An Introduction to Bayesian Statistics

PSYC 757 - Section 002

Patrick E. McKnight

2024-01-14

Table of contents

Welcome

Course Location & Time

George Mason University
Fairfax campus
Planetary Hall, Room 127
Tues/Thurs 9-10:15am

Thank you for your interest in quantitative methods. We hope you enjoy the course.

Sincerely,

Patrick E. McKnight, Ph.D.

Professor

Patrick E. McKnight Office: DK2064/65
Email: pmcknigh@gmu.edu
Office Hours: Tues/Thurs 10-noon (online)

Overview

Psychology 643 is the second course of a two-course sequence that serves to introduce psychology graduate students to statistics, research methodology, research design, and measurement from the vantage point of the general linear model. This course will further student knowledge in the General Linear Model to allow students to evaluate assumptions, deal with group comparisons, work with repeated measures, and address moderation and mediation. The class meetings will consist mostly of lectures and discussions. The lab assignments will focus on providing you with hands-on experience analyzing data and interpreting results.

Objectives

The purpose of the class remains focused on Bayesian statistics, however, unlike many Bayesian devotees, I consider Bayesian statistics a tool. Just like other tools, Bayesian methods have suitable and unsuitable applications. Students enrolled in my course may find my approach more challenging than standard graduate lectures because I emphasize mastery (i.e., learning) over achievement (i.e., grades). I demand more from you (and me) so that we all learn together. If this format is more difficult, why would I implement it? Simple. I want students to learn, not just get grades for taking a class. The more my students learn, the more future professional opportunities await them.

Learning Objectives

Let me...

1. **G**ain the ability to choose the appropriate analytic approach for a specific question
2. **U**nderstand the limitations of inference
3. **I**nterpret results accurately
4. **D**evelop a vocabulary that allows you to read and interpret the literature
5. **E**xpress uncertainty as a probability distribution
6. **U**nderstand the role of prior information in inference

Get it?

1 Workflow

1.1 Initial Setup

Follow each of the steps below when you begin the course OR when you want to either setup another computer or you replace the one you began the course using. Either way, these steps are universal. Follow them in order to avoid any issues.

STEP 1: Install the Tools

We will use R, RStudio, Python, Quarto, Git, and GitHub (Desktop if desired). You need all six entities installed and configured to be functionally enrolled in this course. Thus, we need to get you all setup before I can start distributing things like...this syllabus and my weekly notes. So, let's get started.
First, get R (aka our engine for this course). Go to CRAN and download the latest version of R for your operating system. Install R. Then install RStudio. Go to RStudio and download the latest version of RStudio for your operating system. Then, install RStudio. Once you have a functional RStudio environment, go to the Quarto website and follow the instructions to install Quarto. Finally, install Git. Go to Git and download the latest version of Git for your operating system. Install Git. If you want to use GitHub Desktop, go to GitHub Desktop and download the latest version of GitHub Desktop for your operating system. Install GitHub Desktop - again, optional.

STEP 2: Get a GitHub account (and send to me)

We will use GitHub to store all of our files, including the project reports. Thus, you will need to get comfortable with GitHub. I will help you with this process.
Sign yourself up for a github account. Once you get your username, please send it to me via email. I will add you to the course organization and you will be able to access the course repository (a set of files stored on the cloud). Once you have access to the course repository, you will be able to clone (copy) the repository to your computer. You can even edit these files! Thus, the addition of you to the group repository enables you to participate in the course. Don't worry about using the tool right now, we will build in a routine for each day, week, and quarter(ish) to help you easily get into the flow of things. GitHub allows us to work together in a more efficient manner - compared to Google docs,

for example. We will use GitHub to store all of our files, including the project reports. Thus, you will need to get comfortable with GitHub. Again, I will help you with this process. - Get your PAT setup. See this page for more details. - Add your PAT to your credential manager. See this page for more details. - Test your setup.

