

STAT 140: Introduction to the Ideas and Applications of Statistics

About the Course**Instructor** Evan RayEmail: eray@mtholyoke.edu

Office: Clapp 404C

Office Hours: Mondays from 2:30-3:30pm, Tuesdays from 9:30-10:30am, Fridays 2:30-3:30pm, and by appointment. Please do not hesitate to contact me to set up appointments for additional office hours at any time!! I get paid to teach you, and I take that seriously.

Description The world is awash in data that can be used to address important issues and help make better decisions. The increased availability of information in all fields has made it ever more important for citizens to be able to make sense of these data without relying solely on experts. The primary goal of this course is to give you the technical skills and experience that will enable you to think critically about what we can and can't learn from data, and to be an informed consumer and producer of quantitative and statistical analyses. Through repeated practice and scaffolded exposure to increasingly sophisticated research results, students will be prepared to make sense of key issues to address social and scientific problems.

Introduction to Statistics provides a basic foundation in descriptive and inferential statistics, including constructing models from data. Students will learn to produce meaningful graphical and numerical summaries of data, apply basic probability models, and employ statistical inference procedures using computational tools. Topics include basic descriptive and inferential statistics, visualization, study design, and multiple regression.

Key learning goals:

1. Students should become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.
2. Students should be able to recognize questions for which the investigative process in statistics would be useful and should be able to answer questions using the investigative process.
3. Students should be able to produce graphical displays and numerical summaries and interpret what graphs do and do not reveal.
4. Students should recognize and be able to explain the central role of variability in the field of statistics.
5. Students should recognize and be able to explain the central role of randomness in designing studies and drawing conclusions.
6. Students should gain experience with how statistical models, including multivariable models, are used.

7. Students should demonstrate an understanding of, and ability to use, basic ideas of statistical inference, both hypothesis tests and interval estimation, in a variety of settings.
8. Students should be able to interpret and draw conclusions from standard output from statistical software packages.
9. Students should demonstrate an awareness of ethical issues associated with sound statistical practice.

Textbook We will be using “Statistics: Data and Models” (3rd or 4th edition) by De Veaux, Velleman, and Bock as the text for the class. This is available as a hardcover (ISBN 978-0321986498) or loose leaf (ISBN 978-0321989994). A copy will be on reserve. The book is available for purchase from various sources such as Amazon, The Odyssey Bookshop, or Pearson.

The textbook is not just a reference to use after the instructor has presented new material but a sourcebook to use at every stage of learning. When all students read the text before class, the nature of the class meeting changes to the benefit of everyone. You have thought about the material, and you arrive with your own questions. You’re ready to discuss what you understand, to clarify what you don’t understand, and to hear more on the topic. You need to read ahead in the book prior to class as well as review the material after we’ve discussed it in class. There will often be reading questions due before we discuss material in class, as we will be focusing class time on activities to help understand concepts. Self-study in advance is needed to be successful.

Time commitment While the exact time commitment for the class will vary individually and over the course of the semester, I recommend that you budget approximately three out-of-class hours for every class hour to complete the reading, assignments, and homework. I have designed the class so that it should be feasible to satisfactorily complete the requirements with approximately twelve hours per week of time commitment. If you are spending more time than this on a regular basis I would encourage you to check in with me.

Classes

Section 01: Mon, Wed 11:00am -12:15pm and Fri 11:00am - 11:50am in Clapp 407.

Section 02: Mon, Wed 1:15pm - 2:30pm and Fri 1:15pm - 2:05pm in Clapp 407.

Please plan to bring your laptop to class (we’ll need to have sufficient numbers to allow you to work in pairs and take advantage of the technology available to us). Please contact me if you do not have access to a working laptop; we can arrange for you to borrow one.

Policies

Attendance Your attendance in class is crucial, as is your punctuality. We are all going to learn this material together, so I need to have everyone present and working. I will make my best effort to provide accommodation for an unavoidable absence if you notify me well in advance. One necessary absence during the semester is not unusual; having more than two is uncommon.

Collaboration Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete homework assignments, and prepare for exams. However, every word that you write must be your own.

