# STOR 120-001 Foundations of Statistics and Data Science Summer 2024

### **Course Description**

This course combines three perspectives: inferential thinking, computational thinking, and real-world relevance. Given data arising from some real-world phenomenon, how does one analyze those data so as to understand that phenomenon? The course teaches critical concepts and skills in computer programming and statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, document collections, geographical data, and social networks. It delves into social issues surrounding data analysis such as privacy and design.

## Course Goals and Learning Objectives

#### This course will enable you to:

- think critically about the world around you;
- not take your data for granted;
- use the combination of programming + statistics as a feature, not a bug;
- come away with practical skills;
- know that your inference is sound;
- run your own experiments and plan data collection;
- know the right statistical tools for the job;
- make due with limited data;
- quantify and understand uncertainty in data;
- turn your data analysis into a decision;
- think of ways that you could be wrong;
- consider edge-cases and how this affects your model choice; and
- illustrate the above points with real-world data.

#### **Course Assessments**

Assignments	Percentages	Frequency
Homework	20%	6 Total
Lab	20%	7 Total
Midterm1	17.5%	8-July
Midterm 2	17.5%	16-July
Final	25%	

#### Grading Scale

Your final grade is based on a weighted average according to the previously addressed breakdown. Curving on individual/group assessments should not be expected. A curve may be applied to the final grades depending upon the class average. Conversion to a letter grade will be based on the table below:

А	93 to 100	В	83 to 86.99	С	73 to 76.99	D	60 to 66.99
A-	90 to 92.99	B-	80 to 82.99	C-	70 to 72.99	F	0 to 59.99
B+	87 to 89.99	C+	77 to 79.99	D+	67 to 69.99		

#### Assignment Descriptions

#### Labs:

Attendance to all labs is mandatory. Every week, your lab instructor will take attendance. If you are not present for lab, your lab will not be graded. During this period, students are required to complete a lab assignment. Each lab assignment will be based on the topics discussed in lecture or related to your final project. Students are responsible to turn in their own labs but are encouraged to work in teams and help each other. These assignments are to be completed using Jupyter Notebook and submitted on Canvas by 11:59 PM on the day of the lab. A lab instructor will be provided to help students in the completion of the lab and to facilitate group work. Take heed, no late lab assignments will be accepted. You will need to get a university excused absence to prevent a loss of points in these weekly labs if you miss class.

#### Homework:

You can work with each other on this assignment, but the work you submit should be your own. Any copying of solutions will result in a 0 if caught cheating. These assignments are to be completed by filling in the designated Jupyter notebook and turned into canvas. No late homework assignments will be accepted.

#### Exams:

There are two midterms and one final exam in this course. The midterms will be given the full two-hour lecture time slot, while the final exam will have a full three-hour slot, each completed in person. Each exam will have a coding portion and written portion. The coding portion will be open-book, open-notes, closed internet, while the written portion will be entirely closed-book, closed-notes. The written portion of the exam will not assess your coding abilities (I will not ask you to write code from memory) but your conceptual understanding of topics covered. The final exam will be comprehensive, but if you do better on the final exam than either of the two midterms, the final exam grade will replace the lowest of your two midterm grades.

## Rough Course Outline (updated 6/4/24)

Class Day Date Topics Assignments Due (Ro	ughly)
---	--------

1	М	24-June	Syllabus and Introduction to Python	
2	Т	25-June	Tables, Data Types	
3	W	26-June	Tables, Visualization (Lab 1)	Lab 1
4	Th	27-June	Visualization	HW 1
5	F	28-June	User-Defined Functions (Lab 2)	Lab 2
6	М	1-July	User-Defined Functions	HW 2
7	Т	2-July	Groups and Joins/Conditional Statements and Loops	
8	W	3-July	Conditionals and Loops (Lab 3)	Lab 3
	Th	4-July	No Class – Independence Day	
9	F	5-July	Sampling/Law of Large Numbers	
10	М	8-July	Midterm 1	Midterm Exam
11	Т	9-July	Empirical Distributions	HW 3
12	w	10-July	Empirical Distributions/Decisions and Uncertainty/	Lah 4
		io suly	AB Testing and Types of Errors (Lab 4)	
13	Th	11-July	Bootstrapping/Confidence Intervals	
14	F	12-July	Measures of Center/Spread (Lab 5)	Lab 5
15	М	15-July	Measures of Center/Spread/Normal Distribution	HW 4
16	Т	16-July	Midterm 2	Midterm 2
17	W	17-July	T-Distribution (Lab 6)	Lab 6
18	Th	18-July	T-Distribution	
19	F	19-July	Correlation/Linear Regression (Lab 7)	Lab 7
20	М	22-July	Linear Regression/Least Squares	HW 5
21	Т	23-July	Residuals and Regression Inference	
22	W	24-July	Classifiers and Accuracy (Optional Lab – Review)	
23	Th	25-July	Classifiers and Accuracy	HW 6
			Final Exam	

