MA 354: Data Analysis I – Fall 2019 – 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: Tue 4:00-5:00p in McGregory 201, Wed 1:30-2:30p in McGregory 323, and by appointment.

Purpose: This course aims to connect the beauty of the mathematical foundations of statistics to the power of their application. Students will be introduced to and use statistical software (`Cran R`) to fit generalized linear models that help assess evidence for or against theories about the world. At the end of the course students will be fluent in the language of data analysis: students will understand fundamental and essential data and how to model them, understand what makes them correct choices by assessing assumptions, and interpret these models in the context of research inquiries across disciplines. The comprehensive nature of this course allows students to experience the benefit of balancing mathematics and language.

Handbook Description: Data Analysis I is an applied regression course that involves modeling data with normal models including hands-on Tukey-style data analysis with statistical software. Students explore topics that are widely used today across disciplines in academic research and in business; such topics include inferences for normal parameters, correlation, regression, analysis of variance (ANOVA), model diagnostics, model building, and transformations. Students will start with regression analysis with a single predictor variable, then consider regression analysis where two or more variables are used for making predictions. While applied, this course aims to combine theory and application to emphasize the need for understanding each method's theoretical foundation. This conversation is had through illustrating a variety of inferences, residual analyses and fully exploring the implications of our assumptions.

Course Objectives:
At the completion of this course, students will be able to:

1. Understand sampling distributions and their impact on inference
2. Choose the correct association measurements for data from a research inquiry
3. Select the correct regression methodology for data from a research inquiry
4. Assess regression model fit by evaluating assumptions of the model
5. Ameliorate any transgression of assumptions found when modeling
6. Interpret regression models to answer research questions
7. 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 would like 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 requires us to signal our need for help to each other which necessitates an environment where it is safe to take risks and connect. In-class discussion and curious risk-taking will be celebrated as much as the correct answer. An incorrect response provides the best place to grow. This is how we will *learn* to actively engage and discuss mathematics.

Rules of Engagement:

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

Prerequisite: ECON 375 or BIOL 320 or PSYCH 309 or MATH 260 or COSC 290 or permission.

Technology: It is not assumed that students have much, or any, previous experience with statistical software or programming. This course will introduce students to `Cran R`, which is an extremely useful statistical programming language that has become widely used in recent years. Students shouldn't expect to learn how to code as in an introductory computer science course, but learn to be a consumer of statistical software; that is, to use libraries to complete complicated analysis, not to code all the algorithms from scratch. Coding isn't easy, learning it or teaching it, but I believe that this skill will become more and more important over time. While teaching myself `Cran R`, whenever I was stuck with a particular problem, I searched for a solution in the documentation and tried 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.

Support for Technology: Devices, like laptops, are paramount to success in college. I recognize that these digital devices can be expensive and that students might not have the same access to the latest technology; digital technology changes rapidly and students might rely on older, more problem-prone devices that breakdown 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. This will enable me to assist students in accessing the appropriate resources on campus.

Attendance: Students are expected to attend all classes and to arrive on time. That being said, when a student misses class due to illness, hangovers, interviews, personal crises, deaths in the family (I hope not!), and whatever else, they don't need to let me know but 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 that they will need for assignments and exams later on. In other words, every missed class is a disadvantage, even aside from grade penalties; the obligation is on the student to minimize that effect. That being 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 course as often as possible.

Outside Class Discussion: Students should use the discussion board in Moodle as a safe place to ask questions and be curious about the course material. I expect students to answer such questions and to feed the curiosity of their peers through furthering the discussion; I will monitor the discussion board and chime in from time to time. This is intended to foster students' creativity and curiosity, and prepare them to think critically, ask questions and gain a lifelong value out of their education.

Make-up Policy: Make-ups and extensions will be considered 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 issues which are not acknowledged by the university. All make-up exams must be scheduled to be taken before or on the same day as the exam when possible.

Inclusion: It is my goal and responsibility 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. It is really important to me that we work to make the classroom environment as comfortable and respectful as possible. As a class, we will resolve to listen, learn and act in order to make this classroom proactively welcoming to all students. I encourage all students to see me if they want to discuss their learning process, experience or needs and to point out any blind spots.

Specific Learning Accommodations and Support: I hope that students will feel comfortable in 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 on the basis of disability/medical documentation. Contact info as follows: Lynn Waldman: lwaldman@colgate.edu, (315) 228-7375.

Academic Honesty: Students are expected 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.

