STAT 454: Bayesian Statistics

Professor: Taylor Okonek (she/her)

Email: tokonek@macalester.edu

Office: Olin Rice 126

Section	Meeting Time	Room
01	T/Th 8:00 - 9:30 am	OLRI 241
02	T/Th 1:20 - 2:50 pm	OLRI 241

Office Hours: (Google Calendar Link)

This semester I am teaching Math/Stat 354 *and* Stat 454, so for office hours, I will give priority to students in certain classes. You are welcome and encouraged to come to ***any*** of the following office hours, but please know that if you come to 354 Prio-OH, for example, anyone in 354 will be prioritized first in terms of questions / assistance. Office hours are typically first-come, first-served. However, I will do my best to include everyone present in the conversation.

- Tuesday 10:00 - 11:00 am *(354 Priority)*
- Wednesday 11:00 am - Noon *(454 Priority)*
- Thursday 3:00 - 4:30 pm *(454 Priority)*
- Friday 9:30 - 11:00 am *(354 Priority)*

Note: If the office hour times listed above don't work for you, I am **available by appointment**; email me to set up a time to meet. More expectations on communication are in the Course Communication section.

Textbook

There is no required textbook for this course. We will occasionally reference and use the textbook Bayes Rules! (available for free online), co-authored by Mac's very own Alicia Johnson.

The best resources for this course in addition to the textbook are: attending and participating in class / assignments, and office hours. I am happy to talk about possible additional materials / strategies for effective learning at any time throughout the semester!

Learning Objectives

The overarching goal of this course is to prepare you to be able to *confidently* address an applied statistical question in a Bayesian framework, from start to finish! All course assignments, assessments, and learning objectives were designed with this in mind.

Upon completion of this course, students should be able to…

- distinguish between Bayesian and Frequentist philosophies of statistics, and be able to determine in which scenarios one approach may be preferred over another
- articulate building blocks of Bayesian models, including priors, posteriors, and likelihoods
- analytically derive common posterior distributions from conjugate priors
- computationally approximate posterior distributions using Markov chain Monte Carlo techniques
- conduct a thorough and appropriate Bayesian analysis in an applied context, including:
    - constructing a hierarchical model from a generative model perspective
    - computing appropriate summary measures using modern MCMC techniques
    - evaluating model performance via posterior predictive checks, prior sensitivity analyses, and convergence diagnostics
    - communicating results *and* methodology to a broad audience

----

Tentative Schedule

A detailed, day-to-day schedule will be filled in throughout the semester in the Schedule tab of this website. A rough outline by unit / week is as follows. Each unit will have at least one problem set, and one assessment.

Unit 1: Theoretical Foundations (Weeks 1-4)

- Problem Set 1
- Assessment: Theory Exam

Unit 2: Bayesian Computation (Weeks 5-7)

- Problem Set 2
- Assessment: Technical Report 1 (Peer Review + Oral Exam)

Unit 3: Model Construction & Evaluation (Weeks 8-12)

- Problem Sets 3 & 4
- Assessment: Technical Report 2 (Peer Review + Oral Exam)

Capstone Project (work time and presentations Weeks 13-15)

Course Communication

Course Website: The course syllabus and course resources will be posted on the course website. The Schedule tab of this website will contain information about due dates, in-class activities, and general day-to-day going-ons, and the Documents tab contains any and all files we'll use throughout the course (updated frequently). Please check the course website *every day* to make sure that you aren't missing anything!

Moodle: Moodle will be used *exclusively* for assignment dropboxes, so make sure to check Moodle around the time when things are due.

Slack: There is no preceptor for Stat 454. As such, it is going to serve *everyone* best if there's a relatively quick way to communicate with one another throughout the semester. I expect you to post course content questions on Slack, as opposed to asking me via email, since (1) others probably have the same questions! and (2) so that we can *collectively* answer questions and communicate about the course as a class. Slack conversations / "discussion board" engagement will be a component of your course grade.

Feedback: Feedback on course assignments, grades, etc. will be available in your individual feedback spreadsheets, which I will share with you at the start of the semester.

Email: Outside of office hours, email is the best way to discuss topics of a personal nature (e.g., grades, missing class) or to ask other questions you don't feel comfortable asking in a group setting. **Note that I will not send same-day responses to messages sent after 6:00pm.**

Please communicate respectfully with your classmates and myself, and let me know if there are ways classroom communication can be made more accessible to you.

