MATH 354: DATA ANALYSIS I
TLDR SYLLABUS OVERVIEW

Meeting Times (p. 1)

- Lecture Tuesday/Thursday 9:55a-11:10a
- Office Hours: Tuesday 11:30a-12:30p, Wednesday 11a-12p
- Schedule on Page 9

What you will learn (p. 2)

1. evaluate ethical issues associated with statistical practice
2. critically consume statistically-based results
3. choose and interpret appropriate graphical displays and numerical summaries of real data
4. recognize and explain the central role of variability and randomness in statistical analyses
5. demonstrate an understanding of, and ability to use, ideas of statistical inference in a variety of settings
6. interpret and draw conclusions based on standard output from statistical software

Materials and Technology (p. 2)

- I will provide textbook-style notes for the course.
- We will learn to conduct data analysis using R.

Policies (p. 2-3)

- You need to attend class as often as possible.
- Outside class discussion is encouraged outside of class (Moodle + Hypothes.is)
- Make-ups must be taken on the same day or before.
- It is my goal to make this course accessible and inclusive. Let me know what you need.

Important Dates (p. 8)

- Midterm: 11/05/24 (9:55a-12:55p in TBA)
- Midterm Review: Due by 11/12/24 5p
- Midterm Revision: Due by 11/21/24 5p
- Final (12/19/24 12p-2p in McGregory 210 101)

Grades (p. 5-7)

- Weekly Check-Ins and Homework
- The Midterm Exam (two parts) and the Final Exam

    Course Total $= 0.30(\text{HW}) + 0.30(\text{Midterm Part I}) + 0.10(\text{Midterm Part II}) + 0.30(\text{Final})$.

Contact

- Students should post all questions on course contents or technology on Moodle or Hypotheses.is.
- We can handle other inquiries during office hours or via email to wcipolli@colgate.edu.

MA 354: Data Analysis I — Fall 2024 — Section A
Meeting Time: Tuesday, Thursday 9:55a to 11:10a in McGregory 210

Professor: Will Cipolli — wcipolli@colgate.edu — https://cipolli.com — McGregory Hall 323
Office Hours: Tuesday 11:30a-12:30p, Wednesday 11a-12p, and by appointment.

Purpose: To introduce students to statistical analyses for assessing evidence for or against theories about the world. Students will learn the language of data analysis, which will help them understand how to evaluate assumptions, fit a model, and interpret the results in the context of research inquiries across disciplines.

Course Objectives: After this course, students will be able to:
1. Calculate and evaluate various point estimators
2. Understand sampling distributions and their impact on inference
3. Choose the correct association measurements for data from a research inquiry
4. Select the correct regression methodology for data from a research inquiry
5. Assess regression model fit by evaluating assumptions of the model
6. Ameliorate any transgression of assumptions found when modeling
7. Interpret regression models to answer research questions
8. Articulate the differences in modeling continuous and discrete response variables

In 20 years, I want students to remember that the juice is worth the squeeze. If they struggle with a puzzle,they *can* solve it. We're working toward a holistic understanding and not mindless mimicking.

Productive Failure: I want to recast failure as a learning tool. Realizing mistakes in practice provides a rich time for learning if we complete the hard work of helping each other to the point of epiphany. This approach requires us to signal our need for help, necessitating an environment where it is safe to take risks and connect. In class discussion, we will celebrate curious risk-taking as much as the correct answer. An incorrect response provides the best place to grow – we will *learn* to engage with mathematics.

Rules of Engagement:
1. All humans are accepted members of our classroom.
2. Be aware of others' identities in the room.
3. Assume positive intent.
4. Share talking time.
5. Listen to understand.
6. Be present.
7. Critique ideas, not people.
8. Everyone has expertise. We can learn something from everyone.
9. Share a feeling of mutual responsibility for each other.
10. Encourage others to succeed.

Prerequisite: MATH 240 or ECON 375 or BIOL 320 or PSYC 309 or (MATH 105/CORE 143S and MATH 260) or (MATH 105/CORE 143S and COSC 290) or permission.

