DS 5740 | Advanced Statistics for Data Scientists

Fall 2023

Updated 08.21.2023

This is a *living* document. Updates may occur during the semester.
You will be notified of any changes and receive an amended syllabus when changes occur.

Jump to: Course Schedule Grading Grade Appeal Policy Terms and Agreement

I. Course Information

Class Times: Monday and Wednesday, 3:45pm-5:00pm

Classroom: 19th and Grand, Room 168

Instructor: Alexander Christensen (he/him/his), alexander.christensen@vanderbilt.edu

Office Hours: by appointment through Calendly

TA: Danni Shi, danni.shi@vanderbilt.edu

II. Communication Expectations: The instructor and TA will strive to acknowledge emails within 24 hours of receipt. Some emails may require thoughtful responses and therefore are not always feasible to respond to in 24 hours. Slack is available to discuss more immediate questions. Even if Slack status appears as "online" for the instructor or TA, we may not be available or able to give thoughtful feedback to your request. Whether over email or Slack, an immediate response should not be expected, but a timely response is the goal. Your TA may cc the instructor on any emails they feel necessary.

III. Course Description: Predicting future outcomes is fundamental to human activity. Understanding how past events lead to current conditions can enables us to make predictions about the future. This course will focus on how we can predictions about the future based on past evidence. Applied uses include forecasting market outcomes (e.g., stocks, company earnings) to human behavior (e.g., emotions, public health policy).

Top

The majority of the course covers time series forecasting to use past performance patterns to make predictions about the future. Key topics in this component includes regression, decomposition, smoothing, ARIMA models, and dynamic covariation. A second component will focus on applying dimension reduction techniques such as exploratory graph analysis (EGA) to survey data. A third component of this course will focus on A/B testing (also referred to as experimental design). In this component, we'll cover how we can manipulate different features of the stimulus/environment to uncover predictions about behavior.

This course will use R statistical software and packages. All examples in the course will be provided through R code and all modeling can be performed using open-source R packages. Accompanying books are free and based in R. We will not cover all of the material in any one book, but they are valuable resources for you to hone your skills and continue your career path.

Although this course primarily uses R, you are welcome to code in your language of choice. So long as you can complete the analyses successfully and demonstrate mastery of the concepts, then the course objectives are fulfilled. You are also encouraged to use generative AI – my perspective is that what is learned in this course is only the *start* of your learning and generative AI is a tool that can take your learning to the next level.

IV. Prerequisites: DS 5620

V. Course Objectives:

- Understand how forecasting can be applied and what model is most appropriate
- Identify when and how to apply data reduction and clustering techniques
- Know how to construct experimental designs that can identify variables that affect outcomes

VI. Tentative Course Schedule

This is your one-stop guide to this course.

Content, assignments, and due dates are subject to change based on progress through the

Top

materials. We will cover, if only briefly, all content but time constraints may limit coverage of certain topics in favor of other topics.

VII. Course Materials

All books necessary for the course are provided in the schedule (see VI. Tentative Course Schedule).

VIII. Grading

There will be 12 assignments throughout the semester but only your 10 best grades will count toward your final grade. There are 2 half-semester projects that will *both* count toward your final grade.

Assignments: 10 assignments (6 points each; 60% of your final grade)
Projects: 2 projects (20 points each; 40% of your final grade)

Grading Scale (rounded to the nearest percentage):

A	94-100%	B	83-86%	C	73-76%	D	63-66%
A-	90-93%	B-	80-82%	C-	70-72%	D-	60-62%
B+	87-89%	C+	77-79%	D+	67-69%	F	<60%

Late Work Policy

Assignments turned in a second after 5pm on the Sunday an assignment was due (i.e., 5:00:01pm according to Brightspace) will be considered late. **NO exceptions will be made.** Late work will be worth 80% of your overall grade for the assignment. You will have **one week** to turn in your assignment late after its due date (i.e., until 5:00:01pm, according to Brightspace, on the next Sunday).

Brightspace and Technology Errors

If working up to the last minute of a deadline, **submit** whatever you have to Brightspace at least **fifteen** minutes before the deadline. Technological errors happen. If Brightspace is not cooperating, send me an email (alexander.christensen@vanderbilt.edu) with an image depicting

Top

the issue and **include a timestamp**. Claims of technological errors without substantial evidence will not be considered. If a technological error happens more than one, then we will set up a one-on-one meeting to determine the source of the error.

IX. Grade Appeal Policy

There will be no grade haggling at the end of the semester. Any grade appeals are subject to the conditions below. General, non-specific grade haggling will not be entertained, and I will send an email directly linked to this portion of the syllabus.

The instructor makes all final decisions on grades. The instructor is willing to revisit a grade only in the instance of a **suspected error**. Do **not** request an appeal based on a disagreement with the instructor's judgment. If you disagree with a grade that you've received on an assignment or project, then you are welcome to formally submit an appeal to Dr. Christensen. The appeal should be submitted via email (alexander.christensen@vanderbilt.edu) and must include: (1) The subject header "Formal Grade Appeal", (2) clear and specific reference to the part of the assignment or project in question, and (3) justification for why more credit is earned, citing specific material or evidence. The appeal should be no later than **7 days** after the grade has been posted on Brightspace. If, after the appeal you still have concerns, you should raise them with the DGS (Dr. Blocher; jesse.blocher@vanderbilt.edu or Associate DGS Dr. Kang; h.kang@vumc.org).

X. Vanderbilt Honor Code

In accordance with the undergraduate Vanderbilt University Honor Code:

I pledge to pursue all academic endeavors with honor and integrity. I understand the principles of the Honor System, and I promise to uphold these standards by adhering to the Honor Code in order to preserve the integrity of Vanderbilt University and its individual members.

