200 Reasoning with Data Spring 2017

Instructor: Rebecca Nugent
Baker Hall 232C
rnugent@stat.cmu.edu
<http://www.stat.cmu.edu/~rnugent>
Office Hours: Wednesday 11:30am-12:30pm or by appt

Teaching Assistants: Philipp Burckhardt, Kayla Frisoli
FMS 317, FMS 323
pburckhardt@cmu.edu, kfrisoli@andrew.cmu.edu
Office Hours: (PB) by appt, (KF) Thurs 1:30-2:30pm Porter 117, by appt

Class Meetings: Mondays, Wednesdays, Fridays 10:30-11:20am, Baker 235B
Thursday lab 10:30-11:20am, Baker 140C

Website: <http://www.cmu.edu/blackboard>

Prerequisites: none

Textbook: none required, recommended readings will be provided

Recommended Reference: Introduction to the Practice of Statistics by Moore & McCabe

General Course Plan:

This course will serve as an introduction to learning how to "reason with data". While still an introductory-level course in the Statistics Department, the focus will be more on thinking about the relationship between the application and the data set and extracting useful statistical information rather than taking primarily a methods-driven approach. There will be an emphasis on thinking through an empirical research problem from beginning to end. Types of data will include continuous and categorical variables, images, text, networks, and repeated measures over time. Applications will largely be drawn from interdisciplinary case studies spanning the humanities, social sciences, and related fields. Methodological topics will include basic exploratory data analysis, elementary probability, hypothesis tests, and empirical research methods. There is no calculus or programming requirement. There will be a weekly computer lab for additional hands-on practice.

This course is the credit-equivalent and substitute for 36-201 and will be honored appropriately as a pre-requisite for downstream Statistics courses. As such, this course is not currently open to students who have received credit for 36-201, 36/70-207, 36-220, 36-247, or any 300- or 400-level Statistics course.

Course Objectives:

1. Learn the empirical research process including data collection and design methods
2. Develop and use methods for summarizing and evaluating numerical data.
3. Learn and apply the basic concepts of probability and hypothesis tests
4. Develop skills in the applications of statistical methods to problems in the sciences and the social sciences, including interpretation and communication of results.

Course Work: Your grade in this course will be determined by homework assignments, labs, two midterm projects, and a final project.

- Weekly homework assignments will be due at the beginning of class (10:30am) on Fridays. Assignments should be submitted in class. Late homeworks are not accepted (exceptions may be made depending on circumstances; instructor permission required in advance). Note that the HW deadline is the beginning of class. There is a grace period of 10 minutes to account for printer mishaps, etc. HWs received after the grace period will not receive any credit (but you can request grading for feedback).

Homework Format: name on front page; questions answered in order; all answers marked and labeled. *If you do not staple (or paper clip) your homework, we will take off points.* You must show ALL WORK. You will not receive credit for just writing down a numerical answer or mathematical expression. How you arrived at your answer is more important.

If the HW requires output or graphs, just circling the answers or attaching graphs with no labels or explanation is not acceptable. Answers should be written up appropriately. Any required graphs should be as close to the corresponding problem as possible.

- Weekly labs will involve hands-on practice with the concepts learned in previous classes. Lab assignments will consist roughly of exploring data sets, applying data science and statistical tools to answer questions, and providing short responses. They will be graded primarily for completeness.
- The two midterm projects will cover material from (roughly) the previous five weeks. Students will use case studies and real research examples to demonstrate and apply concepts from class. The projects will include creating websites with summary information and giving presentations. More details will follow.
- The final project will be similar in structure as the midterm projects but cover material from the entire course. Class time will be devoted to project work. Students will give their final presentations during the final exam time period assigned by the registrar.

Grading policy: You are encouraged to discuss homework problems with your fellow students, however the work you submit must be your own. Acknowledge any help received on your assignments. Copied work will receive no credit. Your lowest homework grade will be dropped. **Please come talk to me if there are difficulties; problems/conflicts must be discussed IN ADVANCE.** Cheating/copying on exams results in a zero for the exam and a letter to your dean. Do your own work. Final grades will be computed with the following weights:

Homeworks	0.25
Labs	0.10
Midterm Projects (0.20 each)	0.40
Final Projects	0.25

You have one week from the day an assignment, exam, etc is handed back in class to bring any grading issues, comments, complaints, etc to the attention of the instructor. Please note that if you are absent the day something is handed back, you will not receive an extension unless arrangements have been made in advance with the instructor.

