George Mason University

STAT 634 – Case Studies in Data Analysis

Fall 2022

Instructor: David Kepplinger

Version 1 (January 19, 2022)

Administrative

Course dates: Tuesdays, January 25 – May 3 (Final exam period: May 17)
No class on March 15 (spring recess)

Important dates: January 31: last day to add
February 7: final drop deadline (no tuition penalty)
March 1: end of self-withdrawal period

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

> **Email:** dkepplin@gmu.edu
>
> **Office:** Room 1711, Nguyen Engineering Building (ENGR)
>
> **Virtual office hours:** Friday, 2 – 3pm, via Zoom: https://gmu.zoom.us/j/94368731850
> ?pwd=UjlYdmE3WmdOb1pnYVd2ejZMVzhyZz09 (Meeting ID: 943 6873 1850; Pass-code: 473082).
> For in-person office hours or to schedule an office hour at a different time, email the instructor.

Blackboard course page: https://mymasonportal.gmu.edu/ultra/courses/_448573_1/cl/outline

Class time: Tue, 4:30 – 7:10pm

Class meeting location: Innovation Hall 131

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). Similarly, all learners in face-to-face and hybrid courses must also complete the Mason COVID Health Check daily, seven days a week. The COVID Health Check system uses a color code system and students will receive either a Green, Yellow, or Red email response. Only students who receive a "green" notification are permitted to attend courses with a face-to-face component. **If you suspect that you are sick or have been directed to self-isolate, please quarantine or get testing. Faculty are allowed to ask you to show them that you have received a Green email and are thereby permitted to be in class.**

Learners are required to follow Mason's current policy about facemask-wearing. As of August 11, 2021, all community members are required to wear a facemask in all indoor settings, including classrooms. An appropriate facemask must cover your nose and mouth at all times in our class-room. If this policy changes, you will be informed; however, learners who prefer to wear masks either temporarily or consistently will always be welcome in the classroom.

If the campus closes or class is canceled due to weather or other concern, you should check Blackboard for updates on how to continue learning and 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,
- assignments,
- changes to the posted office hours,
- handouts and readings.

Moreover, general topic forums and forums specific to assignments are available via the Microsoft Teams group at https://teams.microsoft.com/l/team/19%3aHklwRXOzjKbd_2oz_HHx1ejWwPUnIesO8JwBhUiEKH01%40thread.tacv2/conversations?groupId=34018d93-d1fb-4a12-8ab9-28c350693482&tenantId=9e857255-df57-4c47-a0c0-0546460380cb Please check the group channels regularly. You are strongly encouraged to post questions about coursework on Microsoft Teams, as you are likely not the only one with the same question. Consider the Microsoft Teams group as an extra resource for getting help with projects and assignments.

E-mail communication is restricted to questions relating to personal or 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 634` 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.)

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 my office hours or after class. We can discuss alternative arrangements that suit your needs.

Course requirements

Prerequisites: STAT 544 and STAT 654; working knowledge of R.

Recommended readings: The main textbooks for this course are:

- Roger D. Peng and Elizabeth Matsui (2018). *The Art of Data Science*. Leanpub. Available online for free at https://leanpub.com/artofdatascience.

- Roger D. Peng (2019). *Report Writing for Data Science in R*. Leanpub. Available online for free at https://leanpub.com/reportwriting.

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

The following textbooks are optional, but can be useful for guidance on writing about statistical analyses:

- Jane E. Miller (2013). *The Chicago Guide to Writing about Multivariate Analysis*. 2nd edition. University of Chicago Press.

- Rae R. Newton and K.E. Rudestam (2013). *Your Statistical Consultant: Answers to Your Data Analysis Questions*, 2nd Edition. SAGE Publications.

- Gerald van Belle (2002). *Statistical Rules of Thumb*. 2nd Edition. Wiley.

Internet access: Activities and assignments in this course will regularly use the Blackboard learning system, available at https://mymason.gmu.edu. You are required to have regular, reliable access to a computer with an updated operating system (recommended: Windows 10 or Mac OSX 10.13 or higher) and a stable broadband Internet connection (cable modem, DSL, satellite broadband, etc., with a consistent 1.5 Mbps [megabits per second] download speed or higher).

Software requirements: This class will use R (version 4.0 or higher) to perform data manipulation, analysis, and visualization. The R package `rmarkdown` will be used to generate reproducible reports. For data manipulation and analysis you are allowed to use SAS instead of R. The focus of using R (or SAS) in this course is on writing and managing code in a consistent and professional style, interpreting outputs, and ensuring reproducibility. The course will be taught using RStudio Desktop, accessible for free as download, but you can use any interface you are comfortable with. For details on accessing and using RStudio Desktop, as well as SAS Studio if you choose to use it, see Blackboard.

You will also need a Git client on your computer. This course will be taught using the Git client built into RStudio Desktop. Alternatively, you may choose to use the basic Git console, or a dedicated user interface for Git, such as Sourcetree or GitHub Desktop.

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. In an emergency, you can connect through a telephone call, but video connection is the expected norm.

Course description

This course serves as a capstone experience for the MS in Statistical Science and the MS in Biostatistics programs in preparation for employment as an applied statistician. You will not be learning new statistical methodologies in this class. Instead, this course provides an opportunity to gain practical experience in consulting and data analysis through lectures, in-class activities and various projects.

Much of education in statistics is focused on learning how to do statistical analysis, but not necessarily on the other skills needed by an applied statistician. The goal of this course is to explore the phases that precede and follow statistical analyses, as well as the statistical analysis itself.

Participation is an important component of this course. You are expected to contribute regularly to in-class discussions, participate in in-class activities, and critique/evaluate other learners' written reports and oral presentations.

