STA302H1S: Methods of Data Analysis I

Larry Dong

2025 Summer Sub-session S

E-mail: sta302@course.utoronto.ca
Class hours: Mondays 6-9pm (PB B250)
Wednesdays 6-9pm (PB B250)
Office hours: TBD

1 COURSE OVERVIEW

*** This is an in-person course. Please note that lectures and evaluations will take place during the lecture times above and you are expected to be available during those times.**

Course description

The course provides an introduction to data analysis with a focus on the theory and application of linear regression. Topics to be covered include: initial examination of data, correlation, simple and multiple regression models using least squares, inference for regression parameters for normally distributed errors, confidence and prediction intervals, model diagnostics and remedial measures when the model assumptions are violated, interactions and dummy variables, ANOVA, and model selection and validation. Statistical software will be used throughout and required for the completion of various assessments during the term. The development of strong written communication skills will be emphasized. All code and slides presented in the lectures will be available on Quercus.

Learning outcomes

In this course, students will learn how to:

1. Recognize the importance of assumptions and limitations of linear regression models to gauge when linear models are appropriate to use and to be critical of their results.
2. Interpret the results of an analysis involving linear models for technical and non-technical audiences.
3. Apply methods of linear models and data cleaning to new datasets correctly using statistical software in a reproducible way.
4. Explain statistical concepts and theory of linear models to various audiences as would be required in the job market or collaborative environment.
5. Outline the correct use of linear models in a coherent and reproducible analysis plan.

Prerequisites

Prerequisites are **strictly enforced by the department, not the instructor**. Students should have completed the following:

- A second year statistics course;
- A computer science course;
- A linear or abstract algebra course.

Please see the course webpage for accepted pre-requisite courses. If you do not have the equivalent pre-requisites, you will be un-enrolled from the course.

2 COURSE MATERIALS

Course Content

All lecture slides and materials will be posted on the Quercus course page before each lecture. Further, any important announcements will also be posted in Quercus. Please make sure to check it regularly so you don't miss anything.

Required materials

This course will be closely the following textbook:

- Sheather, S.J. (2009). *A Modern Approach to Regression with R*. Springer.

We will cover Chapters 1-7, with suggested practice problems also primarily chosen from this book. However, the other following textbooks may also be helpful:

- Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to Linear Regression Analysis*. Wiley.
- Kutner, M.H., Nachtsheim, C.J., & Neter, J. (2004). *Applied Linear Regression Models*. McGraw-Hill/Irwin.

All textbooks are freely available via the university's online library; there is no requirement to purchase physical copies. All course content will mostly be self-contained in lecture materials.

Statistical Software

We will be using R via RStudio to perform statistical analyses. R is a free software that can either be downloaded onto your personal computer or used in the cloud. If you choose to work with R on your personal computer, then installation will be a two step process:

1. The base R framework is available for download at `http://cran.r-project.org/` for Windows, Mac and Linux operating systems.
2. RStudio is a good integrated development environment to R, i.e. making it simpler to work in R, and can also be downloaded for free at `https://www.rstudio.com/products/rstudio/download/`.

If you don't want to download the program or run into problems with installation, you may want to consider University of Toronto JupyterHub with RStudio selected which only requires you to login with your `utoronto` email and connect to our course project via the link provided. In lectures, examples with R syntax will be provided, which should be sufficient for you to learn how to apply the statistical methods.

3 COURSE COMPONENTS

Lectures

Lectures will be given in-person, in PB B250. During lectures, we will cover important course materials, as well as cover a number of examples illustrating the uses of these methods. Lecture slides will contain some R code and output to show how to perform these methods in practice. Each lecture builds on the material from previous weeks, so it is recommended that you attend lectures regularly/keep on top of the material.

Office Hours

The instructor and TAs will hold office hours through a combination of Zoom and in person. The office hour schedule will be posted on Quercus when it is finalized. It is recommended that you visit office hours whenever you have a question about the material. It is more important than ever in an accelerated class to have material clarified as quickly as possible.

