

APSTA-GE 2006 Fall 2025

Statistics, Math, and Computing Bootcamp

Instructor: eric.novik@nyu.edu

29 Jul 2025

1 Course description

This course aims to prepare students for the [Applied Statistics for Social Science Research](#) or an equivalent program at NYU. We will cover basic programming using the R language, including data manipulation and graphical displays; some key ideas from Calculus, including differentiation and integration, basic matrix algebra, including vector and matrix arithmetic; some core concepts in Probability, including random variables, discrete and continuous distributions, and expectations; and a few simple regression examples. We will also discuss the use of modern AI and LLMs in computational statistics.

2 Where and when

The course will run from August 18 through August 29 from 9 am to 12 pm. We will meet on weekdays from 9 AM to 12 PM at 7 East 12th Street, Room 329, and for those who can't attend in person on Zoom. Students are strongly encouraged to attend in person, as this course will be highly interactive.

3 Course prerequisites

The students should be fluent in basic algebra and have seen exponential and logarithmic functions. Some programming experience would be helpful but not required. Students should be prepared to write many small R programs during the course.

4 Course materials and references

Students are expected to have access to R and RStudio. Use [these](#) instructions on downloading and installing the required software. In addition, you may find the following free resources helpful as you continue your studies.

4.1 Programming and Data Visualization

- [Hands-On Programming with R](#), Grolemund (2014)
- [R for Data Science](#), Wickham, Çetinkaya-Rundel, and Grolemund (2023)
- [Data Visualization, A practical introduction](#), Healy (2018)

4.2 Calculus

- [YouTube: Essence of Calculus](#), Sanderson (2018a)
- [Calculus Made Easy](#), Thompson (1980)
- [Calculus](#), Herman, Strang, and OpenStax (2016)

4.3 Probability

- [YouTube: Probability Animations](#)
- [YouTube: Statistics 110 @ Harvard](#)
- [Introduction to Probability](#), Blitzstein and Hwang (2019)
- [Introduction to Probability Cheat sheet v2](#), Chen (2015)

4.4 Statistics

- [Regression and Other Stories](#), Gelman, Hill, and Vehtari (2020)

4.5 Linear Algebra (for those who need it)

- [YouTube: Essence of Linear Algebra](#), Sanderson (2018b)
- [Introduction to Linear Algebra](#), Boyd and Vandenberghe (2018)
- [Matrix Cookbook](#), Petersen and Pedersen (2012)

5 Tentative schedule

The following tentative schedule assumes ten 3-hour sessions, during which we plan to cover the following material.

- **(1) The Big Picture and Introduction to R**

We will briefly discuss the motivation behind the topics in the class and give some examples of the types of regression problems you may encounter. We will use Shuttle O-Rings damage and John Snow's (the other one) Cholera datasets as motivating examples. We will introduce programming principles in R, including RStudio IDE, R objects, built-in functions, and how to write your own functions.

- Hands-On Programming with R, [Part 1](#).
- R for Data Science, [Introduction](#).
- YouTube: Essence of Calculus, Videos [1](#), [2](#), [3](#).

- **(2) Plotting, Exponentials, Logs, and Derivatives**

We will review linear, exponential, and logarithmic functions and develop some intuition using compound interest. We will explain and code a softmax function using the log-sum-exp trick. Furthermore, we will introduce a derivative geometrically and symbolically. We will cover basic rules for differentiating functions and use noisy measurements of motion in a straight line as a motivating example. We will also practice plotting lines and curves using base plot and ggplot2.

- Hands-On Programming with R, [Part 2](#)
- R for Data Science, [Data visualization](#)
- YouTube: Essence of Calculus, Videos [4](#), [5](#), [6](#).

- **(3) Reshaping Data, Loops, and Maps; Introduction to Integration**

We will practice data wrangling using dplyr and tidyr packages. We will cover the intuition behind integration, approximating areas, the definite integral, and review the basic integration rules. We will learn how to do numerical and symbolic integration (in one dimension) using R.

- Hands-On Programming with R, [Part 3](#).
- R for Data Science, [Data transformations](#).
- YouTube: Essence of Calculus, Videos [7](#), [8](#), [9](#).

- **(4) Introduction to Probability 1**

We will review sample spaces, counting, and introduce the axiomatic definition of probability. We will discuss how to solve probability problems using simulations and will go through a few famous paradoxes. We will extensively use R's sample and replicate functions.

- R for Data Science, [Data Tidying](#)
- [Introduction to Probability](#), Chapter 1

– YouTube: Essence of Calculus, Videos [10](#), [11](#), [12](#).

