MATH 240: COMPUTATIONAL STATISTICS

TLDR SYLLABUS OVERVIEW

Meeting Times (p. 1)

- Lecture Monday/Wednesday 01:20p-02:35p
- Lab A: Tuesday 01:20p-04:00p; Lab B: Thursday 01:20p-04:00p
- Office Hours: 10:00a-11:00a Mon.; 10:00a-11:00a Thurs. (McGregory 225 w/ coffee)

What you will learn (p. 1)

1. complete basic programming tasks in `R`
2. clean and manage data using `dplyr` and `tidyverse`
3. choose, create, and interpret appropriate graphical displays and numerical summaries of real data using `ggplot2` and `tidyverse`
4. understand probability distributions and compute probabilities of interest
5. perform point estimation computationally using multiple approaches
6. complete simulation studies to assess statistical procedures
7. demonstrate an understanding of, and ability to compute and interpret, key values of statistical inference in a variety of settings

Materials and Technology (p. 2)

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

Policies (p. 2-3)

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

Activities (p. 4-5)

- Class time will consist of a mix of lecture and activities
- Weekly homework and labs (primarily due Mondays at 5p)
- Standards (three exams periods)
- Final

Grades (p. 6-7)

$$\text{Course Total} = 0.25(\text{HW}) + 0.50(\text{Standards}) + 0.25(\text{Final}).$$

Contact

- All questions on course contents or technology should be posted on the Moodle or Hypothes.is.
- Other inquiries can be handled in office hours or email to wcipolli@colgate.edu.

MA 240: Computational Statistics — Spring 2024
Lecture: Monday, Wednesday 01:20p to 02:35p in McGregory 226
Lab A: Tuesday 1:20p to 04:00p in McGregory 210 | **Lab B:** Thursday 1:20p to 04:00p in McGregory 210

Professor
Will Cipolli — wcipolli@colgate.edu — www.cipolli.com — McGregory Hall 323
Office Hours: 10:00a-11:00a Mon.; 10:00a-11:00a Thurs. (McGregory 225 w/ coffee); and by appointment.

Purpose: To give students a computational introduction to probability and statistics. Students will learn how statisticians answer questions. First, they will learn the statistical programming language `R`. Then, students will use `R` to explore topics in probability, mathematical statistics, and data analysis.

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

1. complete basic programming tasks in `R`
2. clean and manage data using `dplyr` and `tidyverse`
3. choose, create, and interpret appropriate graphical displays and numerical summaries of real data using `ggplot2` and `tidyverse`
4. understand probability distributions and compute probabilities of interest
5. perform point estimation computationally using multiple approaches
6. complete simulation studies to assess statistical procedures
7. demonstrate an understanding of, and ability to compute and interpret, key values of statistical inference in a variety of settings

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

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

Rules of Engagement:

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

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

Prerequisites: None.

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

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

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

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

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

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

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

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

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

How to Succeed in this Class

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

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

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

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

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

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

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

Course Grade:

Homework (25%): The purpose of homework is to practice concepts introduced in the lecture. Students can expect 10-14 homework assignments. These assignments will consist of two pieces:

- **Homework:** I will grade these assignments for correctness and will not accept late submissions unless previously agreed upon. I encourage students to discuss the homework opportunities with each other and me. This can be done in various ways – on the Moodle discussion board or during office hours. Students should judiciously review posted solutions to homework in preparation for exams.
- **Weekly Diagnostic Check-points:** I will ask students to reflect on what we've discussed by seeing what they thought was important and guide them to engage with the course material so they can ask any lingering unanswered questions.

Standards (50%): We will cover roughly 13 *standards*, which include statistical concepts such as finding the mean of a dataset. Students can expect two cumulative exam periods where they will be evaluated on any of the 13 standards recently covered and, perhaps, be re-assessed on previous standards. These exams 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 what they are doing well and what may need to be revisited.
- **How?:** Exams 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 successes as a chance to hone that knowledge on the subject 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, modeling the process more closely by 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?:** The expected dates for the exam periods can be found in the schedule below.
  - **Exam One** During lab in Week 6
  - **Exam Two** During lab in Week 10
  - **Exam Three** During lab in Week 12
  - **Final Exam:** A portion of the final exam will cover any remaining retesting. **TBA**.

Final Exam (25%) The other portion of the final exam will be a comprehensive exam: **TBA**.

Lab Grade:

Labs (100%) Each lab will require a written submission outlining the solution to the lab exercise and what the results mean. These assignments will be graded on correctness, clarity, conciseness, and polish.