Technology: Students will have the opportunity to learn and use R for this course. I assume that students have some previous experience with R or another programming language (e.g., Python). Coding isn't easy, learning it or teaching it, but this skill will become more and more critical over time. I heavily used documentation while teaching myself R and still do, even now. When tackling a problem, I search for a solution in the documentation, work to understand the code, and tweak it according to my requirements. This is the level of work students should expect – using the course resources, understanding the solutions, and altering them to answer new questions. If students feel like they're reinventing the wheel from scratch or need to be an expert software developer to complete an assignment, they should stop and see me.

Support for Technology: Devices like laptops are paramount to success in college. I recognize that these devices can be expensive and that students might not have the same access to the latest technology. Further, technology changes rapidly, and students might rely on older, more problem-prone devices that break down or become unreliable. These technology issues can become a significant source of stress for students. Given these challenges, students should contact me if they experience a technology-related problem that interferes with their learning in this course. Doing so will enable me to assist students in accessing the appropriate resources on campus.

Attendance: I expect students to attend all classes and to arrive on time. When a student misses class due to illness, hangovers, interviews, personal crises, deaths in the family (I hope not!), and whatever else, they do not need to let me know. Students should talk to classmates and check the Moodle page to see what they missed. All students are responsible for all assignments due or assigned in the class they miss. I want students to attend all classes. Every class they don't attend isn't just discussion and material they missed; it's also thinking they didn't do – thinking they will need for assignments and exams later on. In other words, every missed class is a disadvantage. The obligation is on the student to minimize that effect. That said, there is no penalty or benefit for attendance, as deflating or inflating grades with any percentage of a student's score coming from attendance would make a poor measure of an individual's competency in the course. Students seeking high grades will quickly learn that they need to attend class as often as possible.

Outside Class Discussion: Students should use the discussion board in Moodle and the Hypothes.is annotation platform as safe places to ask questions and be curious about the course material. I expect students to answer such questions and feed their peers' curiosity by furthering the discussion; I will monitor activity and chime in often. Through this design, I intend to foster students' creativity and curiosity, preparing them to think critically, ask questions, and gain lifelong value from their education.

Make-up Policy: I will consider make-ups and extensions on a case-by-case basis. Students who feel they are in an extreme circumstance must notify me at least two days **before** the regularly scheduled deadline or as soon as possible. Students should feel welcome to reach out to discuss any due dates or exam dates that conflict with their religious observations or other dates that the University does not acknowledge. We will schedule all make-up exams on the same day as the exam when possible, and **before** if not.

Inclusion: My goal and responsibility is to make this course and our classroom as accessible and inclusive as possible. I understand that students have different styles and paces of learning and accessing information and that each student comes with their own, sometimes difficult, experiences with learning. I acknowledge the persistence of discrimination and exclusion in mathematics based on race, gender, socioeconomic status, and other factors. I take responsibility for lowering barriers so that access is practical and equitable. We must work to make the classroom environment as comfortable and respectful as possible. As a class, we will resolve to listen, learn, and act to make this classroom proactively welcoming to all students. I encourage all students to contact me to discuss their learning process, experience, or needs and point out any blind spots.

Specific Learning Accommodations and Support: I hope students will feel comfortable notifying me at the start of the course if they require specific learning accommodations or support. I am here to help! This information will remain confidential. In many cases, students requesting accommodations must also contact the Office of Academic Support and Disability Services to receive help determining and coordinating a specific accommodation based on disability/medical documentation. Contact Evelyn Lester: elester@colgate.edu, (315) 228-6955.

Academic Honesty: I expect students to follow Colgate's academic honor code. If a student feels stressed about exams or deadlines, they should come to see me as soon as possible so we can review their options to avoid any academic honesty issues. See Colgate's Academic Honor Code.

