Syllabus

Course description

This course introduces students to the discipline of statistics as a science of understanding and analyzing data. Themes include data collection, exploratory analysis, inference, and modeling. Focus on principles underlying quantitative research in social sciences, humanities, and public policy. Research projects teach the process of scientific discovery and synthesis and critical evaluation of research and statistical arguments. Readings give perspective on why in 1950, S. Wilks said, "Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write."

In this course, students learn how to effectively make use of data in the face of uncertainty: how to collect data, how to analyze data, and how to use data to make inferences and conclusions about real world phenomena. Critiquing data-based claims and evaluating data-based decisions is at the core of this course. Throughout the course students acquire a conceptual understanding and mastery of statistical and quantitative reasoning tools in order to be able to make such critiques and evaluations.

In addition, students are presented with novel data sets and application examples on a daily basis, and they use these data to model outcomes and make inferences about unknown population characteristics. Students learn that the first step of any analysis is identifying the assumptions and conditions necessary to apply the statistical technique(s) required to answer the research question at hand. Students not only learn the mechanics of the quantitative analysis, but also how to interpret conclusions based on quantitative evidence in context of the data and the research questions as well as identifying limitations due to data collection and study design.

For the lab component of this course students prepare weekly lab reports presenting statistical analysis of real data. In addition, students complete two independent data analysis projects where they answer significant research questions via the analysis of real data using statistical inference and modeling tools.

Learning objectives

The course goals are as follows:

1. Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
2. Use statistical software to summarize data numerically and visually, and to perform data analysis.
3. Have a conceptual understanding of the unified nature of statistical inference.
4. Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
5. Model numerical response variables using a single or multiple explanatory variables.
6. Interpret results correctly, effectively, and in context without relying on statistical jargon.
7. Critique data-based claims and evaluate data-based decisions.
8. Complete a research project demonstrating mastery of statistical data analysis from exploratory analysis to inference to modeling.

Prerequisites

This course has no pre-requisites.

Workload

You are expected to put in ~6 hours of work / week outside of class. Some of you will do well with less time than this, and some of you will need more.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, your TAs, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.

- Start your project early and and allow adequate time to complete it.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Textbooks

Readings for the course will come from the following textbooks. They are **freely available online** and you do not need to purchase a physical copy of either book to succeed in this class.

1. `[ims]`: Mine Çetinkaya-Rundel and Jo Hardin. Introduction to Modern Statistics. (in progress) 2nd edition. OpenIntro, 2023.
2. `[r4ds]`: Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund. R for Data Science. 2nd edition. O'Reilly, 2022.

Course community

Duke Community Standard

All students must adhere to the Duke Community Standard (DCS): Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, and accountability. Citizens of this community commit to reflect upon these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity.

To uphold the Duke Community Standard, students agree:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised.

Inclusive community

It is my intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength, and benefit. It is my intent to present materials and activities that are respectful of diversity and in alignment with Duke's Commitment to Diversity and Inclusion. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally, or for other students or student groups.

Furthermore, I would like to create a learning environment for my students that supports a diversity of thoughts, perspectives and experiences, and honors your identities. To help accomplish this:

- If you have a name that differs from those that appear in your official Duke records, please let me know! You'll be able to note this in the Getting to know you survey.
- If you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to come and talk with me. If you prefer to speak with someone outside of the course, your academic dean is an excellent resource.
- I (like many people) am still in the process of learning about diverse perspectives and identities. If something was said in class (by anyone) that made you feel uncomfortable, please let me or a member of the teaching team know.

Pronouns

Pronouns are meaningful tools to communicate identities and experiences, and using pronouns supports a campus environment where all community members can thrive.

Please update your gender pronouns in Duke Hub. You can learn more at the Center for Sexual and Gender Diversity's website.

Accessibility

If there is any portion of the course that is not accessible to you due to challenges with technology or the course format, please let me know so we can make appropriate accommodations.

The Student Disability Access Office (SDAO) is available to ensure that students are able to engage with their courses and related assignments. Students should be in touch with the Student Disability Access Office to request or update accommodations under these circumstances.

