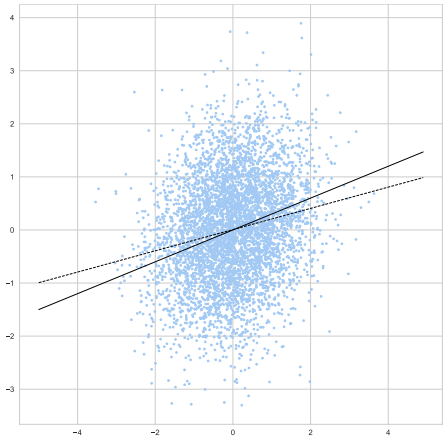
STA302H1F: Methods of Data Analysis 1
LEC5101 — Fall 2023

Instructor Information

Name: Austin Brown
Office: Department of Statistics 9183
Email: sta302@utoronto.ca
Office Hours: TBA



Course Information

Time: Wednesday 18:00 – 21:00
Classroom: AH 100

Course Description

The course provides a solid 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 will be required for the completion of various assessments during the term. The development of strong written communication skills will be emphasized.

Course Goals

By the end of this course, all students should be able 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.
6. Apply and extend linear model theory through completion of problem-solving questions

Prerequisites

Pre-requisites are strictly enforced by the department, not the instructor. If you do not have the equivalent pre-requisites, you will be un-enrolled from the course. Students should have a second year statistics course, such as STA238, STA248, STA255, or STA261, a computer science such as CSC108, CSC120, CSC121, or CSC148 and a mathematics course such as MAT221(70%), MAT223, or MAT240 or equivalent preparation as determined by the department.

Course Materials

Course Content: We have a common Quercus course page. All lecture slides, any recordings and materials will be posted on this Quercus course page. Further, any important announcements will also be posted in Quercus. Please make sure to check it regularly.

Textbook This course does not strictly follow any particular textbook, but rather merges material from a number of sources. All of the below recommended textbooks are freely available as an electronic copy through the University of Toronto Library.

- A Modern Approach to Regression with R, by Simon J. Sheather (Springer)
- Linear Models in Statistics, 2nd edition by Alvin C. Rencher and G. Bruce Schaalje (Wiley)

Statistical Software We will be using the R Statistical Software for performing statistical analyses in this course. R is a free software that can either be downloaded onto your personal computer or used in a cloud environment. We encourage all students to use RStudio through the JupyterHub for University of Toronto. This will allow you to login with your official UofT credentials and use RStudio without the need for a local installation and can be run on any device that has access to an internet connection. More information about using RStudio in JupyterHub will be provided early in the term. R code shown in class will be available on the course page and, along with any additional resources, should be sufficient to complete any assessment involving data analysis.

- University of Toronto Jupyter Hub

Course Components

Lectures: Lectures will be conducted in person. Slides will be available on Quercus. Class time each week will comprise of a combination of lecturing, in-class activities, and code-along sessions. Where possible, you are encouraged to bring a laptop or tablet to follow along with the code.

Office Hours: Instructor and TA's will hold office hours in a combination of online and in-person formats. The office hour schedule and mode of delivery will be posted on Quercus once finalized. It is recommended that you visit office hours whenever you have a question about the material. It is always important to have material clarified as quickly as possible. Don't wait until the last minute to ask your questions!

Discussion Board: We will be using the ED-STEM Discussion Board as an online discussion forum, which can be accessed through the Quercus course page. **All questions about course material should be posted here or asked during TA/instructor office hours.** The instructor and TAs will monitor the board and will help answer questions but students are encouraged to answer posts and help their fellow classmates.

Communication

How your instructor will communicate with you: All communication will be made through Quercus announcements or during lectures. Please ensure that you check Quercus regularly so you don't miss anything important.

Where to send content questions: We will be using the Discussion board to collect student questions regarding course content, assignments, etc. All questions should be posted here.

When to email the instructor: The instructor will only respond to emails of a private or sensitive nature. If you email the instructor with content related questions, you will be asked to repost your question on the content board so the answer may benefit all students. Should you need to email the instructor about a sensitive or personal nature, please use your official mail.utoronto.ca email, include your full name, student number and lecture section in the text. Send all course related emails to sta302@utoronto.ca. The subject (title) of the emails must start with "LEC5101 -". Please allow up to 48 hours for a reply. Emails will not be monitored on evenings and weekends.

