STA304: Surveys, Sampling and Observational Data

Delivery: Fully online   **Online class times:**   L0101   9:10 a.m. ET (Friday)
                                                                      L0301   2:10 p.m. ET (Friday)

Course webpage: All materials will be posted on Quercus (https://q.utoronto.ca)

Instructor Information

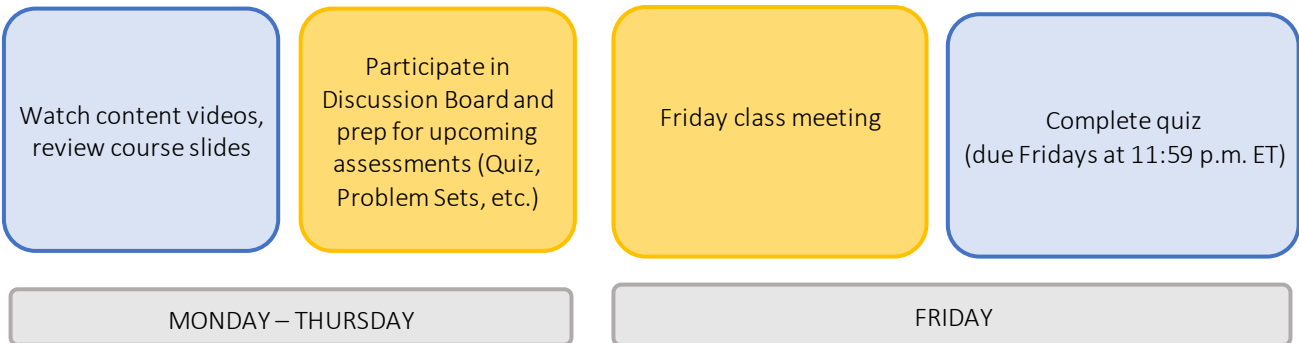|	**Office Hours**
Professor: Samantha-Jo Caetano **How do you pronounce that?** Like the English words "kyte", "a", "no". **Please call me:** Sam (pronouns: she/her)	Wednesdays 11:10 a.m.–1:00 p.m. ET (Toronto time)

Course Information

Statistics is about how we can learn from data. In order to analyze this data, it needs to be collected first. This course will focus on how to appropriately collect data through surveys and samples, as well as some of the pitfalls of collecting data inappropriately. Throughout the course we will look at how to design a survey (including both experimental designs and observational studies) and how to label (and potentially avoid) sources of bias. We will learn about different sampling techniques, including random sampling, stratification, clustering, systematic, and unequal probability selection. Lastly, we will cover how to make inference based on the data collected from a survey/sample. This inference will include estimating population parameters of interest (mean, proportions, variation, correlation, etc.).

Course Format and Organization

This term, we will be using an asynchronous (i.e. a flipped format) for STA304. Each week, lecture videos will be posted, you'll watch videos, read course slides, and complete a quiz ahead of our Friday class meeting. I will then focus on some hands-on, applied, and discussion based problems in our scheduled class time on Fridays. Hopefully, you will join these sessions live so that you can participate, but a video recording of the sessions will be posted on Quercus for revision and to support members of the class in difficult time zones. In addition to these weekly activities, there will be a discussion board, one test, three problem sets and a final project.

Watch content videos, review course slides	Participate in Discussion Board and prep for upcoming assessments (Quiz, Problem Sets, etc.)	Friday class meeting	Complete quiz (due Fridays at 11:59 p.m. ET)

MONDAY – THURSDAY	FRIDAY

Glossary of Terminology and Course Tools

Below, you'll find a list of some of the terms you will encounter in this course and more broadly at the University. You can find a more complete glossary at https://future.utoronto.ca/newly-admitted-students/checklist/glossary-of-terms/

