

## Stat 343: Mathematical Statistics

**About the Course****Instructor** Evan L. RayEmail: [eray@mtholyoke.edu](mailto: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! I get paid to teach you, and I take that seriously.

**Course Website** We have a course website at [http://www.evanlray.com/stat343\\_s2018/](http://www.evanlray.com/stat343_s2018/). I will update this website regularly with a schedule and information about homework assignments, quizzes, and exams.

**Piazza** We have a Piazza page for this course at <https://piazza.com/mtholyoke/spring2018/stat343>. 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**.

**Classes** Class meets Tuesday, Thursday 1:15PM - 2:30PM and Friday 9:00AM - 9:50AM in Clapp 407.

**Description** The purpose of this class is to provide you with the theoretical understanding and computational skills necessary to select and implement appropriate methods for statistical inference in new settings.

We will examine the theory behind statistical inference procedures including point and interval estimation and hypothesis testing, as well as the computational methods needed to implement these inference procedures. In terms of theory, our goals are to understand common methods for deriving inference procedures for statistical model parameters, how and why these methods work, and how we can evaluate the relative performance of different inference procedures. We will focus primarily on inference for parametric models, but will also discuss inferential techniques that can be used when parametric assumptions are suspect. Computational tasks will include implementation of estimation procedures such as Newton's method for maximum likelihood, MCMC methods for Bayesian inference, and the non-parametric bootstrap. We will use simulation studies extensively to compare the performance of different approaches to inference.

A tentative schedule and topic list is below. That this outline is ambitious; we may not actually get to everything outlined here. 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	Weeks
Point Estimation	<b>Introduction to point estimation:</b> Maximum likelihood via analytic maximization; Bayesian inference with conjugate priors; loss and Bayes estimates	1, 2
	<b>Computational Issues:</b> Maximum likelihood via Newton's method; Bayesian inference via Metropolis-Hastings	2, 3
	<b>Sampling Distributions, Comparing Estimators:</b> Sampling Distributions; Bias, Variance, Mean Squared Error; Sufficient Statistics, Admissibility, and the Rao-Blackwell Theorem	3
	<b>Shrinkage:</b> Estimation of the mean of a multivariate normal with known covariance via Maximum Likelihood, Bayesian methods, and James-Stein estimation	4
	<b>Large-Sample Theory:</b> The asymptotic distribution of the MLE; Fisher Information	4
Interval Estimation	<b>Frequentist Confidence Intervals:</b> Large-sample approximate intervals; frequentist intervals for a univariate normal with unknown variance	4, 5
	<b>Bayesian Credible Intervals:</b> The general setup; Bayesian intervals for a univariate normal with unknown variance	5, 6
	<b>The Bootstrap:</b> Bootstrap estimation of a sampling distribution; bootstrap intervals	6
Decisions and Tests	<b>Decision Theory and Bayesian Decision Procedures:</b> More on utility, loss, and risk; Bayes decision rules; examples including comparison of normal distributions and possibly classification	7
	<b>Spring Break:</b> Safe travels!	8
	<b>Frequentist Tests:</b> The frequentist set-up; p-values, errors, power, and power functions; likelihood ratio tests; $t$ and $F$ tests	9, 10
	<b>Connections between hypothesis tests and confidence intervals</b>	10
Regression	<b>Simple Linear Regression:</b> Maximum likelihood and Bayesian inference	11
	<b>ANOVA:</b> Maximum likelihood and Bayesian inference for 1-way ANOVA; Shrinkage for ANOVA; Bayesian Hierarchical Models	12, 13
Special Topics	<b>Project Presentations</b>	14

**Textbook** We will be using “Probability and Statistics” (4th edition, ISBN 978-0321500465) by DeGroot and Schervish as the primary text for this class. A copy will be on reserve. I will assume that you are familiar with most of the material in the first 6 chapters of this text, other than the details of the Central Limit Theorem. We will begin with Chapter 7, and occasionally review material from earlier chapters as necessary.

I couldn’t find a book that I was completely satisfied with for this class. In order to fill in the gaps in DeGroot and Schervish, I will also be drawing on material from other sources. You don’t need to purchase these other texts, but in case you are curious – additional readings will likely come from the following texts, among others:

- Mathematical Statistics and Data Analysis by Rice
- Mathematical Statistics with Resampling and R by Chihara and Hesterberg

I will provide copies of all readings that are not found in DeGroot and Schervish.

## 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 nearly every day, 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.

We will use R via Jupyter notebooks. We have a class set up on <https://gryd.us/> which you can use to edit and compile Jupyter notebooks. You are also welcome to work locally on your own computer if you have Jupyter and R set up; however, please make sure you have installed at least version 3.4.2 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.

**Version Control with Git and GitHub** Git is a version control system that facilitates working on coding and writing projects collaboratively, and allows you to revert your code to a previous version if you realize that you made a mistake. Version control systems such as git are used in most modern data science and statistics positions in industry. Part of my goal as an educator in the statistics program is to ensure that you are prepared to enter the work force, and for that reason the basic use of git is a learning objective for this course. This means that all labs and the computational portion of homework assignments will be distributed to you in git repositories and submitted by committing and pushing the completed assignment to GitHub. I will provide further details and walk through this process, as well as basic interaction with git, in class.

## Assignments

**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 more theoretical component. 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, as well as a final exam 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.

**Project** There will be a final group project in which you will implement an inference procedure for a setting not otherwise discussed in this class, and conduct a simulation study to evaluate the performance of that procedure. You will present this work in a brief in-class presentation during the last week of class. I will give you more details about this project part way through the semester. A component of your grade on the group project will be determined by an assessment from your teammates of your relative contribution to the end result.

**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
Homework	35%
Project	15%
Quizzes and Midterms	35%
Final	15%