STEP 3: Open RStudio and set up Git

Once you have RStudio installed, open it. Then, go to Tools > Global Options > Git/SVN. Click on the box next to "Enable version control interface for RStudio projects". **Windows Users** (Then, click on the box next to "Use Git from the Windows Command Prompt", click on the box next to "Use a custom Git executable" and navigate to the Git executable on your computer). **Mac/Linux Users** select the git executable that RStudio found in your path. If there is no executable, go back and install git. Finally, for all users, click on the box next to "View project status in the RStudio project". Click "OK". You are now ready to use Git with RStudio.
If you want a more, in depth, tutorial on how to use Git with RStudio, go to Happy Git and GitHub for the useR and follow the instructions. I highly recommend this outstanding resource for learning how to use Git with RStudio.

STEP 4: Clone the course repository

This is the easiest part. Once your system is setup and configured to use all the tools, you can now issue the following commands:

```
# Clone the course repository
install.packages("usethis")
library(usethis)
usethis::use_course("pem725/BayesianSyllabus_S2024")
```

STEP 5: Render the syllabus & read (often)

Open ANY of the qmd files - probably index.qmd is the most logical choice. Then, click on the "Render" button that will create a wonderful document to your right. If you click on the button that says "Show in new window", you will have a lovely full-screen version of this beauty. Throughout the semester, I will update this book. Simply do a fresh pull with the same code in the same directory and you will have the most up-to-date version of the syllabus and notes. The *usethis* (Wickham et al. 2023) package is quite handy.

1.2 Daily (or every other day)

Pull the latest changes from GitHub (5 mins max)

Open RStudio and open the course syllabus repository (STEP 4 above). Then, click on the "Pull" button in the Git tab; doing so will refresh your local files and enable you to render the full set of notes. You will need to do this **before** you start working on the weekly assignment.

Read the assigned readings (before Tuesday each week; 4 hours max/week)

Yep, read the assigned readings BEFORE class. Remember, you get to choose. I have two books. One is a light read with great examples (Kurt 2019); the other is a bit dense with tons of code (McElreath 2020). Either one suffices for the course. If you have a keen interest in applying these tools to your own work, I strongly recommend the latter. You choose. But please read.

Write down your questions from the assigned readings

As you read, note any questions you have regarding terminology, concepts, or applications. These questions will be used for discussion during our Tuesday sessions. Your questions will also help me address any issues you may have with the material. I will do my best to answer your questions in a timely manner. If I cannot answer your question, I will find someone who can. I will also use your questions to help me improve the course. Thus, your questions are important. Please write them down and ask them during class.

Work on the weekly assignment (before Thursday each week; 4 hours max/week)

Open up RStudio every day and get accustomed to the workflow. There is a lot you learn simply by repetition. So, lather, rinse, repeat. You can do it. These initial workflow steps are ones that translate to every area of your life. Break things down into repeatable chunks, commit them to a memorable sequence, and stick to it. You will be amazed at what you can accomplish.

1.3 Every 5 Weeks (Once per Project)

i Check your assignments for the project

Via GitHub, each of you will receive a set of tasks to accomplish so that we - together - address the project aims and objectives. You will be assigned a task and a deadline. Please do your best to meet the deadline. If you cannot, please let the team know as soon as possible. The team is all of us. I will do my best to accommodate reasonable needs. But, please, do not wait until the last minute to let me know you cannot meet the deadline. The team will suffer and I will not be happy. And, you will not be happy. So, let's be happy. Communicate early and often.

i Submit all contributions to GitHub

Once you have completed your assigned task, submit your work to GitHub. In fact, once you finish every update to the project, submit the work via a push on GitHub. Don't worry. If you make a mistake, we can easily fix it. Think of this as a "track changes" edit to the group project file. I will review your work, accept as necessary, and provide feedback. If you have questions, please ask. If you have concerns, please voice. If you have suggestions, please suggest. Thank you.

i Review GitHub project and compile **on your computer**

Once you have submitted your work, review the GitHub project and compile the document on your computer. Doing this final step ensures that every member of the team can render the document. Rendering the document also means that everyone has their respective computer setup and configured appropriately, can use these tools, and has the wherewithall to navigate the text and code for a group project on Bayesian statistics.

i Ensure project lead can render the final report

Check with the project lead. Report any problems right away. The team's success depends upon each of us. Do your part. Be ready to help. Be ready to learn. Be ready to teach. Be ready to lead. Be ready to follow.

i Submit and celebrate!