Copying and pasting sentences, paragraphs, or large blocks of R code from another student is not acceptable and will receive no credit or a penalty. No interaction with anyone but the instructor is allowed on any exams or quizzes. All students, staff and faculty are bound by the Mount Holyoke College Honor Code.

To sum up: On homeworks and labs, **I want you to work together.** *But, you must write up your answers yourself.*

Cases of dishonesty, plagiarism, etc., will be reported.

Resources

Course Website The course website is at <https://mhc-stat140-2017.github.io/>. I will update it regularly with lecture notes, lab assignments we will work through in class, and homework assignments.

Computing Modern statistics can't be done without computation. We will use the R statistical programming language in this course. R is one of the most commonly used programming languages in academic statistics, and I use it daily; it's also very commonly used in statistics and data science positions in industry. Knowing R is a marketable skill. In this class, you will learn enough about R to conduct basic data analysis tasks. However, this is not a computer science course. We will focus on learning the minimal amount of R needed to conduct our analyses, and the main focus of the class is on the statistical methods. If you have never programmed before, you do not need to be intimidated. You will learn in a scaffolded structure, starting with complete examples of working code, then making small changes to existing code, and finally writing your own small snippets of code.

We will use RStudio as a development environment for R. You can access RStudio on the web at <http://rstudio.mtholyoke.edu>. However, if you'd like to install RStudio on your personal computer, that is also a possibility; you can download it from <http://rstudio.org>.

As mentioned earlier, it will be important to *bring your laptop to class*; we will be using R and RStudio nearly every day. Much of this work will be done in pairs, but we need to ensure that there are sufficient machines. Please let me know if this presents any issues, as there are laptops available.

You will need to bring a scientific calculator to each midterm exam. This calculator may not be shared with another student, and you may not use your phone (which must be stored in your bag). You do not need a graphing calculator: any simple scientific calculator will suffice.

Writing Your ability to communicate results, which may be technical in nature, to your audience, which is likely to be non-technical, is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

Extra Help In addition to my office hours (which I encourage you to attend – and don't forget that I'm available outside of my formally scheduled office hours!!), or emails to me, other resources are available.

There will also be drop-in student help hours during several evenings each week. The times and locations for these will be posted on the course web site.

Your fellow students are also an excellent source for explanations, tips, etc. *You should work on assignments with others!*

Assignments

Homework Homework is the most effective way to reinforce concepts learned in class. There will be regular homework assignments. Often, questions will relate to material in the reading that will be covered in class. Homework is due at the *start* of class, and unless indicated otherwise, will be accepted with a 25% penalty if turned in within the shorter of 48 hours or the next class period (and no credit otherwise). Extensions may be possible, but need to be requested well before the deadline.

Exams There will be two "midterm" exams and occasional quizzes to be taken during class, which will be worth 50% of your grade. In addition there will be a final exam (worth 20% of your grade) during the exam period. All exams are closed book, while some may include an open-notes take home component. You may bring a calculator and a specified number of pages of paper with notes on both sides to the exams (these will be turned in with each exam). No communication with anyone besides the instructor is allowed on these assessments.

Extra Credit Extra credit is available in several ways: attending an out-of-class lecture (as will be announced) and writing a short review of it; pointing out a substantial mistake in the book, a homework exercise or exam solution; drawing my attention to an interesting data set or news article; etc. The extra credit is applied when a student is near the boundary of a letter grade.

Grading When grading your written work, I am looking for solutions that are technically correct and reasoning that is clearly explained. *Numerically correct answers, alone, are not sufficient* on homework, tests or quizzes. Neatness and organization are valued, with brief, clear answers that explain your thinking. If I cannot read or follow your work, I cannot give you full credit for it.

Your grade for this course will be a weighted average of several components:

| | | |
|--|---------------------------|-----|
| | Homework | 30% |
| | Midterm Exams and Quizzes | 50% |
| | Final Exam | 20% |

In order to extract maximum information from assessment, I do not work with a system where 90% means A, 80% B, and so on. I use the following approach to grading:

- A: all assignments completed, active participation in class, extraordinary and brilliant work
- B: all assignments completed, active participation in class, superior work
- C: almost all assignments completed, average work
- D: missing assignments, quality of work is less than satisfactory, attendance and participation unsatisfactory
- E: missing assignments, poor work, attendance and participation unsatisfactory