STA 304H1 F/1003H, SUMMER 2020
SURVEYS, SAMPLING AND OBSERVATIONAL DATA

Meeting Information

- **Instructor:** Dr. Fode Tounkara
- **E-mail:** f.tounkara@utoronto.ca.
- **Lecture:** Tuesday, Thursday; 2 PM-5 PM.
- **Location:** online-Blackboard Collaborate (Bb Collaborate) in Quercus.
- **Office Hours:** Wednesday 3:00-5:00 PM, Online-Blackboard Collaborate.
- **TA Office Hours:** TBD

Course Description

This course presents mathematical and statistical reasoning behind sampling, aspects of inference from surveys, and the interplay with observational studies.

Course Objectives

At the end of this course, students should be able to:

1. design and implement surveys with the following sampling designs: simple random, systematic, stratified, and cluster;
2. estimate population mean and total, population proportion, and ratio,
3. understand Observational data; correlation vs. causation, missing data, sources of bias.

Pre-requisite

Students should have the following statistics courses:

- ECO220Y1/ECO227Y1/GGR270H1/PSY201H1/SOC300H1/SOC202H1/STA220H1/STA255H1 STA261H1/STA248H1/STA238H1/STA288H1/EEB225H1/STAB22H3/STAB57H3 STA220H5/STA258H5/STA260H5/ECO220Y5/ECO227Y5
- **Exclusion:** STAC50H3, STAC52H3, STA304H5

Resources

1. Course webpage

- Lectures notes are available through the learning portal at **https://q.utoronto.ca**.

2. Texbooks

1. *Elementary Survey Sampling*, **7th edition**, by Scheaffer, Mendenhall, Ott and Gerow.

(We will cover most of Chapters 1 through 5, and selected parts of Chapters 6 through 11.)

- **An online version is available at:**

https://www.nelsonbrain.com/shop/isbn/9780840053619

2. *Sampling: Design and Analysis*, **2nd edition** by Sharon Lohr. (Useful but not required)

3. Additional help

Need extra help with the coursework? Here are some options:

- **Have a question about STA304 online course content or administration?**
  - Review the questions already posted on the Quercus discussion board, and if your question hasn't already been addressed, post your question there.
  - Visit the instructor's and TA's office hours. Instructor and TAs office hours will be held online. TAs hours will be posted on Quercus.
- Post your **course content** question on the class discussion forum on Piazza. you can sign up for the discussion forum here -http://piazza.com/utoronto.ca/summer2020/sta3041003. This forum will be monitored regularly by TAs and instructors.

- **Need to reach the instructor about a personal matter (e.g., illness, grades)?** Email your instructor (f.tounkara@utoronto.ca). Note that e-mail should only be used for emergencies or personal matters. If you email a question to the instructor about course material or course administration, then you will be asked to refer to the course syllabus or to post your question on the Quercus discussion board – these types of questions will not be answered via email. Note also that the TAs are not available for contact over email, etc., outside of their scheduled office hours or online forums or sessions.

Evaluation

Assessment	Weight	Due data	Time	Location
Attendance/Participation in online class sessions	5 %	Weekly	lecture time	PollEverywhere
Assignment 1	10 %	Friday, May 15	11:59 pm	Submit online (Crowdmark)
Assignment 2	10 %	Friday, May 22	11:59 pm	Submit online (Crowdmark)
Assignment 3	10 %	Friday, May 29	11:59 pm	Submit online (Crowdmark)
Term Test	25 %	Thursday, June 4	14:10-15:40	TBA (Crowdmark or Quercus)
Assignment 4	10 %	Friday, June 12	11:59 pm	Submit online (Crowdmark)
Final Assessment	30 %	Between June 17-25	TBA	TBA (Crowdmark or Quercus)

Participation (PollEverywhere)

There will be twelve online classes. Attendance and participation is mandatory for these sessions. **PollEverywhere** will be used to promote engagement and provide feedback on your understanding during classes as well as to provide you with credit on your lecture participation. **PollEverywhere** questions will be asked during lecture. The proportion of questions you answer (starting the week of May 10) will determine the fraction of the available participation grade (5%) that you earn. There will be around 3 **PollEverywhere** questions per class. However, some classes will have more or fewer questions. You earn

- 0 % if $0\% <$ % Answering poll $< 50\%$
- 1 % if $50\% \leq$ % Answering poll $< 60\%$
- 2 % if $60\% \leq$ % Answering poll $< 70\%$
- 3 % if $70\% \leq$ % Answering poll $< 80\%$
- 4 % if $80\% \leq$ % Answering poll $< 90\%$
- 5 % if $90\% \leq$ % Answering poll $\leq 100\%$

You can register for the poll here: https://PollEv.com/fodetounkara043/register. **You MUST register with the email that shows up on ROSI/ACORN (mail.utoronto.ca address)**. Once the poll is active, you must response at: **PollEv.com/fodetounkara043**.

