Syllabus

STAT 6730

**Introduction to Computational Statistics**

* Autumn 2025
* 2 Credit hours
* In-person (P) lectures: MW 1:50pm – 2:45pm in CH 228

## Course overview

### Instructor

* Vincent Q. Vu, Ph.D.
* Email: [vqv@stat.osu.edu](mailto:vqv@stat.osu.edu)
  + My preferred method of contact is the CarmenCanvas Inbox.
* Office hours on Zoom on Tuesdays TBD, except holidays
  + Mondays, 9:00am — 10:00am. See CarmenZoom for schedule and meeting links.

### Course Description

Computational statistics is an area within statistics that encompasses computational and graphical approaches to solving statistical problems. Students will learn how to manipulate data, design and perform simple Monte Carlo experiments, and be able to use resampling methods such as the Bootstrap. They will be introduced to technologies that are useful for statistical computing. Through creating customized graphical and numerical summaries students will be able to discuss the results obtained from their analyses.

### Course expected learning outcomes

By the end of this course, students should successfully be able to:

* Produce dynamic and reproducible reports with Quarto
* Visualize various types of data in R using the ggplot2 package
* Import, manipulate and summarize data in R
* Write and execute R functions that involve iterations or conditional statements
* Apply computational methods including Monte Carlo, smoothing and density estimation, the bootstrap and permutation methods

|  |
| --- |
| Generative AI tools and this course |
| Generative AI tools, such as ChatGPT, Gemini, and Claude, can be very powerful aids, but they are not a substitute for learning, and relying on them from the beginning can hinder your development. Importantly, these tools often produce incorrect or misleading information, and they can be overly confident in their answers. This is easy for skilled R coders and data scientists to spot, but it can be very difficult for beginners like you. So it is important that you learn how to do things yourself first, and then, after completing this course, you can use AI tools to help with tedious/non-creative aspects of your work once you become more proficient.  I do permit students to use AI tools to help with understanding concepts and debugging errors. For example, you can ask an AI tool about a specific function and its usage, or you use it help you debug your code and decipher error messages. Basically, you can use AI tools like an enhanced search engine.  **However, you are not permitted to use AI tools to write code or produce text for your assignments, except for the final project.**  If I suspect that you have used an AI tool on an assignment for this course, I will ask you to explain your process for completing the assignment in question. The unauthorized use of AI tools will result in referral to the [Committee on Academic Misconduct](https://oaa.osu.edu/academic-integrity-and-misconduct).  The assignment and academic policies sections of the syllabus has more details on the use of AI tools in specific assignments and policies. |

## Prerequisites

This course is intended to be taking during the second year of the MAS program. It is expected that students will have exposure at a mathematical level to foundational concepts in probability and statistics including random variables, estimation, hypothesis testing, and linear regression. The formal prerequisites for this course are: STAT 6301 and 6302 or equivalent; STAT 6410 and 6450, or STAT 6910 and 6950, or permission of the instruction. Previous programming experience is not required, but familiarity with computer systems is expected.

## Course materials and technologies

### Textbooks

#### Required

* [**R4DS2E**] Wickham, Çetinkaya-Rundel, and Grolemund (2023): *R for Data Science*, 2nd Edition. Electronic version: <r4ds.hadley.nz>. This web version of the book can be accessed freely from any web browser.
* [**ADVR**] Wickham (2019): *Advanced R*, 2nd Edition. Electronic version: <adv-r.hadley.nz>. This web version of the book can be accessed freely from any web browser.

### Course technology

#### Required equipment

* Computer: current Mac (macOS) or PC (Windows or Linux), capable of running R and RStudio (described below).
* You will need to bring a laptop to class on the days of labs and exams. If you do not have a laptop, please request a Microsoft Surface loaner device from the [Student Technology Loan Program](https://it.osu.edu/student-technology-loan-program).

#### Required software

You will need to install or have access to the **latest** versions of the following software. Even if you have installed these programs before, it is a good idea to check for updates. I will not provide support for outdated software.

