

Colgate University
CORE 143S: Introduction to Statistics[†]
Fall 2020 (8/27 - 12/18)

Professor:	Pierce Donovan	pdonovan@colgate.edu
Lectures:	Clark Room, JC Colgate	MW 3:35 - 4:50 pm
Office Hours:	Clark Room, JC Colgate	MW 4:50 pm - 5:50 pm
	Remote Via Zoom	F 1:00 pm - 2:00 pm

Course Description and Objectives

Statistics provides us with a way to deal with variation and uncertainty, which are generally things humans try to avoid. But if we learn about some of the ways to collect, analyze, and draw conclusions from data—which is inherently *noisy*—we can begin to recover some knowledge, or, more precisely, recognize patterns that may not be by chance. In this course, you will build an understanding of basic concepts in probability theory and statistical inference—but more importantly, you’ll begin to grasp the “feel” of good statistical inference and sense when its being used correctly.

How does polling work? Are there “hot hands” in basketball? How accurate does a SARS-CoV-2 test have to be before we deploy it? How does one build evidence for evolution, or any other scientific theory? We will answer each of these questions in turn. I aim to provide an exciting take on an interesting field while supplying you with the tools to become successful statistical thinkers. My goal is to augment your understanding of the world, how you interpret the news, and shape your interactions with others. Ideally, what you learn here will help you ask questions about things that interest you, and dig into them in a convincing manner.

This course lays out a useful mathematical foundation for many other courses, especially ECON 375: *Applied Econometrics*. Later in the course we’ll encounter the mechanics of regression, which econometricians use to control for confounding factors and isolate causal effects. Success in future courses will in no small part be due to your mastery of the concepts learned here.

Prerequisites (a healthy attitude)

Don’t expect to understand everything immediately. Learning is supposed to be uncomfortable! Not understanding something is a precursor to truly understanding something. I mean, it’s great if everything comes naturally for you, but eventually you’ll find something that’s pretty hard to grasp right off the bat (and will take some real effort outside of the classroom^{††}). If you acknowledge that this is to be expected from time to time, there’s suddenly no reason to feel frustrated with not getting something, and that makes it easier to enjoy the process of figuring something out.

[†]As the semester goes on, I may change the contents of this syllabus regarding the schedule, grading, or other details.

^{††}Google is very useful here. Wikipedia is a great starting point for reviewing certain ideas. StackExchange provides invaluable discussion on some of the harder concepts. Youtube has hundreds of videos that could augment my lectures.

Problem Sets

There will be six problem sets, five of which will be due on Mondays throughout the semester. The last assignment will cover asynchronous material that comes out after Thanksgiving break. I will upload each assignment (and related material) to Moodle two weeks in advance. Assignments will be submitted online—via Gradescope—and will be graded within one week of submission. I encourage you to work together, although you may only *submit* each assignment in groups of two (or solo). Late homework will be accepted for three days, with a 10% penalty per day.

Examinations

I design exams to evaluate how well you keep up with the assigned material and re-emphasize important points. I hope to make you think critically about problems you haven't faced before. Each exam has an in-class and take-home portion, although each piece is designed to take one class period (and you will have two class periods dedicated to both exams). The course is naturally cumulative, but the two exams roughly cover the 1st and 2nd halves. You'll find that the exams are highly correlated (in content and style) with my lectures and problem sets.

Grading

I don't *give* grades, you *earn* them. Further, I don't judge your performance relative to your peers (i.e. curve your grades) during the term in order for you to have the clearest signal about your performance. I look for a proven understanding of the material via the problem sets (60%), midterm (20%) and final (20%).

Covid-specific Things

- Our final exam will be on the last day before Thanksgiving break. In December, I'll have several recorded lectures that cover one last module.
- Missing lecture isn't the end of the world. If you feel sick, please do not come to class. I don't need an excuse, but please be in touch if you will be out for multiple days.
- Lectures will be streamed (audio+slides) for those who are remote or those who end up momentarily remote due to quarantine. In the case that we all have to return home before Thanksgiving, I will continue streaming from home.

