

MATH 354: Data Analysis I

TLDR SYLLABUS OVERVIEW

Meeting Times (p. 1)

- Lecture Monday/Wednesday 01:20p-02:35p
- Office Hours: Tuesday 11:15a-12:15p; Thursday 12:15p-1:15p

What you will learn (p. 1, 5, 8)

1. Effectively summarize data through numeric summaries and data visualization
2. Understand sampling distributions and their relationship to statistical inference
3. Select the correct methodology to answer research questions
4. Assess conditions for using selected methods and adjust where necessary
5. Implement, interpret, and convey results in the context of the research question

Materials and Technology (p. 1)

- I will provide textbook-style notes for the course.
- We will learn to use R for conducting data analyses.

Policies (p. 2)

- 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 the original quiz.
- It is my goal to make this course accessible and inclusive. Let me know what you need.

Tips for Success (p. 3)

- Engage with material in and out of class by reading, practicing consistently, and collaborating.
- Focus on conceptual understanding, clearly communicate work, and seek help when needed.

Activities (p. 4)

- Class time will consist of a mix of lectures and activities
- Weekly homework
- Standards
- Final Exam

Grades (p. 4-7)

$$\text{Course Total} = 0.10(\text{HW average}) + 0.60(\text{Standards}) + 0.30(\text{Final})$$

Schedule (p. 8)

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

Contact

- Students should post all questions on the course contents or technology on Moodle or Hypothes.is.
- Visit office hours or email me for personal or off-topic inquiries.

MA 354: Data Analysis I — Fall 2025

Lecture: Monday, Wednesday 01:20p to 02:35p in McGregory 201

Professor: Will Cipolli — wcipolli@colgate.edu — www.cipolli.com — McGregory Hall 323

Office Hours: Tuesday 11:15a-12:15p; Thursday 12:15p-1:15p.

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 enable them to understand how to evaluate assumptions, fit a model, and interpret the results in the context of research inquiries across various disciplines.

Course Objectives: After this course, students will be able to:

1. Effectively summarize data through numeric summaries and data visualization
2. Understand sampling distributions and their relationship to statistical inference
3. Select the correct methodology to answer research questions
4. Assess conditions for using selected methods and adjust where necessary
5. Implement, interpret, and convey results in the context of the research question

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 with others. In class discussions, 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: I appreciate students who are courteous, on time, thoughtful, interested in learning, honest, and responsible. The following provides a list of ten tips for meeting these expectations.

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.

Prerequisites: Computational Statistics (MATH 240) with lab (MATH 240L).

Technology: Students will have the opportunity to learn to conduct data analyses in R during this course. I assume that students have some familiarity with R through our MATH 240 prerequisite or a similar experience, including working with the `tidyverse` and `ggplot` packages. This course will build on this knowledge, enabling students to conduct more complex analyses. Students will be given ample examples and resources to complete this work. If students feel overwhelmed or like they 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. Furthermore, technology changes rapidly, and students may 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. 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 review the Moodle page to see what they missed. All students are responsible for all assignments due or assigned in the class they miss. 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 quizzes 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 on Moodle and the Hypothes.is annotation platform as safe places to ask questions and be curious about the course material. All requests for help should be made on Moodle, so everyone has access to the same resources. Such posts should include the problem, the attempted solution, and any other necessary information. A toy example should be included where debugging is needed. I expect students to answer such questions and feed their peers' curiosity by furthering the discussion; I will monitor activity and chime in often.

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 quiz dates that conflict with their religious observances or other dates that the University does not recognize. We will schedule all make-up quizzes on the same day as the quiz when possible, and before if not.

Inclusion: My goal 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 challenging, 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. As a class, we will commit to listening, learning, and taking action to create a welcoming environment for all students in this classroom. I encourage all students to contact me to discuss their learning process or needs, and to point out any areas that may need improvement.

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 their 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 quizzes 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](#).

Generative AI and Large Language Models: We will take part in a research project this semester. In exchange for only using PingPong (an AI developed at Harvard), we will all receive a complimentary subscription. If students have questions about how to do something, I have no problem with them "looking it up" on Google (I do this frequently) or using large language models (e.g., PingPong). However, online searches and large language models may provide incorrect content or content beyond the scope of this course. Further, you may find *a* correct answer that is not *the* correct answer in the context of our course. While these tools can help determine how to approach something, I expect student solutions to be their own work, which 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 immediately. 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, call, or walk to the Counseling Center together.