Support: College life can sometimes get bumpy; if you are experiencing emotional or personal difficulties, seek help right away. The counseling center offers completely confidential and highly professional services, and can be contacted at 315-859-4340. If this seems like a difficult step, come 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 in 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.

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 lecture.

3. For every hour in class, you are expected to spend 2-4 hours outside of class on 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 make sure you understand it and don't have major questions. Then break down the assignment into manageable pieces and work on them over the course of 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 welcomed, but not well-informed and will almost certainly not get you the answer you are looking for. 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. If you start to struggle, make plans to see me right away (even if it has to be outside of office hours).

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

Grading:

Homework (25%): The purpose of homework is to practice concepts introduced in lecture. Students can expect six assignments. These assignments will be posted throughout the semester and students will submit their answers in class on specified due dates, including weekly diagnostic check-points.

- **Completing Homework:** I encourage students to discuss the homework exercises with each other and me. This can be done in a variety of ways – on the Moodle discussion board, in office hours, etc.
- **Submitting Homework:** All assignments should be typeset in LaTeX and submitted in pdf format via email by the communicated due date.
- **Weekly Diagnostic Check-points:** Weekly, I will ask students to reflect on what we've discussed by submitting what they thought was important and any unanswered questions. Each diagnostic will also have some questions aimed to help guide homework completion. Diagnostics will open Thursday after class and must be submitted by noon on Monday.
- **Late Homework:** I will not accept late homework. I will, however, happily discuss problematic deadlines before homework is due to allow for the time needed to complete the thoughtful application of the course content that the homework guides students through.

Exams (25% each): Exams will involve foundational questions, as well as the analysis of real data using methods, learned up to that point. The expected dates for the exam periods can be found in the schedule below, though they may change as our course progresses. **Oct. 8, Nov. 19**. Below I delineate the graded pieces of each exam.

During Exam Period

- **Foundational Questions (20%):** There will be several questions on a wide variety of topics. This is meant to serve students by providing clear, specific, and actionable feedback on concepts they understand well and what may need to be revisited.
- **Data Analysis (50%):** A research inquiry and a dataset will be provided and students will write a preliminary consultation that provides an initial full data analysis and interpretation, which helps answer the research inquiry. This should be considered a rough-draft analysis.

After Exam Period

- **Peer Review (5%):** All students will receive another student's exam, with all identifying information removed from the submission. Students will have to provide *constructive* criticism about the solutions that will lead to a better second draft.
- **Revision (20%):** Students, after receiving their peer-reviewed papers, will submit a revised copy of their exam submission. This allows for the opportunity to *improve* work after receiving feedback. The aim is to more accurately measure learning and model the process more closely by allowing students to be evaluated again after receiving feedback and revisiting the material to show their best effort. Students should take an extensive peer review as an invitation to try again – revisit the notes, past solutions, ask for help so that the revision is a success and take a light peer review as a chance to hone their knowledge on the subject by providing a highly formal solution in the revision. This should be considered a final-draft analysis.
- **Response to Peer Review (5%):** Students should provide a document that addresses the concerns and questions that come up in the peer reviews they received and how they were or why they weren't addressed in the final draft. Students should also provide information about any self-motivated changes to their analysis.

Final Exam (25%) A comprehensive final exam will be given on the last day of class and is due **5pm Tuesday, December 17**, submitted via email. Students can use the internet and all course materials as they complete the final exam, but should not receive any outside help from other sources; e.g., students, tutors, faculty, freelancers, etc. There is no peer review or revision opportunity for the final exam, students should self-review before submission.

Scoring

Rubric: Each question asked on homework or during an exam period is scored on the following rubric:

- **A**: Mastery, the solution contains no non-trivial errors and clearly communicates understanding.
- **B**: Sufficient, the solution meets expectations and contains an easily correctable mistake like notation.
- **C**: Progressing, the solution contains correct work and a serious error in understanding or communication of the concept. Revision is needed.
- **D**: Developing, the solution does not contain the correct answer or doesn't fully answer the question and does show some reasoning in the correct direction. Revision is needed.
- **F**: Needs attention, the solution does not contain work in the correct direction and a reasonable attempt was made
- **Z**: No Response, there is no reasonable attempt to provide the correct solution. Not assessable.

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.

Assignment Grade: While each question receives a letter grade and each student will receive a "report card" on each question/concept, each assignment will receive a percentage grade as calculated from the GPA of the assignment's "report card."