Quercus Discussion Board

We will be using the Quercus Discussion Board as an online discussion forum. **All questions about course material should be first posted there or asked during TA/instructor office hours.** Other course-related questions can be sent to the course email. The instructor and TAs will monitor the board and will help answer questions but students are also encouraged to answer posts and help their fellow classmates.

Note on etiquette: *Please make sure that you communicate politely and respectfully with all members of the teaching team and your fellow classmates. Written communications can sometimes take a tone other than what was intended (e.g. can come off as dismissive, rude or insulting), so make sure you re-read your email/post before sending it to make sure it conveys the tone that you intended. Any posts that detract from the learning goal of the board will be removed to keep the board a safe space.*

Communicating with your instructor

If you have a question about course material, please post it on the Quercus discussion board. Likely, another student has the same question. If you have a more personal question or want to appeal a grade (see section below), send me an email at `sta302@course.utoronto.ca`. I will do my best to respond within two business days.

Appealing a grade

Assignment grades may be appealed. There will be a time frame (communicated to you after grades have been released) during which you may appeal your grade. To do so, you should send

complete the following form: https://forms.office.com/Pages/DesignPageV2.aspx?origin=ShareFormPage&subpage=design&m2=1&id=JsKqeAMvTUuQN7RtVsVSENzKM__cCCJDuaI3yf-4OGxUMlZNTllSTVBYOU
In the form, please include a complete justification of why you believe the grade given is in error. If you are requesting a reread about the midterm or final, please include the question that you are concerned about. My decision about grade appeals is final. Appeals will be accepted for one week after grades have been released.

Late Penalty Policy

Late submissions will receive an automatic 5% penalty per day late (including weekends and holidays) following the deadline. If the assignment has not been submitted one week after the deadline without an approved extension, you will receive a **zero**.

Accommodation and extensions

If you wish to be accommodated for an assessment, please fill out the following form: link.

Course Structure & Grading Scheme

Assessment	% of grade	Details	Due/exam date(s)
Project proposal	12%	Dataset & research question proposal, EDA, preliminary analysis	Friday, July 18, 2025, 11pm EST
Midterm	25% or 0%	Materials included are from lectures 1 to 5 inclusively	Monday, July 21, 2025, 6pm – 7:30pm EST in PB B150 or PB B250
Project report	23%	Full report on data analysis	Friday, August 15, 2025, 11pm EST
Final exam	40% or 65%	All lectures	Sometime during exam period of August 16 – 25, 2025

Your final grade will be calculated according to the following formula:

$$\begin{aligned}\text{Final grade} = 12\% \times \text{Project proposal} \\ + 23\% \times \text{ Project report} \\ + \max\left\{25\% \times \text{Midterm} + 40\% \times \text{Final exam} \ , \ 65\% \times \text{Final exam}\right\} .\end{aligned}$$

4 Policies & Statements

Accessibility Statement

Students with diverse learning styles and needs are welcome in this course. If you have an acute or ongoing disability issue or accommodation need, you should register with Accessibility Services (AS) at the beginning of the academic year by visiting https://studentlife.utoronto.ca/department/accessibility-services/. Without registration, you will not be able to verify your

situation with your instructors, and instructors will not be advised about your accommodation needs. AS will assess your situation, develop an accommodation plan with you, and support you in requesting accommodation for your course work. Remember that the process of accommodation is private: AS will not share details of your needs or condition with any instructor, and your instructors will not reveal that you are registered with AS.

Equity, Diversity, and Inclusion Statement

The University of Toronto is committed to equity, human rights and respect for diversity (see https://research.utoronto.ca/equity-diversity-inclusion/equity-diversity-inclusion for additional details). All members of this course must abide by the university's mandate in fostering a welcoming and inclusive space for all. The University does not condone any harassment-related nor discriminatory behavior that may affect any individual or community. As a Course Instructor, I will neither condone nor tolerate harassment-related behavior that undermines the dignity or self-esteem of any individual in this course. and wish to be alerted to any attempt to create an intimidating or hostile environment.

