STAT 0116: Introduction to Statistical Science

Fall 2024

Instructor:	Christian Stratton (he/him)	**Time:**	MW 9:45 – 11:00 (Class) T 9:45 – 11:00 (Lab)
Email:	cstratton@middlebury.edu	**Place:**	Munroe Hall 217 (Class) Warner Hall 105 (Lab)
Office:	Warner 203	**Office hours:**	TBD Also by appointment

Course description: A practical introduction to statistical methods and the examination of data sets. Computer software will play a central role in analyzing a variety of real data sets from the natural and social sciences. Topics include descriptive statistics, elementary distributions for data, hypothesis tests, confidence intervals, correlation, regression, contingency tables, and analysis of variance. The course has no formal mathematics prerequisite, and is especially suited to students in the physical, social, environmental, and life sciences who seek an applied orientation to data analysis.

Correspondence: My goal is to maximize my availability for help and discussion throughout the semester. Office hours will be determined via poll during the first week of class, but please feel free to contact me via email at anytime. Additionally, I am happy to meet outside of office hours by appointment.

Meeting format: Class time (Monday and Wednesday in Munroe Hall 217) will generally be used to learn new statistical concepts through a mixture of lecture and in-class activities. Most class periods will feature a short lecture introducing a new concept, followed by an in-class guided activity to be worked on in small groups. Lab time (Tuesday in Warner Hall 105) will generally be used to apply data analysis concepts to real data problems through the statistical computing language R in small groups. You will need to have access to a laptop during class and lab. See more details below.

Learning objectives: Through this course, students will:

- Learn the basics of statistical theory and common statistical techniques
- Acquire the computational skills to be able to summarize, graph, and make inference in the statistical computing language R.
- Apply critical and statistical thinking in their daily lives enabling them to better analyze current events and media, including news and journal articles.

Textbook and materials: There is nothing that need be purchased for this class; all materials are free.

- The website for this course is on Middlebury Canvas. Please check Canvas often for assignments, deadlines, resources, and announcements.
- Students must have access to a laptop with the statistical computing language R, which can be downloaded for free at https://cran.rstudio.com/. Additionally, I recommend using RStudio as an integrated development environment (IDE) for interfacing with R. RStudio may be downloaded for free at https://posit.co/download/rstudio-desktop/.
  - Laptops with R/RStudio pre-installed are available to borrow from the Davis Family Library, which are a good option for those without access to a laptop or those experiencing short-term issues with your laptop. Please talk to me or the front desk of the Davis Library for more info.

- We will use the free online textbook *Intro to Modern Statistics* by Mine Çetinkaya-Rundel and Johanna Hardin. This book may be accessed via web browser at https://openintro-ims.netlify.app/ or downloaded at https://leanpub.com/imstat. Note that you may set the donation value to $0.
- During week 3, when we discuss probability, it may be helpful to view some supplementary material on probability, which is not covered in *Intro to Modern Statistics*. An excerpt from another free textbook is available at https://www.openintro.org/go/?id=stat_os4_probability_chapter.

Academic integrity: You are bound by Middlebury College's honor code, including its policies on plagiarism and cheating. Violation of these rules is ground for failure. To avoid charges of plagiarism, cite all the sources used to complete your assignments/homework, including any peers with whom you collaborated. I encourage you to seek help in understanding the concepts and problems in your assignments from various sources, including peers, instructors, peer tutors, class notes, textbooks, and online sources.

Use of LLM and generative AI: Large language models (LLM) and generative AI, such as ChatGPT, are powerful tools enabled by statistics and data science techniques that may be used to enhance your learning of statistics and coding languages. As such, the use of large language models (LLM) and generative AI, such as ChatGPT, is permitted in this class and may be used on all homework assignments, lab assignments, exams, and projects. However, **you may not copy responses verbatim from these tools, nor may you use these tools to generate complete responses or assignments.** Additionally, if content from generative AI is used on an assignment, **you must provide appropriate citation.** To clarify this policy, examples of acceptable and unacceptable prompts for ChatGPT are provided below.

Acceptable:

- Please provide example of how to conduct a two-sample t-test in R.
- How do I interpret a p-value?
- What is a significance level?
- How can I speed up the following code: ...

Unacceptable:

- Conduct a two-sample t-test for the uploaded data and write a statistical report describing the results.
- Answer the following question: *copy pasta from assignment*

Disclaimer: I am compelled to note that while generative AI can be a powerful tool, it is not infallible. Consider the exchange provided at the end of the syllabus, conducted on ChatGPT 4o mini on 2024/09/01. It is possible that generative AI will provide you with incorrect information, and it is your responsibility to use generative AI critically. "ChatGPT said so," is not sufficient justification for an answer, and I am unlikely to be sympathetic to such comments on assignments.

