Methods of Data Analysis II

University of Toronto
Department of Statistical Sciences
STA303H1 S – Summer 2023

Instructor: Emily Somerset (she/her)
Course Email: sta303@utoronto.ca
Lectures: Mondays and Wednesdays 1PM - 4PM
Location: BA1160
Office hours: Schedule posted on Quercus homepage

Course Information

Course administrative email

Use the email address sta303@utoronto.ca for all administrative inquiries or personal matters (e.g., notifying the teaching team of a personal circumstance). Please note that the sta303@utoronto.ca email will not be monitored after August 31, 2023.

Course web page

All course materials will be posted on Quercus. Course materials provided on Quercus are for the use of students currently enrolled in this course only. Distributing course materials to anyone outside of the course is considered unauthorized use.

Piazza

Piazza is an online discussion board that you can access through Quercus. In general, you should avoid emailing the instructor or the TAs regarding course material. If you have a question pertaining to course material, you are encouraged to check Piazza to see if someone has asked a similar question. If not, you are encouraged to post it. You can also come to instructor or TA office hours.

Teaching assistants

Information about TAs and their office hours are posted on the course Quercus page.

Course description

This course is focused on generalized linear models (such as logistic regression, Poisson regression, gamma regression, binomial regression), and models for dependent data (multilevel and longitudinal). You will learn the application and implementation of these models in R, as well as the theory underlying them. This course will also cover additional topics in regression analysis including missingness.

Prerequisites

STA302H1/ STAC67H3/ STA302H5
Prerequisites are strictly enforced by the department, not the instructor. If you do not

have the necessary prerequisites, you will be un-enrolled from the course. If you have questions about this, please contact the Statistics Undergraduate Office at ug.statistics@utoronto.ca.

Learning outcomes

After completing this course, you will be able to

- Develop an intuitive understanding of when to apply each of the models listed above in the course description
- Implement these models in R and understand R code that implements these models
- Be able to prove/derive mathematical properties of these models
- Interpret the results of these models in non-technical language
- Further develop your critical thinking skills regarding statistical modelling

How to succeed

- Come to all lectures and check the course website regularly.
- Do not leave assignments to the last minute. When material is presented in lectures, do the questions on the assignment that pertain to that material as soon as possible.
- If you have a question about the course material, either post it on Piazza or ask it during office hours.
- When studying, focus on the content covered in lectures and assignments, rather than textbooks or past exams you have found.

Reference materials

This course follows no single textbook. Some of the lecture content, examples, and recommended readings come from the following texts. These texts are not required to excel in this class but may serve as useful references.

1. Faraway, Julian J. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press, 2016.
2. Faraway, Julian J. *Linear models with R*. CRC press, 2014.
3. Van Buuren, Stef. *Flexible imputation of missing data*. CRC press, 2018.
4. Harrell, Frank E. "Regression modeling strategies." *Bios* 330, no. 2018 (2017): 14.
5. Dunn, Peter K., and Gordon K. Smyth. *Generalized linear models with examples in R*. Vol. 53. New York: Springer, 2018.
6. Gelman, Andrew, and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.
7. Shumway, Robert H., David S. Stoffer, and David S. Stoffer. *Time series analysis and its applications*. Vol. 3. New York: springer, 2000.

1, 2, 3, 4, 5, and 7 are available online through the UofT Library website. 6 is available as hard copy from the library.

Software

We will be using R and RStudio (including R Markdown) in this course. To get started,

1. First download R from https://www.r-project.org/.
2. Then download RStudio: https://www.rstudio.com/products/rstudio/download/.

You can also use a cloud-based version of RStudio at the University of Toronto (JupyterHub). You can access it at https://jupyter.utoronto.ca by logging in with your UTORid.

Assessments

The assessments for STA303 are listed in Table 1. Your final grade will be calculated as the maximum of Weight Scheme A and Weight Scheme B.