Bb Collaborate (aka Blackboard Collaborate)	BbCollaborate is a video conferencing tool that we will use for synchronous class meetings and office hours. You can access it through Quercus (look for Bb Collaborate in the sidebar). You can join the Course Room at any time to update your profile picture. Instructions and other getting started tips here: https://help.blackboard.com/Collaborate/Ultra/Participant/Get_Started
Quercus	The online teaching and learning system where you will find your course homepages (including course materials, syllabi, announcements, and grades) and other resources.
Piazza	Piazza is a free Q&A platform which is used in many courses. In this course, you'll use Piazza to post questions about course content or logistics (see "Getting your questions answered" for more guidance on how to join and use Piazza).
R / RStudio	Free software environment for statistical computing. In this course, you'll learn to use R to produce visualizations, manipulate data, and conduct analyses. *No prior programming experience is assumed.*
UTORID	Your username that gives you access to ACORN, Quercus, etc. (e.g. caetano2)

Minimal technical requirements

All students should consult the minimum technical requirements for participation in online learning. If you are facing financial barriers to obtaining the required technology, please contact your College Registrar's Office to obtain information regarding your potential eligibility for a need-based bursary. If you anticipate having difficulty connecting to University websites (e.g., Quercus), please submit your question here: https://www.utoronto.ca/covid19-contact.

Throughout this course we will be using R/RStudio to work through applied problems. There is no requirement to know how to program in R/RStudio at the start of this course, but familiarity will help the learning curve. Discussions will take place on Piazza (or Quercus).

Learning Objectives and Assessment

By the end of the course, you should be able to:

Design a survey or sample that is appropriately gathering information of interest.

Carry out a variety of statistical analyses in R to make inference on the data collected from a survey/sample.

Identify and implement different sampling techniques and different study designs and the trade-offs involved in each.

Identify sources of bias within a study and comment on a study's design, including it's weaknesses, strengths, and appropriate analyses.

Clearly communicate results of statistical analyses to technical and non-technical audiences.

How will your success be measured?

WEIGHT	WHAT	DATES
9%	Weekly quizzes	Due Fridays at 11:59 p.m. ET Starting: Week of Sept 14 (Friday Sept 18) Ending: Week of Nov 30 (Friday Dec 4) Exceptions: Week of the Test (Fri Nov 20) & Reading Week (Fri Nov 13)
17%	Problem Set 1	Due Thursday October 1
17%	Problem Set 2	Due Thursday October 15
17%	Problem Set 3	Due Monday November 2
10%	Test	Thursday November 19; details TBA on Quercus
30%	Final project	Multiple due dates (Final Project due in Final Assessment Period); details TBA

Quizzes (due Fridays at 11:59 p.m. ET)

There will be a quiz each week apart from the week of the tests and reading week for a total of 10 quizzes. Quizzes are intended to be completed while looking over the lecture material/videos for that week. So it is recommended to go through the quiz immediately after viewing the weekly posted lecture video(s). Quizzes will consist of a combination of multiple choice and fill-in the blank questions. The quiz will cover material presented in that week's module (videos, slides, etc.). Each quiz will be available on Quercus from the time a module is posted (Monday afternoon at 6:00 p.m. ET) until the following Friday at 11:59 p.m.

ET; no late submissions will be accepted. The only exception to this is if you miss the quiz for a valid medical reason; see the "Missed Work" section for the course policy.

Discussion Board (ongoing)

We will be using the Piazza on Quercus to facilitate discussion. I will post questions and topics on the Quercus discussion board in which your participation and responses will be *not* be graded, but I strongly encourage you to use the discussion boards. The TAs and professor will monitor the discussion boards throughout the week, so if you have a general question regarding the course structure, topics or something interesting that you would like to share with the class please post it here. Please post all general questions regarding the course on the discussion board.

Please note that emails and the message board are not checked or responded to by either the TA or me after hours or on the weekend.

Missed Work

You do not need to reveal your personal or medical information to me. I understand that illness or personal emergencies can happen from time to time. The following accommodations to assessment requirements apply in these situations.