A note on email and discussion board 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 or read out loud your email/post before sending it to make sure it has the tone you intended. For more tips on respectful communication, see professional communication tips. The ED discussion board is a teaching and learning tool and therefore should only be used as such. Any posts that detract from the learning goal of the board will be removed to keep the board a safe space.*

Grading Scheme

Each student's final grade will be computed according to the below grading scheme. No special rounding rules or individual grade adjustments (e.g. to meet GPA cut-offs, minimal requirements for programs, etc.) will be used to calculate course grades. No special reweighting of assessments or extra work will be accepted to account for perceived poor performance, nor to account for any assessment(s) that have been missed without accommodation. There are no exceptions to these policies.

Assessment	**Date Due**	**Grade**
Enagement activities (2 parts)		
Syllabus quiz	Sept. 26 6PM	1%
In-class questions / activities		4%
Term test	Oct. 18 6PM	20%
Final project (3 parts)		
Part 1: Research question / proposal	Oct. 11 6PM	5%
Part 2: Analysis flowchart	Nov. 1 6PM	5%
Part 3: Final draft / revision	Dec. 6 6PM	10%
Final exam (during finals period)	Scheduled by FAS	55%

Please note that the last day to drop the course without penalty is November 6, 2023.

Grading Scale

Final grades are assigned according to the following Univerity of Toronto grading scale: University of Toronto grading scale

Evaluation Breakdown

Engagement Activities:

- **Syllabus and Pre-requisite Quiz:** There will be 1 short multiple choice quiz early in the term to ensure that students are prepared for the course in terms of their knowledge of prerequisite material and the syllabus content. This quiz will be conducted on Quercus and will be open for students to take at any time until the deadline. Students will get 2 attempts and the highest score will be counted towards their final grade. On each attempt, students will be given 1 hour to complete the quiz, and each question will show up one at a time and will be locked once the question has been answered.
- **In-class Questions / Activities:** There will be questions asked in class and in class activities throughout the semester. Credit will be assigned for participation.

Term Test: The term test will be conducted in person during the scheduled Wednesday class time. The test will be approximately 1 hour and 40 minutes long. More details will be communicated closer to the test date. The term test will take place during each section's scheduled lecture time. The test will cover all material from Modules 1 to 5. All students will be required to write the test in the lecture time in which they are enrolled to ensure sufficient test papers are available.

Final Project: The final project will consist of a data analysis on a dataset. Students will be required to demonstrate their understanding of the methods taught in the course by developing a reasonable regression model that addresses a valid research question using the techniques from the course. The students will be responsible for choosing the correct methods to apply and providing appropriate justifications defending their choices. The final project is a scaffolded assessment involving three parts:

- Part 1- Research question: Students will be tasked with defining a research question that can be answered with a dataset using linear regression. This portion of the project will require students to provide their research question, explain why linear regression would be a reasonable method to answer this question, and highlight important characteristics of their dataset.
- Part 2 - Analysis Plan Flowchart: Students will be asked to put together a flowchart outlining the steps that they plan to take in their data analysis for the final project on their chosen dataset. This will help in developing a consistent analysis flow and in structuring the methods section of their final report.
- Part 3 - Final Project Report: Students will put together a scientific report that outlines the relevance of their proposed research question, the process of their analysis, the results of the performed data analysis, and a discussion of the meaning of the results as well as limitations of the analysis with respect to the statistical tools used/decisions made or the data used.

All parts of the final project must be done in groups of three to four students. All group members are expected to contribute to the project equally and provide an outline of their involvement in the project. More detailed instructions for each part will be provided on Quercus at a later date.

Final Exam: The final exam will take place during finals week.

Late Assessment and Extension Request Policy

Late Submission Policy:: No late submissions will be accepted throughout this course.

