Statistical Sciences
UNIVERSITY OF TORONTO

STA 304H1S / 1003HS Winter 2021
Surveys, Sampling and Observational Data / Sample Survey Theory
Section L0101 (January 11 to April 30) (except February 15-19)

Land Acknowledgement

We wish to acknowledge the land on which the University of Toronto operates. For thousands of years it has been the traditional land of the Huron-Wendat, the Seneca, and most recently, the Mississaugas of the Credit River. Today, this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land.

Delivery:	Online Synchronous according to Eastern Standard Time (EST)
Course website:	Available through https://q.utoronto.ca (UofT Quercus)
Classes:	Mondays 3-4pm and Wednesdays 3-5pm, EST
Instructor:	Dr. Shivon Sue-Chee (*she, her*)
Teaching Assistants :	Miaoshiqi, Alex, Kevin, and Robert
Office hours:	Drop-in times will be posted in our website
Course email:	sta304@utoronto.ca

1 SYLLABUS CONTENT

2 COURSE OVERVIEW

Course content
This course teaches mathematical and statistical reasoning behind sampling, aspects of inference from surveys, and the interplay with observational studies. In addition to the topics listed in the calendar description, I will include discussion of current studies reported in the news.

Undergraduate calendar description

> Design of surveys, sources of bias, randomized response surveys.
> Techniques of sampling; stratification, clustering, unequal probability selection.
> Sampling inference, estimates of population mean and variances, ratio estimation.
> Observational data; correlation vs. causation, missing data, sources of bias.

Pre-requisite
ECO220Y1/ECO227Y1/GGR270Y1/PSY201H1/SOC300Y1/SOC202H1/ STA220H1/STA255H1/ STA261H1/ STA248H1/ STA238H1/ STA288H1/ EEB225H1/STAB22H3/ STAB57H3/ STA220H5/ STA258H5/ STA260H5/ ECO220Y5/ ECO227Y5

Exclusion: STAC50H3, STAC52H3, STA304H5

Learning Outcomes
By the end of the course, students should be able to do the following, using R/RStudio where appropriate:

- design and analyse data from sample surveys
- derive statistical inference based on different sampling design
- identify and critique survey designs
- clearly communicate survey theory and analytical results to technical and non-technical audiences

Textbooks

- *Elementary Survey Sampling, 7th edition* by R. L. Scheaffer, W. Mendenhall, R. L. Ott and K. G. Gerow (Brooks/Cole 2012)
  We will cover most of Chapters 1 through 5, and selected parts of Chapters 6 through 11.
- *Sampling: Design and Analysis, 2nd edition* by S. L. Lohr (Duxbury 2010). UofT link to electronic copy: `http://go.utlib.ca/cat/13421569`
  This text is highly recommended.

Additional Readings

1. Sampling Theory and Practice by C. Wu and M. Thompson (Springer 2020)
2. Survey Methodology, 2nd edition by Groves, Fowler, Couper, Lepkowski, Singer and Tourangeau (Wiley 2009)
3. Designing Surveys: A Guide to Decisions and Procedures, 3rd edition by J. Blair, R. Czaja and E. A. Blair (Sage 2014). UofT link to 2nd edition: `http://go.utlib.ca/cat/13494253`
4. Design of Observational Studies by P. R. Rosenbaum (Springer 2010). UofT link to electronic copy: `http://go.utlib.ca/cat/7890274`
5. Mathematical Statistics and Data Analysis, 3rd edition by J. Rice (Brooks/Cole 2007)

Computing

We will use `R` and `RStudio` for statistical computing. The main advantage of `R` is that it is a freeware and there is a lot of available help resources online. `RStudio` is an integrated development environment (like SAS) to `R`, which makes it easier to work in `R`. `R`/`RStudio` can be downloaded onto your personal computer or used via our university's web server or in the `RStudio` Cloud. If you would use `R` on your personal computer, then installation is via a two-step procedure:

1. Download the base `R` framework at http://cran.r-project.org/ for Windows, Mac and Linux operating systems.

2. Then download `RStudio` for free at https://www.rstudio.com/products/rstudio/download/.

No prior experience with R/RStudio is expected. Support for downloading and learning `R` and `RStudio` will be provided by the teaching team (Instructor and TAs). Additional resources will be given in our website. In lectures, examples with `R` syntax will be provided, which should be sufficient for you to practice and do your assignments.

3 ASSESSMENT AND GUIDELINES

Assessment Undergraduate students will be evaluated based on the following scheme.