Conduct

These are just a few more things to keep in mind that I put in all my syllabi:

- When asking for help outside of class, please be able to show how you have approached your problem. Simply asking for an answer is not a productive use of our time. I hope to facilitate critical thinking, and that takes effort on everyone's part.
- While I'll be accessible by email, I strongly prefer communicating during class/office hours. Regarding boundaries, I do not plan on answering emails late at night or on Sundays.
- I will not tolerate academic dishonesty. Colgate University's Academic Honor Code ([here](#)) requires instructors to report any suspected cheating, plagiarism, or other misconduct.
- You do not have permission to publish my course materials (online or otherwise). I don't want to see my work hosted somewhere like CourseHero (see the Academic Honor Code above).
- Please be respectful to your classmates. Refrain from talking during class if it is not relevant to lecture or discussion. Cell phone or tablet use should not detract from your ability to follow along with class. No activity on your part should undermine the efforts of other students.
- If you have any problems with this course or any other matters that may affect your work in this course, or you simply need someone to talk to, please contact me sooner rather than later. If you have a learning disability or a physical disability that requires accommodation, please let me know as soon as possible.
- Colgate University is a diverse community of individuals with many perspectives and identities. In order to create an inclusive and intellectually vibrant community, we must understand individual differences and common ground. Colgate University's report on Academic Freedom and Freedom of Expression ([here](#)) reflects the ideals I seek to uphold in this class.

Course Outline

Thinking About Probability

- Randomness is Random Monday, 8/31
Concepts: data generating process, information, jointly-distributed variables
- Statistics and Journalism Wednesday, 9/2
Concepts: numeracy, data presentation, forecasting
- Intro to Bayesian Statistics Monday, 9/7
Concepts: priors, posteriors, likelihood, evidence, beliefs
- Intro to Frequentist Statistics Wednesday, 9/9
Concepts: random variable, sampling, population, law of large numbers, p-values

Statistical Inference

- Statistics is Data Reduction Monday, 9/14
Concepts: expectation, variance, probability mass/density functions
Problem Set #1 due.
- The Central Limit Theorem Wednesday, 9/16
Concepts: distribution of averages, entropy, magic
- Some Initial Inferences Monday, 9/21
Concepts: p-values, null/alternative hypothesis, the normal distribution
- Some Initial Inferences II Wednesday, 9/23
- Student's t—A Versatile Instrument Monday, 9/28
Concepts: representative-ness of a [small] sample
Problem Set #2 due.

Midterm Exam

- Midterm Prep Wednesday, 9/30
- Midterm, Part 1 Monday, 10/5

Midterm, Part 2 Wednesday, 10/7

Makeshift Holiday–No Class Monday, 10/12

Problem Set #3 due.

Linear Regression

Testing Your Foundation Wednesday, 10/14

Concepts: the hypothesis testing process, CLT, t-distribution, sampling, notation

Random Experiments Monday, 10/19

Concepts: random assignment, counter-factuals, confounding factors, difference in means

Regression Anatomy Wednesday, 10/21

Concepts: curve-fitting, optimization, the least-squares estimator

Regression Physiology Monday, 10/26

Concepts: Gauss-Markov assumptions, distribution of regression parameters

Regression in Practice Wednesday, 10/28

Concepts: interpretation of regression results, intro to causality

Causal Thinking and Regression Monday, 11/2

Concepts: omitted variables bias, control variables, identifying variation

Research Design Wednesday, 11/4

Concepts: natural experiments, internal and external validity

Makeshift long weekend (no class) Monday, 11/9

Final Exam

Final Exam Prep Wednesday, 11/11

Problem Set #4 due.

Final Exam, Part 1 Monday, 11/16

Final Exam, Part 2 Wednesday, 11/18

Problem Set #5 due (11am). Thursday, 12/17