Communication

All lecture notes, assignment instructions, an up-to-date schedule, and other course materials may be found on the course website: sta101-f23.github.io.

I will regularly send course announcements via email and Sakai, make sure to check one or the other of these regularly. If an announcement is sent Monday through Thursday, I will assume that you have read the announcement by the next day. If an announcement is sent on a Friday or over the weekend, I will assume that you have read it by Monday.

Where to get help

- If you have a question during lecture or lab, feel free to ask it! There are likely other students with the same question, so by asking you will create a learning opportunity for everyone.
- The teaching team is here to help you be successful in the course. You are encouraged to attend office hours to ask questions about the course content and assignments. Many questions are most effectively answered as you discuss them with others, so office hours are a valuable resource. Please use them!
- Outside of class and office hours, any general questions about course content or assignments should be posted on the course Slack. There is a chance another student has already asked a similar question, so please check the other posts on Slack before adding a new question. If you know the answer to a question posted on Slack, I encourage you to respond!

Check out the Support page for more resources.

I want to make sure that you learn everything you were hoping to learn from this class. If this requires flexibility, please don't hesitate to ask.

- You *never* owe me personal information about your health (mental or physical) but you're always welcome to talk to me. If I can't help, I likely know someone who can.
- I want you to learn lots of things from this class, but I primarily want you to stay healthy, balanced, and grounded during this crisis.

Lectures and lab

The goal of both the lectures and the labs is for them to be as interactive as possible. My role as instructor is to introduce you new tools and techniques, but it is up to you to take them and make use of them. A lot of what you do in this course will involve writing code, and coding is a skill that is best learned by doing. Therefore, as much as possible, you will be working on a variety of tasks and activities throughout each lecture and lab. Attendance

will not be taken during class but you are expected to attend all lecture and lab sessions and meaningfully contribute to in-class exercises and discussion.

You are expected to bring a laptop to each class so that you can take part in the in-class exercises. Please make sure your laptop is fully charged before you come to class as the number of outlets in the classroom will not be sufficient to accommodate everyone. See Duke LIFE loaner laptop program if you need a loaner laptop.

Assessment

Assessment for the course is comprised of four components: attendance and participation, reading quizzes, homework assignments, and projects.

- **Attendance and participation** is required throughout the semester. Students who attend at least 80% of the lectures and participate regularly during lecture will receive full credit for this portion of their grade. Participation in labs as well as on Slack will also count towards this component.
- **Reading quizzes** (6), due every other week (roughly), completed individually. Each quiz is worth 2% of the grade. Lowest quiz score is dropped.

  Reading quizzes will be linked from the course schedule. They always cover reading that is due since the previous quiz and up to and including the deadline for the given quiz. They're due by 12pm ET (beginning of class) on the indicated day on the course schedule.
- **Homework assignments** (6), due every other week (roughly), completed individually. Each homework assignment is worth 8% of the course grade. Lowest homework assignment score is dropped.

  Homework assignments are due by 12pm ET (beginning of class) on the indicated day on the course schedule.
- **Projects** (2), mid-semester and end of semester, completed in teams.
  - Project 1: Teams will be given a dataset to visualize. Project 1 is worth 15% of the course grade.
  - Project 2: Teams will pick a dataset of interest to them and/or build an R package that implements a new type of data visualization in R. Project 2 is worth 25% of the course grade.

  The deliverables for each project will include a data visualization, a write up of the process and findings, and a presentation. For the second project, you will be encouraged to think beyond a traditional two-dimensional data visualization (e.g. interactive web apps/dashboards, data art, generative art, physical/tangible visualizations, **ggplot2** extensions, etc.).

Each project will have a peer review component to provide at least one round of feedback during the process of development. Teams will provide periodic peer feedback to their teammates while working on the projects as well as upon completion. The scores from the peer evaluations, along with individual contributions tracked by commits on GitHub, will be used to ensure that each student has contributed to the teamwork.

All team members must take part in the presentation. Presentations can be given in person in class, or via Zoom if the team prefers. My preference is that the team stick to one method of delivery (all presenters in person or all presenters on Zoom), but I realize a lot can change throughout this semester, and we'll adjust accordingly.