* **R** <https://cloud.r-project.org>
* **RStudio** Desktop IDE <https://posit.co/download/rstudio-desktop>
* **Quarto** <https://quarto.org/docs/get-started>

|  |
| --- |
| Important |
| Note that as of August 2025, Windows ARM-based PCs, e.g. Microsoft Surface, are not fully compatible with Quarto, and users may encounter issues rendering Quarto documents when using these systems. If you are using a Windows ARM-based PC, please contact me as soon as possible so that I can provide an alternative, cloud-based solution for rendering Quarto documents. |

## Grading and instructor response

### How your grade is calculated

| Assignment Category | Points and/or Percentage |
| --- | --- |
| Homework | 15% |
| Labs | 15% |
| Exam 1 | 20% |
| Exam 2 | 20% |
| Final Project | 30% |
| **Total** | **100%** |

### Description of major course assignments

#### Homework

* Description
* Homeworks will generally be assigned on a biweekly basis. Students are required to use Quarto for their homework submissions. The homeworks should be written in a style that smoothly integrates prose, code, tables, and graphics. It should be human-readable, and it should not simply contain computer output with no explanation. Submit both the Quarto document (.qmd) and HTML output (.html) to CarmenCanvas.
* Academic integrity and collaboration guidelines
* You may collaborate with classmates on your homework, but ultimately the code that you write and submission that you make must be your own work. For example, I encourage you to discuss strategies for solving problems, but the actual code and explanations that you write must be your own. Moreover, keep in mind the university policies on plagiarism. Do not copy or plagiarize anything you may find on the Internet or anything produced by an AI tool.
* Permitted uses of AI tools
* You may use generative AI tools, e.g. ChatGPT, Claude, Gemini, etc… to help you understand concepts, like. For example, you can ask an AI tool about a specific function and its usage. However, you are not permitted to use of AI tools to write code or produce text for you. In particular, you should not use AI tools to write code for your homework assignments.
* Required disclosures
* Any collaboration or use of AI tools must be acknowledged in the disclosures section of your homework submission.

#### Labs

* Description
* Labs will consist of exercises to be completed during Wednesday meetings. Please bring a laptop computer with R and RStudio installed to those meetings. As with homework, students are required to use Quarto and the lab report should be written in a style that smoothly integrates prose, code, tables, and graphics. Labs are due within 24 hours of the end of the class meeting and should be submitted as Quarto document (‘.qmd’) to CarmenCanvas.
* Academic integrity and collaboration guidelines
* Lab assignments will be collaborative, however each student must make their own individual submission. These assignments will be graded based on completion rather than correctness.
* Permitted uses of AI tools
* You may use generative AI tools, e.g. ChatGPT, Claude, Gemini, etc… to help you understand concepts, like. For example, you can ask an AI tool about a specific function and its usage. However, you are not permitted to use of AI tools to write code or produce text for you.

#### Exams

* Description
* There will be two exams. Both exams are open book/internet access, but absolutely no communicating with other humans. Each exam is cumulative.
* Academic integrity and collaboration guidelines
* The exam is an individual assignment. You should complete the exam on your own and your submission should be your own original work. You should not discuss it with anyone else.
* Permitted uses of AI tools
* The use of AI tools is absolutely not permitted for the exams.

#### Final Project

* Description
* Students will work on a final data analysis project. The instructor will provide a list of topics. Each group will cooperate on the data analysis, report writing, and making a presentation on the project in class. Details will be announced on CarmenCanvas and during one of the lectures.
* Permitted uses of AI tools
* You may use generative AI tools, e.g. ChatGPT, Claude, Gemini, etc… to help with all aspects of the project, however you agree that your total input (tokens in the prompt) to total output (tokens in the output) ratio that you use in producing your project should be greater than 1:1.

|  |
| --- |
| Late assignments |
| Late homework assignments will automatically receive a 10 percentage point deduction each day that they are late. After the 5th day submissions will no longer be accepted. All other assignments (discussion and final exam…) will not be accepted after the due date. Please pay attention to the exact date and time that an assignment is due. For example, if an assignment is due at 11:59:00 PM, and you submit it at 10 seconds after at 11:59:10 PM, then it will be considered late.  If you absolutely need to turn in an assignment late and have a valid excuse, please contact me for the necessary arrangements. However, you must notify me **in advance** in such a situation. Exceptions to this policy will be permitted only in extreme situations such as serious injury immediately prior to an assignment being due or severe illness requiring hospitalization. |