The final project will be a group effort. We will collectively review the final project and rate our product. Our ratings will NOT be relevant to the course grades. The product may be limited for reasons beyond our control. What we can control is our effort. We

> will do our best. We will learn. We will grow. We will celebrate. We will be happy.

2 Instructional Resources

OPTION 1: For those who want the gentler version of the course, please use this text:

"Bayesian Statistics the Fun Way" (2018)

OPTION 2: For those who want the more technical version of the course, please use this text:

McElreath (2020)

The course website contains tons of videos, links, and resources.

3 Grading and Evaluation

3.1 COURSE REQUIREMENTS

Bayesian statistics requires you to learn many things but not all of you begin the course at the same stage of learning. Thus, I plan to create a collaborative learning environment that mimics the real world. We will all be part of a team and our job is to complete three projects. You will be assigned tasks to complete each week. Everyone will have a different task. The projects will be graded and the grades will be based on the quality of the work and the timeliness of the work.

3.1.1 Attendance (50%)

Yes, attend. You cannot learn if you do not attend the class sessions. Woody Allen once famously said "80% of success is showing up." So why not make success your goal and show up. I offer you half of your grade is predicated on your attendance. Each day, I intend to take a picture of the participants at 9:10a. Please make every effort to attend and attend on time. Thank you.

3.1.2 Active Participation (25%)

Yes, participate. You cannot learn if you do not participate in the discussions and activities. So, given that you plan to attend, you might as well participate. I do intend to call on those who do not contribute often. We need to make this course more interactive. Please come prepared to talk about the material in the assigned readings.

3.1.3 Assignments (25%)

Each of you will have assigned duties for each project. The projects require you to work in one large group (entire class including me). Your responsibility is to complete the tasks you were assigned. If they are too burdensome at any point, it is **your** responsibility to communicate with the team. If you do not complete your tasks, you will not receive credit for the assignment and, as a result, the project will not be completed on time so the entire team suffers. Thus,

do your assigned task. All task assignments will be publicly made and recorded in the github repository.

3.2 GRADING

Final grades will be assigned according to the following percentages: $> 93\%$ = A, 90 – 92% = A-, 87 – 89% = B+, 83 – 86% = B, 80 – 82% = B-, 70 – 79% = C, 60 – 69% = D, $< 60\%$ = F Note: A grade of C or lower denotes an unsatisfactory level of achievement for a graduate student.

4 Projects

I borrowed the idea of these projects from a published paper (Berg 2021) suggesting that a reasonable pedagogical approach to teaching Bayesian statistics is to use practical examples. We deviate a bit from his suggested topics but the idea remains.

4.1 Project 1: Stein's Paradox

First, you must read. I suggest you begin with Bradley Efron and Carl Morris' (Efron and Morris 1977) excellent article in Scientific American. Another more detailed blog post (Antognini 2021) provides depth - perhaps more than you want or need. Second, you need to start thinking about the problem. Look at this resource: Keil et al. (2018) and Clark (2019) for some ideas about how to demonstrate the effect. We will tackle this problem with a variety of tools including (but not limited to) Python, R, Quarto, and Stan.

> Project 1 Aims
>
> Find, create, simulate, or imagine (and then simulate) a dataset that has repeated measures for at least three people. Compute the gain in accuracy and precision from using a shrunken estimator in predicting future scores.

4.2 Project 2: Measurement Error

Most of us think of measurement error as the result of a mistake (accident during data collection), omission (relevant variable, coherent model, range, etc.), or bias. Some of us know better. To capture the essence of what measurement error is (and is not), we shall undertake a deep dive into a simple measurement model. The actual data source will be revealed in time but the data represent a simple instantiation of a common problem. We have XXX unique participant data who completed a measure that was administered across 8 or 16 vignettes.

But, in fact, measurement error is a fundamental part of the process of measurement. We will explore this idea in the context of a simple linear regression model. We will use the same tools as in Project 1.

