STA238H1: Probability, Statistics and Data Analysis II
Winter 2021

Sections L0101 & L0201

Instructor: Prof. Samantha-Jo Caetano

Email: s.caetano@utoronto.ca

Office hours: Wed 11am - 12pm ET, Thur 4pm - 5pm ET (second hour of the lecture)

Preferred pronouns: she/her

Course administrative email

Use the email address sta238@utoronto.ca for all administrative inquiries, including missed assessments and re-mark requests. Please note that this email address will not be monitored after April 30, 2021.

Course web page

All materials will be posted on Quercus https://q.utoronto.ca. 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.

Teaching assistants

See the course Quercus page for information about TAs, office hours and contact.

Calendar description

An introduction to statistical inference and practice. Statistical models and parameters, estimators of parameters and their statistical properties, methods of estimation, confidence intervals, hypothesis testing, likelihood function, the linear model. Use of statistical computation for data analysis and simulation.

Required prerequisites

A course in probability: STA237H1/STA247H1/STA257H1/STAB52H3/STA256H5

It is also assumed that you are comfortable with the calculus prerequisites of the required probability prerequisite.

Please note that all prerequisites for all STA courses are strictly enforced and your instructor cannot waive them. Any questions about prerequisites should be directed to ug.statistics@utstat.utoronto.ca.

Class format

The course is scheduled for 4 hours per week. We will be using a flipped class for most lessons in this class. Most weeks, there will be a set of lecture videos available early in the week (posted Monday morning) to watch prior to the Wednesday/Thursday synchronous class. It is recommended that you use the lecture times on Mondays/Tuesdays to watch these asynchronous lecture videos. The second 2 hour time slot (on Wednesdays/Thursdays) will be held synchronously through BB Collaborate (in our Quercus page). Typically the 2-hour class meeting on Wednesday/Thursday will be one hour of synchronous lecture, followed by a second hour scheduled for extra help/office hour.

Suggested Weekly Routine

Monday & Tuesday	Wednesday & Thursday	Friday
Watch weekly asynchronous videos and work on assigned homework.	Attend synchronous lecture and office hours and extra help.	Work on weekly assigned homework and/or upcoming assessment.

Note: All synchronous lectures will be recorded and made available to all students in the course.

Accessibility needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, 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://www.accessibility.utoronto.ca.

Computing

Computational work is a central part of developing statistical thinking and developing facility in the use of computational tools for carrying out simulations and data analysis is a core objective of this course. We will use R, the R Studio IDE, and R Markdown. All of these are freely available. You need to first install R, and then R Studio. R can be downloaded for free from http://cran.r-project.org. R Studio can be downloaded for free from http://www.rstudio.com/products/rstudio/download/. Additionally, you can also use R Studio through the U of T Jupyterhub, by selecting the RStudio option and logging in with your utorID and password, available here: https://jupyter.utoronto.ca

Some resources for using R and R Markdown:

- The course supplementary notes give guidelines on installing and getting started with R and R Studio.
- A short intro to R workshop is available here: https://awstringer1.github.io/ssu-r-workshop/ssu-r-workshop.html
- Hands-On Programming with R by Garrett Grolemund, available here: https://rstudio-education.github.io/hopr
- R for Data Science by Hadley Wickham and Garrett Grolemund, available here: https://r4ds.had.co.nz
- An R Markdown Cheat Sheet is available at https://rstudio.com/resources/cheatsheets

Reference Materials

We will be relying on material from the following three textbooks. All books are freely available, and we expect all students to have access to each.

[MIPS] F.M. Dekking, C. Kraaikamp H.P. Lopuhaä and L.E. Meester (2005). A Modern Introduction to Probability and Statistics: Understanding How and Why. Springer-Verlag.
This is the primary reference for the course.
Available for free as in pdf through the UofT library here.

[E&R] M.J. Evans and J.S. Rosenthal (2003). Probability and Statistics: The Science of Uncertainty. W.H. Freeman and Co.
Available in pdf here: http://www.utstat.toronto.edu/mikevans/jeffrosenthal.

[ISL] G. James, D. Witten, T. Hastie and R. Tibshirani (2013). An Introduction to Statistical Learning with Applications in R. Springer.

Supplementary Materials

[STA238supplement]
Supplementary materials have been written to correspond to the sections we will be covering in each of the above textbooks. These materials emphasize the use of computation and include sample code and exercises. You should first read the relevant chapters in the textbooks, before reading the corresponding material in STA238supplement. The supplementary materials are available at
https://awstringer1.github.io/sta238-book/index.html.