### Grading Scale

* 93-100: A
* 90-92: A-
* 87-89: B+
* 83-86: B
* 80-82: B-
* 77-79: C+
* 73-76: C
* 70-72: C-
* 67-69: D+
* 60-66: D
* Under 60: E

|  |
| --- |
| Instructor feedback and response time |
| Please use the discussion board in Carmen for questions about the course material and assignments. If you have a question that is personal or that you would like to discuss privately, please email me. I will respond to questions posted on the discussion board or by email within 24 hours (except for weekends and university holidays). If you would like to meet with me over Zoom, please message me to set up an appointment. |

#### Grading and feedback

For weekly assignments, you can generally expect feedback and grades within 10 days.

## Academic policies

The following are important academic policies to be aware of.

* Academic Misconduct; Artificial Intelligence and Academic Integrity; Religious Accommodations; Disability Statement (with Accommodations for Illness); Intellectual Diversity; Grievances and Solving Problems; Creating an Environment Free from Harassment, Discrimination, and Sexual Misconduct.
  + [Policy statement](https://ugeducation.osu.edu/academics/standard-syllabus/standard-syllabus-statements)
* Copyright; Counseling and Consultation Services / Mental Health Statement; Content Warning Language
  + [Policy statement](https://ugeducation.osu.edu/academics/standard-syllabus/optional-syllabus-statements)

## Course Schedule

Refer to our Carmen course page for up-to-date assignment due dates.

| Week | Date | Topics/Readings/Assignments |
| --- | --- | --- |
| 1 | 8/27 | Overview; Introduction to R |
| 1 | 8/27 | Reading: *R4DS2E 2* |
| 2 | 9/1 | **No class**: *Labor Day* |
| 2 | 9/3 | Data types in R |
| 2 | 9/3 | Reading: *ADVR 3.2* |
| 3 | 9/8 | Lists and data frames in R |
| 3 | 9/8 | Reading: *ADVR 3.5–3.6* |
| 3 | 9/10 | Data manipulation and summarization with base R and dplyr |
| 3 | 9/10 | Reading: *ADVR 4* |
| 4 | 9/15 | Grammar of graphics with ggplot2 |
| 4 | 9/15 | Reading: *R4DS2E 1, 9* |
| 4 | 9/17 | Lab 1: ggplot2 |
| 4 | 9/17 | **Homework 1** due |
| 5 | 9/22 | Smoothing |
| 5 | 9/24 | Lab 2: Smoothing |
| 6 | 9/29 | Functions |
| 6 | 9/29 | Reading: *ADVR 6* |
| 6 | 10/1 | Lab 3: Functions |
| 6 | 10/1 | **Homework 2** due |
| 7 | 10/6 | Density estimation |
| 7 | 10/8 | **Exam 1** |
| 8 | 10/13 | Functions |
| 8 | 10/13 | Generating random variables |
| 8 | 10/15 | Lab 4: Conditionals and iteration |
| 8 | 10/15 | Reading: *ADVR 5* |
| 9 | 10/20 | Monte Carlo integration |
| 9 | 10/22 | Lab 5: Monte Carlo |
| 9 | 10/22 | **Homework 3** due |
| 10 | 10/27 | The Bootstrap |
| 10 | 10/29 | Lab 6: The Bootstrap |
| 11 | 11/3 | Functional programming with purrr |
| 11 | 11/5 | Lab 7: Functional programming with purrr |
| 11 | 11/5 | **Homework 4** due |
| 12 | 11/10 | Pipes and dplyr |
| 12 | 11/10 | Reading: *R4DS2E 4.3* |
| 12 | 11/12 | **Exam 2** |
| 13 | 11/17 | Permutation tests |
| 13 | 11/19 | Lab 8: Permutation tests |
| 14 | 11/24 | rsample |
| 14 | 11/26 | **No class**: *Thanksgiving* |
| 15 | 12/1 | Cross-validation |
| 15 | 12/3 | Lab 9: Cross-validation |
| 15 | 12/3 | **Homework 5** due |
| 16 | 12/8 | Final project presentations |
| 16 | 12/10 | Final project presentations |