MA 354: Data Analysis I – Fall 2021 – Living Syllabus
Section A: TR 2:45–4:00p in McGregory Hall 201

Professor
Will Cipolli — wcipolli@colgate.edu — www.cipolli.com — McGregory Hall 323
Office Hours: Tuesday 4:00p-5:00p (McGregory Hall 217) and By appointment.

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

Course Objectives:
At the completion of 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 the juice is worth the squeeze. If they struggle with a puzzle, they *can* solve it. We're working toward a holistic understanding of thought 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, which necessitates an environment where it is safe to take risks and connect. In-class discussion, and 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: 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 `Cran R` for this course. I will not assume that students have much, or any, previous experience with `Cran R` or any other programming experience. Students shouldn't expect to learn how to code as in an introductory computer science course. Instead, they will learn to use statistical software. That is, we will *use* available libraries and scripts to complete analyses, not to code all the algorithms from scratch. Coding isn't easy, learning it or teaching it, but this skill will become more and more critical over time. While teaching myself `Cran R` and even now, I heavily use documentation. 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 resources of the course, understanding the solutions, and altering them to answer new questions. If a student feels 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 to use. These technology issues can become a significant source of stress for students. Given these challenges, students can and 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 for what they missed. All students are responsible for all assignments that are due or assigned in the class they miss. I'd rather students not miss any class. 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 the 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 through furthering the discussion; I will monitor activity and chime in often. Through this design, I intend to foster students' creativity and curiosity and prepare them to think critically, ask questions and gain lifelong value out of 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 are 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, and sometimes difficult, experiences with learning. I acknowledge the persistence of discrimination and exclusion in mathematics based on race, gender, socio-economic status, and other factors. I take responsibility for the work of 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 reach out to me to discuss their learning process, experience, or needs and point out any blind spots.

Specific Learning Accommodations and Support: I hope that 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 will also need to contact the Office of Academic Support and Disability Services to receive help determining and coordinating a specific accommodation based on disability/medical documentation. Contact info as follows: 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 see me as soon as possible to review their options to avoid any academic honesty issues. See Colgate's Academic Honor Code.

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 not only get you past an immobilizing issue in understanding but help me understand where students are in their learning process. There are many times I have a discussion during office hours 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, make plans to see me right away (even if it has to be outside of 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 on your own. Try practice exercises, run the sample code on your own, 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 find yourself '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, and 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?" Answers to these questions will be much more informative and more likely to help us get closer to meeting 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 six assignments. These six homework assignments will consist of weekly diagnostic completion, foundational questions, and data analyses.

- **Completing Homework:** I encourage students to discuss the homework opportunities with each other and me. Students can do this in various ways – on the Moodle discussion board or Hypothes.is, during office hours, or active homework sessions. Students should judiciously review posted solutions to homework in preparation for exams.
- **Submitting Homework:** All assignments should be typeset in LaTeX and submitted in .pdf format via email by 5:00p the day it is due.
- **Weekly Diagnostic Check-points:** Weekly, I will ask students to reflect on what we've discussed by seeing what they thought was essential and prompt them to submit any unanswered questions. Each diagnostic will also have some questions aimed to help guide homework completion. Diagnostics will open Thursday after class and close at 5:00p on Friday.
- **Late Homework:** I will not accept late homework. I will, however, happily discuss problematic deadlines before the assignment is due to allow for the time needed to complete the thoughtful application of the course content that the homework guides students through.

Midterm Part I (30%): The midterm will involve foundational questions, and the analysis of real data using methods learned up to that point. The expected date for the midterm can be found in the schedule below in week 9, though it may change as our course progresses. Below, I delineate the graded pieces of Part I.

- **Foundational Questions (30%):** There will be several questions on various course topics to inventory the concepts students understand well and which topics need revision.
- **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 period 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.
- **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, 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.
- **Response to Peer Review (15%):** Students should provide a document that addresses the concerns and questions that come up in the peer reviews. 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 it is important to respond to *all* feedback.

Final Exam (30%) A comprehensive final exam will be posted on the last day of class and is due **TBA** (according to the University's exam schedule). Students can 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. There is no peer review or revision opportunity for the final exam, 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 (Proficient)	• Contains no non-trivial errors and clearly communicates understanding • 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)	• Meets expectations and contains an easily correctable mistake • Makes correct decision(s) toward solution • Justifies decision(s) toward solution • Effectively communicates solution and support • A slight error, confused reasoning, or notation mistake • Refinement is needed	0.85
C (Progressing)	• Contains correct work and a serious error in understanding or communication • 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 • Revision is needed	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, and/or notation mistake • Revision is needed	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, and/or notation mistake • Revision is needed	0.25
Z (Not assessable)	• No Response, there is no reasonable attempt to provide the correct solution. • Not assessable	0.00

Solutions between categorizations can earn plus or minus grades. For example, a solution with a trivial error that isn't critical to understanding or the correct completion would earn an A-. Alternatively, an easily correctable issue that may show a small misunderstanding would earn a B+, for example.