5 Academic Policies

5.1 EXAM & HOMEWORK POLICY

Without prior arrangement, there will be no extensions or make-ups without penalty except in instances such as the following:

- hospitalization or illness that has been documented and judged by your instructor as preventing you from a) preparing adequately for an exam or quiz or b) completing an assignment
- death or serious illness in your family
- court appearances
- religious observances

Documentation must be provided by health officials (e.g., a physician or member of the student health center staff) in the case of illness; an immediate family member in the case of death or serious illness in the family; and official paperwork in the case of court dates. Decisions regarding extensions and make-ups under these circumstances will be made on a case-by-case basis. In general, quizzes are not eligible for make-up work; instead, your two lowest quiz grades are automatically dropped. Lab homework is penalized at 5% per day for late submission (i.e., starting when lab time begins) and receives no credit if it is submitted over a week late. When possible, I encourage students to reach out as soon as you know that you may miss class and/or related work; it is much easier to develop a plan in advance than it is to do so later.

5.2 ACADEMIC INTEGRITY & HONOR CODE

All students in this course are to become familiar with and follow the University's honor code, which does not tolerate any form of cheating and attempted cheating, plagiarism, lying, and stealing. Exams and assignments are expected to be individual efforts unless otherwise noted by the instructor or teaching assistant. Violations of the GMU Honor Code can result in failure of an assignment or exam, depending on the severity of violation. All violations will be reported to the Honor Committee. The instructor for this course reserves the right to enter a failing grade to any student found guilty of an honor code violation. For more information on the Honor Code please visit this webpage.

5.3 GENERATIVE-AI AND LARGE LANGUAGE MODELS

Use of Generative-AI tools should be used following the fundamental principles of the Honor Code. These principles include being honest about the use of these tools for submitted work. Cite your sources - including citations when using the work of others, whether individual people or Generative-AI tools. Students may use Generative AI tools whenever they believe it would be useful to their learning of course material. Students will be directed if and when citation or a statement-of-usage is required. All academic integrity violations will be reported to the office of Academic Integrity.

For many learning to program, the use of these tools can greatly speed up your learning; their use may also inhibit your growth. If you find "learning" easier with these tools, you may be missing out on the learning that comes from the struggle of learning to program. If you find yourself using these tools, you should consider whether you are using them to learn or to avoid learning. If you are using these tools to avoid learning, you should consider whether you are using them to avoid the struggle of learning. If you are using these tools to avoid the struggle of learning, you should consider the net effects of such avoidance.

Struggle to learn. Such struggles bear more fruit than merely accommodating me and my course requirements. The future you will thank me for keeping your focus.

5.4 ACCOMMODATION OF DISABILITIES

Disability Services at George Mason University is committed to providing equitable access to learning opportunities for all students by upholding the laws that ensure equal treatment of people with disabilities. Note that this provision includes the range of disabilities, including physical, psychiatric, and learning disabilities. If you are seeking accommodations for this class, please first visit here for detailed information about the Disability Services registration process. Then please discuss your approved accommodations with me. All academic accommodations must be arranged through Disability Services. Disability Services is located in Student Union Building I (SUB I), Suite 2500. Email: ods@gmu.edu | Phone: (703) 993 – 2474.

5.5 SEXUAL HARASSMENT, SEXUAL MISCONDUCT, AND INTERPERSONAL VIOLENCE

As a faculty member and designated "Responsible Employee," I am required to report all disclosures of sexual assault, interpersonal violence, and stalking to Mason's Title IX Coordinator per university policy 1412. If you wish to speak with someone confidentially, please contact the Student Support and Advocacy Center (703-380-1434) or Counseling and Psychological

Services (703-993-2380). You may also seek assistance from Mason's Title IX Coordinator (703-993-8730; titleix@gmu.edu).

5.6 STUDENT SUPPORT SERVICES

George Mason offers services to support students' academic and emotional development. Counseling and Psychological Services, located in SUB I room 3129 (caps.gmu.edu), offers workshops in academic skills, stress management training, and virtual counseling for students who would like some help with social, emotional, or educational concerns. Consider taking advantage of these resources if you need them. For additional information about other student support services offered, please visit the Stearns Center Knowledge Center.

6 Technology Policies

6.1 Official Communications via GMU E-mail

Students must use their GMU 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.

6.2 Class cancellation policy

Classes may be canceled by either the University or the Instructor (via email). In the event of a canceled class, the instructor will either switch to an online class modality, try to make up the material during other meetings of the class, or provide a supplementary assignment/video.