All work is expected to be submitted by the deadline and there are no make ups for any missed assessments. See [Late work policy] for policies on late work.

Grading

The final course grade will be calculated as follows:

Category	Percentage
Attendance and participation	5%
Labs	25%
Exam 1	20%
Exam 2	20%
Project 1	10%
Project 2	20%

The final letter grade will be determined based on the following thresholds:

Letter Grade	Final Course Grade
A	>= 93
A-	90 - 92.99
B+	87 - 89.99
B	83 - 86.99
B-	80 - 82.99
C+	77 - 79.99
C	73 - 76.99
C-	70 - 72.99
D+	67 - 69.99
D	63 - 66.99
D-	60 - 62.99
F	< 60

These are upper bounds for grade cutoffs, depending on the class performance the cutoffs may be lowered but they won't be increased.

Attendance and participation

You are expected to be present at class meeting and actively participate in the discussion. Your attendance and participation during class, as well as your activity on the course Slack will make up a non-insignificant portion of your grade in this class. While I might sometimes call on you during the class discussion, it is your responsibility to be an active participant without being called on.

Labs

In labs, you will apply the concepts discussed in lecture to various data analysis scenarios. Labs will focus on both computation and conceptualization. Lab assignments will be completed using Quarto and submitted as PDF for grading in Gradescope. While you may collaborate with others on lab assignments, your final solution should be your own.

Lowest lab score will be dropped.

Exams

There will be two exams. Each exam will be comprised of two components:

- In class: 75 minute in-class exam. This exam is closed book, however you are allowed to use one sheet of notes ("cheat sheet") to the midterm and the final. This sheet must be no larger than 8 1/2 x 11, and **must be prepared by you**. You may use both sides of the sheet.
- Take home: Following the in class exam, you'll have 48 hours to complete the take home portion of your exam. The take home portion will follow from the in class exam and focus on the analysis of a dataset introduced in the take home exam.

Through these exams you have the opportunity to demonstrate what you've learned in the course thus far. Each exam will include small analyses and computational tasks related to the content in application exercises and labs. More details about the content and structure of the exams will be discussed during the semester.

See Section for dates and times of the exams. Exam dates cannot be changed and no make-up exams will be given. If you can't take the exams on these dates you should drop this class.

Projects

There will be a mid-semester prediction project and a final project. The prediction project will introduce you to conducting independent analyses and writing a formal report using a pre-specified data set. The final project allows you to explore a question and data set of your own. More details about the projects will be provided during the semester. Projects will be completed in teams.

You will be assigned to a different team for each of your two projects. You are encouraged to sit with your teammates in lecture and you will also work with them in the lab sessions. All team members are expected to contribute equally to the completion of each project and you will be asked to evaluate your team members after each assignment is due. Failure to adequately contribute to an assignment will result in a penalty to your mark relative to the team's overall mark.

Course policies

Academic honesty

TL;DR: Don't cheat!

Please abide by the following as you work on assignments in this course:

- **Collaboration:** Only work that is clearly assigned as team work should be completed collaboratively.
  - The labs must also be completed individually and you are welcomed to discuss the assignment with classmates at a high level (e.g., discuss what's the best way for approaching a problem, what functions are useful for accomplishing a particular task, etc.). However you may not directly share answers to homework questions (including any code) with anyone other than myself and the teaching assistants.
  - For the projects, collaboration within teams is not only allowed, but expected. Communication between teams at a high level is also allowed however you may not share code or components of the project across teams.
  - On individual assignments you may not directly share code with another student in this class, and on team assignments you may not directly share code with another team in this class.
- **Online resources:** I am well aware that a huge volume of code is available on the web to solve any number of problems. Unless I explicitly tell you not to use something, the course's policy is that you may make use of any online resources (e.g., StackOverflow) but you must explicitly cite where you obtained any code you directly use (or use as

inspiration). Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.