Final letter grades will be determined as usual: [90,100] = A, [80,89] = B, [70,79] = C, [60,69] = D, [< 60] = R. Grades may be curved at the instructor's discretion.

Computing: For this class, we will be using a new software platform being designed by collaborators in the Department of Statistics and Heinz College. The platform will allow students to interact with data and case studies without requiring them to learn details of a new programming language.

The statistical computing package used “behind the scenes” is R. If students would like to learn more about R, they are welcome (but again, not required). R is available on many campus computers, and you may download a free version from www.r-project.org. You can obtain a free temporary version from [myandrew](http://myandrew.com). This version is good for 1 year; you can keep renewing the license as long as you are a CMU student.

R References: manuals available on R website;

<http://www.stat.cmu.edu/~rnugent/teaching/introR>

Introductory Statistics with R, Peter Dalgaard; Springer-Verlag

Modern Applied Statistics with S-Plus Venables, Ripley; Springer

Laptop Policy: Students are expected to be participating in class; any laptop use during class should pertain directly to the class. Instructor reserves the right to not allow laptop use during class. When the class has a guest speaker, laptops must be turned off and put away.

Cellphones/Pagers, etc: All cellphones, pagers, and anything else that makes noise should either be turned off or silenced during class. Texting is not allowed nor is it acceptable professional behavior.

Communication: Assignments and class information will be posted on Blackboard. Help with using Blackboard is available at www.cmu.edu/blackboard/help/. Emails will be sent to the class from stat.cmu.edu; email filters should be set to receive emails from that server.

Email: Sending email to your professor or teaching assistants should be treated as professional communication. Emails should have an appropriate greeting and ending; students should refrain from using any kind of “shortcuts”, abbreviations, acronyms, slang, etc. in the email text. Emails not meeting these standards may not be answered.

Email questions must be sent a reasonable amount of time before a deadline. Students should not assume that their emails will be answered right away. Ask questions early.

Academic Integrity: All students are expected to comply with the CMU policy on academic integrity. This policy is online at www.studentaffairs.cmu.edu/acad_integ/acad_int.html

Cheating, copying, etc will not be tolerated; please ask if you unsure of whether or not your actions are complying with assignment/exam instructions. Always ask if you are unsure; always default to acknowledging any help received.

Video/Audiotaping: No student may record or tape any classroom activity without the express written consent of the professor. If a student believes that he/she has a disability and needs to record or tape classroom activities, he/she should contact the Office of Equal Opportunity Services, Disability Resources to request an appropriate accommodation.

Disability Services: If you have a disability and need special accommodations in this class, please contact the instructor. You may also want to contact the Disability Resources office at 8-2013.

TENTATIVE SCHEDULE: *subject to change*

Date	Topic	Due
1/18-1/20	Introduction to Course; Examples	
1/23-1/27	Data Collection/Design, Exploratory Data Analysis	HW 1 (F), Lab 1 (R)
1/30-2/3	Exploratory Data Analysis	HW 2 (F), Lab 2 (R)
2/6-2/10	More EDA & Experimental Design	HW 3 (F), Lab 3 (R)
2/13-2/17	Probability	HW 4 (F), Lab 4 (R)
2/20-2/24	Probability, Discrete Distributions	HW 5 (F), Lab 5 (R)
2/27-3/3	Midterm Project Presentations	Projects due Friday 5pm
3/6-3/10	no Fri class	HW 6 (F), Lab 6 (R)
3/13-3/17	Spring Break	
3/20-3/24	Continuous Distributions	HW 7 (F), Lab 7 (R)
3/27-3/31	Sampling, Confidence Intervals	HW 8 (F), Lab 8 (R)
4/3-4/7	Significance/Hypothesis Tests	HW 9 (F), Lab 9 (R)
4/10-4/14	Midterm Project Presentations	Projects due Friday 5pm
4/17, 4/19	Inference for Categorical Variables (no lab or Fri class)	HW 10 (F)
4/24-4/28	Linear Regression	Project Work
5/1-5/5	Final Project Work	

Note: Final Exam scheduled by the registrar; we will use that slot for project presentations