STAT 0116A: Introduction to Statistical Science

Fall 2026

Instructor:	Christian Stratton	**Time:**	MW 11:15 – 12:30 (Class) T 11:15 – 12:30 (Lab)
Email:	cstratton@middlebury.edu	**Place:**	WNS 101 (Class) MNR 317 (Lab)
Office:	Warner 203	**Office hours:**	TBD Also by appointment

Course description: This course focuses on a practical introduction to statistical methods and analysis of data. We will use real data from the natural and social sciences to explore the concept of sampling variability, and understand why the presence of sampling variability demands that we introduce statistical methodology when answering questions from data. Topics include descriptive statistics, an introduction to probability, sampling distributions, hypothesis tests, and confidence intervals, with computing in R throughout.

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: Both class and lab 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. 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.
- Develop the ability to communicate statistical results orally and in writing.

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 two free online textbooks: *Intro to Modern Statistics* by Mine Çetinkaya-Rundel and Johanna Hardin and *Modern Dive* by Chester Ismay and Albert Kim. These books may be accessed via web browser at https://openintro-ims.netlify.app/ and https://moderndive.com/v2/index.html. The IMS book may also be 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 and Claude, 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 is permitted in this class and may be used on all assignments, **unless explicitly prohibited by the assignment.** 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?
- 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-paste from assignment*

Disclaimer: I am compelled to note that while generative AI can be a powerful tool, it is not infallible. 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. Additionally, in-person exams will be completed without the use of generative AI. Thus, these tools should be used to aid in your learning, and not used to replace your learning.

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 receive no credit.**

Course assessment: Your grade will be determined by in-class activities, readiness assessments, statistical reports, and oral exams. Each category is loosely defined as follows:

10%	**Practice problems**	Most weeks will feature a set of practice problems that must be completed by the following Tuesday. **Graded on effort**.
10%	**Readiness assessment**	Prior to most classes, a short assessment based on the assigned reading for the day must be completed on canvas. These assessments are designed to be very easy if you completed the reading. **Graded on accuracy**.
40%	**In-person exams**	There will be two in-person exams in this class: the midterm and the final. Exams will be proctored in the evenings and announced by the second week of class.
40%	**Statistical reports**	Throughout the semester, you will be asked to write a number of statistical reports. Full details will be provided as each report arises.

Letter grades: Letter grades will be assigned according to the following scale. Note that I may adjust thresholds at the end of the semester, but they will only ever be adjusted *down*.

F	**D**	**C-**	**C**	**C+**	**B-**	**B**	**B+**	**A-**	**A**
$[0, 60)$	$[60, 70)$	$[70, 74)$	$[74, 77)$	$[77, 80)$	$[80, 84)$	$[84, 87)$	$[87, 90)$	$[90, 94)$	$[94, 100)$

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).

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

Monday	Tuesday	Wednesday
Sep 14th **1** Supp - detecting AI	15th **2** Intro to R and RStudio	16th **3** Intro to data and study design
21st **4** Intro to the Tidyverse	22nd **5** Data visualization with ggplot2	23rd **6** dplyr and ggplot2 continued
28th **7** Supp - Probability I	29th **8** Supp - Probability II	30th **9** Supp - Probability III
Oct 5th **10** Normal random variables	6th **11** The Central Limit Theorem	7th **12** The CLT cont. and sampling distributions
12th **13** Intro to hypothesis testing	13th **14** Intro to confidence intervals	14th **15** Simulation-based inference for one proportion
19th **16** Theory-based inference for one proportion	20th **17** One proportion wrap-up	21st **18** Power and types of error
26th **19** Power and types of error cont.	27th **20** Relative risk	28th **21** Relative risk cont.
Nov 2nd **22** Intro to Chi-squared tests	3rd **23** Chi-squared tests cont.	4th **24** Simulation-based inference for multiple proportions
9th **25** Simulation-based inference for multiple proportions cont.	10th **26** Simulation-based inference for a single mean	11th **27** Theory-based inference for a single mean
16th **28** Simulation-based inference for two means	17th **29** Theory-based inference for two means	18th **30** Theory-based inference for two means cont.
23rd *Thanksgiving break*	24th *Thanksgiving break*	25th *Thanksgiving break*
30th **31** Intro to linear regression	Dec 1st **32** Correlation and R-squared	2nd **33** Assumptions of regression
7th **34** Theory-based inference for regression	8th **35** Simulation-based inference for regression	9th **36** MLR primer
14th *No class - hanging Friday*	15th Reading days	16th Reading days
21st Finals week	22nd **37**	23rd **38**