<div align="center">

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

</div>

Professor:	Pierce Donovan	pdonovan@colgate.edu
Lectures:	Clark Room, Colgate Hall	MW 3:35 - 4:50 pm
Office Hours:	*Remote* (Via Zoom)	TR 1:00 pm - 2:30 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.

[†]This syllabus is not a contract. As the semester goes on, I may change the contents of this document 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 certain Fridays 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%).

LaTeX[†] - an extra credit opportunity

I want to provide a little incentive for you to learn a powerful tool for typesetting [mostly scientific] documents. I will add 2% per assignment handed in that is typeset with LaTeX. That's basically a letter-grade. The easiest way to get started is definitely `Overleaf` (link), a cloud-based editor.

The main reason I value learning LaTeX is not that it makes for nice research papers, but that it helps you develop an important habit: working things out *by yourself* when things go wrong.[††] You'll occasionally need to spend a bit of time searching for a fix for what seems like the simplest thing. But you'll be better for it. Want to italicize text? Want to add a figure? Want to add a little bit of spacing between characters, but not *too* much spacing that it starts to look w e i r d? Google it. Read about how other people solved similar problems. You'll become a different person if you give this an honest attempt. It'll change how you go about solving those trickier statistics problems too.

[†]Pronounce it however you want.

[††]In grad school, I saw countless undergrads simply give up after the tiniest amount of effort, regardless of the difficulty of the problem at hand, because they wanted to be spoon-fed an answer. I got questions all the time from students asking things that were clearly on the syllabus, or would be obvious if they did the reading, or would be easier if they took a second to think about their problem before emailing about it. That's where this extra credit assignment is coming from.

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 not be streamed/recorded (except for the first two weeks). For those who are remote or those who end up momentarily remote due to quarantine, I plan to use the first few minutes of office hours to recap lecture. Like in any other year, it would be prudent to know someone else in the class that you can get the notes from.

- In the case that we all have to return home before Thanksgiving, I will record lectures asynchronously and try to expand office hours via Zoom.

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

Bayesians .... Monday, 9/7

Concepts: priors, posteriors, likelihood, evidence, beliefs

Frequentists .... Wednesday, 9/9

Concepts: random variable, sampling, population

Problem Set #1 due. .... **Friday, 9/11**

Statistical Inference

Statistics is Data Reduction .... Monday, 9/14

Concepts: expectation, variance, data visualization

When Samples Represent the Population .... Wednesday, 9/16

Concepts: degrees of freedom, law of large numbers, consistency, unbiasedness

Probability Distributions .... Monday, 9/21

Concepts: histograms, probability mass/density functions, central limit theorem

Simple Inference: Binomial Distribution .... Wednesday, 9/23

Concepts: test statistics, p-values

Problem Set #2 due. .... **Friday, 9/25**

Principles of Inference .... Monday, 9/28

Concepts: confidence intervals, type I/II errors

Developing Statistical Tests .... Wednesday, 9/30

Concepts: t-test, χ^2 tests, other tests of undetermined letter

Midterm Exam

Linear Regression

Concepts: random assignment, counterfactuals, confounding factors, paired t-test

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

Concepts: Gauss-Markov assumptions, distribution of regression parameters

Concepts: exogeneity, i.i.d. errors, heteroskedasticity, multicollinearity

Concepts: omitted variables bias, control variables, identifying variation

Concepts: natural experiments, internal and external validity

Maximum Likelihood Estimation

Concepts: maximum likelihood, categorical variables, odds, predicting probabilities

Concepts: logistic function, logit function, latent variable, likelihood ratio test

Concepts: resampling, sampling with replacement, approximate distributions

Final Exam

Machine Learning (Online)