Weekly Quizzes & Problem Sets

If a **quiz or a problem set is missed for a valid reason**, you may ask to be *excused* from the assessment. Extensions will *not* be given. If approved, the weight of the missed assessment will be shifted to all remaining assessments of the same type (i.e. the weight of a missed quiz will be shifted to the remaining quizzes and the weight of a missed assignment will be shifted to the remaining assignments). To request to be excused from an assessment, **you must report your absence through the ACORN Absence Declaration Tool** (https://www.acorn.utoronto.ca/) AND **send an email to your instructor** at s.caetano@utoronto.ca. For consideration, your email:

-   must be received within 1 week of the missed assessment,
-   must include your full name and student number,
-   must specify the assessment missed including the date, 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 remaining assessments of the same type."

NB: No more than <u>two</u> of the quizzes can be accommodated in this way and no more than <u>one</u> problem sets can be accommodated in this way. Missed quizzes and problem sets beyond this limit will be recorded as 0%.

Tests

If the **test** is missed for a valid reason, you may ask for an accommodation. If approved, you will be offered an alternate assessment (which may be an oral exam or a test cumulative on all material in the course or a combination of some sort).

To request to be excused from an assessment, **you must report your absence through the ACORN Absence Declaration Tool** (https://www.acorn.utoronto.ca/) AND **send an email to your instructor** at s.caetano@utoronto.ca. For consideration, your email:

- must be received within 1 week of the missed assessment,
- must include your full name and student number,
- must specify the assessment missed including the date, 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 I will need to complete an alternative assessment to gain this grade and that I will be given further instructions in response to this email (if approved)."

Final project

The final paper is a critical piece of assessment. Extensions for valid reasons may be granted for a maximum of three days (i.e. through to the close of the university). The exact extension granted will be at the discretion of the instructor.

To be considered, an extension request MUST be sent s.caetano@utoronto.ca by 11:59 a.m. E.T. (midday) on the business day prior to the due date. Note: For Monday deadlines, this will be 11:59 a.m. E.T. (midday) the previous Friday.

Where possible, alert me to potential issues as early as you can. This will allow us to work together with you to find a suitable solution.

Important note

If too much work is missed, even for valid reasons, a make up test (including all work covered during the entire term) and/or an oral exam may be required to calculate a fair mark, at the discretion of the instructor. Please ensure you and/or your registrar get in touch with me as early as possible if this may be the case for you.

(Tentative) Weekly Assessment Schedule

Week	Quizzes	Other Assessment
Sept 7-11		
Sept 14-18	Quiz 1	
Sept 21-25	Quiz 2	
Sept 28-Oct 2	Quiz 3	Problem Set 1
Oct 5-9	Quiz 4	
Oct 12-16	Quiz 5	Problem Set 2
Oct 19-23	Quiz 6	
Oct 26-30	Quiz 7	
Nov 2-6	Quiz 8	Problem Set 3
Nov 9-13	READING WEEK	READING WEEK
Nov 16-20		Test
Nov 23-27	Quiz 9	
Nov 30-Dec 4	Quiz 10	
Dec 7-9		
Dec 11-22		Final Project

Marking Concerns

Any request to have an assessment remarked must be emailed to s.caetano@utoronto.ca within one week of the grades being posted; your request will be reviewed by the course instructors and head teaching assistant. Your request must include:

- your name and student number,
- a **detailed written justification** referring to your answer and the relevant course material to be considered; it is **not enough** to simply say that you believe your answer deserves higher credit, rather you must **support your request** with specific reference to relevant course materials.

Please note that we reserve the right to review the grading of all questions or parts when you re-submit an assessment for reconsideration (i.e., your grade could go down).

Writing

Communication and especially writing is a critical aspect of the statistical workflow. Papers, assgnments, short answer questions, etc. should be well-written, well-organized, and easy to follow. They should flow easily from one point to the next. They should have proper sentence structure, spelling, vocabulary, and grammar. Each point should be articulated clearly and completely without being overly verbose. You will be heavily penalised for papers that do not meet these basic requirements.