Large Language Models: If students have questions about how to do something, I have no problem with students "looking it up" on Google (I do this frequently) or large language models (e.g., ChatGPT). However, online searches and large language models may provide incorrect content or content beyond the scope of this course. While these tools can help figure out how to approach something, I expect student solutions to be their own work that they fully understand in the context of our course.

Support: College life can sometimes get bumpy; if you are experiencing emotional or personal difficulties, seek help right away. Colgate offers wholly confidential and highly professional counseling and psychological services. You can reach the Counseling Center at 315-228-7385. If this seems like a difficult step, find me – we can talk and call or walk to the Counseling Center together.

How to Succeed in this Class

1. Go to office hours. These conversations get you past an immobilizing issue in understanding and help me understand where students are in their learning process. During office hours, I often have a discussion that completely changes how or what I teach the next class. These meetings help us get on the same page. Come to my office hours regularly, even if you aren't struggling with the current material. If you start to struggle, plan to see me immediately (even if it has to be outside office hours).

2. Come to class prepared to discuss the material for that day's lecture. Being prepared means actively reading and thinking about past material by investigating the concepts independently. Try practice exercises, run the sample code on your own, and try a problem from class without consulting the answer. When you come to class with questions based on the concepts from these activities, we can strengthen and expand our knowledge in lectures.

3. For every hour in class, you are expected to spend 2-4 hours outside of class reading, working on assignments, and studying for exams. Be sure this time is productive – seek advice if you are 'spinning your wheels.'

4. Invest a small amount of time immediately after an assignment is given to ensure you understand it and don't have significant questions. Then, break down the assignment into manageable pieces and work on them over the week. If you wait until the last minute, seemingly insurmountable problems will undoubtedly arise; by then, it's too late to get assistance. Remember, it takes no more time to complete an assignment if you spread it out, not to mention research shows you'll retain more if you do.

5. Ask well-informed questions. Questions such as "I don't understand X; can you explain X to me?" are welcome but not well-informed and will almost certainly not get you the answer you want. Instead, ask questions that reveal your current knowledge of the topic, similar to the following: "I understand how Y works, and I see that X is different from Y in way Z. What is it about X that causes this difference?" Answering these questions will be much more informative and help us reach our goals.

6. Form study groups as soon as possible and actively read, study for exams, and work on homework assignments together.

7. Understand and remind yourself that performance on homework or exams does not represent your capability or intelligence. These assessments are snapshots of where we are and diagnostic tools for where we need to go. We are not proving our intelligence but developing it. The goal is to grow; mistakes are not evidence of a lack of capability but the illumination of places to improve.

Grading:

Homework (30%): The purpose of homework is to practice concepts introduced in the lecture. Students can expect 10-14 homework assignments. These assignments will consist of two pieces:

- **Homework:** I will grade these assignments for correctness and will not accept late submissions unless previously agreed upon. I encourage students to discuss the homework opportunities with each other and me. This can be done in various ways – on the Moodle discussion board or during office hours. Students should judiciously review posted solutions to homework in preparation for exams.

- **Weekly Diagnostic Check-points:** I will ask students to reflect on what we've discussed by seeing what they thought was important and guide them to engage with the course material so they can ask any lingering unanswered questions.

Delayed Work: Homework is due at 5p, so you don't feel pressure to stay up late. Still, each homework will have a grace period and can be handed in until noon the following day without penalty. This is for small unexpected delays (e.g., a compiling issue).

Late Work: If you find yourself in a position where you know you won't be able to complete the homework, notify me as soon as possible. Please send me an email (a) explaining the situation, (b) a complete current draft of your homework demonstrating progress, and (c) a new proposed date that is within 72 hours of the original due date.

Midterm Part I (30%): The midterm will involve foundational questions and data analysis using methods learned up to that point. Below, I delineate the graded pieces of Part I, **which is scheduled for 11/05**.

- **Foundational Questions (30%):** There will be several questions on various course topics to inventory the concepts students understand well and which topics need revision.

