

## 36-200 Reasoning with Data Summer 2017

**Instructor:** Gordon Weinberg, Rebecca Nugent  
Wean 3719, Baker Hall 232C  
gordonw, rnugent@andrew.cmu.edu  
<http://www.stat.cmu.edu/~gordonw>, [~rnugent](http://www.stat.cmu.edu/~rnugent)  
Office Hours: TBA

**Teaching Assistants:** Philipp Burckhardt, Frank Kovacs  
pburckhardt@cmu.edu, fkovacs@andrew.cmu.edu  
Office Hours: TBA

**Class Meetings:** Mondays, Wednesdays, Fridays 12:00pm-1:20pm, Baker 235A  
Tuesday, Thursday lab 12:00pm-1:20pm, Baker 140EF

**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 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 two weekly computer labs 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, quizzes, a final in-class exam, and a group project.

- Weekly homework assignments will be due at the beginning of class (12:00pm) on Tuesday/Thursday (Week 1 exception due to holiday). 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.

- 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.
- The final in-class exam will be cumulative and will assess knowledge of concepts and formulas learned in class. Only calculators will be required; no laptops or software needed.
- For the group project, 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 two lowest homework grades 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	0.25
Labs	0.10
Quizzes	0.25
Group Project	0.20
Final Exam	0.20

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](http://www.r-project.org). You can obtain a free temporary version from [myandrew](http://myandrew). 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/](http://www.cmu.edu/blackboard/help/). Emails will be sent to the class from [stat.cmu.edu](http://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](http://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 are 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*

<b>Date</b>	<b>Topic</b>	<b>Due</b>
M 7/3	Introduction to Course; “Big Picture”; 1-Var EDA	
T 7/4	No lab; 4th of July	
W 7/5	1-Var and 2-Var EDA	
R 7/6	Lab 1	
F 7/7	2-Var EDA	HW 1; Quiz 1
M 7/10	Difficulties with Variable Relationships	
T 7/11	Variable Relationships; Lab 2	HW 2
W 7/12	Experimental Design; Matched Pairs	
R 7/13	Elementary Probability; Lab 3	HW 3
F 7/14	Elementary Probability	Quiz 2
M 7/17	Discrete Distributions	
T 7/18	Binomial Distribution; Lab 4	HW 4
W 7/19	Continuous Distributions (incl. Normal)	
R 7/20	Normal Distribution; Lab 5	HW 5
F 7/21	Sampling Distributions; CLT (mean)	Quiz 3
M 7/24	Sampling Distributions; CLT (proportion)	
T 7/25	Sampling Distributions; Lab 6	HW 6
W 7/26	Confidence Intervals	
R 7/27	Confidence Intervals; Lab 7	HW 7
F 7/28	Intro to Significance Testing	Quiz 4
M 7/31	Significance Testing; t-distribution	
T 8/1	t-distribution; Lab 8	HW 8
W 8/2	Significance Testing for Two Groups	
R 8/3	One-Way Anova; Lab 9	HW 9
F 8/4	Chi-Square Distribution	Quiz 5
M 8/7	Linear Regression; Inference for Matched Pairs	
T 8/8	Lab: Project Presentations	
W 8/9	Review; Statistical Lessons Moving Forward	
R 8/10	Lab: Project Presentations	
F 8/11	Final Exam	