Overall Grade: A student's overall grade will be a weighted average of their **percentage** scores on homework, exams, and the final exam. The overall grade that will be assigned to each student will align with the standard scale.

$$\text{Overall Grade} = 0.25(\text{HW}) + 0.25(\text{Exam I}) + 0.25(\text{Exam II}) + 0.25(\text{Final})$$

- **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. Good is more common than excellent and should be celebrated as a success.
- **C** range signifies adequate but there are clear areas for growth in performing or communicating several basic functions of the course. Adequate 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 the basic functions of the course. I do recognize that a D may also mean that a student truly does not understand what is expected. Students, in either 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. Course warnings are submitted to the appropriate Administrative Dean for students earning a D in this course at any point during the semester.
- **F** range represents a clear 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; course warnings are submitted to the appropriate Administrative Dean for students earning an F in this course at any point during the semester.

A Pedagogical Note:

As a first-generation student, I've come to realize that I have a point of view about education that differs from many students, where I am perceived as a bit of an individualist. While many things affect performance and learning, our resolve to do well is most 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 revisit our exams.

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.

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 revisiting exams. The course is designed for students to 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.

To facilitate revisiting previous work we will revise our exams which involves receiving, providing and addressing feedback on well-defined course objectives. Previous research (Hattie & Timperley 2007; Hattie 2009) suggests that a better understanding of essential principles is needed for grades to accurately reflect students' achievement. Providing feedback about how well a student has met the objectives of each assessment allows for a more nuanced conversation about where they are strong and where they can make plans to improve their knowledge by using retesting as a diagnostic tool.

When I was considering the qualitative difference between grades of 89.9 and a 90.1, I couldn't find a satisfying solution, so I developed a focused idea of what I want grades in this course to mean (see above). By basing students' achievement on regular assessments and their revisions, as opposed to just points earned in a timed setting in the classroom, 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 (O'Connor, 2007; Tomlinson & McTighe, 2006).

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, in terms of serving students, helps me become a better professor for current and future students. I take all feedback seriously and very often take constructive criticism as an invitation to make changes for the better.

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
08/29/19	First Day of Classes (Half-Day Schedule) 2–2:20pm
08/30/19	Syllabus and Association
09/03/19	Association
09/05/19	Association
09/06/19	NASC Colloquium – Ana Jimenez
09/10/19	Linear Models – Model Specification and Assumptions
09/12/19	Linear Models – Case Study
09/17/19	Linear Models – Model Assessment
09/19/19	Linear Models – Model Assessment
09/20/19	NASC Colloquium – Josh Tebbs
09/24/19	Linear Models – Model Selection
09/26/19	Linear Models – Model Selection
10/01/19	Linear Models – Model Selection
10/03/19	Linear Models – Inference
10/08/19	Linear Models – Inference
10/08/19	**Exam I**
10/10/19	Linear Models – Inference
10/12/19	**Exam I Peer Reviews Due**
10/12/19-10/15/19	Mid-Term Recess (Uninterrupted HW Time)
10/17/19	Linear Models – Remedial Measures
10/22/19	Linear Models – Remedial Measures
10/22/19	**Exam I Revisions Due**
10/24/19	Linear Models – Remedial Measures
10/29/19	Linear Models – Remedial Measures
10/30/19	Full-Term Courses: Withdrawal Deadline (with a grade of W)
10/31/19	Flex Day – Catch Up or Snow Day
11/05/19	GLM – Logistic Regression
11/07/19	GLM – Logistic Regression
11/12/19	GLM – Ordinal Logistic Regression
11/14/19	GLM – Multinomial Logistic Regression
11/15/19	NASC Colloquium – Robert Smith?
11/19/19	GLM – Poisson Regression
11/19/19	**Exam II**
11/21/19	GLM – Poisson Regression
11/23/19	**Exam II Peer Reviews Due**
11/23/19-12/01/18	Thanksgiving Recess
12/03/19	GLM – Negative Binomial Regression
12/05/19	GLM – Negative Binomial Regression
12/06/19	**Exam II Revisions Due**
12/10/19	GLM – Zero-Inflated Models
12/12/19	GLM – Zero-Inflated Models
12/13/19	Last Day of Classes
12/17/19	**Take Home Final Exam Due 5pm**

* The exams will be held in McGregory 201 from 5p–8p.
Note: Dates will likely change as I largely let the class dictate the speed of the course through asking questions and completing extra exercises in class. This is a *very* optimistic schedule that includes extra topics I don't necessarily expect to get to.