Table 1:

Assessment	Weight Scheme A	Weight Scheme B	Date
Weekly Quizzes	Best 5 out of 10 5% (1% each)	Best 5 out of 10 5% (1% each)	Generally due Wednesdays and Fridays at 11:59pm ET (Some extensions, see Table 2)
Assignment 1	10%	10%	Due Friday July 21 at 11:59pm ET
Midterm Test	30%	0%	Monday July 31 (in-person) 1:30-3:30pm ET
Assignment 2	10%	10%	Due Friday August 4th at 11:59pm ET
Final Exam	45%	75%	TBA

Policy for Quizzes:

- There will be no accommodations for missed weekly quizzes beyond the flexibility already built into the grading scheme (i.e., best 5 of 10). These quizzes are meant as a quick way to check your understanding of the material presented in the most recent lecture. For example, the quiz due on Friday July 14th will cover materials presented on Wednesday July 12th. For all quiz due dates and material covered on quizzes, see the weekly schedule in Table 2.
- Quizzes will be short (no longer than 15 minutes) and will be each worth a total of 1% of your final grade. Weekly quizzes will take place on Quercus. The quizzes may consist of multiple choice, tick all that apply, numerical answers, or matching. Please

note that most of these questions will be auto graded by Quercus in its default manner. For example, for tick all that apply questions, it will deduct partial points if you select incorrect options and add partial points for selecting correct answers.

Policy for Assignments:
- Assignments will be posted on Quercus at least one week in advance of the due date.
- Assignments must be submitted through Crowdmark. Assignments submitted by email will **not** be accepted.
- Assignments must be completed entirely in R Markdown. Both the .Rmd file and .pdf file should be submitted. Assignment submissions without both components will not be graded. Come to office hours or post on Piazza if you are having difficulties knitting your .Rmd file. **If the code in your .Rmd file does not match the knitted .pdf file you submit, you may be accused of academic misconduct.**
- Due to the accelerated nature of this course**, no extensions** will be given on assignments. The assignment deadlines have been planned to spread your work out and allow the TAs to return assessments/solutions to you before midterms and finals. It is crucial that you meet the assignment deadlines.
- Assignments received after the deadline will receive a grade of 0 unless your College Registrar or the Accessibility Office has made alternative arrangements with the STA303 teaching team.

Policy for the Midterm Test:
The midterm test will be held from 1:30 - 3:30 pm on Monday July 31st. The coverage and location of the midterm will be announced on Quercus in advance. If you miss the midterm, the weight will be shifted to the final exam (i.e., Weight Scheme B in Table 1). There is no need to contact the instructors/teaching team in this event.

Policy for the Final Exam:
This will be a 3-hour exam, covering everything in the course. The final exam will be scheduled by the Faculty of Arts and Sciences during the August final exam period. Any scheduling conflicts or petitions to write a deferred exam must go through the Faculty of Arts and Sciences standard procedures; see https://www.artsci.utoronto.ca/current/faculty-registrar/exams-assessments. Be sure not to make travel plans until the final exam schedule has been officially released by the Faculty of Arts and Sciences.

Practice Problems
You are encouraged to use the weekly quiz problems for practice. Some practice problems may be recommended from the reference materials, but these will not be graded.

Regrade Requests:

Marking schemes/solutions will be provided for all assessments except the final exam. You must review these before requesting a regrade. If you still have concerns about your grading, complete the MS Form provided on Quercus to request a regrade **within one week** of the date that the marks for the assessment became available. **Regrade requests will NOT be accepted by email.** Your request should include a detailed explanation of exactly what you think was graded unfairly and why, with reference to appropriate course material. Note that your grade may increase, decrease, or stay the same as a result of a regrade request.

The regrade process for the final exam is handled by the Faculty of Arts and Sciences.

All regrading will be done by the instructor.