Classroom Environment

I am committed to helping you learn and succeed. Part of this means facilitating environments that support you in the ways that you need. Everyone comes from a different path through life, and it is our duty to listen to each other without judgment, and to respect one another. There will be **no tolerance** for discrimination based on race, ethnicity, gender, religion, sexual orientation, disability, or other identities.

Bayesian statistics has a fraught history in this regard. Statisticians and philosophers have argued (sometimes without much respect for others' views) about the importance and usefulness of the Bayesian inferential framework since its inception. The subjective nature of Bayesian statistics has not always been met with open arms by the broader scientific community. More modern recognition that all statistical frameworks are inherently subjective has helped with this mindset, but regardless, people (and statisticians) will disagree. Such ethical and historical considerations come into play in any applied Bayesian analysis that we do. If you feel that you may not be able to complete a certain activity, do a reading, or partake in a discussion due to the nature of the topic at hand, please reach out to me so that you can help me understand your situation, and so that we can discuss the possibility of replacing the item with an alternative. I am more than happy to work with each of you to ensure that you feel comfortable in your learning.

Grades

Evaluations / Assessments

Problem Sets:

There will be four problem sets throughout the semester, roughly every three weeks. Problem sets are intended to serve as your opportunity to practice the material covered in class and demonstrate your individual understanding of the material. Two of the problem sets will serve as a jumping off point for your technical reports. You are encouraged to work together on Problem Sets, but the work submitted must be your own.

You will receive qualitative feedback on all questions on these problem sets. Each question on a problem set will be scored on the following scale:

High Pass	Pass	Low Pass	No Credit
All work was shown, you arrived at the correct answer, and you demonstrated mastery of the material.	You arrived at the correct answer, and most work was shown and/or sufficient understanding of the material was demonstrated.	You may not have arrived at the correct answer, but you demonstrated some understanding of the material and showed your work.	You did not attempt the question OR you showed no work, and instead only stated the answer.

Problem sets must be submitted on Moodle as .pdf files, rendered from .qmd files. Handwritten work will not be evaluated for credit.

For the first problem set (which will not include code) you may instead submit a .pdf file rendered from a .tex file if you prefer. For problem sets that include code, your code must be visible in your rendered file and well-documented. Rendered files that are tens of pages long with extraneous code output will not be evaluated for credit; you *must* review your file before submitting it.

Assessments:

1. Theory Exam. The first unit of the course (Theoretical Foundations) will conclude in a theory exam. This is intentionally placed early in the semester to ensure that we all have the mathematical language and tools necessary to proceed with computational and applied topics. Theory underpins everything we will do in this course, and sufficient mastery of the material must be demonstrated in order to pass this course.

   Therefore, you will be given the opportunity to take a version of the theory exam up to 3 times until you are able to pass. The first attempt will be in-class, on paper, closed note. All future attempts must be scheduled with the instructor, as a 30-minute, individual oral examination. The types of questions asked will be similar to the in-class exam, but you will instead present your answers on the whiteboard and describe your steps outloud. Subsequent attempts can be scheduled no closer than 3 days apart.

   I will make sure that you pass the exam on the third attempt. This is not intended to be a barrier to your success or completion of the course, rather a way for me to ensure that every student has the foundations we need to move forward.

2. Technical Reports. The assessment structure for Units 2 & 3 of the course will be in the form of Technical Reports (TRs). They will be due roughly one week following the second and third Problem Sets, respectively. These will be *substantive* analyses done in pairs, using the material covered in both coursework and problem sets in an open-ended applied analysis. You will be given a prompt and data roughly two weeks prior to the TR due date, and we will choose partners in-class. Guidelines on the technical reports

are given in more detail on the Technical Reports tab of the course website. The TR will involve a peer-review process as well as an oral exam with the instructor.

Capstone Project:

An important skill for any statistician (and humans in general) is the ability to learn and understand a new-to-them concept at a deep level, and to be able to effectively communicate that concept to others. In your capstone project, you will be expected to build on and go *beyond* the material covered in this course in such a way. Projects will be done in pairs of groups of three, and you will have a list of possible topics to choose from for a final project. More details will be provided on the capstone project in late February / early March.

As in all 400-level MSCS courses, you must pass the capstone project to pass the class.

