

STA 238 - Probability, Statistics, and Data Analysis II
L0101: M 10-12, TUT: W 10-12 // L0201: T 3-5, TUT R 3-5

Instructor: Karen H. Wong - sta238@utoronto.ca

Meeting Access: All course meetings, office hours and online tutorials will take place on MS Teams. In-person tutorial sections begin after Jan. 31 (tentatively), locations will be communicated directly to students in these sections. You will only be able to access the tools covered under the [U of T License](#) by logging into Office 365 with your U of T credentials.

Course 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.

Prerequisites: See course description [here](#). Prerequisites are strictly enforced by the department and cannot be waived.

Textbooks:

1. *A Modern Introduction to Probability* by Dekking et. al., available through the library [here](#) (second online access easiest)
2. *Modern Mathematical Statistics with Applications*, 3rd ed. by Devore and Berk available through the library [here](#)
3. *R for Data Science* by Grolemund and Wickham is a recommended reference text for students new to using R. Chapters 3-9 recommended to reinforce skills in R. Available [here](#)

Course Materials: All lecture slides, video recordings, meeting links, office hours, and other materials can be navigated from our course home page on Quercus. Course materials provided are for the use of students *currently enrolled in this course only*. **Distributing course materials is considered unauthorized use.**

Course Structure: Due to changing conditions, this course will be delivered virtually, with the exception of a few in-person tutorials which will be subject to public health guidelines.

- Asynchronous video lectures primarily to introduce new concepts
- Synchronous lectures will be used for: guided problem solving, introduce new topics, demonstrations in R (including R instruction) and debrief any common questions about the material.
- Tutorial times will alternate between R Labs and online quizzes, and cannot be completed asynchronously. **Note: All tutorials are virtual until at least Jan. 31.**

Everyone learns at a different pace! In addition to our course delivery times (lectures, videos, and tutorial times totalling ~ 3 -4 hours per week), it is recommended and expected that students will need to commit *at minimum 4 hours throughout the week* in the form of self-study, or a total of 7 hours per week.

Self-study includes: creating course notes for yourself (using your own words!), working on practice problems/assignments, reviewing concepts, explore R demos, and posting on the discussion board (~ 45 mins/day). See the next page for a suggested way to structure your week.

Suggested Weekly Routine

Monday	Tues	Wed/Thurs	Friday	Sat/Sun
<p>L0101 Synchronous Lectures: Have your notes and questions handy, and R Markdown files downloaded and opened!</p> <p>Last day to contribute towards the week's discussion boards.</p>	<p>L0201 Synchronous Lectures: Have your notes and questions handy, and R Markdown files downloaded and opened!</p> <p>Video lectures for this week's content are posted. Weekly discussion boards are opened.</p>	<p>R lab or online quiz during your registered tutorial time. See posted schedule.</p>	<p>Assignments are due on Fridays at 8 PM!</p>	<p>WEEKEND</p>
<p>Revisit any questions you had about the material and flag them for class.</p>	<p>Watch posted video lectures. Jot down notes in your own words (as if you were explaining it to yourself!).</p> <p>Attempt 5-10 practice problems. Post your questions/explanations to problems on discussion boards.</p>	<p>Watch posted videos.</p> <p>Daily practice from suggested problems for topics the async videos you've watched.</p>	<p>Finish video lectures, and complete your notes.</p> <p>Review R work (lab or sync session). Post on discussion boards.</p>	<p>REST DAYS or short review, as needed.</p>

Grading Scheme: The course is designed to give you plenty of opportunities to demonstrate your learning, make mistakes, track your progress, and receive feedback for improvement. The grading scheme is listed below:

Weekly Discussion	8%	Throughout Winter Term
Syllabus & Pre-Course Quiz	2%	Available from Jan. 11 to 25
R Lab #0	1%	Available from Jan. 19/20 to Jan. 25
Individual Assignments (2)	8% each	Feb. 18, Apr. 1 @ 8 PM
Quizzes (Best 3 of 4)	7% each	Synchronous - See schedule on the last page as well as posted on Quercus
R Labs (4)	3% each	
Term Test	20%	TUT01** : Mar. 16 @ 10 AM - 12 PM, TUT02** : Mar. 17 @ 3 - 5 PM
Final Project	20%	Proposal (5%): Mar. 21 Final Report (15%): Apr. 19

Discussion Board: There will be weekly topic threads on the Quercus Discussion Board as a space for students to discuss textbook problems, course topics, and exchange peer support. The instructor and TAs will be monitoring and contributing regularly on the discussion board. Students are expected to contribute by working collaboratively with each other and build upon your understanding of concepts through answering each others' questions. The grading scheme will be provided in the Welcome Slides in Week 1.