Academic integrity:

You, the student, are responsible for understanding the [University of Toronto's Code of Behaviour on Academic Matters](#). Here are some tips to help you avoid committing an academic offense:

- You may use any materials provided to you in the form of code or lecture notes.
- You may discuss assignment questions with other students, but you may not share/copy your answers or code in any way. **It is an academic offense to copy work or allow your own work to be copied.**
- When submitting your assignments via Crowdmark, the code you submit in your .Rmd file must generate the output in the .pdf. If it does not, you may be accused of academic misconduct.
- You must not misrepresent your identity at any point during this course.
- You may be contacted by a private tutoring company trying to sell you services. These companies put you at risk of committing an academic offence. If you hire a tutor, they should help you with understanding course concepts and not completing your assignments.

General Course Policy

Accessibility

The University of Toronto offers academic accommodations for students with disabilities. If you require accommodations, or have any accessibility concerns about the course, the classrooms, or course materials, please contact Accessibility Services as soon as possible at accessibility.services@utoronto.ca or http://accessibility.utoronto.ca.

Communications

Announcements and other course information will be posted on Quercus.

We will use Piazza as the platform for discussions related to the course material and assessments. You can find our Piazza page at https://piazza.com/utoronto.ca/summer2023/sta303. Students can post anonymously to classmates on Piazza, but the identity of the author of all posts is viewable by instructors.

The sta303@utoronto.ca email should be used for inquiries about administrative matters (such as missed assessments) or for personal matters that cannot be shared with the rest of the class. To be fair to all students, we are not able to answer questions about the course material by email; these should be asked on Piazza or during office hours.

Please note that the course email and Piazza will not constantly be monitored by the teaching team. Thus, students are encouraged to respond to each other's questions on Piazza to allow for quicker response times. Depending on the number of emails please allow a reasonable amount of time for email and Piazza responses.

Course Schedule

Table 2

Lecture #	Lecture Date	Topics	Due Dates
1	Wed, July 5	Syllabus Review Prelude to Modelling/ggplot Review of Linear Regression Regression Models	Quiz 1 due on Tuesday July 11 at 11:59pm ET (Covers Lecture 1)
2	Mon, July 10	Missing Data Maximum Likelihood Logistic Regression	Quiz 2 due on Wednesday July 12 at 11:59pm ET (Covers Missing Data)
3	Wed, July 12	Logistic Regression (continued)	Quiz 3 due on Friday July 14 at 11:59pm ET (Covers Maximum Likelihood and Logistic Regression)
4	Mon, July 17	Generalized Linear Models: Exponential families Count data	Quiz 4 due on Wednesday July 19 at 11:59pm ET (Covers TBA)

			Assignment 1 due on Friday July 21 at 11:59 ET
5	Wed, July 19	Generalized Linear Models: Binomial Regression	Quiz 5 due on Friday July 21 at 11:59pm ET (Covers TBA)
6	Mon, July 24	Generalized Linear Models: Gamma Regression Latent Variable Formulation **Midterm Review**	Quiz 6 due on Wednesday July 26 at 11:59pm ET (Covers TBA)
7	Wed, July 26	Mixed Level Models	Quiz 7 due on Wednesday August 2nd at 11:59pm ET (Covers TBA)
	Mon, July 31	**Midterm (2 hours) – location TBA**	
8	Wed, Aug 2	Mixed Level Models Longitudinal Data	Quiz 8 due on Friday August 4th at 11:59pm ET (Covers TBA) Assignment 2 due on Friday August 4th at 11:59 ET
	Mon, Aug 7	**NO CLASS**	
9	Wed, Aug 9	Time Series	Quiz 9 due on Friday August 11th at 11:59pm ET (Covers TBA)
10	Mon, Aug 14	Time Series	Quiz 10 due on Tuesday August 15th at 11:59pm ET (Covers TBA)
11	Tues, Aug 15	Final Exam Review Study space Space to complete Quiz 10	
	Aug 17-25	Final exam (3 hours, scheduled by FAS)	