

Stats 506: Computational Methods and Tools in Statistics

- Location: [Weiser 296](#)
- Time: Tuesday/Thursday, 10-11:30 am
- Website: <https://dept.stat.lsa.umich.edu/~jerrick/courses/stat506>, Canvas
- [Course calendar](#) can be subscribed to for lectures, midterm, and office hours, including any adjustments to the regular schedule.

Contact Information

- Instructor: Josh Errickson (jerrick@umich.edu)
- Instructor office hours: Mon 10am-12pm, Fri 1pm-2pm, starting 9/1.
 - Office hours will occasionally be rescheduled. Please check [Course calendar](#) for up-to-date information.
- Instructor office location: 3560 Rackham (Enter the CSCAR suite, ask for my office)
 - All office hours will be hybrid: <https://umich.zoom.us/j/95430614786>
 - * Priority will be given to in-office students, but I'll try and accomodate remote as much as I can.
- GSI: Mengqi Lin (lemonkey@umich.edu)
- GSI office hours: Wed 3-5pm (Zoom only), Thurs 4-5pm (in-person only)
- GSI office location: G219 Angell Hall, <https://umich.zoom.us/j/97508222458>

Key Dates

- Midterm Exam: Oct 26
- Final Project
 - Assigned: Oct 31
 - Proposal due: Nov 9

– Report due: Dec 8

- Final Class: Dec 5
- No Class: Oct 17, Nov 23
- [Office of the Registrar](#)

If you have a conflict with the midterm date or either of the project deadlines, or a conflict with religious observation, please let me know prior to October 1 to work out alternative arrangements.

Grading information

- Problem sets (6, roughly every 2 weeks): 42%
- Midterm Exam: 29%
- Final Project: 29%

Problem sets may have different total point values, but each will account for 7% of your final grade. Problem Sets will be due *before* the start of class on the Tuesdays they are due (every 2 weeks). To accommodate unexpected circumstances you may turn in problem sets *before* the start of the class on the following Thursday (up to 48 hours late) for a 10% penalty on that set.

The project proposal is *ungraded* however, must be submitted so I can approve it. It is due *before* the start of class on the date indicated above. **Any** late submission of the project proposal incurs 2% penalty on your final project grade.

The final project will not be accepted late. It is due *at noon* on the date indicated above.

Extraordinary situations may warrant relaxing the dates/penalties, if you feel you have an extraordinary situation, please reach out to me as soon as you are able.

Final Project

For the final project, I will provide a few options - one (or more) will be data sets from which you will generate a hypothesis and produce a report addressing it; one (or more) will be more focused on examining a particular statistical methodology.

The Proposal will be a statement of your research question and a brief (1/2 page) description of your analysis plan. This will not be graded, but will need to be approved by me. If necessary, we may iterate on the proposal.

The final report should be approximately 2 pages (no more than 3 pages, excluding graphs) describing your results.

Full details for the project will be provided after the midterm.

Piazza

We'll use Piazza for course discussions. I encourage everyone to post questions there so your classmates may be able to help you, and everyone will benefit from the answers.

Attribution of sources

Addressing academic integrity for code submissions has challenges that written assignments don't - namely that there may be a limited, or even single, correct way to write code, whereas there is more room for individuality in writing.

In addition, in a professional or academic setting, code-writing is rarely a solitary activity. Most professionals or academics have colleagues around to discuss code with, and of course the vast resources available online.

It would be unrealistic of me to insist that you do not take advantage of these resources when doing assignments for this class, however, I will introduce the following rules:

1. Taking complete code from another source (whether it is another student or person, or a resource online) in its entirety is not allowed. It is fine to use other resources to generate ideas of how to approach a problem, but the code you write must be your own.
2. If you substantially* collaborate with another student, you should both indicate on the submission the name(s) of the other people you collaborated with. **Note that this is not endorsing treating the problem sets as group projects**, you must still each do your own work in entirety.
3. Any online resources (excluding the ones I provide in the course) from which you gain substantial* benefit from should be cited in your submission, with a brief statement of what you gained from the resource. (E.g., "xyz.com: helped me figure out a more efficient way to code my loop for problem 2 by doing ZZZ and QQQ.")
4. Be prepared to explain your code as a way to prove to me that you wrote it - if any questions arise, I may schedule a conference where you talk me through your submission to show me that you understand what you wrote.

* What is substantial? I'm leaving it vague, but better to be safe than sorry. Googling "what's the name of the function to do XXX" isn't substantial; asking someone to explain how their entire code works is - the line between insubstantial and substantial is somewhere in the middle there.

Text books

There are no required texts for this course. There are three recommend texts.