Course Engagement:

As an upper-level course in statistics, you will need to consistently engage with the material both in and outside of class in order to effectively learn the content. To this end, there are multiple components of your grade which will be related to course engagement:

1. **Slack** **discussion boards**: One of the best ways to engage with course material is talking about it in conversation with others, whether that be asking and answering each others questions or prompting deeper discussion on readings ahead of in-class discussions. To this end, you are expected to post **at least twice per week** (roughly once per class period) to our course Slack channel. These posts could include, but are not limited to:

   - A question about *content* that came up in class, or while reading pre-class material

   - A response to another students' question, or a question posed by the instructor. If your response is related to a Problem Set or Technical Report, you should *not* simply provide the answer but instead provide *assistance* in approaching the problem

   - A reflection on course material that goes beyond what was discussed in class

   - A question about *code*, particularly if you have an error that you are struggling to debug

   A discussion post that is a meme, or something similarly unhelpful (even if fun), will not "count" towards your expected number of discussion posts.

2. **In-class engagement**: I expect you to be present and engaged in class. I additionally expect you to be *on time*, so as to not be a distraction to your fellow students and myself. Throughout the semester, I will keep track of whether you are absent (for any reason), late and/or not on-task during class, or present and engaged. If you are *not* consistently present and engaged, your final course grade *may* be modified by **one third** of a letter grade (i.e. A -> A-, B+ -> B). This is not intended to be a strict policy, but rather serve as a guideline. I recognize that, particularly in our current political climate,

attendance may be an issue. If you find yourself in extenuating circumstances, please have a conversation with me about alternative ways to engage with the course.

3. **MSCS Capstone Reflection**: Every student in a T/Th MSCS course in the Spring semester is expected to attend a minimum of 3 capstone talks during MSCS Capstone days (March 5th and 6th). These can be during our usual class time (which be cancelled for Capstone days) or any other time you are free and interested in attending a talk! If you are presenting during Capstone days, you may count your own talk as one of the 3 required reflections for this course. I will provide you with a feedback form to fill out in advance of Capstone days, along with the presentation schedule once it is made public.

4. **MSCS Seminar Reflection**: Every student in a 400-level MSCS course is expected to attend a minimum of two seminars throughout the semester. I will announce seminars as they come up, and provide you with a feedback form (linked in the title of this item!) to fill out in advance of any talk.

Late Policy

Throughout the quarter, you may use up to **three**, three-day extensions on Problem Sets. Note that this means you can use extensions on *almost every* Problem Set. This level of flexibility is in place for Problem Sets because it *cannot* be used on any other assessments in the course, including the technical reports and the preparation needed for peer-review days. The purpose of deadlines (and extensions) are thus to keep you accountable for your own learning, to keep you on track with the pace of the course (which builds upon itself throughout the semester), and to provide myself with the ability to provide you with timely feedback on assignments.

Extensions can be used automatically, without letting me know in advance. If you do not hand in a Problem Set at the beginning of class when it is due, I will assume you are using an extension. To hand in an assignment you are using an extension on, you must submit it on Moodle within 72 hours of the deadline. If you do not submit an assignment on time or within the extension period, I *may* provide you with feedback, but you will **not receive credit** for the assignment.

I expect you to keep track of how many extensions you've used. I will email you a reminder if you have used all three of your extensions and have none remaining.

If you have run out of extensions and/or an extenuating circumstance occurs that impacts your ability to submit assignments on time, please email me to discuss the situation. I will try to be flexible, within reason, as long as you communicate!

Academic Integrity

As a globally- and community-oriented institution, Macalester College expects respectful exchange of ideas. Students are expected to be familiar with the college standards on academic integrity (link). You are encouraged to work with your classmates on assignments and activities, but in order to receive individualized feedback on your own learning, **all submitted work (including code) must be written in your own words**. Issues of academic integrity will be taken very seriously, and any concerns about this policy being violated will be followed up on individually.

AI Use

I *strongly* encourage you *not* to use AI at all throughout this course. While it might feel helpful in the moment, it certainly won't help you with your long-term retention of the material, and it won't help you be prepared for the in-class or oral assessments. Remember that learning comes from *you* actually thinking through and coming to a solution, not a machine producing the correct answer for you. If you do use AI on a problem set or technical report, I expect you to cite when/how/where you used it, in accordance with Macalester's academic integrity policies. If you use AI and *do not* cite it, that will be considered an academic integrity violation.