How to Succeed in this Class:

1. **Work toward conceptual understanding.** The goal of college-level mathematics is to understand the material in a way that helps students apply it in new ways and to new situations, not to reproduce what is done in class. The emphasis is on developing judgment that can only be attained through conceptual understanding. Memorization and a cursory understanding will not be sufficient, and students may need to adjust their study methods and approach to completing homework assignments.
2. **Be active at home.** Mathematics is like juggling – students cannot learn to juggle by listening to someone lecture about it. Time spent in class is intended to introduce students to the material, providing them with the tools to learn and become comfortable with it. During this process, students will experience some discomfort as they grapple with the material independently by reading the textbook, reviewing class notes, and completing the homework.
3. **Be active in class.** Students should come to class prepared to discuss the material. When students come to class with a cursory understanding of the topics for discussion, they can better grasp the lecture and ask productive questions that clarify any confusion.
4. **Read the textbook.** It is impossible to learn *everything* in class. The textbook provides detailed descriptions of the course material, offering more depth and examples than can be covered in a lecture. I recommend reading the material before class to have the best chance of absorbing the introduction to the material. Then, students can use the textbook to answer any questions while revising the material and completing homework.
5. **Pace yourself.** Students should allocate 2-4 hours outside of each class meeting to read, work on assignments, and study for quizzes. Students should invest a small amount of time immediately after an assignment is given to ensure they understand it and don't have significant questions. Then, divide the assignment into manageable pieces and work on them over the week. Waiting until the last minute leads to overwhelming problems and limited access to assistance. Remember, it takes no more time to complete an assignment if a student spreads it out over time, not to mention research shows they'll retain more if you do.
6. **Communicate effectively.** Solutions are not just numbers. A comprehensive understanding of the proposed solution requires the full process, including all steps and necessary justification. Students should ensure their solutions are clearly communicated as they complete their work. Doing this will help them reinforce understanding and make connections between concepts.
7. **Work together.** Form study groups and actively read, study for quizzes, and work on homework assignments together. Discussing the material will help students familiarize themselves with the concepts in their own terms and provide them with memorable moments of clarity.
8. **Ask well-informed questions.** If students do not understand the material, they should talk with me right away to avoid falling behind. Mathematics is annoyingly cumulative, so any lapse in understanding will compound with time. When students are confused, they should ask well-informed questions right away. Students should avoid vague questions such as "I don't understand X; can you explain X to me?" Instead, they should ask specific questions that reveal current knowledge of the topic. For example, "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?" Answers to these questions will be much more informative and helpful.
9. **Go to office hours.** These sessions get students past an immobilizing issue in understanding and help me understand where they are in their learning process. During office hours, I frequently engage in discussions that inform subsequent lectures. If students start to struggle, they should plan to see me immediately (even if it has to be outside office hours).
10. **Be kind to yourself.** Students should understand and remind themselves that performance on homework or quizzes reflects a specific point in the learning process, not their capability or intelligence. These assessments provide snapshots of our current state and serve as diagnostic tools for identifying areas for improvement. We are not proving our intelligence but developing it.

Course Grade:

Homework (10%): The purpose of weekly homework is to practice concepts introduced in the lecture. I will grade these assignments based on their correctness. I encourage students to discuss the homework opportunities with each other and me.

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.

Standards (60%): We will cover roughly 11 *standards*, which include statistical concepts such as creating a visual display of data. Students can expect three standards quiz periods, during which they will be evaluated on recently covered standards and reassessed on previously assessed standards. These quizzes do not have a percentage score. Instead, questions are graded and tracked individually.

