**Land Acknowledgement**

McGill University is on land which has long served as a site of meeting and exchange amongst Indigenous peoples, including the Haudenosaunee and Anishinabeg nations. We acknowledge and thank the diverse Indigenous peoples whose presence marks this territory on which peoples of the world now gather.

# LING XYZ: Bayesian Data Analysis for Linguistics

## Course Information

| | |
|---|---|
| Instructor: | Roger Yu-Hsiang Lo (`roger.lo@xxx.yyy`) |
| Credits: | 3 |
| Time: | Wed 9:00 AM–12:00 PM |
| Location: | TBD |
| Online discussion forum: | TBD |
| Office hours: | By appointment |

## Course Overview

Bayesian statistics is an increasingly influential analytical framework in linguistics, offering a nuanced, probabilistic perspective beyond traditional hypothesis testing in frequentist approaches. This course introduces the philosophy and computational techniques underlying modern Bayesian statistical methods, with a specific focus on applying these methods to common linguistic data types.

We will begin by reviewing probability—a branch of mathematics essential to the Bayesian framework. From there, we will explore both the conceptual and formal foundations of Bayesian inference, transitioning into how linear models are case in this framework. Through linear models, we will compare Bayesian and frequentist approaches and demonstrate a typical Bayesian workflow. We will then examine the Markov Chain Monte Carlo algorithms that make modern Bayesian modelling possible. In the second half of the course, we will apply Bayesian methods to widely-used statistical models, including multiple linear regression, generalized linear models, and hierarchical models. Alongside these applications, we will address important concepts such as prior selection, model checking, and model comparison.

## Learning Objectives

Upon completion of this course, you will be able to:

- Describe the philosophy behind Bayesian statistics and contrast it with frequentist approaches;
- Explain key terms associated with Bayesian statistics, such as *prior*, *likelihood*, *posterior*, and *MCMC*;

- Conduct Bayesian analyses on typical linguistic data using `brms`, including model specification, setting priors, and drawing samples;
- Understand and address warning and error messages that arise during model fitting;
- Interpret model output and effectively communicate results in a report.

## Prerequisites

This course has no formal prerequisites; however, students should be comfortable programming in R, particularly with data wrangling and visualization. Familiarity with linear (mixed-effects) models and hypothesis testing will be beneficial, although not required.

## Course Materials

### Required

- McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and STAN*. CRC Press, 2nd edition. (The first two chapters are available online)
  - Refer to A. Solomon Kurz's eBook for implementations of in-text examples using `brms`
- van de Schoot, Rens, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G. Tadesse, Marina Vannucci, Andrew Gelman, Duco Veen, Joukje Willemsen, and Christopher Yau. 2021. Bayesian statistics and modelling. *Nature Reviews Methods Primers* 1. (Publisher link)

### Optional

- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian data analysis*. CRC Press, 3rd edition. (Full text)
- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and other stories*. Cambridge University Press. (Full text)
- Johnson, Alicia A., Miles Q. Ott, and Mine Dogucu. 2022. *Bayes rules! an introduction to applied Bayesian modeling*. CRC Press. (eBook)
- Kruschke, John K. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2nd edition. (Full text)
- Kruschke, John K. 2021. Bayesian analysis reporting guidelines. *Nature Human Behaviour* 5:1282–1291. (Publisher link)

## Course Format

Each class meeting will typically consist of two parts: a lecture and a live-coding session. In the lecture portion, we will cover the theoretical, mathematical, or technical foundations of that week's topics. In the live-coding segment, we will run code examples together to see firsthand how these concepts are implemented in practice. You are expected to complete assigned readings prior to class and to actively participate in discussions. Lecture notes will be posted in advance, and you should bring laptops to follow along with coding demonstrations.

## Assessment

- **Homework assignments** (56%; 8% per assignment): There will be seven homework assignments designed to help you apply course concepts and techniques to practical problems. Each assignment will guide you through performing statistical analyses on provided datasets and interpreting or visualizing model outputs. Assignments are **due before class** (see the schedule below for specific dates) and should be submitted electronically via TBD. Late submissions will incur a 10% deduction for every 24 hours past the deadline. However, you have a five-day "grace period" for **ONE** assignment, allowing for a late submission without any point deduction (Life happens sometimes!). If you choose to use this option, please clearly indicate it on your assignment. If you have an emergency, please reach out to the instructor as soon as possible.

- **Final project** (44%; 10% proposal + 34% final report): For the final project, you will perform and report on a Bayesian statistical analysis using a real dataset. You are strongly encouraged to use your own research data, but if not available, you may also select publicly accessible data, as long as it has not been previously analyzed using Bayesian methods, or if it has, that you apply a different model. This project aims to provide you with a reusable script and report template for your future research. Evaluation will focus on adherence to best practices in structuring code, justifying priors, assessing model fit, and interpreting model output (e.g., as recommended in Kruschke, 2021).