Academic Integrity Statement

The University treats cases of cheating and plagiarism very seriously. The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic dishonesty and the processes for addressing academic offences. Potential offences in papers and assignments include using someone else's ideas or words without appropriate acknowledgement, submitting your own work in more than one course without the permission of the instructor, making up sources or facts, obtaining or providing unauthorized assistance on any assignment. On tests and exams, cheating includes using or possessing unauthorized aids, looking at someone else's answers during an exam or test, misrepresenting your identity, or falsifying or altering any documentation required by the University.

***Statement on the Use of Generative Artificial Intelligence*:** *Use of generative AI is permitted to check grammar and sentence structure, but it is prohibited to generate answers or complete sentences in lieu of your own writing. If I suspect that you have used AI tools to write significant portions of your assessments, an investigation may be conducted.*

Mental Health and Well-Being

Your mental health is important. Throughout university life, there are many experiences that can impact your mental health and well-being. As a University of Toronto student, you can access free mental health and well-being services at Health & Wellness (https://studentlife.utoronto.ca/department/health-wellness/) such as same day counselling, brief counselling, medical care, skill-building workshops, and drop-in peer support. You can also meet with a Wellness Navigation Advisor who can connect you with other campus and community services and support. Call the mental health clinic at 416-978-8030 ext. 5 to book an appointment or visit https://uoft.me/mentalhealthcare to learn about the services available to you. You can also visit your College Registrar to learn about the resources and supports available: https://www.artsci.utoronto.ca/current/academic-advising-and-support/college-registrars-offices.

If you're in distress, you can access immediate support: https://uoft.me/feelingdistressed

Intellectual Property Statement

Course material – including but not limited to lecture slides, assessment questions, and other supplementary course material available on Quercus – is the intellectual property of the teaching team and is made available to you for your personal use in this course. Sharing, posting, selling, or using this material outside of your personal use in this course is not permitted under any circumstances and is considered an infringement of intellectual property rights.

Recording of Classroom Material by Students Statement

Recording or photographing any aspect of a university course – lecture, tutorial, seminar, lab, studio, practice session, field trip, etc. – without prior approval of all involved and with written approval from the instructor is not permitted.

Acknowledgement of Territory

I wish to acknowledge this land on which the University of Toronto operates. For thousands of years it has been the traditional land of the Huron-Wendat, the Seneca, and most recently, the Mississaugas of the Credit River. Today, this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land together.

Tentative Schedule

Here is a tentative of the topics that will covered in class. This schedule is subject to change.

Lecture	Date	Content
1	Wed, July 2	**Introduction**: communication and exploratory data analysis.
2	Mon, July 7	**Simple linear regression**: functional and statistical relationship, least square estimation, and interpretation.
3	Wed, July 9	**Inference for simple linear regression**: estimation properties, hypothesis testing, confidence interval and prediction interval
4	Mon, July 14	**Regression assumptions**: residual plots, influential & leverage points outliers)
5	Wed, July 16	**Multiple linear regression**: vector and matrix operations, multiple linear regression model, least squares, polynomial regression. *Project proposal due: Friday, July 18, 2025 at 11pm EST*
6	Mon, July 21	**Midterm test** followed by short lecture on multicollinearity
7	Wed, July 23	**Inference for multiple linear regression**: statistical properties, hypothesis testing, best linear unbiased estimator (BLUE)
8	Mon, July 28	**Variance decomposition** (part 1): sum of squares decomposition, ANOVA, F tests, coefficients of determination.
9	Wed, July 30	**Variance decomposition** (part 2)
*	Mon, Aug. 4	*Civic holiday - University closed; no classes.*
10	Wed, Aug. 6	**Writing workshop**: how to write reports, final project teamwork session
11	Mon, Aug. 11	**Model building and variable selection**: interaction terms, forward and backward variable selection, LASSO
12	Wed, Aug. 13	Make-up lecture if necessary, review or Q&A *Project report due: Friday, August 15, 2025 at 11pm EST*