Practice Problems

There are several opportunities to practice the course material. Practice problems from the textbooks will be assigned for each topic. It is strongly recommended that you complete all of the "quick exercises" in MIPS and all of the assigned practice problems. Solutions for many of the assigned practice problems are available in the back of MIPS. In addition, you should complete the exercises provided in STA238supplement.

Course Materials

All course materials are copyrighted. If they are from the textbook, the copyright belongs to the textbook publisher. If they are provided by an instructor (for example, lecture notes, computer code, assignments, tests, solutions) the copyright belongs to the instructor. Distributing materials online or sharing them in any way is a copyright violation and, in some situations, an academic offence.

Communication

Tentatively, we will be using Black Board Collaborate (BB Collaborate) for most synchronous meetings in the course.

We will be using Piazza as the platform for discussions related to the course material and assessments. You can find our course page at:
https://piazza.com/utoronto.ca/winter2021/sta238h1slec0101lec0201.
Students can post anonymously to classmates on Piazza, but the identity of the author of all posts is view able by instructors.

Be sure to read Piazza's Privacy Policy and Terms of Service carefully. Take time to understand and be comfortable with what they say. They provide for substantial sharing and disclosure of your personal information held by Piazza, which affects your privacy. When you use Piazza, only provide content that you are comfortable sharing under the terms of the Privacy Policy and Terms of Use. With that being said Piazza will still be considered a part of our class and thus all posts and conduct on Piazza must remain professional. Posts regarding personal matters such as inquiries about grades, reporting absences, regrade requests, etc. should be communicated via email (at sta238@utoronto.ca) and *NOT* be posted on Piazza. Piazza is intended for students to receive support regarding course information and content and thus should be an overall positive and professional environment.

Again, email is appropriate only for personal matters that can not 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 questions should be asked on the discussion forum or during office hours.

Inquires about administrative matters, such as missed tests and re-mark requests, should be sent to sta238@utoronto.ca. Any emails sent to this address must include in the subject line STA238. Please note that email will not be monitored on evenings or weekends (Toronto time), as well as Piazza. Depending on the amount of emails please allow a reasonable amount of time for email (and Piazza) responses. Please note that this email address will not be monitored after April 30, 2020.

Announcements and other course information will be posted on Quercus.

Course Content

The course will consider three overarching themes in statistical theory and data analysis:

- Understanding Data: methods for describing data numerically and graphically, error and statistical models
- Making Inferences from Data: how data can be used to explain phenomenon
- Using Data for to Make Predictions

We will consider various perspectives on these themes, including Bayesian, frequentist, and likelihood approaches. We will consider methods that rely on mathematical thinking and methods that rely on computational thinking, with particular emphasis on computational approaches to analyzing data and understanding statistical methods.

This course also will have emphasis on estimation and analyzing data. Through this, there will be a communication component. Being able to interpret and communicate your findings from a statistical analysis, is just as important as being able perform a statistical analysis. Thus, there will be components of the course where you will be expected to communicate ideas.

The following topics will be covered:

Topic	Reference Materials
Exploratory Data Analysis and Limit Theorems	Chapters 15, 16, 13, 14 of MIPS and corresponding sections of STA238supplement
Statistical Models and Estimators	Chapters 17, 18, 20 of MIPS Section 7.1 of E&R and corresponding sections of STA238supplement
Statistical Inference	Chapters 18, 21, 23, 24 of MIPS Sections 7.2.1 and 7.2.2 of E&R and corresponding sections of STA238supplement
Prediction	Section 7.2.4 of E&R and corresponding sections of STA238supplement

Assessment

Assessment	Weight	Date
Assignments	Best 3 out of 4 15% each	Friday January 29 at 11:59pm ET Friday February 12 at 11:59pm ET Friday March 5 at 11:59pm ET Friday March 19 at 11:59pm ET
Test	25%	Friday March 26 (timed test in a 24 hour window)
Final Project (Proposal)	5%	April 2, 2021
Final Project (Final Report)	25%	April 16, 2021

Notes:

- Assignments will be posted on Quercus at least one week in advance of the due date.
- To accommodate for students writing in different time-zones, the test will be held within a 24hour window, once the test is started you will have a time limit within that window.
- Extensions for assignments may be granted. Requests must be made in advance of the assignment due date via the course email. Requests can also be made the same day an assignment is due, but the request must be made before the time that the assignment is due. Please note that requests made on the due date may not be granted. Thus it is recommended to make a submission before the due date regardless of the extension being granted or not. Requests for extensions made after the due date will not be considered.
- Extensions for up to 5 days for the final report of the final project may be granted. Requests must be made in advance of the final report due date via the course email. Requests can also be made the same day, but the request must be made before the time that the assignment is due, and the request may not be granted. Thus it is recommended to make a submission before the due date regardless of the extension being granted or not. Requests for extensions made after the due date will not be considered.