The Art of R Programming, by Norman Matloff, is recommended for those with little to no previous experience in R.

[Advanced R](#), by Hadley Wickham, is recommended for those who would like to develop a deep understanding of R and its inner workings.

[R for Data Science](#), by Hadley Wickham and Garrett Grolemund, is a helpful bridge between these two.

Pre-requisites

You should have taken or be currently enrolled in an intermediate applied statistics course such as Stats 500, and you should have some familiarity programming in at least one programming language or scripting in a statistical software. If you have no experience with any programming or scripting, please speak to me ASAP.

Course Description

Stats 506 covers a variety of topics related to the use of computers for analyzing, managing, and presenting data. The goals of the course are for students to:

1. gain fluency with common computing tools, methods, and concepts used by data scientists and statistical analysts;
2. develop good habits for coding, documentation, and workflow;
3. demonstrate growth in presenting and communicating data and analysis.

The first half of the course will represent a survey of the statistical software landscape. The midterm will (ideally) cover that material. The second half of the class will be devoted solely to R, building up to the final project.

The topics below represent an approximate plan for the course. *I reserve the right to modify this plan.*

- R: vectors, arrays, objects, I/O, functions, programming.
- git & version control
- RMarkdown & Quarto
- R: style guidelines
- R: vectorization, Monte Carlo studies
- Stata: syntax, I/O, data management, aggregations by group
- Stata: regression, macros, iteration
- SAS: syntax, I/O, data management, aggregations by group
- SAS: regression, macros
- SQL
- Regular expressions

- **Midterm**
- Other Statistical software we didn't cover
- R: the tidyverse, emphasizing the dplyr and tidyr packages
- R: graphics using ggplot2
- R: the bootstrap, permutation testing, cross validation
- R: understanding and managing memory utilization
- R: R's object oriented systems
- R: parallel computing and asynchronous computing with futures
- Batch computing in an HPC environment
- **Final Project**

Ideally, the midterm exam will cover through "Regular expressions".

Computing Resources

All of the software in this course is available without charge for UM students. Some software, such as R, is free and open source and can be installed on your personal machine. You should install both R and RStudio. Links to install both can be found at <https://posit.co/download/rstudio-desktop/>. If you installed R at a much earlier date, please make sure you versions are up-to-date. As of the last revision of this document, R's current version is 4.3.1.

Details about accessing SAS and Stata can be found here:

- SAS: <https://software.umich.edu/titles/sas-analytical-suite>
- Stata: <https://software.umich.edu/titles/statase>

It is **not** necessary to purchase your own licenses for these, but you can if you wish of course.

SAS also offers a free [SAS OnDemand for Academics](#). It's a web-based approach but is very useful for shorter SAS sessions with small data. It *should* be sufficient for everything we do during this class.

Accommodations for students with disabilities

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, such as assignments, in-class activities, and recorded lectures may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; <http://ssd.umich.edu>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

Academic Integrity

The University of Michigan community functions best when its members treat one another with honesty, fairness, respect, and trust. The college promotes the assumption of personal responsibility and integrity, and prohibits all forms of academic dishonesty and misconduct. All cases of academic misconduct will be referred to the Office of the Assistant Dean for Undergraduate Education. Being found responsible for academic misconduct will usually result in a grade sanction, in addition to any sanction from the college. For more information, including examples of behaviors that are considered academic misconduct and potential sanctions, please see <https://lsa.umich.edu/lsa/academics/academic-integrity.html>.

Mental Health and Well-being

The University of Michigan is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at 734.764.8312 and caps.umich.edu during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at 734.764.8320 and <https://www.uhs.umich.edu/mentalhealthsvcs>, or for alcohol or drug concerns, see <https://www.uhs.umich.edu/aodresources>.

Mandatory Reporting and Sexual Misconduct

Title IX prohibits sex discrimination to include sexual misconduct: harassment, domestic and dating violence, sexual assault, and stalking. If you or someone you know has been harassed or assaulted, you can receive confidential support and academic advocacy at the Sexual Assault Prevention and Awareness Center (SAPAC). SAPAC can be contacted on their 24-hour crisis line, 734-936-3333 and online at sapac.umich.edu. Alleged violations can be reported non-confidentially to the Office for Institutional Equity (OIE) at institutional.equity@umich.edu. Reports to law enforcement can be made to University of Michigan Police Department at 734-763-3434.

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a mandatory reporting responsibility. I am required to share information regarding sexual misconduct or information about a crime that may have occurred on U-M's campus with the University. Students may speak to someone confidentially by contacting SAPAC's Crisis Line at (734) 936-3333.