STAT 215: STATISTICS AND MODELING (FALL 2017)

Instructor. Colin Reimer Dawson (*he/him/his*)
Office. King 204
Email. `cdawson@oberlin.edu`
Website. `http://colindawson.net/stat215/`
Office Hours. M 1:30-2:30, T 4-4:50, Th, 4-5:30, F 9-9:50 and 11-11:50. *If your schedule prevents you from attending any of the posted hours, please let me know as soon as possible.*
Locations and Times. T and Th 3-3:50 in King 239 and F 2:30-4:20 in King 137 (lab)

Course Description

Goals. The general goals of the course are to review the foundational ideas developed in AP/introductory statistics, sharpening your ability to reason from data, and then to become familiar with *statistical models*, which quantify and simplify relationships among variables. Specifically, we will dicuss linear and logistic regression models, and Analysis of Variance (ANOVA) models, with arbitrary numbers of quantitative and/or categorical explanatory variables.

Note that this is a *statistics* course, *not* a math course. The focus will be statistical reasoning, and the use and interpretation of statistical models, not mathematical derivations. Some mathematical detail will be glossed over, and we will rely on software to do the nitty gritty calculations. Computational thinking is an indispensible skill for doing statistics in the 21st century, and constitutes an important part of the course.

Prerequisites / Who This Course is For. This course is designed as an "accelerated introduction" to statistics, combining the material from STAT 113 and 213 into a single semester. Students taking this course should meet *one or more of* the following criteria:
  (1) You took AP Statistics in high school and scored 3 or above on the AP exam
  (2) You took (and did reasonably well in) a college level intro statistics course at some other institution

---

Date: Last Revised September 9, 2017.

(3) You have a math and/or computer science background which is solidly above that of the typical incoming Oberlin student (talk to me if you're not sure what qualifies)

. I will assume fluency with basic descriptive statistics (mean, meadian, standard deviation, histograms, scatterplots), and some exposure to basic ideas in inferential statistics (sampling distributions, confidence intervals, P-values). We will review the basic inferential ideas, but you will at least need to be able to pick them up quite quickly.

Materials

Textbooks. We will use two textbooks for this course. The textbook for the introductory material (first 4-5 weeks) is *Introductory Statistics with Randomization and Simulation* by Diez, Barr, and Cetinkaya-Rundel. This book is open source and freely available in PDF at `http://www.openintro.org/stat/textbook.php?stat_book=isrs`. Be sure to get the one with "Randomization and Simulation" in the title, as there are three different intro stat books at openintro.org. The textbook for the modeling material (last 8-9 weeks) is *STAT2: Building Models for a World of Data*, by Ann Cannon, et al., which is available at the bookstore in physical or electronic formats.

Computing. We will use the free statistical computing environment RStudio, which is an interface to the language R. You may either install R and RStudio on your personal machine (`www.r-project.org` and `www.rstudio.com`, respectively), or use Oberlin's RStudio server via a web browser (`rstudio.oberlin.edu`). The R language has become the standard computing tool used by practicing statisticians and data scientists, and so although statistical reasoning is the main goal of the course, competence in R and written presentation of results is a learning objective unto itself as well.

Laptops. It will be relatively uncommon that I will stand and lecture for an entire class. Many classes will consist entirely of working on problems in groups, often using the computer (even on non-lab days, so if you have a laptop you might want to bring it to class — it will be sufficient if one member per group of 2-3 has one). At first assignment of students to groups; later in the semester I may hand select groups based on complementary strengths and compatible styles.

Outside of these activities, I strongly discourage the use of electronic devices. For note-taking, take hand-written notes if you are physically able to do so: there is evidence that writing by hand improves your cognitive processing of the material.

Logistics

Email. Email is the best way to reach me outside of a face-to-face meeting. You are welcome to address me by my first name, which is generally what I will use when signing emails. I do not consistently respond to email after about 5:30 P.M., so plan sufficiently ahead if you need a response from me to make progress with an assignment. I try to respond to most email within 24 hours, but may not manage this 100% of the time when volume is high.

Accommodations. If you have a disability of any sort that may require accommodations in order for you to do your best work in this class, please let me know as early as possible, and consult as well with the Office of Disability Services (ODS). By college policy, *all requests for accommodation require documentation from ODS.*

Honor Code. The Oberlin College Honor Code formalizes the idea that all work that you submit is your own and that you have given credit to the ideas and work of others when you incorporate them. You will be asked to write and sign the honor pledge on each graded assignment that you hand in. The honor pledge reads: "I have adhered to the Honor Code in this assignment."

What it means to adhere to the honor code depends on context. For each assignment type, I describe what it means to follow the honor code on that assignment below.

More information about the honor code can be found on the web here Dean of Students site:
http://new.oberlin.edu/students/policies/honor-system-charter

Your responsibilities

Reading Assignments. There will be reading assignments each day (posted on the website), which should be completed before the start of class. We will often begin new topics with basic questions about the reading to discuss in pairs before we discuss new material as a class, so it is important to come prepared and ready to contribute. I recognize that there will be occasional days when you may not have

done the reading. When this happens, just be up front about it. As long as it doesn't happen too often, no one will hold it against you.