	Weight	Due Date	Time
Assignment 1	15%	Thursday, February 4	by 8 pm
Engagement Activities	5%	(various dates)	
Term Test 1	15%	Monday, February 22	3:10-4pm
Assignment 2	10%	Thursday, March 4	by 8 pm
Term Test 2	15%	Monday, March 22	3:10-4pm
Assignment 3	15%	Thursday, April 1	by 8 pm
Final Assessment	25%	Between April 13-23	(2 hrs)

Students must complete the final assessment, at least one assignment, and at least one test in order to pass this course.
Graduate students will be given the opportunity for extended learning and will be alternatively assessed. More information will be given in lecture or by email communication.

Engagement Activities Students will be required to participate in various activities such as data collection surveys, course evaluation surveys, exit polls and peer reviews.

- They will be held on a roughly bi-weekly basis to foster consistent student engagement and will be designed to be completed during the last 30 minutes of class time for the week. Activities will be announced at the start of the week in which they will be held.
- Late submissions will not be allowed.
- Students will be graded out of 80% of all activities.
- Participation must be done individually.
- We will use Quercus and/or peerScholar for these activities.

Assignments The assignments will each be of a practical nature, for which the use R will be required.

- Assignments are to be submitted online into Quercus by 8pm on the due dates. Late assignments will be accepted but subject to a 1% penalty of the total assignment marks per hour late. Late submissions will not be allowed beyond 48 hours of the due date.
- Students who would like additional accommodations should email the instructor at least 48 hours before the assignment is due.
- There are no make-up assignments. A missed assignment will be given a grade of 0.

Tests There will be two (2) tests. They will be held online in Quercus under Quizzes.

- Each test will be for 50 minutes and will be available from 3pm to 4pm on the test days.
- The test will have multiple choice and/or short-answer questions. Short-answers may require that you upload R output, short videos and/or hand-written answers. More details on the test coverage will be posted in our website.
- Students who are eligible for special test accommodations will be facilitated through the university's Accommodated Testing Services.
- If a test is missed, for any reason, you need to send an email to sta304@utoronto.ca as soon as possible and no later than 5 business days. You may use the following sentences "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. I will be available to do a makeup after (specify the date here)." The makeup test will take the form of a video assignment and will typically be held within 7 to 10 days after the original test date. You will be given 24 hours to complete the task.

Final Assessment The final assessment is cumulative. The date and time will be fixed by the Faculty of Arts and Science. Further details will be announced later in the course.

Homework
Practice problems from the textbooks will be posted in the lecture notes and some may be discussed in class. They are to help you prepare for the tests and final assessment and are not to be handed in. Solutions will not be posted. However, additional guidance will be available through TA and Instructor office hours and the class Piazza discussion forum.

Re-grading Policy

- Any requests to have graded work re-evaluated must be made in writing by email to sta304@utoronto.ca within one week of the date the work was returned to the class.
- The request must contain a justification and the corresponding marks you think you deserve.
- Be sure to include your official name and student number for identification purposes, and give a suitable label such as "Test 1", "Assignment 2" in the subject line.
- Regrading requests should be processed by the teaching team within two weeks of the request date. Please note that the teaching team reserves the right to review a part or the whole of your work. Hence, your marks may go down or up. **Unsuccessful regrades requests will be will result in a deduction of two assessment points.**

Missed Final Assessment Policy

If the final assessment is missed, for any reason, a makeup written and oral assessment will be offered at the earliest possible time.

Academic Integrity

The University supports acting in honesty, trust, fairness, respect, responsibility, and courage in all academic matters. Students are responsible for knowing the content of the University's Code of Behaviour on Academic Matters. All suspected cases of academic dishonesty will be investigated following procedures outlined in the Code of Behaviour above. If you have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, you are expected to seek out additional information on academic integrity from your instructor or from other institutional resources (`http://academicintegrity.utoronto.ca/`). Here are a few guidelines that apply to this course:

- Students may consult course materials during tests, however, sharing or discussing questions and answers is an academic offence.
- Instructions for each assessments should be strictly followed. All assessments must be completed individually.
- Do not personate another person, or have another person personate at any assessment.
- It is acceptable to get help with your assignments from someone outside the course, but the help must be limited to general discussion and examples that are not the same as the assignments. As soon as you get some else to actually start working on one of your assignments, you have committed an academic offence!
- You must not copy mathematical derivations, computer output and input, or written descriptions from anyone or anywhere else, without reporting the source within your work. This includes copying from solutions provided to previous semesters of this course. Please read the UofT Policy on Cheating and Plagiarism, and don't plagiarize.

