Stat 242: Intermediate Statistics

About the Course

Basic Information

- Meeting Time and Location: Mon, Wed, and Fri 11:00AM - 12:15PM in Clapp 402.

- Course Website: `http://www.evanlray.com/stat242_f2019/`

- Email: `eray@mtholyoke.edu`

- Office: Clapp 404C

- Office Hours: I will hold regularly scheduled office hours each week at times to be selected by you. These times will be posted on the course web site. Please do not hesitate to contact me to set up appointments for additional office hours at any time!

Textbook We will be using "The Statistical Sleuth" (3rd edition, ISBN 978-1133490678) by Ramsey and Schafer as the text for this class. A copy will be on reserve at the library.

This is a challenging book. I have chosen to use it because it places an emphasis on working with real data and it gives a thorough treatment of the subject. We will not have time this semester to cover all of the material in the book in depth - but this is the sort of book that you can refer back to as you conduct data analyses in the future.

Piazza We have a Piazza page for this course at `https://piazza.com/mtholyoke/fall2019/stat242`. I ask that you **please submit all questions about course content as questions on the Q&A forum at Piazza**. This will allow other students to answer your questions if they see them before me, and will allow other students to benefit from the answers to your questions. Additionally, **you can post questions and answers to Piazza anonymously**.

Description If you are taking this course, it is because you may need to conduct a data analysis to answer a scientific question some day. This course will introduce you to the proper use of the most commonly used statistical models in applied data analyses, including ANOVA, multiple regression, and (if time allows) logistic regression. It includes methods for choosing, fitting, evaluating, comparing, and interpreting statistical models.

A tentative schedule and topic list is below. This outline is ambitious; we may not actually get to everything outlined here. I may also change the order of some topics. An up-to-date list of topics covered so far and a time line for upcoming classes will be kept on the course website.

Unit	Topic	Chapter	Weeks
Introduction	Conceptual foundations of inference	1	1
t-based Inference and Extensions	t based inference for one or two groups	2	1, 2
	Sensitivity of analyses to assumption violations	3.1 - 3.3	2
	Approaches when assumptions are not met: transformations, bootstrap	3.5, other source	3
One-Way ANOVA	The model; t and F tests	5	4
	Multiple comparisons and linear combinations	6	5
Simple Linear Regression	The model; inference	7	6
	Checking conditions and addressing problems	8	7
Multiple Regression	Model statement and interpretation	9	8
	Inference for multiple regression	10	9
	Checking conditions and addressing problems	11	10
	Strategies for variable selection	12	11
Two-Way ANOVA	Model statement, interpretation, and inference	13	12, 13
Logistic Regression	Model statement, interpretation, and inference	20	14, 15

Policies

Attendance   Your attendance in class is crucial, unless you are sick. If you are sick, please let me know and stay home and rest; I hope you feel better!

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

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

Technology

Computing with R   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 use R most days, and for many homework problems. I expect that you are familiar with R from previous classes, but I do not expect that you are an expert at R yet. That said, it is imperative that you let me know if you are confused about anything we are doing in R as soon as possible.

We will use R via RStudio; Mount Holyoke's version of RStudio Server can be accessed at `https://rstudio.mtholyoke.edu/`. You are also welcome to work locally on your own computer if you have RStudio set up; however, please make sure you have installed at least version 3.5.0 of R and the latest versions of any R libraries we use.

It will be important to **bring your laptop to class**; we will be using R nearly every day. Much of this work will be done in pairs, but we need to ensure that there is a sufficient number of computers. Please let me know if this presents any issues, as there are laptops available for you to borrow.

Assignments

Participation and Labs   The best way to learn statistics is to do statistics. For that reason, you will have an opportunity to put the methods we are discussing to use nearly every day in worksheets and labs. I will ask you to complete and submit some of these labs for credit. I may also occasionally ask you to complete small assignments at home that are not large enough to merit being counted as separate homework assignments. These labs and small assignments will be graded for completion only.

Homework   Homework is the most effective way to reinforce concepts learned in class. There will be regular homework assignments. Homework assignments will generally include a computational component and a written component. Late assignments are a headache, and I'd rather not deal with them. I may accept a homework assignment within 24 hours of the due date for a 25% penalty. If you turn in an assignment late, there's a decent chance that it won't be graded until the end of the semester. Extensions may be possible, but need to be requested well before the deadline.

Quizzes   We will have a **short quiz two days each week** (other than weeks with exams). The questions for these quizzes will be selected at random from a bank of quiz questions that you will have access to. We will add to this bank of questions every week. One quiz each week will be on material covered that week or the previous week; the other will be drawn from the full list of possible questions. At all times, you will have a list of all possible quiz questions and their solutions available to you. I suggest that you review this bank of questions for a few minutes every day. The purpose of these quizzes is to encourage you to study continuously throughout the semester, not to intimidate or scare you. The lowest 3 quiz grades will be dropped.

Exams   There will be one or two midterm exams, as well as a final exam during the exam period. I haven't decided yet whether the exams will be in class, take home, or a mixture of in class and take home.

Mini-Project   You will complete a mini-project in the last third of the semester. This will not be an extended project; it will probably require about as much work as 2 homework assignments. It will be more self-directed than the homework assignments: you will choose a data set you are interested in and conduct a statistical analysis of that data set to answer a scientific question. You will summarize your findings in a written report; depending on time constraints, I may also require to give a brief presentation of your analysis for the class.

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. That said, we are all constantly improving at writing. Your classmates and I are here to help you improve as a writer.

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, an exam solution, or something I present in class; 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 the following components:

Item	Weight
Participation and Labs	5%
Quizzes	20%
Homework	20%
Exams	45%
Mini-Project	10%