- **Initial Data Analysis (70%):** I will provide students with a research inquiry and a dataset. Students will write a preliminary consultation during an in-person exam that includes data analysis and interpretation, which helps answer the research inquiry. The expected result is a rough draft analysis.

Midterm Part II (10%): The midterm will also have a take-home aspect where students will engage in peer review and revise their data analysis from Part I. Below, I delineate the graded pieces of Part II.

- **Peer Review (15%):** All students will receive two anonymized exams to review. For each exam, the student must create a document containing praise, *constructive* criticism, and remaining questions about the analyses that will aid the anonymous students in creating a better second draft. **This is due 11/12 at 5p.**

- **Revision (70%):** Students will submit a revised copy of their data analysis after receiving their peer-reviewed exams. That is, students have the opportunity to *improve* work after receiving feedback. The aim is to measure learning more accurately and model the process more closely by re-evaluating students along the learning process. Students should take a critical peer review as an invitation to try again – revisit the notes and past solutions and ask for help so that the revision is a success. A light peer review is a chance to hone their knowledge on the subject by providing a highly formal solution in the revision. The expected result is a highly polished final draft **due 11/21 at 5p**.

- **Response to Peer Review (15%):** Students should provide a document addressing the peer reviews' concerns and questions. This document should provide point-by-point responses to comments in the peer reviews. It is okay not to follow *every* piece of advice, but responding to *all* feedback is essential. **This is due with the revision on 11/21 at 5p.**

Final Exam (30%) A comprehensive final exam will be posted on the last day of class and is due **December 19th at 2p**, as scheduled by the registrar. Students may use the internet and all course materials to complete the final exam but should not receive outside help from other sources, e.g., students, tutors, faculty, freelancers, etc. The final exam has no peer review or revision opportunity, so students should self-review and revise before submission.

Scoring

– *Rubric:* Each question asked in an exam period is scored on the following rubric:

Designation	Required Objectives	Points
A (Mastery)	• Perfect for the standard being assessed • Achieves a correct solution • Justifies decision(s) toward solution • Effectively communicates solution and support • Notation used is appropriate and clearly shows all steps	0.95
B (Sufficient)	• Essentially contains the correct answer but contains a slight error • Makes correct decision(s) toward solution • Justifies decision(s) toward solution • Effectively communicates solution and support • A slight error, confused reasoning, or notation mistake	0.85
C (Progressing)	• Does not contain the correct answer but does show work in the correct direction • Makes some correct decision(s) toward solution • Some justification of decision(s) toward solution • Attempts to communicate solution and support • A wrong decision, confused reasoning, or notational mistakes	0.75
D (Developing)	• Does not contain the correct answer but shows some correct work • Incorrect decision(s) toward solution • Insufficient or incorrect justification for decision(s) toward solution • Little or no communication of solution and support • Several wrong decisions, confused reasoning, or notation mistake	0.65
F (Needs Attention)	• Does not contain the correct answer or work in the correct direction • Missing or incorrect decision(s) toward solution • Little or no justification for decision(s) toward solution • Little or no communication of solution and support • Several wrong decisions, confused reasoning, or notation mistake	0.25
Z (Not assessable)	No Response: There has been no reasonable attempt to provide the correct solution.	0.00

Plus or minus grades will be decided for solutions between categorizations. For example, a solution with a trivial error that isn't critical to the understanding or the correct completion of the exercise would be an A- whereas a B+ would be, for example, an easily correctable issue that may show a small, tangential misunderstanding – a solution that shows a clear understanding of the material but some area for growth in a pre-requisite or a small part of the concept being assessed.