Assignments: Submission on Crowdmark or Quercus (will be specified on assignment). Only select problems on the assignments will be evaluated. Late submission receive a penalty of 5% for every hour past unless prior arrangements have been made. E.g., if an assignment is submitted 10 minutes late, there would be a penalty of 5% on the assignment grade (i.e. 90% → 85%).

R Labs: TAs will lead you through using tools in R to perform data cleaning and analysis, and to use R Markdown to present your source code and discussions neatly. The remaining tutorial time is dedicated to independent or pair (recommended) work to complete some exercises on your own. All R Labs are worth 3%, **with both .rmd and knit .pdf files submitted on Quercus**. Late submissions are not accepted and will receive a grade of 0.

Term Test: Test information will be posted on Quercus approximately 1 week prior to the date of the assessment. The midterm will be a timed online assessment. Any **time conflicts** with the midterm must be communicated as soon as you are aware, and **no later than 2 weeks prior to** the midterm. Accommodations cannot be provided for late notice of time conflicts.

Minimum Passing Requirement: Students must complete and submit at minimum:

- 2 R Labs and 2 quizzes
- 1 assignment
- Both the term test and the final project

If you miss more than these minimum passing requirements, even with accommodation, you will not be eligible to pass this course. There are no make-up tutorials. If an absence extends beyond 14 consecutive days, or if you have a non-medical personal situation preventing you from completing your academic work, you should connect with your College Registrar. They can provide advice and assistance reaching out to instructors on your behalf. If you get a concussion, break your hand, or suffer some other acute injury, you should register with Accessibility Services as soon as possible.

Accommodation Requests: Accommodation requests for assignments must be made by email (sta238@utoronto.ca) at least 24 hours before the assessment due date. Extensions will be arranged where possible.

Specific Medical/Personal Circumstances For 2022 S-term, a Verification of Illness (also known as a “doctor’s note”) is temporarily not required. Students who are absent from academic participation for any reason (e.g., COVID, cold, flu and other illness or injury, family situation) and who require consideration for missed academic work should report their absence through the online absence declaration. The declaration is available on ACORN under the Profile and Settings menu. **Students should also advise their course teaching team of their absence.**

Accessibility Services: 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 [here](#). 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.

Missed Work: Missed term work (quizzes and R labs up to a maximum of 2 each, and midterm) due to legitimate reasons (e.g., injury, illness, or other exceptional circumstances) should be communicated to the instructor (sta238@utoronto.ca) within 1 week following the assessment with the absence declaration form on ACORN completed and attached to the email. Missed work without prior communication or appropriate absence documentation will receive a 0. For documented missed work:

- First missed quiz will be treated the lowest grade dropped. If a second quiz is missed, the second quiz will be redistributed among the two remaining quizzes.
- Missed R labs will be distributed equally between R labs (1.5%) and final project (1.5%).
- Missed assignment will be distributed among the other assignment, midterm, and final project only in the event that an extension cannot be arranged.
- Missed midterm: A make-up midterm will be scheduled for students who missed the midterm with documentation.

Reread Requests: Marking schemes/solutions will be provided for all assessments except the final project. These must be reviewed before requesting a reread. If you still have concerns about your grading, complete this [form](#) to request a reread **no later than 1 week** after the assessments have been released back to you. Requests after this window, or without reviewing the provided solutions, or made in any other form will not be considered.

