STAT 336: MATHEMATICAL STATISTICS (SPRING 2016)

Instructor. Colin Reimer Dawson (*he/him/his*)
Office. King 204
Email. `cdawson@oberlin.edu`
Website. `http://colinreimerdawson.com/stat336/`
Office Hours. General: M/W 4-5:15; Class-specific: T 3:30-4:30 and Th 5-6
Locations and Times. T/Th 9:30-10:50 in King 235

Course Description

Goals. The goal of the course is for you to develop understanding of the most important ideas in modern statistical inference and the mathematical theory underpinning these ideas. The statistical ideas, not the proofs, are the most important part of the course, and in the interest of getting exposure to a breadth of modern statistical ideas, some results will be stated without proof (although you are of course encouraged to prove them on your own). However, the course will be fairly mathematically rigorous.

Prerequisites / Who This Course is For. This is primarily intended as a course for math majors interested in statistics, but it is also appropriate for computer science majors interested in AI / Machine Learning, or anyone else with the requisite probability background. The main prerequisite for the course is MATH 335 (Probability). I assume that this also entails a solid grounding in multivariable calculus. Linear algebra is not strictly speaking a prerequisite, but will be used sometimes. Previous coursework in applied statistics (any of STAT 113, 114, 213, 215, 237) would give you a clearer picture of some of the things the theory in this course can be used for, but I will not assume this background.

Textbook and Course Outline. The textbook is *All of Statistics: A Concise Course in Statistical Inference*, by Larry Wasserman. Part I (Chapters 1-5) cover probability theory, most of which you should already be familiar with, and so we will move rapidly through this material over about two weeks. We will spend about six weeks covering nearly all of Part II (Statistical Inference), and the remainder (about four weeks) on selected topics in Part III (Statistical Models and Methods) and,

Date: Last Revised February 1, 2016.

time-permitting, one or two additional topics from my area of expertise (Bayesian statistics and machine learning) that do not appear in the text. For a more detailed (tentative) schedule, see the course website (link at the top of this syllabus).

Computing. This is mainly a theory course, however it will sometimes be instructive to use computer simulations to get a better feel for what is going on mathematically. Occasional homework problems will require you to do simple simulations. You may use any language that you are familiar with to do these. I would recommend using a high-level language with well-developed prob/stat libraries, like R, Python, or MATLAB/Octave. I will provide reference material for R, as a default choice. In any case, you will not need to write sophisticated programs; scripts of at the very most a few tens of lines will suffice.

Structure of Class. I will assign you to small groups that will change in composition every two or three weeks. Each day I will identify a set of readings (mainly from the textbook) that you should do before the next class, along with a few homework problems that deal with a mix of material from the last class and basic ideas for the next class. Each group will be responsible for presenting a subset of the problems.

Most days I will give mini-lectures setting up key ideas, but we will spend a majority of the typical class period working on problems in groups, with you presenting your solutions to each other.

There will be four major individual written assignments, a textbook project, a final exam, and various in-class activities and exercises. More detail on these, and the grading scheme, is given below.

Logistics

Office Hours and Open Door Policy. I will hold some "generic" office hours, and some specific to this class. You are of course welcome to come to either, but may want to opt for the specific hours if possible, since it is more likely that the other students there will have similar questions to yours. You are also welcome to make an appointment, or simply drop by, outside these times (I usually will be in my office during the day Monday through Thursday).

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., but I sometimes check in once later in the evening.

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 written 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 at the Dean of Students site:
`http://new.oberlin.edu/office/dean-of-students/honor/students.dot`

Graded Work

Written Assignments / Takehome Exams (40%). There will be four major assignments (think of these as either big homeworks or takehome exams) at roughly three week intervals, to be done individually. These will consist of a mix of conceptual questions, calculations, and proofs, and will cover, respectively, (1) probability and foundations of inference, (2) frequentist estimation (parametric and nonparametric), (3) Bayesian inference, decision theory, and computation and (4) testing, prediction, and classification. Collectively, these make up 40% of the final grade (10% each). *Honor Code: Unlike the day-to-day group problems, these large assignments must be done individually.*

Textbook Project (30%). Over the course of the semester, you will write your own "textbook" for the course, consisting of verbal explanation of the key ideas, selected worked examples, theorems and proofs. Essentially, it consists of the notes you are taking anyway, but cleaned up and typed in LaTeX(I will provide a template). You are encouraged to use the Wasserman text as a skeleton, but it is terse; your presentation should be more fleshed out, and written in your own style. The audience is you before you took this course (or a student coming into a course like this one, with similar background). At two points throughout the semester, you will exchange parts of your work-in-progress with peers, who will read and give feedback. You will then have the opportunity to revise. Each peer editing step is worth 5%; the final product is 20%. *Honor Code: Each person is responsible for their own work, but discussions with peers are encouraged. There will undoubtedly be some overlap with the Wasserman book, and it is okay to use examples from the book or from other sources, but wording must be your own, and equations must be typeset by you.*

Final Exam (20%). There will be an in-class final exam during the finals period **on Thursday, May 12, from 2-4 P.M.**, during which you may consult the textbook that you developed. This constitutes 20% of the grade. *Honor Code: The exam must be done individually. Communication with peers about exam problems during the exam is not permitted.*

In-Class Activities (10%). The remaining 10% of the grade comes from in-class work — a combination of presentations and short writing prompts. *Honor Code: Free collaboration is encouraged, and in fact necessary.*