- **Why?:** This is to provide an iterative process of learning. Initial grades are meant to serve students by providing clear, specific, and actionable feedback on their strengths and areas for improvement.
- **How?:** Quizzes will be returned quickly and accompanied by detailed feedback on the progress made on each assessed standard, which provides a blueprint for future success.
- **Using Feedback:** Make attempts productive. Take a struggle with a particular topic as an invitation to try again – revisit the notes and past solutions, and ask for help so that the second time is a success. Take success as an opportunity to retain and refine that learning by providing a highly formal solution the second time.
- **What's the Benefit?:** This allows for the opportunity to *improve* understanding after receiving feedback. The aim is to measure learning more accurately by modeling the process more closely, allowing students to be evaluated after receiving feedback and revisiting the material. Additionally, the course topics are implicitly cumulative, and repeated retrieval leads to deeper learning and easier access to new material.
- **Weighting:** The first and second attempts account for 25% and 75%, respectively, unless equal weighting benefits a student.
- **When?:** These assessments will occur roughly every three weeks.

Final Exam (30%) The final exam will be a comprehensive exam: **12/16/25 3-5p.**

Standards Scoring

– List of Standards:

ID	Standard/Objective	First Attempt	Second Attempt
1	Data Visualization	Period 1	Period 2
2	Probability Models	Period 1	Period 2
3	One Sample Inference	Period 1	Period 2
4	k Sample Inference	Period 1	Period 2
5	k Sample Post Hoc Tests	Period 1	Period 2
6	Regression Assumptions	Period 2	Period 3
7	Regression Interpretation	Period 2	Period 3
8	Regression Interactions	Period 2	Period 3
9	Regression Model Selection and Comparisons	Period 3	—
10	Robust Regression Methods	Period 3	—
11	Count Regression Methods	Period 3	—

– *Rubric:* Each question asked in a quiz is scored on the following rubric:

Designation	Required Objectives	Points
A (Mastery)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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, and/or notational mistakes 	0.75
D (Developing)	<ul style="list-style-type: none"> 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, and/or notation mistakes 	0.65
F (Needs Attention)	<ul style="list-style-type: none"> 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, and/or notation mistakes 	0.25
Z (Not assessable)	No Response: There has been no reasonable attempt to provide the correct solution.	0.00

A Note on Grading Code:

Learning extends beyond the classroom, but submitted work must demonstrate understanding. While I expect students will use resources available to them to complete the work of the class, I further expect that students *fully* understand any work they complete and submit.

Scripts using statements, structures, or libraries not covered in class must include both: (a) a citation to the learning resource and (b) a narrative explaining the code's function, methodology, and why it was necessary to go beyond what we learned in class. Submissions lacking either will receive an "F."

– Grading –

Typical testing and grading techniques focus on evaluative work at the end of the semester, when such feedback is too late to make meaningful changes. This leads to a situation where students feel over-tested but are left under-assessed – at least when such assessment can be corrective. The standards-based grading scheme, on the other hand, tracks student progress throughout the semester – from when they first learn the material to after they have received meaningful feedback about their learning.

The rubric we will use aims to provide nuanced feedback after every assessment by using a rubric that considers evidence indicating a clear understanding of the concept. Simple mistakes that lead to incorrect answers still receive a “B” range grade, and solutions that show progress but don’t reach the correct answer still receive a grade in the “C” range. This emphasizes understanding by recognizing a student’s ability to show they know what needs to be done by better rewarding students that show they understand what resources they would need to solve a new problem and truly requiring mastery for a grade in the “A” range.

Students should not feel stressed to review all of their notes the night before a quiz, but are incentivized to *learn* the material over the semester. If students feel stressed, they should visit during office hours to discuss their studying techniques and how they might make changes toward deeply learning concepts and away from memorization. Almost without exception, students end up with far better scores at the end of the semester, as current performances are used as a diagnostic tool to show students which topics they might have more questions about. In other words, initial scores are *not permanent*, and students can change them by taking current performance as an *incentive to learn more*.

Since the scoring of standards is based on achievement, which may be new for some students, calculating this part of the grade may not seem obvious. At the end of the semester, the number of points students earn on standards will be based on the distribution of scores on their first and most recent attempts, with more weight likely given to the most recent score. After each quiz, I will post the solutions to Moodle, and students will receive nuanced feedback via email that lists their progress for each standard and an updated percentage score.

Plus or minus grades will be decided for solutions that fall between categories. 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 prerequisite or a small part of the concept being assessed.