Learning objectives: After successfully completing STAT 634, you will be able to

- identify approaches for analyzing data gathered to address a specific question;
- understand the structure of a consulting meeting and the consulting process;
- create appropriate tables and visualizations and integrate these into a report for effective communication;

- apply your problem solving skills to statistical case studies;

- utilize the statistical computing tool R to generate reproducible reports;

- synthesize statistical information to clearly and effectively communicate your findings to statisticians and non-statisticians (verbally and in writing).

Assessments and grading

Your grade in this course will be based on individual homework assignments of various types (22%) and two case studies (78%). The case studies are comprised of multiple assignments each. There are no tests for this course. However, you will be submitting a final report and doing an oral presentation using visual aids during the final exam period.

Assignment	Teamwork	Tentative due date	Weight
Homework assignments	no	various	22%
Case study 1 (accounts for a total of 30%)			
Data analysis	yes	March 1	8%
Oral presentation	yes	March 1	4%
Peer review (of oral presentation)	no	March 4	3%
Draft report	yes	March 8	5%
Git repository (i.e., analysis/project organization)	yes	March 8	2%
Peer review (of draft report)	no	March 11	3%
Report	yes	March 22	5%
Case study 2 (accounts for a total of 48%)			
Oral presentation of proposed research question	yes	March 29	2%
Draft data analysis	yes	April 26	7%
Draft report	yes	May 3	4%
Peer review (of draft report)	no	May 8	5%
Final report	yes	May 17	12%
Final data analysis	yes	May 17	5%
Final oral presentation	yes	May 17	9%
Git repository (i.e., analysis/project organization)	yes	May 17	4%
Total			100%

Written communication is an integral part of this course, and as such, grammar, style, and spelling are part of grading rubrics applied to all deliverables: homework assignments, reports, peer reviews and oral presentations. You are encouraged to use the resources and tutoring offered by the writing center (https://writingcenter.gmu.edu).

The case studies in this class are designated as team projects (except for the peer reviews). For team assignments, you will work as part of a team consisting of two or three learners. A single submission is made for each team assignment, including names of all participants, and all

members of that team receive the same grade for that assignment. Working in teams is designed to develop teamwork and leadership skills, not to make some learners work harder than others! Each team submission must be accompanied by a short explanation of the contributions by each team member.

Other assignments are designated as individual assignments, which are to be undertaken independently. For these assignments, you may discuss your ideas with others but everything you turn in must be your own work. You may not share analyses, graphs, 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 individual 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 SAS or R?" However, as you gain enough familiarity, you should get in the habit of using online help and trying logical possibilities, then asking for help only if these do not succeed after a reasonable try.
- Using SAS/R code found on the internet to conduct your analysis, if using proper attribution (clearly identifying all code snippets which are not your intellectual product).

The following types of collaboration on individual 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 expected and oral presentations as well as peer reviews of others' oral presentation is part of your final grade. Dates of oral presentations will be posted on Blackboard at the beginning of the term and you need to be present on at least one of those days to give feedback to your peers.

If you miss class, 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, 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 the textbook to be covered in class, and
- reviewing class materials posted in Blackboard to be covered in class.

Homework assignments: Homework assignments will vary in length, content, and weight for the final grade. Some homework assignments are based on in-class activities, others involve solving exercises related to material covered in class. Due dates will be posted in Blackboard and your deliverables are submitted either on Blackboard or GitHub. Homework assignments are generally due on Monday 11:59pm and will not be accepted late.

Written analysis reports: Written analysis reports involve analyzing a data set and submitting a report and the fully functional SAS/R code via GitHub to reproduce the analysis. The binding due dates will be posted on Blackboard. Reports and source code will not be accepted late.

You are expected to address your analysis reports with the same level of preparation and presentation that you would associate with a finished product on your job as an applied statistician. The reports that you write for this course will be graded on both your analysis and your writing. Each report will be no more than 10 pages of text (typically 4–5 pages) plus a few more for tables, graphs, and charts. Reports must not include any SAS/R code or raw SAS/R output. All analyses must be reproducible using the code in the accompanying Git repository.

Oral presentations: You will give two oral presentations with visual aids based on the written analysis reports for the two case studies and one short oral presentation of your proposed research question for case study 2. More details will be provided in class and on Blackboard.

Peer review: You will give feedback on two drafts of written analysis reports and one oral presentation, as assigned by the instructor. For reviewing the oral presentation you will be given an evaluation rubric to follow. Peer reviews of written analysis reports are 1–2 pages each and can be either in paragraph or bullet-point format. More details will be provided in class and on Blackboard.

Project organization: We will use Git to organize the source code and deliverables for each project. You are required to maintain a Git repository with all relevant files (SAS/R script files, reports, figures, etc.) in a clear and consistent structure. Teams will need to agree on a structure and style and follow it throughout the project. You can earn up to 2 bonus percentage points for very well structured and maintained repositories and cleanly structured code in your individual homework assignments.

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 you 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 College of Engineering and 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 class and in writing. I use he/him/his for myself and you may address me as "David", "Prof. Kepplinger" or "Mr. Kepplinger" in writing and verbally.

Individual accommodations: Disability Services at George Mason University is committed to providing equitable access to learning opportunities for everyone 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 4:30 p.m. and end on time at 7:10 p.m., with a 10-minute break around 5:45 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 can be quite disruptive to your classmates. So, please make arriving to class late or leaving early an exception, not a habit. **Regular attendance for the full period of each class is very important for this course!**

- Please 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, Microsoft Teams, 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 assault, sexual harassment, interpersonal misconduct, and stalking: 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 your instructor 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 the instructor 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 the instructor, course, and date. If used beyond citation, permission of the instructor is required.

Version History

Version 1 (2022/01/14) Initial version.