SC212b: Introduction to Statistics and Data Science

Professor: Jerzy Wieczorek
Office: Lovejoy 110 (inside the Lovejoy 101 suite)
Telephone: (207) 859-5841
Email: jawieczo@colby.edu

Office hours: See our Moodle page

Class meetings: Lectures MWF 11:00 – 11:50am in Lovejoy 213

Class materials: Available at our Moodle course site
Textbook: Statistics: Unlocking the Power of Data (**3rd edition!**)
                       by Lock, Lock, Lock, Lock, & Lock
Course description:

> "Statistical thinking will one day be as necessary for efficient citizenship
> as the ability to read and write."   *- H.G. Wells*

> "For me, it is better to grasp the world as it really is than to persist in delusion,
> however satisfying and reassuring."   *- Carl Sagan*

Statistics is the science of "making sense of data".   Virtually *all* fields utilize statistics in some way. Not only that, but you are exposed to statistics (for better or worse!) practically every day – when you browse the internet, watch a sporting event, or listen to the news. *Statistics* are everywhere. This is true – even necessary – because *data* are everywhere, more so now than at any other time in history. How can we use all of this data to yield *useful* information? Extracting useful information requires more than learning a laundry list of methods; it requires careful consideration of context and the ability to *think* critically about data.

In this course, we will explore several tools that will help you make sense of data and draw conclusions, including methods for data collection, displaying and analyzing data, developing and testing hypotheses.   We will use RStudio, a statistical software package, to explore many real sets of data and to analyze them. Over the course of the semester, you will learn how to use this software effectively and how to interpret output from it. Additionally, we will spend considerable time learning and practicing how to communicate results effectively.

Statistics IS NOT just a list of methods. It is not "regression, T-tests, and ANOVA". You will most definitely learn about those topics in this class, but you will learn about the underpinnings of statistics: What is bias? What is confounding? What is a sampling distribution? Why should we care? When do traditional methods work? What should be done when they don't? And much more…

Course objectives:

By the end of the semester:

1. Students will understand the methods of data collection and the importance of utilizing appropriate sampling methods and experimental design procedures.

2. Students will be able to recognize which data analysis and data visualization methods are appropriate for a given situation.

3. Students will learn basic data manipulation and data cleaning skills.

4. Students will understand the limits of statistical inference and understand the assumptions behind common statistical methods.

5. Students will learn the implications of bias, variation and confounding in the context of statistical inference and methods to address and control them.

6. Students will become comfortable with the use of statistical software and the basics of computational thinking.

7. Students will understand the critical role statistics plays in research applications across a variety of disciplines.

8. Students will demonstrate the ability to communicate statistical results effectively.

Topics covered:

Obtaining data/sampling
Observational studies and Experiments
Visualizing data/plotting
Data manipulation and data cleaning
Linear regression
Hypothesis testing and confidence intervals
Bootstrapping and Randomization tests
Density curves, the normal distribution, and the central limit theorem
Parametric hypothesis testing: t-tests, z-tests
Categorical data analysis: chi-square tests, Fisher's exact test
ANOVA: planned and unplanned comparisons
Inference for linear regression
Multiple linear regression
Additional topics, time permitting

Evaluation in SC212

Grades: Your course grade will be determined as follows (subject to change):

	Written components of course grade:	
	Homework	15.0%
	Labs	10.0%
	Reading checks	5.0%
	Exams:	
	Midterm Exam #1	20.0%
	Midterm Exam #2	20.0%
	Comprehensive Exam #3	30.0%
	Total	100.0%

NOTE: IN ORDER TO PASS SC212 YOUR AVERAGE EXAM SCORE MUST EXCEED 60.0%.

Homework: Homework represents an important aspect of the course. Readings and selected problems will be assigned regularly. The assigned problems and due dates will appear in the Course Schedule on Moodle. Homework will require extensive use of the computer package RStudio. The assignments are designed to thoroughly test your knowledge of topics we discuss in class and are definitely not meant to be one-night jobs. Check Moodle for between-class announcements, hints, or modifications in assignments. Although you are encouraged to discuss problems with each other, I expect each person to submit their own work. You may choose to work with a partner on data analyses and thus produce the same computer output – that's okay, but I want written interpretations and responses to be in your own words. And remember, an answer to a statistics question is almost never just a number – I am looking for thoughtful explanations and conclusions!

Homework due dates and times will be posted on Moodle. Turn in assignments by 4pm on the due date. ***Late homework will reduce the student's grade substantially (by 25% of the maximum score for each day late), and very late homeworks will not be accepted,*** *in order to let the graders return everyone's work quickly and give you timely feedback. The professor will not grant extensions except for documented illness or other extreme circumstances.*

Class preparation: Short written check-ins will be given roughly every week (typically on Wednesdays) at the start of class, to help gauge how you are doing with the material. Questions will be based on the week's assigned reading and recently covered lecture material. I expect you to try your best on all the check-ins, but your 2 lowest grades will be dropped.
The purpose of the check-ins is to encourage you to keep up with the textbook readings and preview the material before we discuss it in class. They also provide me with a better idea of what material you understand thoroughly, what material is proving to be difficult, and what material to cover in more depth during class time.

Labs: Problem-based computing labs will be assigned roughly every week. These will be for you to complete on your own time; we will **not** be meeting as a class in a computer lab. The skills you learn from these assignments (and during regular class periods) are essential if you want to perform well on tests and homework assignments.