6.3 Course Materials

Activities and assignments in this course will be made available to students through a learning management system (such as Blackboard) or other websites. Students are required to have regular, reliable access to a computer with an updated operating system and a stable Internet connection suitable for online meetings, streaming videos, completing assignments, etc.

6.4 Online Meetings

Although we are planning on having the class in person, things can change. In the event we need to pivot to an online class, then activities and assignments in this course will use Zoom. In addition to the requirements above, students are required to have a device with a functional microphone and camera.

6.5 Course Materials and Student Privacy

All course materials posted to Blackboard or other course site are private to this class; by federal law, any materials that identity specific students (via their name, voice, or image) must not be shared with anyone not enrolled in this class. Video recordings – whether made by instructors or students – of class meetings that include audio, visual, or textual information from other students 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 students must be viewed privately and not shared with others in your household or recorded and shared outside the class.

6.6 Statistical Packages

You will need to use a statistical package of **your choice** to complete assignments in this course; most students use SPSS or R. Students can access to SPSS through the online Citrix Computer Lab. R and R Studio are also available for free, direct download.

7 Tentative Schedule

Project	Wk	Date	Topic	Kurt (2019)	McElreath (2020)	Notes
1	1	1/15	Overview Probability Theory	c1-2	c1	Bring Computer
1	2	1/22	Uncertainty	c3-4	c2	Project 1 Intro
1	3	1/29	Probability Distributions	c5-6	c3	
1	4	2/5	Bayes Theorem/Stan	c7-8	c4	
1	5	2/12	Priors & Causality	c9-10	c5-6	Big Week
2	6	2/19	Parameterization	c11-12	c7	Project 2 Intro
2	7	2/26	Estimation	c13-14	c8 & c15	
	8	3/4	**SPRING BREAK!!**	get ahead	get ahead	or catch up
2	9	3/11	Testing/MCMC	c15	c9	
2	10	3/18	Bayes Factor/Entropy	c16	c10	
3	11	3/25	Linear Models/Forecasting	c17	c11	Project 3 Intro
3	12	4/1	Inference	c18	c12	
3	13	4/8	Estimation to Inference	c19	c13	
3	14	4/15	Project Wrap-up			
3	15	4/22	Final touches			
		4/29	**LAST DAY of Classes**			

8 Week 1

Three points for this week

1. Course Overview
2. Learning Objectives (G.U.I.D.E. U.)
3. Probability theory

8.1 Course Overview

Review the syllabus.

Antognini, Joe. 2021. "Understanding Stein's Paradox." https://joe-antognini.github.io/machine-learning/steins-paradox.

"Bayesian Statistics the Fun Way." 2018. https://nostarch.com/learnbayes.

Berg, Arthur. 2021. "Bayesian Modeling Competitions for the Classroom." *Revista Colombiana de Estadística* 44 (2): 243252.

Clark, Michael. 2019. "Michael Clark: Shrinkage in Mixed Effects Models," May. https://m-clark.github.io/posts/2019-05-14-shrinkage-in-mixed-models/.

Efron, Bradley, and Carl Morris. 1977. "Stein's Paradox in Statistics." *Scientific American* 236 (5): 119–27. https://www.jstor.org/stable/24954030.

Keil, Alexander P., Eric J. Daza, Stephanie M. Engel, Jessie P. Buckley, and Jessie K. Edwards. 2018. "A Bayesian Approach to the g-Formula." *Statistical Methods in Medical Research* 27 (10): 3183–3204. https://doi.org/10.1177/0962280217694665.

Kurt, Will. 2019. *Bayesian Statistics the Fun Way: Understanding Statistics and Probability with Star Wars, Lego, and Rubber Ducks*. No Starch Press.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. 2nd ed. Routledge. https://www.routledge.com/Statistical-Rethinking-A-Bayesian-Course-with-Examples-in-R-and-STAN/McElreath/p/book/9780367139919.

Wickham, Hadley, Jennifer Bryan, Malcolm Barrett, and Andy Teucher. 2023. "Usethis: Automate Package and Project Setup." https://CRAN.R-project.org/package=usethis.