- **Use of generative artificial intelligence (AI)**: You should treat generative AI, such as ChatGPT, the same as other online resources. There are two guiding principles that govern how you can use AI in this course:[1] (1) *Cognitive dimension:* Working with AI should not reduce your ability to think clearly. We will practice using AI to facilitate—rather than hinder—learning. (2) *Ethical dimension***:** Students using AI should be transparent about their use and make sure it aligns with academic integrity.

  - **AI tools for code:** You may make use of the technology for coding examples on assignments; if you do so, you must explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism. You may use these guidelines for citing AI-generated content.
  - **AI tools for narrative:** Unless instructed otherwise, you may not use generative AI to write narrative on assignments. In general, you may use generative AI as a resource as you complete assignments but not to answer the exercises for you. You are ultimately responsible for the work you turn in; it should reflect your understanding of the course content.

Regardless of course delivery format, it is the responsibility of all students to understand and follow all Duke policies, including academic integrity (e.g., completing one's own work, following proper citation of sources, adhering to guidance around group work projects, and more). Ignoring these requirements is a violation of the Duke Community Standard. Any questions and/or concerns regarding academic integrity can be directed to the Office of Student Conduct and Community Standards at conduct@duke.edu.

Any violations in academic honesty standards as outlined in the Duke Community Standard and those specific to this course will

- automatically result in a 0 for the assignment,
- can further impact your overall course grade, and
- will be reported to the Office of Student Conduct for further action.

Late work & extensions

The due dates for assignments are there to help you keep up with the course material and to ensure the teaching team can provide feedback within a timely manner. We understand that things come up periodically that could make it difficult to submit an assignment by the deadline.

[1] These guiding principles are based on *Course Policies related to ChatGPT and other AI Tools* developed by Joel Gladd, Ph.D.

Policy on late work depends on the particular course component:

- Labs:
  - Late, but within 24 hours of deadline: -20% of available points.
  - Any later: No credit, and we will not provide written feedback.
  - Note that lowest lab score will be dropped, even if that score is a 0.
- Exams:
  - In class portions of the exams can obviously not be turned in late.
  - Late exams are not accepted.
- Projects: The following three components contribute to your project score.
  - Presentation: Late presentations are not accepted and there are no make ups for missed presentations.
  - Write up: GitHub repositories will be closed to contributions at the deadline. If you need to submit your work late, Slack/email me to reopen your repository.
    - Late, but within 24 hours of deadline: -20% of available points.
    - Any later: No credit, and we will not provide written feedback.
  - Peer evaluation: Late peer evaluations are not accepted. If you do not turn in your peer evaluation, you get 0 points for your own peer score as well, regardless of how your teammates have evaluated you.

Waiver for extenuating circumstances

If there are circumstances that prevent you from completing a lab or homework assignment by the stated due date, you may email the Head TA (Shuo Wang, shuo.wang717@duke.edu) before the deadline to waive the late penalty. In your email, you only need to request the waiver; you do not need to provide explanation. This waiver may only be used for **once** in the semester, so only use it for a truly extenuating circumstance.

If there are circumstances that are having a longer-term impact on your academic performance, please let your academic dean know, as they can be a resource. Please let me know if you need help contacting your academic dean.

Regrade requests

Every effort will be made to mark your work accurately. We are on your side, and want you to receive every point you have worked to earn. However, sometimes grading mistakes happen. If you believe that an error has been made, return the paper to the instructor **within four days**, stating your claim in writing.

The following claims will be considered for re-grading:

1. points are not totaled correctly;
2. the grader did not see a correct answer that is on your paper;
3. your answer is the same as the correct answer, but in a different form (e.g., you wrote a correct answer as 1/3 and the grader was looking for 0.333);
4. your answer to a free response question is essentially correct but stated slightly differently than the grader's expectation.

The following claims will **not** be considered for re-grading:

5. arguments about the number of points lost;
6. arguments about question wording.

Considering re-grades consumes time and resources that TAs and the instructor would rather spend helping you understand material. Please bring only claims of type 1-4 to our attention.

Note that during the regrade process your score could go up or go down or not change.

> ⚠ Warning
>
> No grades will be changed after the project presentations.