Assignments

- There will be Four equally-weighted assignments, each assignment will consist of an individual component to be completed and submitted on your own.
  - A1 : Friday, May 15th, due 23:59
  - A2 : Friday, May 22th, due 23:59
  - A3 : Friday, May 29th, due 23:59
  - A4 : Friday, June 12th, due 23:59
- The assignments will each be of a practical nature, for which the use of R will be required
- Must be written as PDF files using R Markdown and submitted online into **Crowdmark** by 11:59pm on the due dates
- Late assignments will be accepted but subject to a 10% penalty per day late
- Late submissions will not be allowed beyond 48 hours of the due date
- Email submission will not be accepted.
- **Due to the nature of these assignments, there will be no extensions granted under any circumstances**

Additional Instructions about Assignments

- For the assignment questions that involve R computation, you need to use Rmarkdown and include your R source code in the Rmarkdown document.
- For the questions without any R computation, you could choose to hand-write those part. But you need to convert the hand-written part into a PDF file.
- Be carefull when submitting your answers.You will lose points If you submit any assignment question in the wrong location on crowdmark.
- For the assignment questions that involve R computation, you will lose points if the solution just includes R code without proper comments and conclusions.

Term Test

- **90-minutes multiple-choice and short-questions** test
- Thursdays June 4th, from 2:10-3:40 PM
- The term test will be written online (Quercus or Crowdmark).
- The location and information on coverage will be posted on Quercus in advance.

Final assessment

- There will be a **3-hours final assessment**
- Scheduled by the faculty
- The final assessment will be written online (Quercus or Crowdmark).
- The final assessment is cumulative
- More information on coverage and the availability period will be posted on Quercus.

Re-grading Policy

Any requests to have marked work re-evaluated must be made in writing within one week of the date the work was returned to the class. The request must contain a justification for consideration. Be sure to include your official name, student number and/or paper number for identification purposes.

Missed Term Test

- If the term test is missed, then you must report the reason (i.e illness) by email within a week. If the reason is valid, then you will be given an online makeup test. The instructor will schedule the time and date of the makeup test. The online makeup test will be worth 25% of the course grade.
- If both the originally scheduled and makeup tests are missed for valid reasons approved by the instructor, then the midterm grade will be predicted using your assignments and final assessment grades. A multiple linear regression will be fit on the midterm grades of all the students who wrote it with assignment and final assessment grades as predictors, and this fitted regression model will be used to predict your midterm score to calculate your STA304 grade.
- Other reasons for missing a test will require prior approval by your instructor. If prior approval is not received for non-medical reasons, then you will receive a grade of zero for the missed term test.

Important notes about all assessments

- Students must complete a minimum of three assessment out of five (assignments and midterm) to pass. If they complete 2 or fewer assessments during the term then they should not be able to pass, even with accommodation.
- There is no assignment extensions for those who request accommodation. However, there will be accomadation for the midterm test.
- Note that we will offer a make-up for anyone who misses the final assessment.
- There are no deferred exams.

Software

The course includes a lot of numerical calculation. We will use **Statistical programming language and environment R** software to support the class and sampling from real populations (data sets provided by the textbook). **Using R and understanding of R outputs are required**, on the level explained in the class.

R is freely available for download at

- **R** : https://www.r-project.org/

Rstudio is a good integrated development environment to **R**. It is also freely available at:

- **Rstudio** : https://www.rstudio.org/

Good online reference

- https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

For assignment, you can use **Rmarkdown** to write your solution.

- **Good online reference for Rmarkdown** : https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf

University of Toronto academic integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic integrity at http://www.governingcouncil.utoronto.ca/Governing_Council/policies.htm. If you have any question about what is or is not permitted in this course, please do not hesitate to contact your instructor.

Students with Disabilities

If you have a disability and need special arrangements, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Service as soon as possible at https://www.utoronto.ca/accessibility.

Your responsibility

The course is designed to actively engage you in the course material. We hope you'll find the subject of statistics interesting, challenging, and fun, and an excellent opportunity to truly learn the material. In order for these sessions to be effective, preparing by learning about the week's concepts through the notes is essential.

Course outline

Almost all of the course material is covered by the textbook. Related to the basic level of the textbook, some theoretical results will be considered in more detail.

The following is a tentative schedule for the course:

	Date	Topics	Important Reminders
Week 1	May 4-8	Introduction to the course, Review Basic concepts (Ch 3) Sampling problems and notions (Ch 2)	
Week 2	May 11-15	Simple random sample (Chap 4), SRS-Inference, Sample Size Calculation	
Week 3	May 18-22	Ratio, Regression, and Difference Estimations in SRS (Chap 5), Stratified Random Sample (Chap 6)	
Week 4	May 25-29	Stratified Random Sample (Continued), Sample size allocation, Ratio Estimation in SR Sampling (Chap 5)	
Week 5	June 1-5	Systematic Sampling (Ch 7)	June 1-Last day to drop Thursday, June 5
Week 6	June 8-12	Cluster sampling (Ch 8, and Ch 9)	
Week 7	June 15-25	Jun 17-25 – Final assessment period	Monday, June 15-Classes end Tuesday, June 16-Study day

How to communicate with your instructor

- Questions about course material such as:
  - How do I do question 3.7 in the textbook?
  - When is the midterm?

should be posted on the discussion forums on Piazza. Questions can be posted anonymously (so that the author is anonymous to other students but not to the instructors), if desired.

- For private communication, such as:
  - I missed the test because I was ill.

e-mail your instructor (Please use my *@utoronto.ca* email account), and include your full name and student number.

Note: I will only respond to e-mails you send me if they come from your e-mail account *@utoronto.ca*.