Attendance and Participation. I will spend some class time presenting material, but we will devote much of our time to explorations of the material in groups. There is increasingly strong evidence that learning via self-discovery is more effective than passively listening to a lecture. It is expected that you will attend class nearly always and participate actively.

Homework and Labs. There will be (roughly) weekly assignments, posted on the website, involving both written and and computer-based problems, which will generally be due electronically (via Blackboard) by the start of class on Tuesdays. Most homework and labs are graded for completion and "good faith effort" only (exceptions will be clearly noted in advance), and written feedback will be provided by the student grader. Details about the homework policy are in the "Grading Policy" handout (which is also on the website).

Honor Code: I encourage you to work on ungraded problems collaboratively, with no restrictions on the amount of collaboration that is appropriate, but you must accurately convey what degree of responsibility you have for the written solutions or code. Graded problems (if there are any) must be done individually.

Quizzes and Exams. There will be a short ($\sim$10 min) quiz most Fridays at the start of the lab day. There will be an in-class exam at roughly the $1/3$ point in the semester (tentatively October 3rd), which will be a paper and pencil exam, roughly equivalent to a final exam in intro stats (e.g., STAT 113, 114 or AP stats). A second exam, on the STAT2 material, will take place on the last day of classes in the computer lab, and will require you to analyze real data using the computer.

Honor Code: Exams and quizzes must be done individually. Quizzes and the first exam will be closed book, but an 8 1/2" $\times$ 11" note sheet and a calculator will be allowed on the exam. The second exam will likely be open book and open notes.

Data Analysis Projects.

Individual Mini-Projects: There will be two data analysis mini-projects to be done individually (tentative due dates are 11/3 and 11/30, respectively), for which you will be provided a dataset (but you can substitute a dataset of interest to you if you prefer) and asked to do some open-ended modeling to allow you to learn something about the world through the data. The deliverable for these projects will be a

short report written in RMarkdown with all code needed to reproduce the results embedded in the document. (This is the same format that will be used for most lab assignments)

Final Group Project: Instead of a final exam, you will be asked to do a larger scale data analysis project in teams of 2 or 3, for which you will identify and find a dataset that allows you to say something about a research question of interest (collecting the data yourself is also an option, though not required). Ideally you will use at least two of the three major model classes that we cover in the course — linear regression, logistic regression, and Analaysis of Variance — to address different aspects of your research topic, though if there is only one of these that makes any sense in the context of your question, we can talk about how you might do something more sophisticated with it instead. The deliverables will be a written report in the same format as the mini-projects, only longer, and a group presentation given on the scheduled final exam day.

Honor Code: Describing the work that other people did before you is an important part of intellectual inquiry, but you must give credit and cite sources for for any data or ideas that did not originate from you. This includes paraphrases as well as direct quotations. The analysis you do must not have been done previously. All code and model interpretation must be original.

Grades and Policies

Grade Components. The grade is a weighted average of the grades on the various assignments (I know, shocker, right?). The weights for each component, along with rough targets for letter grades, are given below. Thresholds are subject to some modification at my discretion, based on the classwide grade distribution, which will affect everyone equally, and on attendance and participation, which may affect individuals who are near the border differently.

Homework (effort/completion only)	10%
Quizzes	20% total
Exam 1	15%
Exam 2	15%
Mini-Project 1	10%
Mini-Project 2	10%
Final Project	20%

A	93%
A-	90%
B+	86%
B	82%
B-	78%
C+	74%
C	70%
C-	66%
D	60%

Late Policy. Credit is of course guaranteed if homework is submitted by the posted deadline. If homework is turned in late, credit will be granted only if the class submissions for that assignment have not yet been downloaded by the person grading them (i.e., either me or the student grader), which may at times be right after the deadline; at other times it may be a day or two later. Once a set of assignments is downloaded, we will not check again for late submissions.

Replacement Quizzes. You may take up to three "replacement quizzes" (at most one per week) during my office hours, no questions asked, provided that you are current on the homework on the material covered in the quiz you want to make up or retake. If you did not turn in an assignment on time, you will not get homework credit for that assignment, but I will consider homework that is turned in with the following assignment when determining eligibility for a replacement quiz. A fourth replacement quiz is available to replace part (about 1/5) of Exam 1. One of the other three replacement quizzes may optionally be used to replace another approximately 1/5 of Exam 1 instead of a quiz.

Key Dates

Final Presentations. Final presentations will be held on the scheduled final exam day in lieu of an exam. This is Saturday 12/16, from 7-9 P.M. Since this is a Saturday night, I will supply pizza and beverages (non-alcoholic, of course). If you want to eat pizza, try to arrive at 6:30 so we have plenty of time for the presentations.

In-Class Exams. There are two exams scheduled during regular classes. The first one will tentatively be Tuesday, October 3rd. The second one will definitely be Friday, December 9th (the last day of classes).

Holidays. None of the college-wide holidays fall on our class days, so the only days when there is no class are the week of Oct. 16-20 (fall break), and Nov. 23-24 (Thanksgiving).