## Course Policies and Resources

Number	STOR 120.001		
Title	Foundations of Statistics and Data Science		
Course Format	<ul> <li>In-person unless otherwise advised. In the case of a transition to a virtual environment, lectures will continue synchronously via zoom.</li> <li>Lectures will be supplemented with in-class programming and practical discussion. Students will also be required to register for a weekly lab with required attendance in person.</li> </ul>		
Prerequisites	High school algebra		
Target Audience	This course does not have any prerequisites beyond high-school algebra. The curriculum and format is designed specifically for students who have not previously taken statistics or computer science courses. If you have some prior experience in either statistics or computing you are welcome to enroll and will find much of interest due to the innovative nature of the course. If you have taken several statistics or computer science courses should instead take a more advanced course.		
Instructor	Andy AckermanOffice: Hanes B-7Email: <a href="mailto:atacker@email.unc.edu">atacker@email.unc.edu</a> Office Hours: W 9-10, F 8-9:30 am Hanes B-07		
Lab Instructors/Instructional Assistants	Akshay Sakanaveeti: MTH 2-3 pm via <u>zoom</u> Kyung Rok Kim:WF 11:45-12:45 pm via <u>zoom</u> Adrian Allen: ThF 4-5 pm via <u>zoom</u>		
Course Website	https://uncch.instructure.com/courses/61073?for_reload=1		
Lecture Details	MTTh 9:45-11:45 am; WF 10:45-11:45 am Hanes 107		
Course Texts	Computational and Inferential Thinking: The Foundations of Data Science, by Ani Adhikari and John DeNero Freely available at <u>link</u>		

Teaching Philosophy	My general approach aims to make two principles abundantly clear: there are no (sincere) bad questions, and education is only one facet of your life.
	To the first point, if you ask me a question out of genuine curiosity or confusion, I guarantee to respect and attend to it regardless of the level of complexity. Note, I do not claim to always have the answer, and certain questions may require an outside (office hours) conversation. But my guarantee is to provide an answer when I am able and search with you when unable. The flip side of this guarantee is your willingness to keep me accountable. Ask questions. If you are confused, please reach out for help. If in-class questions are too daunting, take advantage of office hours, tutorial sessions, or electronic communication.
	As to the second point, while I am ultimately a very ardent advocate of the merits of education, I want to reiterate that no single facet of your life – perhaps least of which performance in STOR 120 – fully defines you as a person. This is not to say that this, or any class, is unimportant, rather that this class is meant to enable and instruct not to define.
Honor Code	It is my aim to make my expectations for acceptable and original work as clear and explicit as possible. Gray areas frustrated me as a student, so I aim to prevent them as an instructor.
	To that end, students are bound by UNC's <u>honor code</u> in taking exams and in written work, and the submission of said work signifies understanding and acceptance of those requirements.
	More specific to this course, collaboration is not only permitted but <i>encouraged</i> on homework and labs. Exams, by contrast, will be completed individually. I recognize this can be immensely stressful, and I do not intend to add to that burden. That said, my job is to cultivate and assess your <i>personal</i> understanding. Insofar as individual assignments serve this purpose, they will be occasionally used.
	Large language models, such as Chat-GPT, should not be used for any component of this course. If I suspect that such a resource has been used to complete a homework, lab, or exam, I will consider it a breach of the honor code.
	Please consult with me if you have any questions about the honor code. Asking for clarification in advance is far preferred to asking for forgiveness post-facto.

COVID-19	Each of us has a responsibility to know and act on these standards and policies in a way that maximizes a safe and healthy environment for us to teach, work, learn and live. To this end, we are developing a set of community standards and policies for our students, faculty, staff and visitors. We are all in this together, and we believe that together, we can face the challenges presented by COVID-19 with resilience,
	determination and great support for our community. See <u>https://carolinatogether.unc.edu/community-standards-3-2/</u> for a list of guidelines that we all need to follow to reduce the spread of COVID-19.
	See the Carolina Together Roadmap at <u>https://carolinatogether.unc.edu/</u> for more information on the University's plans regarding COVID-19. This website is continuously updated and should be checked weekly.
	More specifically, in accordance with the current iteration of these policies, masks will not be mandatory in class.
Accessibility Resources	UNC-Chapel Hill facilitates the implementation of reasonable accommodations for students with learning disabilities, physical disabilities, mental health struggles, chronic medical conditions, temporary disability, or pregnancy complications, all of which can impair student success. See the ARS website for contact and registration information: <u>https://ars.unc.edu/about-ars/contact-us</u>
Attendance Policy	Lecture attendance is not required. That said, class participation is fundamental to success in this class. You, the student, are responsible for filling information that was missed during absences.
University Testing Center	The College of Arts and Sciences provides a secure, proctored environment in which exams can be taken. The center works with instructors to proctor exams for their undergraduate students who are not registered with ARS and who do not need testing accommodations as provided by ARS. In other words, the Center provides a proctored testing environment for students who are unable to take an exam at the normally scheduled time (with pre-arrangement by your instructor). For more information, visit <u>http://testingcenter.web.unc.edu/</u> .
Counseling and Psychological Services	CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: <u>https://caps.unc.edu/</u> or visit their

	facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.
Title IX	Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Please contact the Director of Title IX Compliance (Adrienne Allison – Adrienne.allison@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (gvsc@unc.edu; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.
Technology Use	Students are required to bring their computer to every class and lab with a working copy of R and RStudio. Directions for free downloads of this software will be provided. The professor or lab assistant will occasionally request computers to be closed for dynamic discussion and guest speakers.
Legal	Andrew Ackerman reserves the right to make changes to the syllabus, including all due dates. These changes will be announced as early as possible so that students can adjust their schedules.