MATH 454: STATISTICAL LEARNING (DRAFT)

TLDR SYLLABUS OVERVIEW

Meeting Times (p. 1)

- Lecture Monday/Wednesday 01:20p-02:35p
- Office Hours: Monday 11:30a – 1p; Thursday 11:30a – 1p

What you will learn (p. 1, 6)

1. fit typical regression models such as the generalized linear models, polynomial regression, and generalized additive model in `R`.
2. Repeated measures analyses for generalized linear models using `lmer`.
3. Supervised learning classification algorithms such as logistic regression, k-nearest neighbors, discriminant analysis, random forests, artificial neural networks, support vector machine, etc.
4. Supervised learning prediction algorithms such as k-nearest neighbors regression, random forests, artificial neural networks, support vector machines, etc.
5. Unsupervised learning – clustering and latent profile analyses.

Materials and Technology (p. 1)

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

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. 6)

- Class time will consist of a mix of lecture and activities
- Weekly homework and biweekly interviews
- A Project
- A comprehensive final exam

Grades (p. 4-5)

$$\text{Course Total} = 0.40(\text{HW}) + 0.20(\text{Project}) + 0.40(\text{Final}).$$

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 454: Statistical Learning — Spring 2026
Lecture: Monday, Wednesday 01:20p to 02:35p in McGregory 226

Professor: Will Cipolli — wcipolli@colgate.edu — www.cipolli.com — McGregory Hall 323
Office Hours: Monday 11:30a – 1p; Thursday 11:30a – 1p.

Purpose: To give students a diverse toolkit for modeling categorical and continuous outcomes using state-of-the-art methods. Students will learn how to use, assess, and interpret foundational models of statistical learning.

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

1. fit typical regression models such as the generalized linear models, polynomial regression, and generalized additive model in `R`.
2. Repeated measures analyses for generalized linear models using `lmer`.
3. supervised learning classification algorithms such as logistic regression, k-nearest neighbors, discriminant analysis, random forests, artificial neural networks, support vector machine, etc.
4. supervised learning prediction algorithms such as k-nearest neighbors regression, random forests, artificial neural networks, support vector machines, etc.
5. unsupervised learning – clustering and latent profile analyses.

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), and Data Analysis (MATH 354).

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 and MATH 354 prerequisites 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 technological 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 check 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 inflating or deflating grades with any percentage of a student's score based on 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 help determine and coordinate 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.

Large Language Models: 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., ChatGPT, Gemini, Grok, etc.). However, online searches and large language models may return incorrect or out-of-scope content. 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 highly confidential, 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 approaches to completing 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 assignments.

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 assignments.

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 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 assignments 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 (40%): 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 or Late Work: Homework is due at 5:00 PM. To account for unexpected technical issues, there is a no-penalty grace period until 12:00 PM the following day. For delays beyond this, please email me as soon as possible with: (a) an explanation, (b) your current progress draft, and (c) a proposed new deadline within 72 hours of the original.

Interviews: Students will have a 10-minute standing appointment bi-weekly to defend their work. I will ask you to explain your logic and interpret your `R` output to ensure conceptual understanding.

- Great/Good: Grade confirmed.
- Developing: Partial resubmission required.
- Poor/No Validation: Failure to explain your own work results in a one-letter grade deduction.

Project (20%): This will require a from-scratch explanation of an advanced method, including 'hand coding' of the method, confirmation that it matches `R` output, simulations for assessing any conditions for use, and a reproduction of a published article that uses the method.

Final Exam (40%) The final exam will be a comprehensive exam: **TBA**.

Rubric:

Each question asked this semester will be 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, and/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, and/or notation mistakes	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 mistakes	0.25
Z (Not assessable)	No Response: There has been no reasonable attempt to provide the correct solution.	0.00

Overall Grade: A student's overall grade will be a weighted average of their scores on homework, project(s), and final exam. The overall grade earned by each student will be determined as follows.

$$\text{Overall Grade} = 0.40(\text{HW}) + 0.20(\text{Project}) + 0.40(\text{Final})$$

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 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.

A Note on Grading Code:

Learning extends beyond the classroom, but submitted work must demonstrate understanding. While I expect students to use the resources available to them to complete the classwork, I also 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."

A Note on Curving:

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.

Schedule:

Week 1	
01/19/26	First Day of Classes (Half-Day Schedule)
01/21/26	Review of MATH 354
Week 2	
01/26/26	Poisson Regression
01/28/26	Poisson Regression
Week 3	
02/02/26	Negative Binomial Regression
02/04/26	Zero-Inflated Models
Week 4	
02/09/26	Logistic Regression
02/11/26	Logistic Regression
Week 5	
02/16/26	Multinomial Regression
02/18/26	Multinomial Regression
Week 6	
02/23/26	Generalized Linear Mixed-Effects Models
02/25/26	No class
Week 7	
03/02/26	Generalized Linear Mixed-Effects Models
03/04/26	Generalized Linear Mixed-Effects Models
Mid-Term Recess 3/9–3/17	
Week 8	
03/09/26	k Nearest Neighbors
03/11/26	k Nearest Neighbors
Week 9	
03/23/26	Linear Discriminant Analysis
03/25/26	Quadratic Discriminant Analysis
Week 10	
03/30/26	Support Vector Machines
04/01/26	Support Vector Machines
Week 11	
04/06/26	Random Forests
04/08/26	Random Forests
Week 12	
04/13/26	Artificial Neural Networks
04/15/26	Artificial Neural Networks
Week 13	
04/20/26	Generalized Additive Models
04/22/26	Generalized Additive Models
Week 14:	
04/27/26	Approaches for Clustering
04/29/26	Approaches for Clustering / LPA
Final Exam Thursday, May 7th 3-5p	

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