Overall Grade: A student's overall grade will be a weighted average of their **percentage** scores on homework, midterm, and final exams. The overall grade earned by each student will be decided as follows.

$$\text{Course Total} = 0.30(\text{HW}) + 0.30(\text{Midterm Part I}) + 0.10(\text{Midterm Part II}) + 0.30(\text{Final}).$$

Letter	Final Grade
A	93-100%
A-	90-92.$\bar{9}$%
B+	87-89.$\bar{9}$%
B	84-86.$\bar{9}$%
B-	80-82.$\bar{9}$%
C+	77-79.$\bar{9}$%
C	73-76.$\bar{9}$%
C-	70-72.$\bar{9}$%
D+	67-69.$\bar{9}$%
D	63-66.$\bar{9}$%
D-	60-62.$\bar{9}$%
F	< 60%

- **A** range represents above and beyond expectations, excellence with distinction. These are not impossible to achieve but are difficult to come by. While there is merit to hard work and long hours, it does not always guarantee success. Excellence refers to the combined results, not just the effort.

- **B** range signifies that a student is meeting the expectations of the course in most or all aspects. Good is more common than excellent and should be celebrated as a success.

- **C** range signifies adequate and at the level of expectation for several aspects of the course. Average is not usually an appealing categorization for those who strive for extraordinary. A grade of C, however, is a respectable point. If students don't want to be categorized as adequate, they must recognize what more is needed, make a plan to achieve that, and execute it; I can help with a plan!

- **D** range represents less than adequately equipped to perform many of the essential functions of the course; just passable. I recognize that a D may also mean a student does not understand what is expected. Students, in this case, should make an appointment with me to discuss how they might make a plan and take action. I will submit course warnings to the appropriate Administrative Dean for students earning a D in this course at any point during the semester.

- **F** range represents an apparent failure to meet the expectations of the class. F represents a lack of effort and interest in the course. This is a cause for deep concern; I will submit course warnings to the appropriate Administrative Dean for students earning an F in this course at any point during the semester.

Remark: I do not curve or round grades at the end of the semester. No matter what policy is followed, some could miss a grade boundary by a minimal amount. I prefer to keep it straightforward by announcing the sharp grade boundary and strictly following it. I find it helps keep the process more objective and does not allow room for subjective grade adjustments, which are almost always unfair. I expect students to use the nature of retesting to "curve" their grades and work with me to *earn* the grade that they want.

Schedule:

Week 1	
08/29/24	First Day of Classes (Half-Day Schedule Meet at 9:30a)
08/30/24	Introduction to R, Sweave, and Knitr
Week 2	
09/03/24	Coding in R
09/05/24	Tidyverse
Week 3	
09/10/24	Tidyverse
09/12/24	No Class
Week 4	
09/17/24	Summarizing Categorical Data
09/19/24	Summarizing Quantitative Data
Week 5	
09/24/24	Discrete Probability Distributions
09/26/24	Continuous Probability Distributions
Week 6	
10/01/24	One-Parameter Maximum Likelihood Point Estimation
10/03/24	Two-Parameter Maximum Likelihood Point Estimation
Week 7	
10/08/24	Resampling
10/10/24	Central Limit Theorem
Mid-Term Recess 10/12–10/15	
Week 8	
10/15/24	No Class
10/17/24	One Sample Inference (t-test)
Week 9	
10/22/24	One Sample Inference (z-test)
10/24/24	Bootstrapping and Randomization
Week 10	
10/29/24	ANOVA and Tukey's HSD
10/31/24	Chisquared and two-sample z-tests
Week 11	
11/05/24	**Midterm Exam (9:55a-12:55p in TBA)**
11/07/24	Multiple Linear Regression: Model
Week 12	
11/12/24	Multiple Linear Regression: Assumptions and Outlier Analysis
11/14/24	Multiple Linear Regression: Transformations
Week 13	
11/19/24	Multiple Linear Regression: Robust Method
11/21/24	Multiple Linear Regression: Untangling Interactions
Thanksgiving Recess 11/23–12/01	
Week 14:	
12/03/24	Multiple Linear Regression: Wrap Up
12/05/24	Logistic Regression
Week 15:	
12/10/24	Logistic Regression
12/12/24	*Time Permitting:* Multinomial Regression, Count Regression
Final Exam 12/19/24 12-2p	

Remark: Dates will likely change as I let the class dictate the speed of the course.