Extreme Situations/Prolonged Illness Extensions: Should a student be experiencing a prolonged illness or other situation that prevents them from turning in their work or contributing to the group project, they should immediately contact their instructor and College Registrar to inform them of their situation. They should also submit an Absence Declaration form on ACORN or a Verification of Illness (VOI) form that lists every day during which they were incapacitated and unable to work. These documentations should be sent to sta302@utoronto.ca. Accommodations will not be considered without a completed declaration or VOI, and will only be considered for extreme circumstances at the request of the College Registrar.

Accessibility-Related Extension Requests: Students registered with Accessibility Services should notify the instructor as soon as possible if additional time is needed on assessments that are eligible for such accommodation. Please notify the instructor by email of your situation and cc your accessibility advisor in the process. The instructor will work with the accessibility advisor to determine an appropriate accommodation for your situation. However, note that group work can generally not be granted further extensions beyond those in the above policy.

Missed Assessment Policy

If you experience a prolonged absence due to illness or emergency that prevents you from completing any number of assessments, please contact your College Registrar as soon as possible so that any necessary arrangements can be made.

Missed Final Project: Due to the nature of this assessment, there will be no extensions on the research proposal or project under any circumstances. Late projects will not be accepted and there are no accommodations available for individuals missed contributions to their groups project

Missed Term Test: If a student is experiencing a serious personal illness or emergency on the date of the test, the student must declare their absence on ACORN and notify the teaching team via email no later than one week after the date of the test. An accommodation will be determined at the discretion of the instructor at a later date. The format of the accommodation could be one of a reweight, a written make-up test, or some other additional piece of work. If a make-up test or additional make-up assessment is missed, the student will receive a 0 on the term test.

Regrade Requests

Regrade requests will be accepted for all assessments. Regrade requests must provide a justification for where there exists a grading error and/or how the work meets the grading rubric. These justifications must further be backed up with concrete references to the course material. All regrade requests will be accepted through a form available on the Quercus course page and will be accepted no later than one week after the grade for that assessment is released. No regrade requests will be accepted by email or after the 1 week deadline. **The instructor/TA will re-grade the assessment in its entirety (i.e. grades can go up or down)** . Please allow a few weeks for regrade requests to be processed by the instructor.

Intellectual Property

Course materials provided on Quercus, such as lecture slides, assessments, videos and solutions are the intellectual property of your instructor and are for the use of students currently enrolled in this course only. Synchronous sessions will be recorded and be made available to other students enrolled in the course. **Providing course materials to any person or company outside of the course is unauthorized use and violates copyright.**

Acceptable Uses of Generative Artificial Intelligence

ChatGPT and other generative AI are freely available tools that can perform a variety of functions for us. However, it's important to understand how such tools are allowed to be used in this course. Acceptable uses of generative AI in this course include:

- Editing or rephrasing written work that has already been written by the student to improve the syntax, grammar and overall readability of the work.
- Synthesizing or explaining course concepts while learning and studying to contribute to their under- standing of the course material
- Looking up appropriate syntax of individual R functions for use in a data analysis or for understanding errors that may arise when running R code.

However, the work turned in by students must ultimately be their own and students will therefore be accountable for the work they turn in. Unacceptable uses of generative AI in this course include:

- Copying from any generative artificial intelligence applications, including ChatGPT and other AI writing and coding assistants, for the purpose of completing assignments in this course.
- Producing an entire data analysis, written report, or any other piece of work meant for grades.

In summary, generative AI like ChatGPT can be really helpful in your learning process and to improve skills valued in the workplace. However, it cannot be used as a substitute for learning and material produced from these tools should not be passed off as your own. This would be considered academic misconduct (see below). The instructor therefore reserves the right to ask students to explain their work and their process for creating their assignment.

