STA238H1-S: Probability, Statistics and Data Analysis II
Summer 2022

Sections L0101
Instructor: Sonia Markes
Office hours: Fri 11am-12pm ET (Zoom)
Preferred pronouns: she/they

Teaching assistants
See the course Quercus page for information about the TA team.

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@utoronto.ca.

Class format

Lectures: Mondays 10am-1pm ET & Wednesdays 10am-1pm ET
Each lecture will have 2 breaks, lasting for 15 minutes each. Typically, new material will be introduced, along with examples showing how it's used to analyze data. Questions are welcome throughout. There are three specific ways you can ask a question:

1. Raise your hand during a lecture.
2. Immediately following each break, there will be a short question period.
3. If you'd prefer to ask privately, the last 5 minutes of each break are a good time.

Tutorials: Mondays 9-10am ET & Wednesdays 9-10am ET
Typically, material based on previous lecture will be reviewed, with a focus on working through solutions to practice problems or challenging examples. Questions that came up during your studies can be addressed here, especially those that may make it more difficult to take in the new material in lecture. Each TA will let you know how they want to receive your questions.

Classes run from July 4 – August 15, 2022. There is no tutorial on July 4 and there are no classes on August 1 (Civic holiday).

Communication guidelines

- Choose the appropriate context. *When* and *how* to communicate appropriately in this course is outlined in this syllabus. Note that appropriateness depends on the content of the message, that is, on *what* the message is about.

- Aim for clarity. All necessary details are included and delivered in as succinctly as possible.

- Show courtesy. Some ways of being courteous include: using a gentle tone, considering the perspective of your audience, and respecting boundaries expressed by others.

- Engage in good faith. Try to participate honestly and regularly, to the best of your abilities given your current situation. Everyone else (including the teaching team) will be trying too, although they have different abilities and situations, and these are private.

Email: sta238@utoronto.ca

Please use this address for individual administrative inquires of the teaching team, such as missed assessments. To be fair to all students, questions about course material will not be answered by email.

Please include your preferred name, your UTORid, and student number in any email sent. Informative subject lines are recommended.

I aim to return all emails within 48hrs, excluding weekends. If you require quicker response, please put [***] ahead of the content of the subject line and I will prioritize your email for a quicker response, if possible.

Please note that this email address will not be monitored after August 30, 2022.

Course web page: https://q.utoronto.ca

All announcements and course materials will be posted on Quercus, including any updates to guidelines for communication. 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.

Discussion: https://piazza.com/utoronto.ca/summer2022/sta238/home

We will be using Piazza as the platform for discussions related to the course material and assessments. Out of respect for classmates, personal or individual matters should not be posted on Piazza.

All course members, including students and teaching team, will be able to see posts on Piazza. Students can post anonymously to classmates on Piazza, but the identity of the author of all posts is view-able by instructors. Students are encouraged to answer each other's questions, however, teaching team members will check-in daily, excluding weekends.

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, as well as meeting the course's guidelines for communication.

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/.

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. This book is available in the University of Toronto bookstore. A pdf version of this textbook is freely available through the University of Toronto library website.

[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.
Available in pdf here: http://faculty.marshall.usc.edu/gareth-james/ISL.

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

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 completing 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.

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 and scientific methods.

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

Assignments: 2 assignments, due July 18 & Aug 8

Midterm: 1.5hrs, on July 27 during tutorial/lecture

Final exam: 3hrs, date TBA by the Faculty of Arts and Science

The final grade will be calculated with the following weightings, under the option that maximizes each student's grade:

Version	Assignments	Midterm	Final Exam
1	15% each	20%	50%
2	Best of 2 = 20%	20%	60%
3	15% each	0%	70%

Notes:

- The compressed timeline of summer courses does not leave much flexibility for extensions or make-up tests, so alternative the alternative weightings in the above table will be the main way that accomodations can be offered. There are no make-ups for the assignments or midterm.
- Late assignment submissions (without a granted extension) will recieve a mark of 0.
- If the midterm is missed, the final grade will be calculated under weighting Version 3.
- If the midterm is missed and one or both assignments are unable to be submitted, an alternative assessment (including a different format e.g. oral exam) or final grade weighting scheme may be considered, at the discretion of the instructor. Under these special circumstances, an email with the following must be received within one week of the midterm date:
  - includes "Reporting Midterm Absence" in the subject line
  - includes your full name, UTORid, and student number
  - contains 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 or weighting scheme may be arranged at the instructor's discretion."
- Mistakes occasionally happen when marking. If you feel there is an issue with the marking of an asssessment, 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 can be submitted using the following form: https://forms.office.com/r/Xim2BWGubz.
  Note that the 'Regrade Request' form requires that submissions:
  - must not be sent within the first 24 hours of the release of the assessment grade,
  - must be received within 8 days of the date that the marks for the assessment became available,

- must include your UTORid 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 assessment may be re-marked if it is submitted for a regrade request, meaning that your grade may remain the same or go up, but it may also go down.

For the final exam, the re-mark process is handled by the Faculty of Arts and 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 your work with your classmates and the teaching team is encouraged, but you may not submit for credit work completed by another student. In particular, you are expected to complete and submit independent work for the computational components of assignments. You may use code provided by your 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.