Course Conduct:

- **Email:** Any administrative or personal concerns regarding the course, or questions about active assessments should be discussed privately with the course instructor, either over email (sta238@utoronto.ca) or during office hours. Private breakout rooms will be provided in the latter case. Questions regarding course material and concepts are best addressed during open office hours or the discussion board.
- **Programming Languages:** We will be using [R statistical software](#) and [R Studio](#), both free to download. R Studio is an Integrated Development Environment for R, so you will need to install both locally for R Studio to operate. An installation guide will be provided at the start of term. You will also need to knit your Markdown documents to .pdf which requires local installation of LaTeX <https://www.latex-project.org/get/>.
 - If you have problems with any of the local installations, we strongly recommend that you access R Studio through the version hosted on [JupyterHub](#) (<https://jupyter.utoronto.ca>)
 - Log in with your U of T credentials, select ‘R Studio’. If that’s unavailable, click ‘New’, and ‘R Studio’ on drop down menu in the upper right hand corner of the landing page

- Any R work required of you will be covered during our synchronous sections or by your TAs during the R Labs. R components are an important part of the course content, and you are expected to participate and practice as you would with any other course content.

It is expected that you have both installed, or have accessed the cloud version by R lab #0, and definitely by R Lab #1. Knitted R markdown files will be required for all R work for the rest of the course, so be sure to address any knit-issues during your R Lab #0 or during office hours before your first R Lab!

By the end of the course, you are expected to be able to conduct basic data analysis, cleaning, simulations, and inferences using the R skills covered in our course. Check the weekly pages on Quercus to see if you will need to have R ready to use during the synchronous meeting times.

Academic Integrity: All students, faculty and staff are expected to follow the University's guidelines and policies on academic integrity. For students, this means following the standards of academic honesty when writing assignments, collaborating with fellow students, and writing tests and exams. Ensure that the work you submit for grading represents your own honest efforts. Plagiarism includes representing someone else's work as your own or submitting work that you have previously submitted for marks in another class or program, and is considered a serious offence that can result in sanctions. Speak to the instructor for advice on anything that you find unclear. To learn more about how to cite and use source material appropriately and for other writing support, see the U of T writing support website at <http://www.writing.utoronto.ca>. Consult the Code of Behaviour on Academic Matters for a complete outline of the University's policy and expectations. For more information, please see <https://www.artsci.utoronto.ca/current/academic-advising-and-support/student-academic-integrity> and <http://academicintegrity.utoronto.ca>

Potential offences include, but are not limited to:

- Obtaining or providing unauthorized assistance on any test/assignment/tutorial/quiz/project including:
 - working in groups on individual assessments, including giving hints to the answer!
 - having someone rewrite, edit, or add material to your independent work
 - researching for inspiration, hints, or answers to any graded problem
 - posting active assessment questions on discussion boards/private tutoring companies for hints/solutions
 - collaborating with members outside of your assigned groups
 - plagiarizing by passing off someone's work or ideas as your own
- Lending your work to a classmate who submits it as their own with or without your permission

The University of Toronto treats cases of academic misconduct very seriously. All suspected cases of academic dishonesty will be investigated following the procedures outlined in the Code. The consequences for academic misconduct can be severe, including, but not limited to, a failure in the course and a notation on your transcript. **If you have any questions about what is or is not permitted in this course, please do not hesitate to contact the instructor.** If you are experiencing personal challenges that are having an impact on your academic work, please reach out to the instructor or seek the advice of your college registrar.

Tentative Schedule:

Week Of	Topics	Tutorial Schedule
1: Jan. 10	Introduction to the Course, EDA: Summary & Graphical Statistics	—————
2: Jan. 17	EDA: Graphical Summaries, Approximations & Limiting Theorems	R Lab #0: Intro to R
3: Jan. 24	Central Limit Theorem, Properties of Estimators, Simulating Data	R Lab #1
4: Jan. 31	Sampling Distributions, Bootstrap Principle, Moment Generating Functions	Quiz #1: Includes Week 3
5: Feb. 7	Parametric Bootstrap, MOM Estimation, Likelihood Functions	R Lab #2
6: Feb. 14	MLE Estimation, Other Common Sampling Distributions	—————
Feb. 21	WINTER READING BREAK	—————
7: Feb. 28	Confidence Intervals: By Hand and Via Bootstrapping	Quiz #2 : Includes Week 6
8: Mar. 7	Hypothesis Testing: By Hand and Via Simulation	R Lab #3
9: Mar. 14	Testing Errors, Goodness of Fit Tests	Term Test
10: Mar. 21	Goodness of Fit Tests, Simple Linear Regression	Quiz #3: Includes Week 9
11: Mar. 28	Simple Linear Regression	R Lab #4
12: Apr. 4	Time Permitting: Intro to Bayesian Statistics and Estimation	Quiz #4: Week 10-11