Academic Integrity

The University treats cases of plagiarism and cheating very seriously. It is the students' responsibility for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters. All suspected cases of academic dishonesty will be investigated following procedures outlined in the above document. If you have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, you are expected to seek out additional information on academic integrity from your instructor or from other institutional resources (see http://academicintegrity.utoronto.ca). Here are a few guidelines regarding academic integrity:

- Using ChatGPT and other generative AI for any purpose not outlined above.
- Being dishonest when reporting an illness or personal emergency to get an extension or accommoda- tion is an academic offence.
- You may consult class notes/lecture slides during assessments, however sharing or discussing questions or answers with other students is an academic offence.
- Students must complete all assessments individually. Working together is not allowed unless otherwise specified.
- Paying anyone else to complete your assessments for you is academic misconduct.
- Completing assessments for another student is academic misconduct.
- Sharing your answers/work/code with others is academic misconduct.
- Using sources external to the course (anything not on Quercus) on an assessment is an academic offence.
- All work that you submit must be your own! You must not copy mathematical derivations, computer output and input, or written answers, etc. from anyone or anywhere else. Unacknowledged copying or unauthorised collaboration will lead to severe disciplinary action, beginning with an automatic grade of zero for all involved and escalating from there. Please read the UofT Policy on Cheating and Plagiarism, and don't plagiarise.

Accessibility Needs

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

CIA's University Accreditation Program and Pathway to Actuarial Credential

This course is one of the mandatory courses under Canadian Institute of Actuaries (CIA)'s University Accreditation Program (UAP). UAP has moved away from the course-by-course accreditation method and towards program accreditation method (the "Pathway 1 of CIA qualification"). Under the new pathway, in order to obtain ACIA (Associate of CIA) professional credential, students need to:

1. Complete a degree from an actuarial program (ACT Specialist or Major) at University of Toronto and pass a list of mandatory courses. No minimum course grade or GPA is required as long as students pass all the mandatory courses. The full list of UofT's 16 mandatory courses are: ACT240, ACT245, ACT247, ACT348, ACT349, ACT370, ACT451, ACT452, ACT466, STA257, STA261, STA302, STA314, ECO101, ECO102, MGT201/RSM219. For transition: CIA will accept an actuarial degree from UofT completed between June 30, 2015 and October 31, 2023 without all the specified mandatory courses.

2. Complete the ACIA Module (administered by CIA, projected Spring 2023). For transition: a student can be exempt from the ACIA Module if they complete SOA exam PA and the 8 FAP Modules and assessments by December 31, 2023.

3. Complete an open-book ACIA Capstone Exam (administered by CIA, projected Fall 2023). For transition: a student can be exempt from the capstone exam by completing any combination of UAP credits or exams for P, FM, IFM, LTAM, STAM and SRM by October 31, 2023. The deadline to apply for UAP credits is September 30, 2023.

Details on the new pathway for students can be found here: https://education.cia-ica.ca/acia/acia-for-accredited-university-students/.

Important Dates and Deadlines

For a complete list of due dates and synchronous class times, see the attached calendar. It is recommended that you save/print the calendar and/or copy the dates to your personal calendar to make it easier to stay on track with the course.

Tentative Schedule of Topics

Below is a tentative schedule of topics to be covered in class. The schedule is subject to change and modification.

Week	Topic
Sept. 11-15	**Simple Linear Regression Basics:** relationships, notation, estimation, inter- pretation.
Sept. 18-22	**Multiple Linear Regression Basics:** relationship, notation, estimation, inter- pretation.
Sept. 25-29	**Assumptions of Linear Regression:** introduction to assumptions, residuals and residual plots, detecting violations.
Oct. 2-6	**Correcting Assumptions:** Transformations, interpretation, role in sampling distributions.
Oct. 9-13	**Inference in Linear Regression:** Hypothesis tests and/or confidence intervals on coefficients and mean responses, prediction intervals
Oct. 16-20	TERM EXAM
Oct. 23-27	**Decomposition Of Variance Part 1:** Sum of squares decomposition, ANOVA F test, Partial F test
Oct. 30-Nov. 3	**Decomposition of Variance Part 2:** coefficients of determination, multicollinearity
Nov. 6-10	READING BREAK
Nov. 13-17	**Problematic Observations:** Outliers, Leverage Points, Influential Points, De- tection and Impact
Nov. 20-24	**Model Building and Variable Selection:** Context, likelihood criteria, auto- mated methods, hypothesis tests.
Nov. 27-Dec. 2	**Model Validation and Wrap-up:** How to validate your models, MLR data analysis process overview, writing workshop on reports
Dec. 4-8	**Review/Office Hour class:** Final exam review and/or questions and answers.
Dec 10-20	FINAL ASSESSMENT PERIOD