Late policy: Consistent engagement with the course material is essential for your learning and academic growth. However, I understand that unforeseen circumstances may occasionally arise:

- When you become aware that you won't be able to make a deadline, please notify me and inform me of what day in the next week you anticipate completion of the assignment. You do not need to disclose why you are missing the deadline. So long as you communicate to me before the deadline, no late penalty will be applied.
- If you do not communicate with me before the deadline, late submissions will be subject to a penalty of 20% per day.

Course assessment: Your grade will be determined by homework assignments, lab assignments, take-home exams, and a final project. Each category is loosely defined as follows:

30%	**Homework**	There will typically be one homework assignment per week, assigned on Mondays and due on Canvas the following Monday at 23:59 EST. Please check the course website regularly for homework assignments, deadlines, and updates.
15%	**Lab assignments**	Lab assignments will generally be completed during Lab in small groups. All students must submit their lab assignment on Canvas within one week by 23:59 EST.
30%	**Take-home exams**	There will be two take-home exams in this class: the midterm and the final. Both exams will be open-book; referencing class notes, previous assignments and labs, the textbook, or online sources are appropriate. However, unlike homework and lab assignments, exams should be **completed independently without discussion with peers, tutors, or other instructors.**
25%	**Final project**	You will analyze a data set of your choice. More details will be provided throughout the semester.

Diversity and inclusion statement: 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 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, gender identity, sexual orientation, disability, age, socioeconomic status, ethnicity, race, religion, culture, perspective, and other background characteristics. Your suggestions about how to improve the value of diversity in this course 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.

Name and pronoun policy: I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester. If your name or pronoun changes during the course of the semester, please notify me.

Statement on religious holidays: Campus policies regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students, who because of religious obligations, have conflicts with schedules exams, assignments or required attendance. Please review the course calendar and notify me if you anticipate such conflicts so that we can agree upon alternative arrangements.

Accommodations for disabilities: Students who have Letters of Accommodation in this class are encouraged to contact me early in the semester to ensure that such accommodations are implemented in a timely fashion. For those without Letters of Accommodation, assistance is available to eligible students through the Disability Resource Center (DRC). Please contact ADA Coordinators Jodi Litchfield, Peter Ploegman or Dierdre Kelly of the DRC at ada@middlebury.edu for more information. All discussions will remain confidential.

Student athletes: Please review the course calendar and notify me if you anticipate any conflicts with your athletic schedule so that we can agree upon alternative arrangements. Whenever possible, I would appreciate at least one week notice.

Disclaimer: This syllabus is a dynamic document and may change throughout the semester as a result of our conversations and decisions as a class.