Overall Grade: A student's overall grade will be a weighted average of their scores on homework, the midterm, and final exam. The overall grade for the course is:

$$\text{Overall Grade} = 0.30(\text{HW}) + 0.30(\text{Midterm Part I}) + 0.10(\text{Midterm Part II}) + 0.30(\text{Final})$$

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 challenging 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 is 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. Students who want to improve from adequate must recognize what more is needed, make a plan, 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 that a student truly does not understand the expectations of the course. Students, in this case, should make an appointment with me to discuss how they might make a plan and take action to achieve at the level they desire to. I will submit course warnings to the appropriate Administrative Dean for students earning a D in this course.
- **F** range represents an apparent failure to meet the expectations of the class. An F represents a lack of effort and interest in the course. This outcome is a cause for deep concern. I will submit course warnings to the appropriate Administrative Dean for students earning an F in this course.

Remark: I do not curve or round grades at the end of the semester. No matter the policy, 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. 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 grade and work with me to *earn* the grade that they want.

A Pedagogical Note:

As a first-generation student, I've realized that I have a point of view about education that differs from many students. While many things affect performance and learning, our resolve to do well is paramount to our success. My goal is to work with all students so that they are *earning* the grade that they want while recognizing that success is hard work, not innate talent. I expect students to grow each semester, and sometimes there are growing pains; I do not expect anyone to get things perfectly the first run through any task; this is why we will employ retesting.

A university-level course should not be easy, no matter the level of the course. I think the best courses seem challenging but provide students with all the resources to succeed. Bjork (1994) coined this type of course as having "desirable difficulty," where the learning tasks require a "desirable" amount of effort, which improves long-term performance. This class will be "desirably" challenging. Students should ask questions, read more sources, really dive into what they're learning. By doing so, students will become more effective thinkers and communicators through this process as they will retain the skills learned this semester and have practice persistence in figuring out the unknown.

In primary and secondary school, achievement and external recognition drove my pursuit of education. During this time, I thought learning was successful imitation. I obtained a practiced skill in memory, but I could not develop my own solutions through critical thinking. Fortunately, my first year of courses at Quinnipiac University, a liberal arts university in Connecticut, deepened my perspective about education. I took challenging courses in both the sciences and humanities that stressed revision and improvement. Due to this focus on intellectual growth, I iteratively became a better reader, writer, problem-solver, and communicator.

When considering how grading systems incentivize learning, I was not satisfied with how the usual system works. To address this, I developed a focused idea of what I want grades in this course to mean (see above). I hope to help students *experience the importance of quality* and attain a higher degree of self-sufficiency by giving clearer, meaningful, and actionable feedback about their academic achievement. In this light, I base student achievement on regular assessment *and* their revision, as opposed to just points earned.
To facilitate this type of learning and feedback, we will use standards-based evaluation and retesting. This approach involves receiving and addressing feedback on well-defined course objectives. Providing feedback about how well a student has met each objective of the course allows for a more nuanced conversation about where students are strong and where they can make plans to improve their knowledge by using retesting as a diagnostic tool – a blueprint for future success.

This course is designed for students of vastly different mathematics experiences to do well – students that need more time to master the material are granted such time through retesting. Students have the opportunity to assess their learning and take it as an invitation to try again – read more sources, review exercises (even ones that aren't assigned), or ask for guidance. Revisiting previous work is a technique for learning how to learn, becoming a self-starter, and conquering a challenging course that rewards long-term performance and discourages temporary memorization.

In this light, students will find that I will incessantly ask for feedback throughout the semester. The reason I survey students and rigorously review their comments about what could be changed models the behavior I expect from them. Having an accurate snapshot of what I am doing well and what I can work on helps me serve students effectively and helps me become a better professor. I take all feedback seriously and very often take constructive criticism as an invitation to make changes for the better. Similarly, retesting is a tool for learning how to learn and conquering a challenging course that rewards long-term performance and discourages temporary memorization.

Please ask for what you need. I want all of you to succeed this semester. Let's make the juice worth the squeeze.

Schedule:

Date	Event
08/26/21	First Day of Classes (Half-Day Schedule)
08/27/21	Introduction to R RSweave and Knitr (All Tuesday Classes Meet)
08/31/21	Objects in R
09/02/21	Summarizing Data
09/07/21	Graphical Displays
09/09/21	Probability Distributions
09/14/21	Probability Distributions
09/16/21	Point Estimation
09/21/21	Sampling Distributions
09/23/21	Sampling Distributions
09/28/21	One Sample Inference
09/30/21	One Sample Inference
10/05/21	Two Sample Inference
10/07/21	Two Sample Inference
Mid-Term Recess: 10/09/21 - 10/12/21	
10/14/21	ANOVA and Mood's Median Test
10/19/21	ANOVA and Mood's Median Test
10/20/21	Full-Term S/U Grade Option and Course Withdrawal (with a W) Deadlines
10/21/21	Linear Regression
10/26/21	Linear Regression
10/27/21	**Midterm Exam***
10/28/21	Linear Regression
11/02/21	Robust Regression
11/04/21	Multiple Linear Regression
11/09/21	Multiple Linear Regression
11/11/21	Flex Day
11/16/21	Multiple Linear Regression
11/18/21	Multiple Linear Regression
Thanksgiving Recess: 11/20/21 - 11/28/21	
11/30/21	Logistic Regression
12/02/21	Logistic Regression
12/07/21	Count Regression
12/09/21	Count Regression
TBA	**Final Exam (TBA)**

* I will provide more details about the midterm exam as we approach it.

Remark: Dates will likely change as I largely let the class dictate the speed of the course through asking questions and completing extra problems in class.