- There will be no extensions granted for the Proposal of the Final Project (April 2, 2021) as it will contain a peer-review component which will require on-time submissions.
-
- Late submissions (without a granted extension) will receive a mark of 0.
- We will be calculating your assignment mark as an average of your top 3 assignments.
- If you miss more than one assignment your grade will recorded as a 0 for those missed assignments.
- If the test is missed you must contact me via the course email within one week of the missed test. For consideration your email must:
  - be received within one week of the test date,
  - must include 'STA238 Reporting Test Absence' in the subject line,
  - must include your full name and student number,
  - must include a screenshot/photo of your self-declared absence on Acorn, and
  - must include the following two sentences:
    1. "I affirm that I am experiencing an illness or personal emergency and I understand that to falsely claim so is an offence under the Code of Behaviour on Academic Matters."
    2. "I understand that an alternative assessment will be arranged at the instructor's discretion (including an oral exam and/or a make-up assessment in April, after the lecture period)."
- If you miss the test and complete the accommodation procedure correctly (described above). An alternative assessment will be arranged at the instructor's discretion, if accommodation for a missed term test is requested as described above. Note that this alternative assessment may have a different format (e.g., oral assessment) and may be scheduled in April after classes end.
- Mistakes occasionally happen when marking. If you feel there is an issue with the marking of a test/assignment, you may request that it be re-marked. The course re-mark policy exists to correct mistakes, and any request should clearly identify the error (for example, a question that was not marked, or a total incorrectly calculated). Requests to correct such mistakes must be sent by email to sta238@utoronto.ca. For consideration, any email for a re-mark request:
  - must not be sent within the first 24 hours of the release of the assessment grade,
  - must be received within two weeks of the date that the marks for the assessment became available,
  - must include 'STA238 Regrade Request ¡Assessment Name¿' in the subject line of the email,
  - must include your full name and student number, and
  - must give a specific, clear, and concise reason for each request, referring to a possible error or omission by the marker. Re-mark requests without a specific reason will not be accepted.

  Please note that your entire test/assignment may be re-marked when submitting a re-marking request.

  For the final project, the re-mark process is handled by the Department of Statistical Sciences.

Academic integrity

Academic integrity is fundamental to learning and scholarship at the University of Toronto. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that the University of Toronto degree that you earn will be valued as a true indication of your individual academic achievement, and will continue to receive the respect and recognition it deserves. Familiarize yourself with the University of Toronto's Code of Behaviour on Academic Matters available at http://academicintegrity.utoronto.ca.

Discussion about lecture materials, textbook concepts and course concepts with your classmates and the teaching team is encouraged, but it is expected that you work independently on all individual assessments. Please note, you may not submit for credit any work that was completed by another student (or person). This includes, but is not limited to, partially or fully completed code, communication of solutions, and plagiarism. In particular, you are expected to complete and submit independent work for assignments (that are not group work), the test, and the final project. Specifically, you are expected to work on individual work, individually. You may discuss lecture materials and general course concepts, but it is expected that you work individually through assessments. You may use code provided by your STA238 instructors without providing a citation. If you use code from any other source, you must provide the source. To protect yourself from potential academic integrity offences, do not share your code and written submissions.

Writing Resources

Again, communication and writing will play a role in this course, thus I wanted to emphasize some of the writing resources that the university has made available to its students:

- https://writing.utoronto.ca/writing-centres/arts-and-science/
- https://www.artsci.utoronto.ca/current/academic-advising-and-support/english-language-learningfall-2020-mini-courses-accordion-1

COVID-19 & Mental Health Resources

This iteration for STA238 will be running during the COVID-19 pandemic, and will be completely online. There may be times where extensions for students are needed, and/or instructors and TAs may take longer than usual to respond to emails and/or marking needs. It is recommended to please stay active in the course as much as possible (attend lectures, visit office hours, post on Piazza, etc.) and please notify us of needs for extensions or other course related content as early as possible.

The Faculty of Arts and Science have put together the following list of Frequently Asked Questions (FAQs) regarding COVID-19:
https://www.artsci.utoronto.ca/covid19-artsci-student-faqs.

Additionally, learning online can be more challenging than learning in-person. If you need help regarding mental health, please do not hesitate to find support. Here are some UofT mental health resources:

- https://prod.virtualagent.utoronto.ca/.
- https://studentlife.utoronto.ca/department/health-wellness/.

- Call Good2Talk. Free, confidential helpline with professional counselling, information and referrals for mental health, addictions and well-being, 24/7/365 1-866-925-5454