Papers should demonstrate your understanding of the material you have learnt and your confidence in drawing on the terms, techniques, and issues you have considered. Your work must be thoroughly referenced.

If you have concerns about your ability to do any of this, then please make use of the writing support provided to students - https://writing.utoronto.ca/. The services are designed to target the needs of both native and non-native speakers and the programs are free. I have used similar services in the past at other universities and always found them very helpful.

Core Texts

There is not strictly one textbook that we will be working through. Instead, we will be calling upon multiple texts to gain insight into surveys and sampling. Here are a list of the main texts that we draw on:
1. Wu, Changbao and Mary E. Thompson, 2020, *Sampling Theory and Practice*, Springer.

This text is currently available for download from the U of T library website: https://onesearch.library.utoronto.ca/

2. Gelman, Andrew, Jennifer Hill and Aki Vehtari, 2020, *Regression and Other Stories*, Cambridge University Press.
3. Kohavi, Ron, Diane Tang, and Ya Xu, 2020, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press.
4. McElreath, Richard, 2020, *Statistical Rethinking*, 2nd Edition, CRC Press.

Getting your questions answered

Question(s) about course logistics? e.g.
- *What is the deadline for the weekly quiz?*
- *Where do I see the discussion board topic?*

Question(s) about course content e.g.
- *I don't understand the difference between stratification and clustering?*
- *Can we make a causal statement here?*
- *My code won't run for the assignment (please include screenshots of your code and the error message!)*

Information / resource to share with classmates e.g.
- *I have a link/resource/opportunity to share with my classmates*

Piazza forum (available on Quercus)
Posts can be anonymous for your classmates, but instructors and TAs will be able to see your name.

Before posting a question, search to see if someone else has already asked a similar question (you can edit the question to add yours or post a follow-up at the bottom)

Try to answer your classmates' questions – this is a great way to reinforce your own understanding while also helping your classmates! *Don't worry if you aren't 100% sure of the answer – all answers will be reviewed / endorsed / completed by TAs and instructors!*

Question(s) related to your personal circumstances (i.e. something which is not appropriate to share with the whole class) e.g.
- *I would like to request an accommodation for yesterday's quiz because I was ill (make sure to include all information listed in the "Missed Work" section)*
- *I would like question 2 on the assignment to be regraded (be sure to include clear justification, as outlined in the "Marking Concerns" section)*

My contact email: s.caetano@utoronto.ca
(only send emails from your utoronto.ca email address to ensure it doesn't automatically go to a Junk folder and be sure to include your full name and student number)

This account will be monitored by the course instructor; if you want to reach a specific instructor or TA please email them directly.

Allow 48 hours for a response during the week (Monday to Friday, ET) and do not expect responses on the weekend.

If you cannot meet a deadline because you are ill, please refer to the "Missed Work" section in this syllabus and submit all required information to this email account.

Questions about course content won't be answered here, but rather redirected to Piazza or office hours.

How to succeed in this course

The course is designed to actively engage you in the course material. We hope you'll find the statistical reasoning and data science interesting, challenging, and fun! In order for you to get the most from the classroom sessions:

- Always watch/read the weekly content before Friday's class meeting (there will be a quiz to help you check your knowledge),
- Complete the assignments,
- Keep up-to-date in the course–do not leave working on discussions or your project to the last minute, and
- **Ask questions!** Post in/watch the course discussion forum on Piazza and attend instructor and/or TA office hours (TA Office hours will be posted on Quercus).

Recognized Study Groups

Recognized Study Groups (RSGs) are small study groups of 3 to 6 students from the same course who meet weekly to learn course content in a collaborative environment.