  This project has two parts:

  - **Proposal** (10%; due 10/09 [Wed]): A one-page, single-spaced document (excluding tables, figures, and references) outlining your research question, chosen dataset, and preliminary analysis plan.
  - **Final Report** (34%; due 12/13 [Fri]): This report, capped at 10 single-spaced pages (including tables and figures but excluding references), should detail your full analysis and include a link to a repository containing your analysis code.

## Grading Scale

Percentage grades will be assigned for all assessments and converted to final letter grades based on the scale published by the university:

| Let. | % grade | Definition | Let. | % grade | Definition |
|---|---|---|---|---|---|
| A | 85–100 | Excellent performance | C+ | 60–64 | Satisfactory performance |
| A− | 80–84 | | C | 55–59 | |
| B+ | 75–79 | Good performance | D | 50–54 | |
| B | 70–74 | | F | 0–49 | Unsatisfactory performance (fail) |
| B− | 65–69 | | | | |

## Communication

For course-related questions, please follow these steps for the quickest response:

1. Consult this syllabus.
2. Post your question on the online discussion forum or ask classmates.
3. Meet with me during office hours.

For personal questions, feel free to email me directly. I aim to respond within 48 hours.

## Accessibility

- **Accommodation for students with disabilities**: Students requiring academic accommodations due to a disability or medical condition should reach out to Student Accessibility & Achievement. More information is available on this page.
- **Well-being**: Being a student at any level can be challenging. You should always prioritize your well-being if you experience physical or psychological difficulties. Please refer to Student Wellness Hub for resources provided by the university.

## McGill Policy Statements

### Academic integrity

McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures. (See McGill's guide to academic honesty for more information.)

L'université McGill attache une haute importance à l'honnêteté académique. Il incombe par conséquent à tous les étudiants de comprendre ce que l'on entend par tricherie, plagiat et autres infractions académiques, ainsi que les conséquences que peuvent avoir de telles actions, selon le Code de conduite de l'étudiant et procédures disciplinaires. (Pour de plus amples renseignements, veuillez consulter le guide pour l'honnêteté académique de McGill.)

### Language of submission

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French written work that is to be graded. This does not apply to courses in which acquiring proficiency in a language is one of the objectives.

Conformément à la Charte des droits de l'étudiant de l'Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté, sauf dans le cas des cours dont l'un des objets est la maîtrise d'une langue.

### Copyright



### Use of generative artificial intelligence (GenAI) tools

You may choose to use GenAI tools as you work through the assignments in this course. However, you should be aware that the code/text generated by GenAI may by inaccurate, biased, or incomplete. You are ultimately accountable for the work you submit, and any content generated or supported by an artificial intelligence tool must be documented appropriately. The documentation should include what tool(s) were used, how they were used, and how the results from the GenAI were incorporated into the submitted work.

### Extraordinary circumstances

In the event of extraordinary circumstances beyond the University's control, the content and/or assessment tasks in this course are subject to change and students will be advised of the change.

## Tentative Schedule & Topical Outline

| Wk # | Date | Topics | Readings | Due |
|---|---|---|---|---|
| 1 | 08/28 (Wed) | - Software set-up<br>- Probability<br>- Bayes' theorem | - McElreath ch. 1, 2<br>- Install `CmdStanR`, `brms` | |
| 2 | 09/04 (Wed) | - Overview of Bayesian stats | - McElreath ch. 3 | HW1 |
| 3 | Tuesday, September 10, is the add/drop deadline | | | |
| | 09/11 (Wed) | - Linear models | - McElreath ch. 4 | HW2 |
| 4 | 09/18 (Wed) | - Bayesian workflow | - van de Schoot (2021) | HW3 |
| 5 | 09/25 (Wed) | - Causal inference | - McElreath ch. 5, 6 | HW4 |
| 6 | 10/02 (Wed) | - Multiple linear regression<br>- Interaction | - McElreath ch. 7, 8 | HW5 |
| 7 | 10/09 (Wed) | - Markov Chain Monte Carlo | - McElreath ch. 9<br>- Optional: Runtime warnings and convergence problems | Proposal |
| | ~~10/16 (Wed)~~ | Fall reading break (no class) | | |
| 8 | 10/23 (Wed) | - Generalized linear model I | - McElreath ch. 10 | |
| 9 | 10/30 (Wed) | - Generalized linear model II | - McElreath ch. 11 | |

[continued on the next page]

| Wk # | Date | Topics | Readings | Due |
|---|---|---|---|---|
| 10 | 11/06 (Wed) | - Hierarchical models I | - McElreath ch. 13<br>- Optional: An introduction to hierarchical modeling | HW6 |
| 11 | 11/13 (Wed) | - Hierarchical models II | - McElreath ch. 14 | |
| 12 | 11/20 (Wed) | - Missing data<br>- Measurement error | - McElreath ch. 15 | HW7 |
| 13 | 11/27 (Wed) | - Final presentations | | Final report due on 12/13 (Fri) |