36-200 Reasoning with Data
Fall 2017

Instructor: Rebecca Nugent, Philipp Burckhardt
Baker Hall 232C
rnugent@andrew.cmu.edu
http://www.stat.cmu.edu/∼rnugent
Office Hours: Tue 2-3pm in Baker 232C (RN); Thu 4-5pm in Baker 132A (PB)

Teaching Assistants: Tanguy Dauphin, Eva Gjekmarkaj, Frank Kovacs
tdauphin, egjergji, fkovacs@andrew.cmu.edu
Office Hours: Mon 3-4pm (EG), Tue 11am-12pm (FV), Wed 2-3pm (TD)

Class Meetings: Mondays, Wednesdays, Fridays 10:30am-11:20am, Baker A51
Labs: Thursday 10:30am-11:20am (BH 140F), 11:30-12:20pm (BH 140E)
Friday 11:30am-12:20pm, 12:30pm-1:20pm (both Wean 5207)

Website: http://www.cmu.edu/canvas

Prerequisites: none

Textbook: none required, recommended readings will be provided

Recommended References: Statistics: The Art and Science of Learning from Data by Agresti & Franklin; Introduction to the Practice of Statistics by Moore & McCabe

General Course Plan:

This course is an introduction to learning how to make statistical decisions and "reason with data". The approach will emphasize thinking through an empirical problem from beginning to end and using statistical tools to look for evidence for/against an explicit argument/hypothesis. Types of data will include continuous and categorical variables, images, text, networks, and repeated measures over time. Applications will largely 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 one weekly computer lab for additional hands-on practice using an interactive software platform that allows student-driven inquiry.

This course is the credit-equivalent to 36-201 and will be honored appropriately as a prerequisite 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 course grade will be determined by homework assignments, labs, exams, and projects.

- Weekly homework assignments will be due at the beginning of class (10:30am) on Wednesdays. Depending on the week, assignments will be submitted either in class or on Canvas. 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 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.

- Discussion labs will involve hands-on practice with the concepts learned in previous classes. Lab assignments will consist roughly of exploring case study data sets, applying data science and statistical tools to answer questions about real scenarios, and providing short responses. They will be graded primarily for completeness.

- There will be two midterms and a final exam. The midterms will primarily cover material from the previous 5-6 weeks. The final exam will be cumulative and will assess knowledge of concepts and formulas learned in class. For the exams, only calculators will be required; no laptops or software needed.

- For the projects, students will use case studies and real research examples to demonstrate and apply concepts from class. The projects will include creating documents/slides with summary information and giving presentations. More details will follow.

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 as follows:

Homeworks	0.15
Labs	0.05
Midterms	0.30 (0.15 each)
Projects	0.35 (0.10, 0.10, 0.15)
Final Exam	0.15

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 a programming language.

From the Eberly Center: For this class, the instructors, Rebecca Nugent and Philip Burckhardt, are conducting research on the impacts of various teaching methods and materials on student learning. You will not be asked to do anything above and beyond the normal learning activities and assignments that are part of this course. You are free not to participate in this research, and your participation will have no influence on your grade for this course or your academic career at CMU. Please note that students choosing not to participate in the research will not be excused from required course activities. Participants will not receive any compensation. The data collected as part of this research will include student grades. All analyses of data from participants' coursework will be conducted after the course is over and final grades are submitted. The Eberly Center may provide support on this research project regarding data analysis and interpretation. To minimize the risk of breach of confidentiality, the Eberly Center will never have access to data from this course containing your personal identifiers. All data will be analyzed in de-identified form and presented in the aggregate, without any personal identifiers. Please contact Professor Rebecca Nugent at rnugent@andrew.cmu.edu, or in person, if you have questions or concerns about your participation.

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 Canvas. **BE SURE TO TURN ON YOUR CANVAS NOTIFICATIONS.** Help is available at www.cmu.edu/canvas.

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; this includes copying material from solutions to old homework, exams, etc. 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 accomodations 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**
M 8/28	Introduction to Course; "Big Picture"	
W 8/30	Variable Types and EDA	Intro Survey
F 9/1	1-var EDA (categorical)	Lab 1 (R,F)
M 9/4	No class; Labor Day	
W 9/6	1-var EDA (quantitative)	HW 1
F 9/8	1-var EDA (quantitative)	Lab 2 (R, F)
M 9/11	2-var EDA	
W 9/13	2-var EDA	HW 2
F 9/15	2-var EDA	Lab 3 (R, F)
M 9/18	Relationships between Variables	
W 9/20	Bias; Lurking Variables	HW 3
F 9/22	Experimental Design	Lab 4 (R, F)
M 9/25	Experimental Design	
W 9/27	Review	HW 4
F 9/29	**Midterm 1**	Lab 5 (R, F)
M 10/2	Elementary Probability	
W 10/4	Elementary Probability	
F 10/6	Elementary Probability	Lab 6 (R, F); **Project 1**
M 10/9	Discrete Distributions; Binomial	
W 10/11	Discrete Distributions; Binomial	HW 5
F 10/13	Density Curves; Normal Distribution	Lab 7 (R,F)
M 10/16	Normal Distribution; Standardization	
W 10/18	Sampling Distribution; Central Limit Theorem	HW 6
F 10/20	No class; mid-semester break	No labs
M 10/23	Sampling Distribution; CLT (means)	
W 10/25	Sampling Distribution; CLT (props)	HW 7
F 10/27	Confidence Intervals (means)	Lab 8 (R,F)
M 10/30	Confidence Intervals (props)	
W 11/1	Significance Tests	HW 8
F 11/3	Significance Tests	Lab 9 (R,F)
M 11/6	t-distribution	
W 11/8	**Midterm 2**	
F 11/10	No class; CMU 50th Anniversary	No labs
M 11/13	Inference for two means/proportions	
W 11/15	Inference for two means/proportions	HW 9
F 11/17	Inference for Linear Regression	Lab 10 (R, F); **Project 2**
M 11/20	One-Way ANOVA	
W 11/22	No class; Thanksgiving Break	
F 11/24	No class; Thanksgiving Break	
M 11/27	Relationship between Categorical Variables	
W 11/29	Chi-Square Test	HW 10
F 12/1	Special Topics	Lab 11 (R, F)
M 12/4	Special Topics	
W 12/6	Special Topics	HW 11
F 12/8	Final Review	**Final Project (R, F)**

Final Exam Day TBA (do NOT make travel plans before the final exam schedule is posted)