Intellectual Property

Course materials provided on Quercus, such as lecture videos and slides, assignments, quizzes and solutions are the intellectual property of your instructor and are for the use of students currently enrolled in this course only. **Providing course materials to any person or company outside of the course is unauthorized use.** Failure to comply can result in legal action against all parties involved.

4 SUPPORT AND ACCOMMODATIONS

Course website

The course website is available through the Quercus Management Engine via `https://q.utoronto.ca`. It will be regularly updated with lecture notes, practice problems, assignments, announcements, readings and grades. Please turn on email notifications to receive announcements, and emails from me if urgent matters arise.

Lecture outlines will typically be posted by midnight on Sundays.

What to expect during live class sessions

Live sessions will be conducted via Bb Collaborate or Zoom mainly by the instructor, with potentially TA moderator support. The main lecture will be recorded and will made available to students after the live class.

A typical 60-minute session, beginning at the 'x'-th hour, will have the following program:

x:00- x:10	(Re-)Entry and setup
x:10- x:15	Ice breaker poll questions
x:15- x:45	Delivery of Lecture (Will be recorded and posted after class)
x:45- x:55	Discussions/ Piazza Live Q&A/ Small group discussion via breakout rooms
x:55- x:60	Wrap-up

Your responsibility

The live classroom sessions are designed to actively engage you in the course material. I hope you'll find them interesting, challenging, and fun, and an excellent opportunity to truly learn the material. In order for these sessions to be effective, coming prepared, by learning about the week's concepts through the textbook, is essential.

Online Discussion Board

This semester you will have the option to use Piazza for class discussion. The Piazza system is highly catered to getting you course material help fast and efficiently from classmates, the TAs, and myself. Rather than emailing questions, I encourage you to post your questions on Piazza. To sign up for the discussion forum go to the link at:

`https://piazza.com/utoronto.ca/winter2021/sta304`

If you decide not to use Piazza, it will not disadvantage you in any way, and will not affect official University outcomes (e.g., grades and learning opportunities). If you choose not to opt-into Piazza, then you can ask questions or discuss course material with the instructor or TAs during online office hours.

Be sure to read Piazza's Privacy Policy and Terms of Use 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. If you decide to participate in Piazza, only provide content that you are comfortable sharing under the terms of the Privacy Policy and Terms of Use.

Note that moderation of the forum is subject to TA availability and further details of how and when the forum will be moderated, will be announced in our website. Please use the forum in accordance with its purpose. Inappropriate posts will not be tolerated and will be dealt with accordingly. **On test days, the forum will be closed temporarily.**

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 at accessibility.services@utoronto.ca or `http://www.studentlife.utoronto.ca`.

Communication

E-mail is appropriate for emergencies or personal matters. Please use your `*utoronto.ca` account. You will not get a response if you email from other email addresses. Please send course correspondence and regrade requests to sta304@utoronto.ca. Your email should contain an appropriate subject line, the addressee, your official name and UTORid for identification purposes. This departmental course mailbox will be shared among the teaching team members to efficiently handle course correspondence. Please expect an answer to your e-mail within two business days.

To contact the instructor directly, you can use the Quercus Inbox via the course website.

Succeeding in our course

Students are encouraged be active learners by learning the course materials, staying connected via our course website, being involved in live classroom sessions and diligently completing assessments. Here are some suggestions for achieving success in our course:

- Connect with the instructor and/or TAs during live sessions and office hours.
- Post and answer questions on the discussion forum.
- Join an online recognized study group (RSG) or get advice on learning (outside of the classroom) from a learning strategist via Sidney Smith Commons at `https://sidneysmithcommons.artsci.utoronto.ca/online-learning/`. RSGs are voluntary, peer-led study groups of up to 8 students enrolled in the same A&S course. The RSG program is designed to increase student engagement within individual courses, support academic skill-building and keep students socially connected throughout the term.
- Send an email to sta304@utoronto.ca in cases of emergencies or personal matters.

5 TENTATIVE SCHEDULE

Week	Topic	ESS Text Chapters
1	Basic sampling concepts and definitions	1, 2
2	Types of samples, Introduction to R	2
3	Questionnaire Design, Good and bad surveys	2
3	Statistics review, Probability samples	3
4	Simple random sampling	4
5	Stratified random sampling	5
6	TEST 1, Stratified random sampling	1-5, 5
7-8	Ratio, Regression and Difference estimation	6
9	Systematic Sampling	7
10	One-stage cluster sampling	8
11	TEST 2, Two-stage cluster sampling	1-8, 9
12	Estimating the population size, Review	10