Standards Scoring

– *List of Standards:*

ID	Standard/Objective	Sections	First Attempt	Most Recent Attempt
1	Check-In I		Exam 1	Final
2	Objects, Loops, Functions, and Conditionals		Exam 1	Exam 2
3	Cleaning Data		Exam 1	Exam 2
4	Summarizing Data I (Categorical Data)		Exam 1	Exam 2
5	Summarizing Data II (Quantitative Data)		Exam 1	Exam 2
6	Check-In II		Exam 2	Exam 3
7	Probability Distributions I (Discrete)		Exam 2	Exam 3
8	Probability Distributions II (Continuous)		Exam 2	Exam 3
9	Point Estimation		Exam 2	Exam 3
10	Resampling and Central Limit Theorem		Exam 2	Exam 3
11	Hypothesis Testing		Exam 3	Final
12	Confidence Intervals		Exam 3	Final
13	Statistical Inference via Bootstrapping		Exam 3	Final

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

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

The standards-based grading scheme allows for meaningful feedback about learning and scores for students that fully address understanding. 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 rubric above aims to provide nuanced feedback after every assessment by using a rubric that considers evidence indicating an understanding of a concept. Simple mistakes that lead to incorrect answers still receive a "B" range grade, and solutions that show progress but don't get to 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 the exam 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 learning 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 what 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*.

– *Grading* – Since the scoring of standards is in terms of achievement, which might 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, likely with more weight on the most recent score. After each exam period, I will post solutions to Moodle, and students will receive nuanced feedback in an email that lists their progress for each standard and an updated percentage score.

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.

– *Conversion to Percentage* – A student's final percentage score considers their average 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 the following ways, and the highest grade is taken:

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

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*.

Though students may want 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 is used to encourage this long-lasting learning and discourage simple memorization of a particular topic for an exam.

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.

$$\text{Overall Grade} = 0.25(\text{HW}) + 0.5(\text{Standards}) + 0.25(\text{Final})$$

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

- **A** range represents above and beyond expectations, excellence with distinction. These are not impossible to achieve but are difficult to come by. While there is merit to hard work and long hours, it does not always guarantee success. Excellence refers to the combined results, not just the effort.
- **B** range signifies that a student is meeting the expectations of the course in most or all aspects. Good is more common than excellent and should be celebrated as a success.
- **C** range signifies adequate and at the level of expectation for several aspects of the course. Average is not usually an appealing categorization for those who strive for extraordinary. A grade of C, however, is a respectable point. If students don't want to be categorized as adequate, they must recognize what more is needed, make a plan to achieve that, and execute it; I can help with a plan!
- **D** range represents less than adequately equipped to perform many of the essential functions of the course; just passable. I recognize that a D may also mean a student does not understand what is expected. Students, in this case, should make an appointment with me to discuss how they might make a plan and take action. I will submit course warnings to the appropriate Administrative Dean for students earning a D in this course at any point during the semester.
- **F** range represents an apparent failure to meet the expectations of the class. F represents a lack of effort and interest in the course. This is a cause for deep concern; I will submit course warnings to the appropriate Administrative Dean for students earning an F in this course at any point during the semester.

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

Schedule:

Week 1	
01/22/24	First Day of Classes (Half-Day Schedule)
01/24/24	Introduction to Coding in `R`
Lab	Installing and Using RStudio
Week 2	
01/29/24	Objects, Loops, Functions, and Conditionals
01/31/24	Objects, Loops, Functions, and Conditionals
Lab	Reverse Engineering a Dartbot
Week 3	
02/05/24	Introduction to the `tidyverse`: Storing and Cleaning Data
02/07/24	Introduction to `ggplot2`
Lab	Creating a Cricket Dartbot
Week 4	
02/12/24	Summarizing Categorical Data
02/14/24	Summarizing Quantitative Data
Lab	Compiling Datasets from Various Sources
Week 5	
02/19/24	Summarizing Datasets
02/21/24	Probability Distributions
Lab	Telling a Data Story
Week 6	
02/26/24	Discrete Probability Distributions
02/28/24	Continuous Probability Distributions
Lab	Exam Workshop
Week 7	
03/04/24	Univariate Point Estimation
03/06/24	Multivariate Point Estimation
Lab	Describing Probability Distributions
Mid-Term Recess 3/9–3/17	
Week 8	
03/18/24	Flex Day
03/20/24	Resampling and Sampling Distributions
Lab	Point Estimators for Parameters
Week 9	
03/25/24	Central Limit Theorem
03/27/24	Central Limit Theorem
Lab	Resampling distributions of Point Estimators and Central Limit Theorem
Week 10	
04/01/24	One Sample Inference for the Mean
04/03/24	One Sample Inference for the Mean
Lab	Exam Workshop
Week 11	
04/08/24	One Sample Inference for the Proportion
04/10/24	One Sample Inference for the Proportion
Lab	Hypothesis Testing for the population mean
Week 12	
04/15/24	One Sample Inference for the Median
04/17/24	One Sample Inference for the Median
Lab	Exam Workshop
Week 13	
04/22/24	k Sample Inference for Means
04/24/24	k Sample Inference for Proportions and Medians
Lab	Simulation for Power, Type I, and Type II Errors
Week 14:	
04/29/24	Association
05/01/24	Regression
Lab	Modeling Linear and Monotone Relationships
Final Exam 05/09/24 3-5p	

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