In addition, I abide by responsibility detailed in the Vanderbilt's Faculty Guide to the Honor System. In general, this course is intended to be collaborative and therefore students may and should work together on assignments and projects. Each assignment and project, however, must be a student's own work and submitted separately.

Top

By taking and continuing to be enrolled in this course, you are agreeing to "pledge on your honor that you have neither given nor received unauthorized aid on the assignments and final project."

Transformers and Large Language Models

Data science is at the forefront of the AI revolution. As teachers, especially of data science, we cannot ignore the impact of the latest models called *transformers*. As emerging data scientists, it is *expected* that you will use generative AI on your homework and projects. That said, I encourage you to share when you've used AI to generate code, perform analyses, or understand material. The goal is transparency. The more the world is filled with AI content, the more transparency will be valued. Share links to output or .pdf documents (if possible).

XI. Classroom Accommodations

Vanderbilt University and Dr. Christensen are committed to equal opportunity for students with disabilities. If you need course accommodations due to a disability, please contact VU Student Access Services to initiate the process. After SAS has notified me of relevant accommodations, we will discuss how these accommodations may best be approached in this class, and I will facilitate the accommodations.

If emergencies or extenuating circumstances keep you from class, please get notes and announcements from a classmate. You're welcome to arrange a meeting with Dr. Christensen or the teaching assistant to ask questions about the missed material (also check the class's Brightspace).

XII. Mental Health & Wellness

If you are experiencing undue stress that may be interfering with your ability to perform academically, Vanderbilt's Student Care Network offers a range of support services. The Office of Student Care Coordination (OSCC) is the central and first point of contact to help you navigate and connect to appropriate resources. You can schedule an appointment with the OSCC at https://www.vanderbilt.edu/carecoordination/ or call 615-343-WELL. You can find a calendar

Top

of services at https://www.vanderbilt.edu/studentcarenetwork/satellite-services/. If you or someone you know needs to speak with a professional counselor immediately, the University Counseling Center offers Urgent Care Counseling. Students should call the UCC at (615) 322-2571 during office hours to speak with an urgent care clinician. You can also reach an on-call counselor after hours or on the weekends by calling (615) 322-2571 and pressing option 2 at any time. You can find additional information at https://www.vanderbilt.edu//ucc/.

XIII. Names and Pronouns

If you would like to use a different name or different pronouns than those provided through YES, please let me know at any time prior to or during the semester. Information is available through the LGBTQI Life offices about how to change either or both in YES.

XIV. Religious Holidays

If you will miss class to observe a religious holiday, please email Dr. Christensen in advance of your planned absence. You may make up (without penalty) any work missed due to the observance of a religious holiday. Please let Dr. Christensen know two weeks prior to the observance so he can make proper arrangements for you. Dr. Christensen has done his best to organize his class schedule around observances outlined here but please let him know if any observances are not listed that apply to you.

XV. Mandatory Reporter Obligations

All University faculty and administrators are mandatory reporters. What this means is that all faculty, including me, must report allegations of sexual misconduct and intimate partner violence to the Title IX Coordinator. In addition, all faculty are obligated to report any allegations of discrimination. I am willing to discuss with you such incidents but can only do so in the context of us both understanding my reporting obligations. If you want to talk with someone in confidence, officials in the Student Health Center, the University Counseling Center, and the Office of the Chaplain and Religious Life (when acting as clergy) can maintain confidentiality. In addition, officials in the Project Safe Center have limited confidentiality, in that they must report the incidents but can do so without providing identifying information. The Project Safe Center serves as the central resource for those impacted by sexual misconduct and intimate

Top

partner violence and can assist with navigating all facets of the University's resource and support network and other processes.

XVI. Class Technology Policy

Please turn the volume off of ALL electronic devices—cell phones, laptops, tablets, etc.—before you come to class. It can be quite distracting to fellow students and to your instructor if your device starts to play music in the middle of class. A computer is a fundamental part of the course and will be required as part of in-class code examples.

XVII. In-class Policy

In alignment with Vanderbilt's academic mission, on-campus academic programs were designed for in-person instruction because it offers an important opportunity for learning, leadership, and the scholarly exchange of ideas. As such, university policy generally requires in-person teaching for undergraduate, professional, and graduate courses (excluding our online programs designed for distance learning). Dr. Christensen will not offer remote or online options unless otherwise mandated by Vanderbilt.

XVIII. Terms and Agreement

This syllabus is a contract between you, me, and the rest of the class. I acknowledge and agree to all terms written in this document. By taking this course and continuing to be enrolled in this course, you also agree to all terms written in this document. Failure to read this document does not grant privilege to ignore what is written in it.

XIX. Resources and Data

Resources

Vanderbilt Library (free): access to R books, R resources, and free R workshops

Posit (RStudio) primer courses (free): basics of {tidyverse}

DataCamp (costs money): many, many R courses in data science

The Software Carpentries (free): courses on programming with R and R for reproducible scientific analysis

Top

R ladies in Nashville offers meetup opportunities to promote gender diversity in the R community

Data

FRED datasets: https://fred.stlouisfed.org/

Kaggle datasets: https://www.kaggle.com/datasets

Over 1000 datasets: https://vincentarelbundock.github.io/Rdatasets/datasets.html

Open Psychometrics: https://openpsychometrics.org/_rawdata/

Journal of Open Psychology Data: https://openpsychologydata.metajnl.com/

Open Science Framework: https://osf.io/search/

UCI Machine Learning Repository: https://archive-beta.ics.uci.edu/ml/datasets

Top