George Mason University

STAT 778– Statistical Computing

Spring 2024

Instructor: David Kepplinger

Version 1 (01/09/2024)

Administrative

Course dates:

Wednesdays, Jan 17 – Apr 3 (final exam period: May 1);
no class on Mar 6 (Spring recess)

Important dates:

Jan 23: last day to add
Jan 30: final drop deadline (no tuition penalty)
Feb 20: end of self-withdrawal period

Instructor: Dr. David Kepplinger (he/him/his)

Email: dkepplin@gmu.edu

Office: Room 1711, Nguyen Engineering Building (ENGR)

Office hours:

Monday, 3 – 4pm in-person in ENGR 1711
Virtual office hours over Zoom by appointment (Zoom link)

Blackboard course page: https://mymasonportal.gmu.edu/ultra/courses/_511079_1/

Class time: Wed, 7:20 – 10:00pm in Innovation Hall 319

The class is scheduled for face-to-face on-campus meetings. All learners taking courses with a face-to-face component are required to follow the university's public health and safety precautions

and procedures outlined on the university Safe Return to Campus webpage (https://www2.gmu.edu/safe-return-campus). If the campus closes, or if a class meeting needs to be canceled or adjusted due to weather or other concern, learners should check Blackboard for updates on how to continue learning and for information about any changes to events or assignments.

Communications The Blackboard site for this course is the primary channel of communication. Please check the Blackboard course regularly for updates! Information posted on the Blackboard site includes

- announcements,
- lecture notes,
- homeworks and assignments,
- changes to the posted office hours,
- handouts and readings.

Any question related to concepts and topics should be asked on the Blackboard discussion board (under *Discussion Board* > *Course Q&A)*. Questions will be visible to all registered students, and everyone should actively participate in answering questions posted by peers. Active participation in answering questions will be counted towards the participation grade.

E-mail communication must be restricted to questions relating to sensitive and confidential information (such as grade concerns, personal circumstances requiring specific accommodations, etc.).

- E-mails will be returned within 2 business days and may not be returned on weekends/holidays.
- When you send an e-mail to me, please put `STAT 778` at the beginning of the subject line.
- E-mails related to this course must be sent and received via your Mason e-mail account. **E-mails sent from other e-mail accounts may not be answered.** (This is a university policy and part of your guaranteed rights under FERPA.)
- E-mails with questions that should be posted to our course Q&A may not be answered.

Should you have concerns that you may not be able to fully participate or engage in any of the activities listed below, please do not hesitate to contact me either by e-mail or speak to me in person during office hours or after class. We can discuss alternative arrangements that suit your needs.

Course requirements

Prerequisites: STAT 652 and STAT 672 or equivalent; you must be comfortable with reading and writing R code and have working knowledge of functional programming.

Recommended readings: The main textbooks for this course are

- Gentle, & Härdle, W. K., & Mori Y. (2012). *Handbook of Computational Statistics (2nd edition).* Springer. Available online.
- Lange. (2010). *Numerical Analysis for Statisticians.* Springer New York. doi:10.1007/978-1-4419-5945-4. Available online.
- Eubank, & Kupresanin, A. (2011). *Statistical Computing in C++ and R (1st edition).* Chapman and Hall/CRC. doi:10.1201/b11538. Available online.
- Wickham. (2014). *Advanced R.* Chapman and Hall/CRC. doi:10.1201/b17487. Available online.
- Givens, & Hoeting, J. A. (2013). *Computational Statistics (2nd edition).* Wiley. Available online.

The following reference books are also helpful resources for assignments and class discussions:

- Phillips. (2018). *Python 3 Object-Oriented Programming (3rd edition).* Packt Publishing. Available online.
- Stroustrup. (2014). *Programming: Principles and Practice Using C++ (2nd Edition).* Addison-Wesley Professional. Available online.
- Meyers. (2014). *Effective Modern C++: 42 Specific Ways to Improve Your Use of C++11 and C++14.* O'Reilly. Available online.
- Eddelbuettel. (2013). *Seamless R and C++ Integration with Rcpp.* Springer New York. Available online.

A number of relevant articles will be posted in Blackboard as different topics are discussed.

Hardware requirements: We will frequently use laptop computers for in-class activities. Please be respectful of your peers and your instructor and do not engage in activities that are unrelated to the class.