Exams: The three exams may involve both in-class and take-home components. These exams will focus on your ability to understand output from statistical software, to interpret results, to express an understanding of statistical concepts, and to engage in statistical thinking. They will *not focus* on plug-and-chug mathematics or elaborate computations. See our Moodle page for the 2 midterm exam dates; our final exam will be scheduled by the Colby registrar. **You must meet with me within the first two weeks of classes if you need to reschedule an exam date or request accommodations.**

FINAL EXAM – TBA (during <mark>Weds May 14 – Mon May 19</mark>)

YOU MUST BE PRESENT – MAKE SURE YOUR TRAVEL PLANS DO NOT CONFLICT WITH THE FINAL!!!

Moodle site: I'll be posting all HW assignments and labs on our Moodle site along with other resources, hints, schedule updates, and important class-related info. Check the Moodle site early and often at https://moodle.colby.edu

Software: We'll be using the software package RStudio to analyze data. Today, virtually all data analysis is done with the aid of at least one computer software package. R is one of the most powerful and most flexible ones – it's also probably the most widely used of all statistical packages. You can access RStudio via the web at: https://rstudio.colby.edu

In addition to using and becoming familiar with this software, you will also need a calculator that can at least do square roots, logs, and exponential functions.
We'll also make extensive use of the free StatKey website that accompanies the book. StatKey is both a visual aid for understanding statistical concepts and a web-based tool that can perform statistical analyses.

Laptop use in class: Please do not use electronic devices (e.g., computers, tablets, smart phones) during class. It is distracting for those around you, for lecturers, and for you. Please consult with me if you have concerns about this as a support for your learning needs.

Masks: Per guidance from the Dean of Faculty, "Regardless of campus-wide policies, faculty and teaching staff may require masks in classrooms and other teaching spaces. More generally, we appreciate everyone respecting the needs of others and wearing masks as requested, and we ask that you have a mask with you when you are on campus." Depending on conditions, I may expect the SC212 classroom to mask in order to protect those of us with compromised immune systems, those of us with high-risk family members, and those of us who simply want to avoid transmitting viruses to others or being re/infected themselves.

Academic accommodations:
I am available to discuss academic accommodations that any student with a documented disability may require. Please note that you'll need to provide a letter from the Dean of Studies Office documenting your approved accommodations. *Please meet with me within two weeks of the start of the semester to make a request for accommodations so that we*

can work together with the College to make the appropriate arrangements for you. Colby's Office for Student Access and Disability Services is the primary contact for accommodations and any questions related to educational testing and documentation: https://life.colby.edu/get-support/access-disability-services/

Sexual misconduct/Title IX statement:

Colby College prohibits and will not tolerate sexual misconduct or gender-based discrimination of any kind. Colby is legally obligated to investigate sexual misconduct (including but not limited to sexual assault and sexual harassment).

If you wish to speak confidentially about an incident of sexual misconduct, please contact Colby Counseling Services (207-859-4490) or the Confidential Resource Advisors (207-509-9122).

Students should be aware that faculty members are considered responsible employees; as such, if you disclose an incident of sexual misconduct to a faculty member, they have an obligation to report it to Colby's Title IX Coordinator, Emily Schusterbauer (207-859-4266). "Disclosure" may include communication in-person, via email/phone/text, or through class assignments.

To learn more about sexual misconduct or report an incident, visit https://life.colby.edu/your-safety/sexual-violence-title-ix/

Academic honesty & consequences for academic dishonesty:

Honesty, integrity, and personal responsibility are cornerstones of a Colby education and provide the foundation for scholarly inquiry, intellectual discourse, and an open and welcoming campus community. These values are articulated in the Colby Affirmation and are central to this course. Students are expected to demonstrate academic honesty in all aspects of this course.

Academic dishonesty includes, but is not limited to: plagiarism (including quoting sources without " " around the borrowed words and a citation); presenting another's work as one's own; buying or attempting to buy papers or projects for a course; fabricating information or citations; knowingly assisting others in acts of academic dishonesty; violating clearly stated rules for taking an exam or completing homework; misrepresentations to faculty within the context of a course; and submitting the same work, including an essay that you wrote, in more than one course without the permission of instructors.

Academic dishonesty is a serious offense against the college. Sanctions for academic dishonesty are assigned by an academic review board and may include failure on the assignment, failure in the course, or suspension or expulsion from the College.

For more on recognizing and avoiding plagiarism, see: https://libguides.colby.edu/avoidingplagiarism

For resources and information on academic integrity, see: https://www.colby.edu/academics/academic-integrity/

A special note on AI and ChatGPT:

Under no circumstances should you be using any type of AI output in your submissions for assignments or exams. This includes, but is not necessarily limited to, using any kind of online AI chat (e.g. ChatGPT). You may find AI useful as a supplementary learning tool (such as searching for code syntax, or rephrasing the explanation of a statistical concept to help you understand it yourself), but you may not use AI to produce answers to the assignments

themselves. All writing that is conducted in this class must be original and be your own. If you have any questions about this, ask me.

The Colby Affirmation:

Colby College is a community dedicated to learning and committed to the growth and well-being of all its members. As a community devoted to intellectual growth, we value academic integrity. We agree to take ownership of our academic work, to submit only work that is our own, to fully acknowledge the research and ideas of others in our work, and to abide by the instructions and regulations governing academic work established by the faculty. As a community built on respect for ourselves, each other, and our physical environment, we recognize the diversity of people who have gathered here and that genuine inclusivity requires active, honest, and compassionate engagement with one another. We agree to respect each other, to honor community expectations, and to comply with College policies. As a member of this community, I pledge to hold myself and others accountable to these values.

https://www.colby.edu/academics/academic-integrity/the-colby-affirmation/