Each group is made up of students from the same course. One student volunteers to be the RSG Leader and helps organize and plan weekly activities. The RSG Leader is a student who is trained in group facilitation and effective learning techniques. RSG Leaders are not tutors – they are learning along with group members.

A student staff member is also assigned to each group to help connect you to academic resources and support your group's goals.

While not compulsory for this course, I would highly recommend you get involved with an RSG.

Meet to complete

Meet to Complete is an online "study with me" space where you can study alongside other students. Each Meet to Complete is hosted by a student to welcome you and provide support & encouragement, if needed.

To join Meet to Complete, enroll in the Meet to Complete course on Quercus! Online learning doesn't need to be lonely.

Academic Integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters.

As a general rule, we encourage you to discuss course material with each other and ask others for advice. However**, it is not permitted to share answers or to directly share R code or written answers for anything that is to be handed in**. For example, "For question 2.1 what R function did you use?" is a fair question when discussing course material with others in the class; "Please show me your R code for question 2.1" is not an appropriate question. If writing or code is discovered to match another student's submission or outside source, this will be reported as an academic offence. When asked to hand in code and a problem set or project document, the code you submit must have been used to generate the document. If it does not (i.e., the submitted code does not match the submitted output), this is also considered an academic offense.

I will not tolerate any academic offenses. This includes (but is not limited to) plagiarism, cheating, copying R code, communication/extra resources during closed book assessments, purchasing labour for assessments (of any kind). Academic offenses will be taken very seriously and dealt with accordingly. If you have any questions about what is or is not permitted in this course, please do not hesitate to contact your instructor.

Please consult the University's site on Academic Integrity http://academicintegrity.utoronto.ca/. Please also see the definition of plagiarism in section B.I.1.(d) of the University's Code of Behaviour on Academic Matters http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppjun 011995.pdf. Please read the Code. Please review *Cite it Right* and if you require further clarification, consult the site *How Not to Plagiarize* http://advice.writing.utoronto.ca/wp-content/uploads/sites/2/how-not-to-plagiarize.pdf.

Note that when an assignment is required to be completed as a team (e.g., project), you may discuss and share answers and code with other members of your team, but not with another team in the class or anyone outside the course.

Intellectual Property Statement

Course material that has been created by your instructor (i.e. lecture slides, term test questions/solutions and any other course material and resources made available to you on Quercus) is the intellectual property of your instructor and is made available to you for your personal use in this course. Sharing, posting, selling or using this material outside of your personal use in this course is not permitted under any circumstances and is considered an infringement of intellectual property rights.

While recordings of class meetings will be made available to you on the course website, these are intended only for students registered in the course. You are not authorized to copy these materials or distribute them to individuals who are not registered in the course. If you would like to record any course activities in this course, you MUST ask permission from your instructor in advance. According to intellectual property laws, not asking permission constitutes stealing.

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: email accessibility.services@utoronto.ca or visit the website at http://accessibility.utoronto.ca

If you have an accommodation letter from your accessibility advisor that is relevant to this course, please do the following:

- *Email your letter to s.caetano@utoronto.ca with "Accommodation letter" as part of the email subject, CC your advisor and let us know anything else you wish us to know/any*

questions you have. Please do this as soon as possible after you enrol in the course/receive this syllabus.

- *Confirm any accommodations for **each** specific assessment **1 week** before the assessment. (I.e. if you receive extra time for timed assessments, confirm this one week prior to the midterm assessment and final assessment, even if we have already discussed this at the beginning of the semester.)*

Covid-19

We are in the middle of a pandemic. This term will be a challenging one for you, but also for everyone involved, including your TAs and me, faculty, staff, and of course the other students in your courses. Nonetheless, I am hoping to make the best of the situation and want to provide you with an opportunity to get the most out of this course (and overall situation), by providing you with a set of resources and infrastructure to help your statistical learning and communication. Some degree of flexibility and good faith is needed from all of us. If you need accommodations, then please be as proactive as possible in asking for them.