Attendance policy

Responsibility for class attendance rests with individual students. Since regular and punctual class attendance is expected, students must accept the consequences of failure to attend. More details on Trinity attendance policies are available here.

However, there may be many reasons why you cannot be in class on a given day, particularly with possible extra personal and academic stress and health concerns this semester. All course lectures will be recorded and available to enrolled students after class. If you miss a lecture, make sure to watch the recording and review the material before the next class session. Lab time is dedicated to working on your assignments and collaborating with your teammates on your project. If you miss a lab session, make sure to communicate with your team about how you can make up your contribution. Given the technologies we use in the course, this is straightforward to do asynchronously. If you know you're going to miss a lab session and you're

feeling well enough to do so, notify your teammates ahead of time. Overall these policies are put in place to ensure communication between team members, respect for each others' time, and also to give you a safety net in the case of illness or other reasons that keep you away from attending class.

Note that attendance and participation is part of your grade as well.

Attendance policy related to COVID symptoms, exposure, or infection

Student health, safety, and well-being are the university's top priorities. To help ensure your well-being and the well-being of those around you, please do not come to class if you have tested positive for COVID-19 or have possible symptoms and have not yet been tested. If any of these situations apply to you, you must follow university guidance related to the ongoing COVID-19 pandemic and current health and safety protocols. If you are experiencing any COVID-19 symptoms, contact student health (dshcheckin@duke.edu, 919- 681-9355). Learn more about current university policy related to COVID-19 at https://coronavirus.duke.edu. To keep the university community as safe and healthy as possible, you will be expected to follow these guidelines. Please reach out to me and your academic dean as soon as possible if you need to quarantine or isolate so that we can discuss arrangements for your continued participation in class.

Inclement weather policy

In the event of inclement weather or other connectivity-related events that prohibit class attendance, I will notify you how we will make up missed course content and work. This might entail holding the class on Zoom synchronously, we may rely on Duke's designated make-up days, or you may be asked to watch a recording of the class.

Policy on video recording course content

All lectures will be recorded and available on Panopto, so students should not need to create their own recordings of lectures. If you feel that you need record the lectures yourself, you must get permission from me ahead of time and these recordings should be used for personal study only, no for distribution. The full policy on recording of lectures falls under the Duke University Policy on Intellectual Property Rights, available at https://policies.provost.duke.edu/docs/faculty-handbook-appendix-m-intellectual-property. Unauthorized distribution is a cause for disciplinary action by the Judicial Board.

Accommodations

Academic accommodations

If you are a student with a disability and need accommodations for this class, it is your responsibility to register with the Student Disability Access Office (SDAO) and provide them with documentation of your disability. SDAO will work with you to determine what accommodations are appropriate for your situation. Please note that accommodations are not retroactive and disability accommodations cannot be provided until a Faculty Accommodation Letter has been given to me. Please contact SDAO for more information: sdao@duke.edu or access.duke.edu.

Religious accommodations

Students are permitted by university policy to be absent from class to observe a religious holiday. Accordingly, Trinity College of Arts & Sciences and the Pratt School of Engineering have established procedures to be followed by students for notifying their instructors of an absence necessitated by the observance of a religious holiday. **Please submit requests for religious accommodations at the beginning of the semester so that we can work to make suitable arrangements well ahead of time.** You can find the policy and relevant notification form here: https://trinity.duke.edu/undergraduate/academic-policies/religious-holidays.

Note: If you've read this far in the syllabus, email me a picture of your pet if you have one or your favorite meme!

Important dates

- **Monday, August 28:** Classes begin
- **Friday, September 8:** Drop/add ends
- **Friday, October 13:** Mid-semester grades reported
- **Saturday, October 20 - Tuesday, October 18**: Spring Break (continuing, but limited course support available)
- **Friday, November 10:** Last day to withdraw with W
- **Friday, December 8**: Classes end
- **Saturday, December 9 - Tuesday, December 12**: Reading period
- **Wednesday, December 4, 2-5pm**: Project 2 presentations

For more important dates, see the full Duke Academic Calendar.