

Syllabus - STA101L Summer I 2022

[Click here](#) to download a PDF copy of the syllabus.

Course info

	Day	Time	Location
Lectures	Mon, Tue & Thu	3:30 pm - 5:35 pm	Old Chemistry Building 003
Labs	Wed & Fri	3:30 pm - 4:45 pm	Old Chemistry Building 003

Overview

Welcome to STA101L Data Analysis and Statistical Inference. The goal of this class is to prepare you to be critical consumers of statistical analyses in your scientific fields of practice and future professions. Our point of departure will be to think about data collection: how to (not) collect data, and how the way in which data are collected impacts the analysis that we conduct. We will then quickly delve into data visualization. Ever heard of a mosaic plot or an inter-quartile range? This is an area of data analysis where creativity and an eye for good design can make a difference. Once we have good grasp of how to visualize data, we will construct statistical models to make predictions and to understand the relations that exist between variables.

While models can be useful (especially if their predictions are accurate!), all models are inherently *wrong*. The statistical tests that we will learn in the second half of the course will help us quantify how (un)certain we are that the models we construct pick up real patterns in the data and not just background noise.

Throughout the class, you will attend hands-on labs in which you will learn to implement all these techniques in the statistical computing software R.

Learning objectives

By the end of the term you will be able to . .

- visualize and summarize data sets with numerical and categorical variables;
- construct and investigate linear regression models for forecasting; this includes fitting, evaluating, comparing and selecting models, as well as interpreting their output;
- conduct hypothesis tests and construct confidence intervals for proportions, differences between proportions, means, differences between means and regression coefficients;
- implement these techniques in R, and use RMarkdown to write reproducible reports;
- plan and complete a statistical analysis of a real-world phenomenon using visual and numerical summaries, hypothesis tests, confidence intervals and regression models.

Prerequisite

There is no prerequisite for this class.

Community

Duke Community Standard

As a student in this course, you have agreed to uphold the [Duke Community Standard](#) as well as the practices specific to this course.

Inclusive community

It is my intent that students from all 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 organize 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 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.

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 at rmorsomme.github.io/website.

I will regularly send course announcements via email, make sure to check your mail box 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!
- **If you email me, please include “STA 101” in the subject line.** Barring extenuating circumstances, I will respond to STA 101 emails within 24 hours Monday - Friday. Response time may be slower for emails sent Friday evening - Sunday.

Check out the [Support](#) page for more resources.

Textbook

The class will closely follow the book [Introduction to Modern Statistics](#) (first edition) by Mine Çetinkaya-Rundel and Johanna Hardin. This is an open-source book freely available online. You're welcomed to read on screen or print it out. If you prefer a paperback version you can buy it at the cost of printing (around \$20) on Amazon. The textbook store will not carry copies of this text.

Chapters 1-7 of the book [R for Data Science](#) by Garret Grolemund and Hadley Wickham (also open-source and freely available) will also be useful.

Lectures and labs

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. 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 may not be sufficient to accommodate everyone. More information on loaner laptops can be found [here](#).

Assessment

Assessment for the course is comprised of three components: homework assignments and two projects.

Homework assignments

To reduce the number of assignments during the summer session, problem sets and lab exercises will be merged. In these homework assignments, you will apply what you've learned during lectures and labs to show conceptual understanding of the content and complete data analysis tasks in R. You may discuss homework assignments with other students; however, homework should be completed and submitted individually. Homework must be typed up using RMarkdown and submitted as a PDF on Gradescope.

The homework assignment with the lowest grade will be dropped at the end of the term.

Projects

There will be two group projects. Through these, you have the opportunity to demonstrate what you've learned in the course thus far. The projects will focus on the two pillars of statistics and data science: *prediction* and *inference*. In the first project, you will construct a regression model for prediction. The goal will be to build a model that provides predictions that are as accurate as possible. In the second project, you will analyze a phenomenon that interests you using real-world data. More detail about the projects will be given during the term.

Grading

The final course grade will be calculated as follows:

Assessment	Percentage
Homework	50%
Prediction project	20%
Inference project	30%

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

Note that the final grade may be curved.

Five tips for success

Your success on this course depends very much on you and the effort you put into it. The course has been organized so that the burden of learning is on you. Your TA and I will help you by providing you with materials and answering questions, but for this to work you need to do the following:

1. complete the reading before the lectures;
2. ask questions *quickly*; don't let a day pass by with lingering questions;
3. do the homework and lab;
4. practice, practice, practice;
5. don't procrastinate; start on the homework assignments and the projects early.

Course policies

Academic integrity

Don't cheat!

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 affirm their commitment to uphold the values of the Duke University community by signing a pledge that states:

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

Regardless of course delivery format, it is your responsibility to understand and follow Duke policies regarding academic integrity, including doing one's own work, following proper citation of sources, and adhering to guidance around group work projects. Ignoring these requirements is a violation of the Duke Community Standard. If you have any questions about how to follow these requirements, please contact Jeanna McCullers (jeanna.mccullers@duke.edu), Director of the Office of Student Conduct.

Collaboration policy

Only work that is clearly assigned as team work should be completed collaboratively.

- The homework assignments must 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 TA.
- For the projects, collaboration between teams at a high level is also allowed; however, you may not share code or components of the project with other teams.

Policy on sharing and reusing code

I am well aware that a huge volume of code is available on the web to solve any number of problems. You may make use of any online resources (e.g. RStudio Community, 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. On homework assignments you may not directly share code with another student in this class, and on the projects you may not directly share code with another team.

Late work policy

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. Given the fast pace of a summer class, **any assignment submitted after the deadline will not be graded**. The homework assignments will be published early to give you ample time to work on them. Note that the lowest homework will be dropped.

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 [Raphael Morsomme](#) and our TA [Rohit Roy](#) before the deadline to obtain a 24-hour extension. 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 a member of the instruction team know if you need help contacting your academic dean.

Regrade request policy

Regrade requests must be submitted on Gradescope within 48 hours of when an assignment is returned. Regrade requests will be considered if there was an error in the grade calculation or if you feel a correct answer was mistakenly marked as incorrect. Requests to dispute the number of points deducted for an incorrect response will not be considered. Note that by submitting a regrade request, the entire question will be graded which could potentially result in losing points.

No grades will be changed after the inference project report is due.

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 term. Lab time is dedicated to working on your lab 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. 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.

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 symptoms related to COVID-19, have had a known exposure to COVID-19, or have tested positive for COVID-19. 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 at 919-681-9355. 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 or watching a recording of the class.

Policy on video recording course content

If you feel that you need record the lectures, you must get permission from the instructor 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 provost.duke.edu/sites/default/files/FHB_App_P.pdf. Unauthorized distribution is a cause for disciplinary action by the Judicial Board.

Learning during a pandemic

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 health crisis.

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

Important dates

- **May 11:** Classes begin (Monday meeting schedule)
- **May 13:** Drop/add ends
- **May 16:** Regular class meeting schedule begins
- **May 30:** Memorial Day holiday, no classes are held
- **June 8:** Last day to withdraw with W
- **June 17:** Classes end
- **June 20:** Juneteenth holiday
- **June 21:** Reading period
- **June 23:** Last assignment

Click [here](#) for the full Duke academic calendar.

Attribution

Some portions of this syllabus are based on the syllabus of [STA 210 - Spring 2022](#) by Prof. Mine Çetinkaya-Rundel.