

Methods of Data Analysis II

University of Toronto
Department of Statistical Sciences
STA303H1 S – Summer 2022

Instructor: Justin Slater
Email: sta303@utoronto.ca
Lecture time: Mon/Wed 1-4pm in BA1160
Instructor office hours: Tues/Thurs 4-5pm on Zoom
TA office hours – schedule posted on Quercus homepage

Course description:

This course is focused on generalized linear models (such as logistic regression, Poisson regression, gamma regression), and models for dependent data (multilevel, longitudinal, time series). You will learn how to apply these models, as well as the appropriate theory. You will also learn how to implement these models in R.

Learning outcomes

- Develop an intuitive understanding of when to apply each of the models listed above
- Implement these models in R, understand R code that implements these models
- Be able to mathematically prove/derive 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, check course website regularly
- Don't 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.
- Pay close attention in labs and in-class examples, it will make assignments much easier.
- If you have a question regarding course or assignment material, post on Piazza or come to office hours
- When studying, focus on the content covered in lectures/labs/assignments, as opposed to the textbooks or past exams you found.

Prerequisites: STA302H1 or STAC67H3 or STA302H5

Pre-requisites are strictly enforced by the department, not the instructor. If you do not have the necessary pre-requisites, 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.

Course Page (Quercus): Important announcements, assignments, lecture materials, and suggested readings will all be posted on the course Quercus page. **It is your responsibility to check it regularly.**

Piazza: Piazza is an online discussion board that you can access through Quercus. In general, you should avoid emailing the instructor regarding course material. If you have a question pertaining to course

material, I encourage you to check Piazza to see if someone has asked a similar question. If not, I encourage you to post about it. You can also come to instructor or TA office hours.

Suggested texts:

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

1. Faraway, J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Non-parametric Regression Models*. CRC press, 2016
2. Dunn, Peter K., and Smyth, Gordon K. *Generalized linear models with examples in R*. Vol. 53. New York: Springer, 2018.
3. Gelman, Andrew, and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.
4. Wood, Simon N. *Generalized Additive Models : an Introduction with R*. Second edition. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2017.
5. Diggle, Peter, Peter J. Diggle, Patrick Heagerty, Kung-Yee Liang, and Scott Zeger. *Analysis of longitudinal data*. Oxford university press, 2002.
6. Shumway, Robert H., and David S. Stoffer. *Time Series Analysis and Its Applications with R Examples*. 3rd ed. New York: Springer. 2010

2, 4, and 6 are available online through U of T's library. The other three are available as hard copies from the library.

Software: We will be using R (RStudio and Rmarkdown) in this course. To get started,

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

The instructor will walk through this on the first day of class.

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

Evaluation details

Labs (10%, best 2 of 3): These are meant to review course material, while improving your computer/R skills. Labs will be done in class on certain Mondays (see course calendar) and will be approximately 1 hour. You should be able to finish these in class, but you will have until Wednesday at 11:59pm to submit them.

Assignments (20%): Completed entirely in Rmarkdown and submitted through Quercus on the assigned due date. Both the .rmd file and .pdf file should be submitted. **The code in your rmd file needs to match the knitted pdf file you submit, otherwise you may be accused of academic misconduct.**

Midterm (25%): The midterm will be held during class time on July 27th and will be two hours long (1:10pm – 3:10pm). The coverage of the midterm will be announced on Quercus in advance.

Final (45%): This will be a 3-hour exam, covering everything in the course. This will be scheduled by the Faculty of Arts and Science for during the August final exam period. The coverage of the final will be announced on Quercus in advance.

Missed/late assessment policy:

If you miss an assessment due to illness, within 2 days after the due date or test date:

1. Fill out the ACORN absence declaration form, AND
2. Email the instructor (at sta303@utoronto.ca) once you have done so with the following 2 sentences:
 - "I affirm that I am ill or having a personal emergency and understand that to falsely claim so is an offense under the Code of Behaviour on Academic Matters."
 - "I understand the weight of this assessment will be reallocated as per the scheme in the syllabus."

If you have followed BOTH steps 1 and 2, then missed assessments will be distributed among the remaining assessments using the following scheme:

1. Your first missed lab will automatically count as your dropped lab. The weight of a subsequent missed lab will be distributed the other lab. At most one lab may be accommodated this way. 0% will be recorded for a 3rd missed lab.
2. Missing assignment 1: 2/3 of weight will be applied to assignment 2, 1/3 to the midterm.
3. Missing the midterm: all the weight will be moved to the final.
4. Missing Assignment 2: 1/2 of weight to assignment 1, 1/2 to the final

If more than one assessment is missed due to illness, then the above scheme will be applied sequentially.

Due to the accelerated nature of this course, **no extensions** will be given on labs or assignments.

Regrade Policy:

All regrade requests must be sent to sta303@utoronto.ca within one week of the grade's release date. 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.

Communication:

If you have a question about course material, please ask this on piazza or during office hours.

If you need to email the instructor (sta303@utoronto.ca) for personal reasons, please:

- use your University of Toronto email address
- Include your UTORid in the email

If you don't follow these instructions, your email may be ignored.

Intellectual Property:

Lecture slides, assignments, labs, tests, and solutions are the intellectual property of the instructor. Do not share these with anyone not enrolled in the course. Posting or **sharing course materials with any person or company without access to our Quercus site is considered stealing.**

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 I provide you in the form of code or lecture notes. You are responsible for the results.
- You may discuss labs and assignment questions with other students but may not share/copy answers or code in any way. **Copying, or allowing your work to be copied are both examples academic offences.**
- When submitting your labs/assignments via Quercus, the code you submit must match the output. If it does not, this is a form of academic misconduct.
- Do 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 concepts and not your assignments/labs.

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: accessibility.services@utoronto.ca or <http://accessibility.utoronto.ca>.

(Tentative) Week-by-week schedule: An up-to-date version of this table will be on the Quercus Homepage. **Visit it regularly.**

	Lecture Date	Topics	Due Dates
1	Mon, July 4	Course introduction Linear regression review Intro to Rmarkdown	
2	Wed, July 6	Maximum likelihood Logistic regression	
3	Mon, July 11	Logistic regression Lab 1: Latex and logistic regression	
4	Wed, July 13	Generalized Linear Models	Lab 1 Due Wed July 13 th at 11:59pm
5	Mon, July 18	Generalized Linear Models Lab 2: Dplyr and GLMs	
6	Wed, July 20	Mixed/Multilevel models	Lab 2 due Wed July 20 th at 11:59pm Assignment 1 due Friday July 22 nd at 11:59pm

7	Mon, July 25	Mixed/Multilevel models Midterm review	
8	Wed, July 27	Midterm (2 hours)	Midterm during class on July 27 1:10pm-3:10pm
	Mon, Aug 1	NO CLASS	
9	Wed, Aug 3	Mixed/Multilevel models Longitudinal models	
10	Mon, Aug 8	Longitudinal models GGplot and GLMMs	Lab 3 due Wed Aug 10 at 11:59pm
11	Wed, Aug 10	Time series	Assignment 2 due Friday Aug 12 th at 11:59pm
12	Mon, Aug 15	Time series Final Exam review	
	Aug 17-25	Final exam (3 hours) to be scheduled by FAS	

****Note:** generalized additive models may be covered, depending on how quickly we get through mixed and longitudinal models

Important Academic Dates:

July 4, 2022	First day of class
Aug 1, 2022	Civic Holiday: no class
Aug 1, 2022	Last day to drop course without academic penalty
Aug 15, 2022	Last Day of classes
Aug 17, 2022 – Aug 25, 2022	Final exam period