- **(5) Introduction to Probability 2**

We will introduce conditional probability, the law of total probabilities, and independence and discuss a few famous paradoxes. We will write R simulations to check analytic solutions.

– [Introduction to Probability](#), Chapter 2

- **(6) Introduction to Probability 3**

We will introduce Random Variables, PDFs, and CDFs. We will cover Bernoulli, Binomial, Uniform, Normal, and Exponential RVs. We will discuss the concept of expectations. We will use R's distribution functions to simulate the realizations of RVs and compute their properties.

– [Introduction to Probability](#), Chapter 3

- **(7) Introduction to Linear Algebra**

We will review the basics of Linear Algebra, including vector and matrix addition and multiplication. We will practice matrix and vector arithmetic in R. Time permitting, we will solve the over-determined system of linear equations for the motion in a straight-line problem from Lecture 2.

– YouTube: Essence of Linear Algebra, Videos [1](#), [2](#), [3](#)

- **(8) Statistical Inference**

We will cover basic topics in Statistical Inference including sampling distributions, standard errors, classical confidence intervals, bias and uncertainty, and problems associated with unconscious use of statistical significance for making decisions.

– Regression and Other Stories, [Chapter 4](#)

- **(9) Analysis Workflow and Linear Regression**

We will introduce an analysis workflow using a dataset containing ratings of red wines. We will use the `rstanarm` package, but the same model can be done in with R's `lm()` function. We will attempt to determine what makes a good wine. After doing some basic EDA, we will fit several regressions and evaluate how well our model performs.

– [Introduction to Probability](#), Chapter 3

- **(10) Review and Discussion**

This is a buffer lecture where we will cover the overflow topics from the previous days. Time permitting, we will also quickly review the material, answer questions, and discuss the contents of the final assessment.

6 Assessment

This is a Pass/Fail class with the following grade distribution.

- Class participation: 50%
- Final assessment (take home): 50%

To earn class participation credit, students must attend all lectures *and* post in the online lecture discussions on Brightspace under the Online Discussion topic. For each lecture, please ask a question about the material, respond to a question asked by another student, or post an observation.

The final assessment will consist of approximately ten questions covering all major topics in the class. The final will be distributed during the last day of class.

7 Academic Integrity

All students are responsible for understanding and complying with the [NYU Steinhardt Policies and Academic Integrity](#).

8 Students with Disabilities Statement

Students with physical or learning disabilities are required to register with the Moses Center for Students with Disabilities at 726 Broadway, 2nd Floor, (212-998-4980) and are required to present a letter from the Center to the instructor at the start of the semester in order to be considered for appropriate accommodation.

9 Mental Health Statement

If you are experiencing undue personal and/or academic stress during the semester that may be interfering with your ability to perform academically, the NYU Wellness Exchange (212 443 9999) offers a range of services to assist and support you. I am available to speak with you about stresses related to your work in my course, and I can assist you in connecting with the Wellness Exchange. Additionally, if you anticipate any challenges with completing the assignments, readings, exams and other work required in this course, I encourage you to register with the Moses Center (212 998 4980) in advance so that you may be granted the proper academic accommodations.

References

- Blitzstein, Joseph K., and Jessica Hwang. 2019. *Introduction to Probability*. Second edition. Boca Raton: crc Press/Taylor & Francis Group.
- Boyd, Stephen P., and Lieven Vandenbergh. 2018. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge, UK ; New York, NY: Cambridge University Press.
- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Analytical Methods for Social Research. Cambridge, England: Cambridge University Press.
- Grolemund, Garrett. 2014. *Hands-on Programming with r*. First edition. Sebastopol, CA: O'Reilly. <https://rstudio-education.github.io/hopr/>.
- Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.
- Herman, Edwin, Gilbert Strang, and OpenStax. 2016. *Calculus Volume 1*. <https://d3bxy9euw4e147.cloudfront.net/oscms-prodcms/media/documents/CalculusVolume1-OP.pdf>.
- Petersen, Kaare Brandt, and Michael Syskind Pedersen. 2012. “The Matrix Cookbook.” <https://www.fretechbooks.com/the-matrix-cookbook-t435.html>.
- Sanderson, Grant. 2018a. “Essence of Calculus,” November. <https://www.youtube.com/playlist?list=PLZHQObOWTQDMsr9K-rj53DwVRMYO3t5Yr>.
- . 2018b. “Essence of Linear Algebra,” November. https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab.
- Thompson, Silvanus P. 1980. *Calculus Made Easy: Being a Very-Simplest Introduction to Those Beautiful Methods of Reckoning Which Are Generally Called by the Terrifying Names of the Differential Calculus and the Integral Calculus*. 3d ed. New York: St. Martin's Press.
- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. O'Reilly Media.