Software requirements: This class will use R (version 4.3 or higher), Python (version 3.8 or newer) and C++ (standard C++17) for illustration and implementation of topics discussed in class. For your assignments you are allowed to use other programming or scripting languages, but you must seek advance permission from the instructor. You must always submit the complete code for reproducibility.

You will also be required to have a Git client on your computer. This course will be taught using SourceTree (https://www.sourcetreeapp.com/) available free as download. Alternatively, you may choose to use a different interface for Git, such as GitHub Desktop, the UI built into RStudio Desktop, or the git command line interface.

In case of campus closures, activities and assignments in this course will use the web-conferencing software Zoom. In addition to the requirements above, you are required to have a device with a functional camera and microphone.

Course description

Learning objectives: After successfully completing STAT 778, you will be able to:

- Conduct efficient stochastic simulations and implementing appropriate methods for inference.
- Understand the sources of errors in statistical computations.
- Approximately compute statistics and other quantities, and quantify the approximation and statistical errors.
- Use parallelization and high-performance clusters for stochastic simulations.
- Adapt and implement resampling methods for statistical inference.

Main topics: You can expect the following topics to be covered in some detail.

- Overview of scripting/programming languages commonly used for statistical analyses.
- Errors due to numerical and computational approximations; computer arithmetic.
- Stochastic simulation (random number generation, Monte Carlo methods, resampling techniques).
- Parallel computing techniques relevant for statistical analyses.
- Numerical integration, optimization and root finding.

Assessments and grading

Your grade in this course will be based on (roughly) bi-weekly homework assignments of various types, a take-home midterm, a written final project report, two paper presentations, two presentation notes, two peer reviews of presentation notes, and in-class participation (including weekly oral quizzes about readings). There is no sit-in final exam, but you will be submitting a final project report due at the final exam period.

Assignment	Tentative due date	Weight
In-class participation	every week	15%
Homework assignments	about bi-weekly	20%
Midterm take-home assignment	March 22	10%
Final project report	May 10	25%
1st paper presentation		
Oral presentation	prior to or on Feb 28	10%
Presentations note	Sunday after presentation	3%
Peer review	Wednesday after receiving notes	2%
2nd paper presentation		
Oral presentation	on or after Mar 13	10%
Presentations note	Sunday after presentation	3%
Peer review	Wednesday after receiving notes	2%
Total		100%

Written and oral communication are an integral part of any statistical work, and as such, grammar, style, and spelling are part of grading rubrics applied to all deliverables. You are strongly encouraged to use the resources and tutoring offered by the writing center (https://writingcenter.gmu.edu).

All assignments in this course are designated as individual assignments, which are to be undertaken independently. You may discuss your ideas with others but everything you turn in must be your own work. You may not share analyses, graphs, code, and other materials. You are responsible for making sure that there is no reason to doubt that the work you hand in is your own. The following types of collaboration on individual assignments are not honor code violations:

- Working on assignments with someone who is at roughly the same stage of progress as you, provided both learners contribute in roughly equal quantity and quality (in particular, thinking) on whatever problem or problem parts they collaborate. This type of collaboration is actually encouraged!
- A moderate amount of asking, “How do I do this in python/R/…?” However, as you gain enough familiarity, you should get in the habit of using online help and trying logical possi-

bilities, then asking for help only if these do not succeed after a reasonable try.

- Using programming code found on the internet or in software libraries, if using proper attribution (clearly identifying and citing all code snippets which are not your intellectual product).

The following types of collaboration on assignments **are** honor code violations:

- Working together with one learner the doer and one the follower.
- Any type of copying. In particular, splitting up a problem so that different learners do different parts is not authorized collaboration on homework. This also includes copying code from the internet without properly identifying the source.

Attendance: Attendance is mandatory and in-class participation, two oral presentation as well as presentation notes of two other oral presentation are part of your final grade. You will be able to choose the two dates of your oral presentation at the beginning of the term, but the date of note taking and peer reviewing will be assigned by the instructor after the presentation dates are chosen.

In case of approved absence, please get notes from your peers. You are responsible for material covered in class and announcements made during class.

Participation: Success in this course requires active participation in in-class activities and discussions, for which you will need to prepare in advance for each class period. Accordingly, you are expected to prepare for class period by

- reading the corresponding sections of textbooks or research articles to be covered in class,
- reviewing class materials posted in Blackboard to be covered in class,
- familiarizing yourself with the use of the covered methods and techniques in a statistical programming system of your choice.

