

STA238H1: Probability, Statistics and Data Analysis II
Winter 2020

Sections L0101 and L0201

Instructor: Prof. Alison Gibbs

Email: alison.gibbs@utoronto.ca

Office: SS 6009

Office hours: Wed 15:00-16:30, Fri 11:00-noon

Preferred pronouns: she/her

Section L0301

Instructor: Alex Stringer

Email: alex.stringer@mail.utoronto.ca

Office: 155 College, Room 380K, Section 388

Office hours: Tuesdays 17:30-19:00

Preferred pronouns: he/his

Course administrative email

Use the email address sta238.winter2020@utstat.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, 2020.

Course web page

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

Teaching assistants

See the course Quercus page for information about TA office hours and other information about your course TAs.

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 meetings

The course is scheduled for 4 hours per week. Most weeks, lectures will take place during the 3 hours of each week, typically the 2-hour class meeting on Tuesday or Wednesday and one of the hours of the 2-hour class meeting on Thursday or Friday. The fourth scheduled hour will be used for tutorial, extra help, and assessments.

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 Golemund, available here:
<https://rstudio-education.github.io/hopr>
- R for Data Science by Hadley Wickham and Garrett Golemund, 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

<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. We strongly recommend 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.

Communication

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/winter2020/sta238h1/home>.

Students can post anonymously to classmates on Piazza, but the identity of the author of all posts is viewable 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.

Your instructor’s email is above. 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_winter2020@utstat.utoronto.ca. Any emails sent to this address must include in the subject line STA238 and the section of the course in which you are enrolled (L0101, L0201 or L0301). 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 and Tentative Lecture Schedule

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.

The following lecture schedule is a tentative plan:

Weeks	Topic	Reference Materials
1, 2, 3	Exploratory Data Analysis and Limit Theorems	Chapters 15, 16, 13, 14 of MIPS and corresponding sections of STA238supplement
4, 5, 6	Statistical Models and Estimators	Chapters 17, 18, 20 of MIPS Section 7.1 of E&R and corresponding sections of STA238supplement
7, 8, 9, 10	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
11, 12	Prediction	Chapter 2 of ISL Section 7.2.4 of E&R and corresponding sections of STA238supplement

Assessment

Assessment	Weight	Date	Location
Assignment 1	5%	Computational and survey component: Wed January 22 at 11:59PM In-class component: Thurs Jan 23 (L0301), Fri Jan 24 (L0101 and L0201)	Submit on Quercus
Assignment 2	5%	Computational and survey component: Wed February 26 at 11:59PM In-class component: Thurs Feb 27 (L0301), Fri Feb 28 (L0101 and L0201)	Submit on Quercus
Assignment 3	5%	Computational and survey component: Wed March 25 at 11:59PM In-class component: Thurs Mar 26 (L0301), Fri Mar 27 (L0101 and L0201)	Submit on Quercus
Test 1	20%	Tuesday February 4 (L0301), Friday February 7 (L0101 and L0201)	TBA
Test 2	20%	Thursday March 12 (L0301), Friday March 13 (L0101 and L0201)	TBA
Exam	45%	Scheduled by the Faculty of Arts and Science	

Notes:

- Assignments will consist of two components: (1) Surveys and work in R that must be submitted on Quercus by the relevant Wednesday at 11:59PM. (2) In-class activities on the following Thursday or Friday that will use the R work.
- No late submissions will be accepted. Be sure to submit your work well before it is due.
- To avoid conflicts, all tests will be held during the scheduled course time. Each section of the course will write a different test.
- Tests will be written in a different room than your lecture room (location TBA).
- All tests and in-class components of assignments must be completed with the section of the class in which you are enrolled.
- If any assessment is missed, you may ask to be excused from the assessment. The weight of the assessment will be shifted to the final exam. To request to be excused from the assessment you must send an email to sta238.winter2020@utstat.utoronto.ca. For consideration, the email:
 - must be received within one week of the assessment date,
 - must include STA238 and your section in the subject line of the email,
 - must include your full name and student number,
 - must specify the assessment missed, 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 the weight of this assessment will be shifted to the final exam.”
- If you write both tests and your exam mark is better than one or both of your test marks, the weight for the relevant test(s) will be reduced by 10% and the weight of the exam will be increased by 10%.
- Mistakes occasionally happen when marking. If you feel there is an issue with the marking of a test, 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.winter2020@utstat.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 and your section 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 may be re-marked when submitting a re-marking request.

For the final exam, the re-mark process is handled by the Faculty of Arts and Science.

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. For the in-class component of assignments, you will submit work jointly as a team, with all team members contributing to the submission.