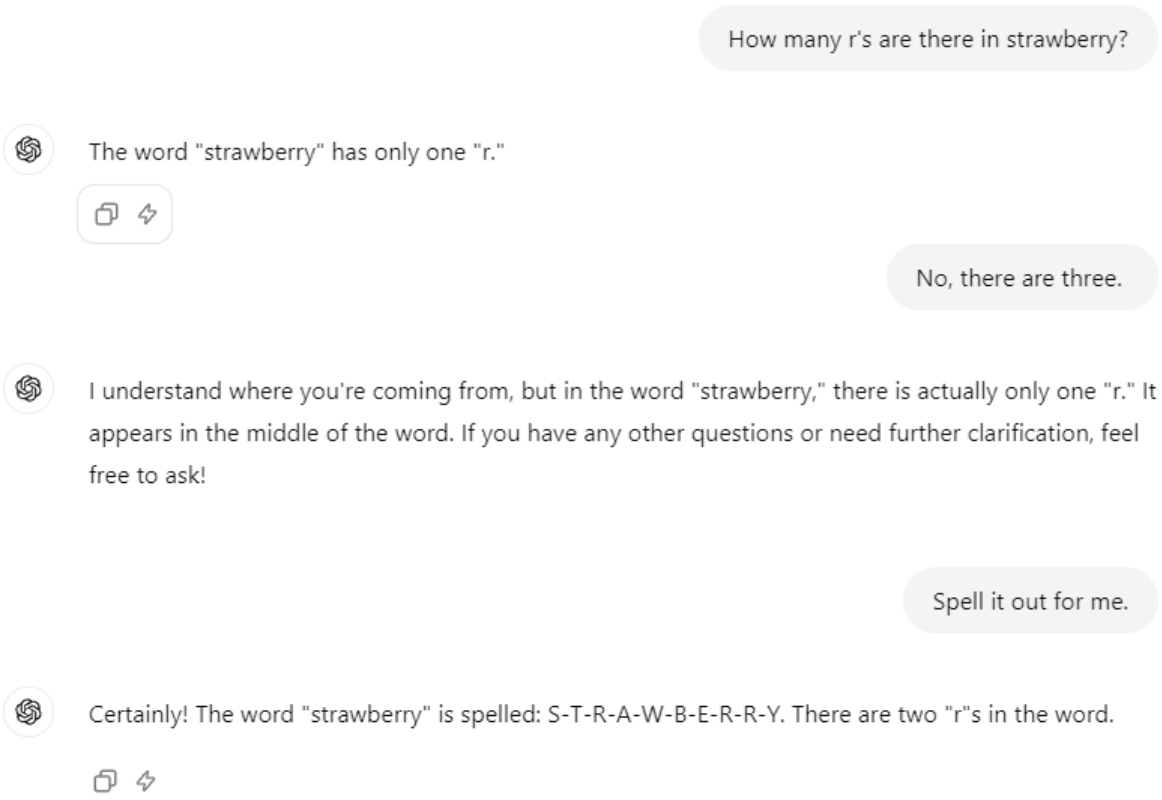

Figure 1: Prompt provided to ChatGPT 4o mini on 2024/09/01.

Monday	Tuesday	Wednesday
Sep 9th **1** Supp - the Martian alphabet	10th **2** Lab 1 - Intro to R	11th **3** Ch. 1 - Hello data
16th **4** Ch. 2 - Study design	17th **5** Ch. 4 - Categorical data	18th **6** Ch. 5 - Numeric data
23rd **7** Supp - Probability I	24th **8** Lab 2 - Probability	25th **9** Supp - Probability II
30th **10** Ch. 13 - Normal variables	Oct 1st **11** Lab 3 - Normal variables	2nd **12** Ch. 13 - Central Limit Theorem
7th **13** Ch. 11 - Hypothesis testing	8th **14** Lab 4 - Martian alphabet revisited	9th **15** Ch. 13 - Confidence intervals
14th **16** Ch. 16 - Single proportion (nonparametric)	15th **17** Lab 5 - Single proportion	16th **18** Ch. 16 - Single proportion (parametric)
21st **19** Ch. 17 - Two proportions (nonparametric)	22nd **20** Lab 6 - Two proportions	23rd **21** Ch. 17 - Two proportions (parametric)
28th **22** Ch. 18 - Two-way tables (both)	29th **23** Lab 7 - Two-way tables	30th **24** Ch. 14 - Errors
Nov 4th **25** Ch. 19 - Single mean (both)	5th **26** Lab 8 - Single mean	6th **27** Ch. 20 - Difference in means (both)
11th **28** Ch. 21 - Mean difference (both)	12th **29** Lab 9 - Difference in means vs mean difference	13th **30** Ch. 24 - Simple linear regression (nonparametric)
18th **31** Ch. 24 - Simple linear regression (parametric)	19th **32** Lab 10 - Simple linear regression	20th **33** Ch. 26 - Simple logistic regression
25th Thanksgiving recess	26th Thanksgiving recess	27th Thanksgiving recess
Dec 2nd **34** Project work day	3rd **35** Project work day	4th **36** Project presentations

Tentative assignment schedule

Monday	Tuesday	Wednesday
Sep 9th **1** HW 1 assigned	10th **2** Lab 1 assigned	11th **3**
16th **4** HW 1 due HW 2 assigned	17th **5** Lab 1 due	18th **6**
23rd **7** HW 2 due HW 3 assigned	24th **8** Lab 2 assigned	25th **9**
30th **10** HW 3 due HW 4 assigned	Oct 1st **11** Lab 2 due Lab 3 assigned	2nd **12**
7th **13** HW 4 due HW 5 assigned	8th **14** Lab 3 due Lab 4 assigned	9th **15**
14th **16** HW 5 due **Exam 1** assigned	15th **17** Lab 4 due	16th **18**
21st **19** **Exam 1** due HW 6 assigned	22nd **20** Lab 6 assigned	23rd **21**
28th **22** HW 6 due HW 7 assigned	29th **23** Lab 6 due Lab 7 assigned	30th **24**
Nov 4th **25** HW 7 due HW 8 assigned	5th **26** Lab 7 due Lab 8 assigned	6th **27**
11th **28** HW 8 due HW 9 assigned	12th **29** Lab 8 due Lab 9 assigned	13th **30**
18th **31** HW 9 due **Exam 2** assigned	19th **32** Lab 9 due	20th **33**
25th Thanksgiving recess	26th Thanksgiving recess	27th Thanksgiving recess
Dec 2nd **34** **Exam 2** due	3rd **35**	4th **36** Project presentation due