How I expect you cite any AI use (not recommended!) for this course: Include a paragraph at the end of your problem set or technical report, explaining what you used AI for and what prompts you used to get the results. Note that you *will* be expected to be able to explain your work *in person* and *in your own words* regardless of your AI usage. This includes all lines of code produced for Problem Sets, Technical Reports, and your Capstone Project. If you cite AI use and are unable to explain your work, you may not receive credit for an assignment or assessment.

Course Grading System

The course grading system is designed to help you achieve the learning objectives while allowing space to make and learn from mistakes along the way. The table below presents the criteria for earning an A, B, C, or D/NC. Depending upon the level of progress towards each goal, half-letter grades (e.g. A-, B+) may be assigned when some but not all expectations for a higher letter grade are met.

Important details & caveats:

- The goal of sharing this specific information is to provide transparency around final grades, hence clear goals to work toward. That said, assigning grades is much more nuanced than *any* grading rubric / framework might suggest (for good reasons!).

- Note that the cutoffs for letter grades above may be altered at the end of the semester to be *lower*, but will not be moved higher than what is listed above (i.e. they would only potentially move in your favor).

Letter Grade	Expectations
A	- Problem Sets: Earn at least a "low pass" score on all PS questions, and $\geq$ 70% "high pass" scores otherwise - Theory Exam: "pass" in under 2 attempts - Technical Reports: $\geq$ 4 "high pass" across all 5 components on each report, and 0 "insufficient / not attempted" "pass" both peer-reviews - Capstone Project: "pass" the project outline, 5-minute talk, and final presentation $\geq$ 3 "high pass" across all 4 components of the final paper - Engagement: - Consistently on-time and engaged in class - Average of 2 discussion posts on Slack per week - Complete MSCS Capstone reflection

B

- Problem Sets: Earn $\geq 60\%$ "high pass" and "pass" scores combined, with *very* few "low pass" or "insufficient / not attempted" scores

- Theory Exam: "pass" in under 2 attempts

- Technical Reports:

  - ≥ 4 "pass" across all 5 components on each report, and 0 "insufficient / not attempted"
  - "pass" both peer-reviews

- Capstone Project:

  - "pass" the project outline, 5-minute talk, and final presentation
  - ≥ 3 "pass" across all 4 components of the final paper

- Engagement:

  - Consistently on-time and engaged in class
  - Average of 2 discussion posts on Slack per week
  - Complete MSCS Capstone reflection

C

- Problem Sets: Earn a majority ($\geq 50\%$) of "high pass", "pass", or "low pass" scores

- Theory Exam: "pass" in 3 attempts

- Technical Reports:

  - ≥ 3 "pass" across all 5 components on each report
  - "pass" both peer-reviews

- Capstone Project:

  - "pass" or "low pass" the project outline, 5-minute talk, and final presentation
  - ≥ 2 "pass" across all 4 components of the final paper

- Engagement:

  - On-time and engaged for $\geq 70\%$ of class periods
  - Average of 1 discussion post on Slack per week
  - Complete MSCS Capstone reflection

> D/NC   D/NC will be given when expectations for a C or higher are not met. To achieve credit for this course, you *must* sit for (and demonstrate effort on) the theory exam, submit all assessments, and pass the capstone project.

Statistical Software

We will use the free, R programming language throughout this course. RStudio (an interface for R) will facilitate our use of R. For this course, you *may not* use the RStudio server. You *must* use the desktop version of R.

- Desktop version: Download for Windows or Mac at https://posit.co/downloads/. **Note**: You first need to download and install R on your computer in order to use the desktop version of RStudio

More detailed instructions on downloading, installing, and getting started with R, RStudio, and an important package called Quarto is available on the R & LaTeX Resources tab. **For this course, you must have an R version 4.2.0 or later installed**!

RStan

Throughout the course we will use the **rstan** package in R, which is a *very* important R package for Bayesian modeling. Install the package by carefully following the directions linked here.

LaTeX

LaTeX is a document preparation system used to create high-quality scientific and technical communications. We will be using LaTeX to typeset mathematical notation. It is freely available, and we will primarily be using it in Overleaf to create collaborative, compiled pdf documents. I will provide .tex templates for Technical Reports and some Problem Sets, and .qmd templates for Problem Sets that involve code. Note that if you have not previously used Overleaf you will need to make an account to get started, but making an account is free.