– Conversion to Percentage –

A student’s final percentage score is based on their average of the first and second attempts. A percentage score is calculated for the first and second attempts as follows:

1. Add points for each standard for the first and most recent attempts.
2. Divide by the number of standards tested for the first and most recent attempts.
3. The standards grade percentage is calculated in the following ways, and the highest grade is taken:

$$\begin{aligned}\text{Standards Percentage} &= 0.25(\text{1st attempt percentage}) + 0.75(\text{most recent attempt percentage}) \\ \text{Standards Percentage} &= 0.50(\text{1st attempt percentage}) + 0.50(\text{most recent attempt percentage})\end{aligned}$$

By the nature of this retesting scheme, the weight on the first attempt is low, thus putting more emphasis on the most recent attempt. This allows students to use previous attempts as *learning experiences* and to incentivize them to *revisit* materials they need to after receiving nuanced feedback. Students who do well on the first attempts are further incentivized to retain and revisit the material, which is important as this course is *very cumulative*.

Although students may want to make their best attempt to count, the most recent score often *is* the best score, as initial attempts are taken as an invitation to revisit a topic more thoroughly. The most recent score encourages this long-lasting learning and discourages simple memorization of a particular topic for a quiz.

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

$$\text{Overall Grade} = 0.10(\text{HW}) + 0.60(\text{Standards}) + 0.30(\text{Final})$$

Letter	Final Grade
A	93-100%
A-	90-92.9%
B+	87-89.9%
B	84-86.9%
B-	80-82.9%
C+	77-79.9%
C	73-76.9%
C-	70-72.9%
D+	67-69.9%
D	63-66.9%
D-	60-62.9%
F	< 60%

- **A** range represents excellence with distinction. This demonstrates work that exceeds expectations, reflecting a combination of exceptional effort and outstanding results.
- **B** range signifies a student is consistently meeting or exceeding the expectations of the course. A “B” is a strong grade that shows solid understanding and a high level of performance.
- **C** range represents a grade that meets the essential requirements of the course. This is a respectable and adequate grade. To move beyond this level, a student needs to identify what's missing and create a plan to improve. I'm here to help with that.
- **D** range represents a passing but inadequate level of performance. This grade indicates that the student is not meeting many of the course's essential learning objectives. It is highly recommended that we meet to discuss a plan for improvement. I will notify your Administrative Dean through a course warning if your grade falls into this range.
- **F** range represents a failure to meet the basic requirements of the course. This grade indicates a significant lack of understanding and effort. We need to meet immediately to develop a new plan for approaching the coursework. I will notify your Administrative Dean through a course warning if your grade falls into this range.

Remark: I do not curve or round grades at the end of the semester. Regardless of the policy, some students may 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	
08/28/25	First Day of Classes (Half-Day Schedule)
Week 2	
09/01/25	Review
09/03/25	One Sample Inference
Week 3	
09/08/25	One Sample Inference
09/10/25	One Sample Inference
Week 4	
09/15/25	ANOVA and Tukey's HSD
09/17/25	ANOVA and Tukey's HSD
Week 5	
09/22/25	Chisquared and two-sample z -tests
09/24/25	Mood's Median and two-sample sign tests
Week 6	
09/29/25	Wrap up k -sample tests
10/01/25	Standards Period 1
Week 7	
10/06/25	Multiple Linear Regression: Model
10/08/25	Multiple Linear Regression: Assumptions and Outlier Analysis
Week 8	
10/13/25	Midterm Recess: No Class
10/15/25	Multiple Linear Regression: Model Fit and Cross Validation
Week 9	
10/20/25	Multiple Linear Regression: Untangling Interactions
10/22/25	Multiple Linear Regression: Untangling Interactions
Week 10	
10/27/25	Standards Period 2
10/29/25	Multiple Linear Regression: Model Selection and Comparisons
Week 11	
11/03/25	Robust Methods: Iteratively Reweighted Least Squares
11/05/25	Robust Methods: Quantile Regression
Week 12	
11/10/25	Poisson Regression
11/12/25	Poisson Regression
Week 13	
11/17/25	Flex Day
11/19/25	Standards Period 3
Thanksgiving Recess 11/22-11/30	
Week 14	
12/01/25	Negative Binomial Regression
12/03/25	Zero-Inflated Models Regression
Week 15:	
12/08/25	Logistic Regression
12/10/25	Logistic Regression
Final Exam 12/16 3p-5p	