Mandatory readings for the following class will be posted by Friday 11:59 PM. An oral quiz will be done at the beginning of each lecture assessing your understanding of the material covered in the readings.

Homework assignments: There will be six, roughly bi-weekly homework assignments throughout the term which will vary in length and content. Some involve in-class activities and continuing problems started in class, others involve solving exercises related to the material covered in class. Only five of these six homework assignments will be grade. It is your responsibility to not submit

a homework assignment if you do not want it to be graded. Once the homework is submitted, you cannot withdraw that homework and it will count towards your final grade. Once you submitted 5 homework assignments over the course of the semester, any following homeworks will not be graded or considered for the final grade. Due dates will be posted in Blackboard and your deliverables are submitted on Blackboard, usually in the form of a link to a Git repository. Late submissions will be penalized by reducing the total number of points possible by 10% of the original total number of points. For example, if a homework assignment is worth a total of 4 points, it will be worth only 3.6 points when submitted within the first 24 hours after the due date, 3.2 points when submitted between 24–48 hours after the due date, and so on. Submissions will not be accepted more than 4 days past the due date.

Oral presentation: You will give two oral presentations with visual aids on papers/readings assigned by the instructor. You will be required to provide an in-depth discussion of the paper, as well as an application of its methods/ideas. The first presentation will be given before Spring recess, and the second presentation after Spring recess. You will be able to indicate your preference for the dates on Blackboard. More details will be provided in class and on Blackboard.

Presentation notes: You will need to take notes for two presentation given by your peers, one in each half of the semester. This will be on a different date than your own oral presentation and assigned by the instructor. You will have until the following Sunday (4 days) to type up your notes and commit them to a shared GitHub repository. You may communicate with the presenter to ask for clarifications and the presenter may share visual aids used for the presentation. Once you push your notes to the shared repository, the peer assigned to review the notes and the instructor will give feedback and possibly request a revision until the following Wednesday (3 days). You will have until the following Sunday (4 days) to incorporate that feedback. Both the original submission and the final revision will be considered when assigning a grade. The revised version of the notes will be shared with everyone enrolled in the class. More details will be provided in class and on Blackboard.

Final project: The final project will involve designing, implementing, and summarizing a computationally intense statistical analysis. The analysis can be chosen by you, but must be approved by the instructor no later than April 10. More details will be provided in class. You must submit a written report on Blackboard and the fully functional, well-structured and efficient code via GitHub. Specific deliverables depend on the chosen project and will be discussed individually. You must ensure that the analysis and all results are fully reproducible. The final report and analysis code will not be accepted late.

You are expected to address your final project report with the same level of preparation and pre-

sentation that you would associate with a finished product on your job as statistician. The report that you write for this course will be graded on both your in-depth analysis and your writing. Each report will be no more than 6 pages of text (typically 4–5 pages) plus a few more for tables, graphs, and charts. Reports must not include a long appendix of outputs from statistical software and no computer code.

Regrading policies: You have at most one week after a score is posted for an assignment to appeal the score. If you want parts of an assignment remarked, send me an email specifying the question/part and the reason for requesting a review of grading. If you do not notify me in writing of any issues with your score within that time, then the posted score stands (whether or not it is correct).

Policies and Classroom Climate

During classes and online you are encouraged to discuss and share ideas with your classmates (see above how this relates to the honor code). To facilitate a respectful and inclusive classroom climate, be open to explore and challenge each other's ideas without criticizing individuals. Diversity is a source of creativity and innovation and I ask that learners appreciate diverse perspectives, that they listen respectfully and let everyone speak. If you have concerns about the dynamics or classroom climate, please do not hesitate to bring them to my attention.

The School of Computing seeks to create a learning environment that fosters respect for people across identities. We welcome and value individuals and their differences, including gender expression and identity, race, economic status, sex, sexuality, ethnicity, national origin, first language, religion, age and ability. We encourage all members of the learning environment to engage with the material personally, but to also be open to exploring and learning from experiences different than their own.

Gender identity and pronoun use: If you wish, please share your name and gender pronouns with me and how best to address you in-person or via email. I use he/him/his for myself and you may address me as "David", "Prof. Kepplinger" or "Dr. Kepplinger" in email and verbally.