For using LaTeX within Quarto, you will need to download TinyTex, and either MacTex (Apple) or MiKTeX (Windows) to be able to knit to pdfs. Additional LaTeX resources are available on the R & LaTeX Resources tab.

Prof. Kristin Heysse has a nice guide on getting started with Overleaf that you are welcome (encouraged!) to use as a resource, linked here. All sections of the document other than the TIKz one are relevant for this course.

Accommodations

I am committed to ensuring access to course content for all students, including those with disabilities. If you have a disability, please meet with me early in the semester to discuss your accommodation plan. If you have not yet obtained a plan or are unsure if you have a disability that requires accommodation, please contact the Center for Disability Resources: disabilityresources@macalester.edu, or call 651-696-6275.

Religious Observance

Students may wish to take part in religious observances that occur during the semester. If you have a religious observance/practice that conflicts with your participation in the course, please contact me before the end of the second week of the semester to discuss appropriate accommodations.

In an effort to respect religious diversity, I request that students who plan to observe a religious holiday during scheduled class meetings/ class requirements talk to me about reasonable consideration by the end of the second week of the course.

Title IX

Macalester is committed to providing a safe and open learning and living environment for all students, staff, and faculty. Any community member experiencing sexual harassment, sexual violence, relationship violence, or stalking, is encouraged to seek help and support.

Please be aware that as a faculty member, it is my responsibility to report disclosure about sexual harassment, sexual misconduct, relationship violence, and stalking to the Title IX Office. The purpose of this report is to ensure that anyone experiencing harm receives the resources and support they need. I will keep this information private, and it will not be shared beyond this required report.

You may also contact Macalester's Title IX Coordinator directly (phone: 651-696-6258; e-mail: titleixcordinator@macalester.edu); she will provide you with supportive measures, resources, and referrals. Additional information about how to file a report (including anonymously) is available on the Title IX website.

If you prefer to speak with someone confidentially, or need 24/7 support, there are resources available on- and off-campus to assist you:

- Counseling Services at the Hamre Center – on-campus counseling resource for students **Free, Urgent, Phone Counseling (Press 2) is available to Macalester students anywhere in the world, 24/7/365.** Speak to a licensed mental health counselor 24 hours a day by calling Hamre Center at 651-696-6275, then press or say option 2 when prompted.

- Center for Religious and Spiritual Life – chaplains may be reached via 651-696-6298 or email religiouslife@macalester.edu

- SOS Sexual Violence Services of Ramsey County 651-266-1000 (24-Hour Crisis Hotline staffed by trained advocates)

- Rape, Assault, and Incest National Network (RAINN) 1-800-656-4673 (24-Hour Hotline and Live Chat)

- Additional local and national resources can be found at https://www.macalester.edu/title-ix/find-support/

Land Acknowledgement

I would like to honor the fact that we are on Dakota land. This is the ancestral homeland of the Dakota people who were forcibly exiled from the land because of aggressive and persistent settler colonialism. I make this acknowledgment to honor the Dakota people, ancestors, and descendants, as well as the land itself.

About Your Professor

Hi! Please call me Taylor or Professor Okonek (never Dr. Okonek, it sounds weird to me). My pronouns are she/her. I was born on Madeline Island, Wisconsin, and grew up partially there / partially in White Bear Lake, Minnesota. I went to St. Olaf College for undergrad, and double majored in Mathematics and Religion, with a concentration in Statistics. I am a first-generation college student from a low-income background and was a part of the TRIO McNair program at St. Olaf. I completed my Ph.D. in Biostatistics at the University of Washington, Seattle in 2023. My research interests are in demography, mortality estimation, spatial and survey statistics, and survival analysis (and their intersections!).

When I'm not thinking about / doing Statistics, I like to hike, cook, read, play video games (recently Silksong and Stardew Valley, and Celeste is my favorite game of all time), and hang out with my three cats, Alice, Ellie, Norra. I'm always happy to talk about research if you have any questions, and happy to chat about your interests/goals as well, be they statistics-adjacent or not!

Note

This syllabus is subject to change at any time! Announcements of changes will be made in class or via email.