Individual accommodations: Disability Services at George Mason University is committed to providing equitable access to learning opportunities for all learners by upholding the laws that ensure equal treatment of people with disabilities. If you are seeking accommodations for this class, please first visit http://ds.gmu.edu for detailed information about the Disability Services registration process. Then please discuss your approved accommodations with me. Disability Services is located in Student Union Building I (SUB I), Suite 2500. Email: ods@gmu.edu | Phone: (703) 993-2474.

Class etiquette: Class will start on time at 7:20 p.m. and end on time at 10:00 p.m., with a 10-minute break around 8:30 p.m. Although situations may arise making it impossible for you to arrive on time and/or requiring you to leave early, please remember that late arrivals and early departures will factor into your in-class participation grade. **Regular attendance for the full period of each class is very important for your success in this course!**

- Mute your phones during class, and keep them stowed away.
- You may eat during class, as long as it is done discreetly, quietly, and odorless.
- Immediately before or after class is not a good time to ask lengthy questions. Please come to office hours (or make an appointment) instead. Questions during class are welcomed and encouraged.

Netiquette: We will often communicate via discussion forums, GitHub issues, and other forms of online communication. To facilitate effective communication via these channels, please adhere to the following:

- *Be relevant and concise:* When posting a message to an online discussion, stick to the topic, make sure that you send enough information, and be concise.
- *Use accurate topic titles:* Each posting should include a topic title (a subject line) that lets the recipient know the posting's content. This allows others to scan their online messages, read the more important messages first, and keep organized.
- *Read before posting:* Read posted questions/answers before asking a new question to avoid repeating points already made, asking questions already answered, or bringing up points that have already been argued and either accepted, rejected, or exhausted. In addition, by "replying" to messages instead of starting a new message, a thread of communication can be kept going.
- *Be polite:* Avoid inflammatory messages and language. Do not send a message that ridicules someone else. Also, be careful when using humor or sarcasm, as most of it gets lost in the medium.
- *Review messages before submitting:* Think before you "speak" electronically. For the most part, electronic communication is a non-visual form of communication; therefore, people are unable to rely on facial expressions, tone of voice, or body language to interpret electronic messages. Misunderstandings can easily occur because of these factors.

Notice of mandatory reporting of sexual or interpersonal misconduct: As a faculty member, I am designated as a "Non-Confidential Employee," and must report all disclosures of sexual assault, sexual harassment, interpersonal violence, stalking, sexual exploitation, complicity, and retaliation

to Mason's Title IX Coordinator per University Policy 1202. If you wish to speak with someone confidentially, please contact one of Mason's confidential resources, such as Student Support and Advocacy Center (SSAC) at 703-380-1434 or Counseling and Psychological Services (CAPS) at 703-993-2380. You may also seek assistance or support measures from Mason's Title IX Coordinator by calling 703-993-8730, or emailing titleix@gmu.edu.

Honor Code: The integrity of the University community is affected by the individual choices made by each of us. Mason has an Honor Code with clear guidelines regarding academic integrity; you are responsible to know your requirements for this course. All violations of these rules will be referred to the Honor Committee; I take the Honor Code seriously and so should you. No grade is important enough to justify academic misconduct. If you have any questions concerning the Honor Code and how it relates to this particular course, please contact me.

Some kinds of participation in online study sites violate the Mason Honor code: these include uploading of any of the course materials or exams; and uploading any of your own answers or finished work. Always consult your syllabus and me before using these sites.

Privacy: Your privacy is governed by the Family Educational Rights and Privacy Act (FERPA) and is an essential aspect of this course. You must use your MasonLive email account to receive important University information, including communications related to this class. I will not respond to messages sent from or send messages to a non-Mason email address.

All course materials posted to Blackboard or other course sites are private to this class; by federal law, any materials that identify specific learners (via their name, voice, or image) must not be shared with anyone not enrolled in this class.

- Videorecordings — whether made by me or learners — of class meetings that include audio, visual, or textual information from other learners are private and must not be shared outside the class.
- Live video conference meetings (e.g., Zoom) that include audio, textual, or visual information from other learners must be viewed privately and not shared with others in your household or recorded and shared outside the class.

Copyrights of course material: This course gives you access to presentations, handouts, and copyrighted material and articles. Please treat them accordingly. All material other than copyrighted material should be regarded as authored materials, which if used or referred to must be fully credited through reference to me, the course, and date. If used beyond citation, my permission is required.

Version History

Version 1 (01/09/2024) Initial version.