STA130H1: An Introduction to Statistical Reasoning and Data Science

L0101
Instructor: Dr. Bethany White (bethany.white@utoronto.ca)
Lectures: Monday 10:10-12:00, PB B150
Tutorials: Friday 10:10-12:00, various rooms (see https://q.utoronto.ca)
Instructor Office Hours: Monday 12 (after class)-2pm, SS 6006

L0201
Instructor: Dr. Nathalie Moon (nathalie.moon@utoronto.ca)
Lectures: Monday 2:10-4:00, MC 102
Tutorials: Friday 2:10-4:00, various rooms (see https://q.utoronto.ca)
Instructor Office Hours: Monday 4:00 (after class)-5:30pm SS 6024A

COURSE DESCRIPTION AND LEARNING OBJECTIVES

Statistics is about how we can learn from data. Data Science is a relatively new interdisciplinary field that also includes the computational aspects of carrying out a data analysis, including acquisition, management, and analysis of data. Statistical reasoning and computing with data play important roles in this emerging discipline. The purpose of this course is to give you a broad introduction to many of the ways statisticians learn from data. In addition to statistical reasoning, learning from data involves computation and communication. We will use the R programming language and environment for statistical computing, and in tutorials, you will gain experience communicating statistical ideas and knowledge.

- **Course Web Page:** https://q.utoronto.ca/courses/115817 All course materials, resources and assignments will be available on this site. We will be sending course announcements from the site from time to time. Please ensure your email settings are up to date so that you receive messages sent to the class.
- **Course Discussion Board (Piazza):** All questions related to the course content must be posted here. The instructors (and TAs) will not answer questions sent to our email addresses. Email should only be used for questions/concerns of as more personal nature (e.g., to let your instructor know that you were ill for the term test).
    - Signup link: https://piazza.com/utoronto.ca/fall2019/sta130
    - Class link: https://piazza.com/utoronto.ca/fall2019/sta130/home
- **RStudio.cloud**: Students in the course will be able to do all computing on RStudio Cloud (https://rstudio.cloud). You can join a shared workspace for this course on RStudio Cloud, where weekly problem sets will be posted. Instructions on how to get set up on our shared space are posted on Quercus.
- **Textbook**: There is no required textbook for this course. Relevant readings have been compiled from multiple sources will be made available on Quercus for each week.

INTENDED LEARNING OUTCOMES

By the end of this course students should be able to:
- Describe how statistical methods can be used to learn from data, including methods for description, explanation, and prediction.
- Carry out a variety of statistical analyses in R and interpret the results of the analyses.
- Implement the computational steps involved in the management and statistical analysis of data using R.
- Identify appropriate uses of statistical methods to answer questions, including their strengths and limitations.
- Clearly communicate the results of a data analysis to both technical and non-technical audiences.

TOPICS

- Data visualization
- Data wrangling and summarizing data
- Statistical testing and estimation
- Statistical models for description and prediction
- Supervised and unsupervised statistical learning
- Ethical issues in data collection and analysis

EVALUATION

Assessment	Weight	Date	Time	Location
Survey completion	2%	Distributed during last week of classes	NA	NA
Mentorship program	3%	Details will be available at your second tutorial (Friday, Sept 20)	NA	NA
Tutorial	20%	Weekly problem sets due on Quercus night before each tutorial + assignments due in Friday tutorial (must attend your tutorial section)	L0101 - 10-12 L0201 - 2-4	Various, see ACORN and https://q.utoronto.ca
Term Test	20%	***Friday, Nov 1*** (must write your section's midterm)	L0101 - 10-12 L0201 - 2-4	TBA
Final Project	20%	***Thursday, Dec 5***	L0101 - 10-12 L0201 - 2-4	Medical Sciences Building (MS) Lobby
Final Exam	35%	TBA (between Dec 7-20)	TBA	Scheduled by Faculty Arts & Science

Important notes about tutorials

1. Your tutorial grade includes weekly problem sets due before tutorial (to be submitted by 11:59PM the night before tutorial) and any work done during tutorial. *Late assignments will <u>not</u> be accepted.* Each week, you will earn a tutorial grade out of 6:

Attendance for the entire tutorial	1
Completed/submitted assigned weekly Problems	1
Tutorial assignment	4
Total	*6*

2. There are **10** tutorials during the term to be held every Friday **starting Sept 13 and ending Nov 29** *except* Nov 8 (Reading Week) and Nov 1 (term test). Your tutorial grade for the semester will be calculated as the average of your 10 tutorial grades. The tutorials are an important part of the course. If you miss a tutorial, 0 will be recorded for your tutorial and there is no way to make up this grade. The only exception to this is if you miss the tutorial for a valid medical reason (see "Missed term tests or tutorials" section below).

3. If you are officially enrolled in the 10-12 lecture section, your tutorials will be Friday from 10-12, and if you are officially enrolled in the 2-4 lecture section, then your tutorials will be on Friday from 2-4pm (*no exceptions*). It is your responsibility to make sure you are registered for a tutorial on ACORN (the

tutorial section on ACORN is the section you must attend) and that tutorial time must match your official lecture section time.

MISSED TERM TESTS OR TUTORIALS

- Late weekly problems (due at 11:59pm on the day before your tutorial) and tutorial assignments (due at end of tutorials) are <u>not</u> accepted.
- If a term test or tutorial work is missed for a valid medical reason, you must submit the University of Toronto Verification of Student Illness or Injury form **to your <u>instructor</u> (not TA)** within one week of the test or missed tutorial.
- The form will only be accepted as valid if the form is filled out according to the instructions on the form.
- The form must indicate that the degree of incapacitation on academic functioning is moderate, serious, or severe in order to be considered a valid medical reason for missing the term test or tutorial. If the form indicates that the degree of incapacitation on academic functioning is negligible or mild then this will **NOT** be considered a valid medical reason and 0% will be recorded for your missed term test or tutorial.
- Note: If you write the term test or attend the tutorial, it will be assumed that you deemed yourself fit enough to do so and your grade will stand as calculated. No accommodation will be made based on claims of medical, physical, or emotional distress after the fact.
- Other reasons for missing the term test or tutorial will require <u>*prior*</u> approval by your instructor. If prior approval is not received from your instructor for non-medical reasons, then 0% will be recorded for your missed term test or tutorial.
- If a tutorial is missed for a reason deemed valid by your instructor, then the weight for that tutorial work will be shifted to the remaining tutorials when your instructor computes your course grade at the end of the semester.
- If the term test is missed for a valid and approved reason, then the weight for the test will be shifted to the final exam. In other words, the final exam will be worth 55% of your final grade.

MARKING CONCERNS

Any request to have a test question or tutorial assignment remarked must be emailed **to your <u>instructor</u> (not TA)** within <u>one</u> week of the grades being posted. Your request must include a detailed written justification referring to your answer and the relevant course material to be considered. 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).

HOW TO COMMUNICATE WITH YOUR INSTRUCTORS

Questions about course material or organization, such as,
- What do I change the colour of my plotting symbol?
- What is the difference between supervised and unsupervised learning?
- When is the term test?

should be posted on the discussion forums on Piazza or asked in person (in class or during office hours). Questions can be posted anonymously on Piazza so that the author is anonymous to other students (but not to the instructors), if desired.

If your communication is private (e.g. I missed the test because I was ill), then e-mail your instructor. Use your utoronto.ca e-mail account to ensure that your message doesn't automatically go to a Junk folder and be sure to include your full name and student number.

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 <u>not</u> permitted to share answers or to directly share R code or written answers for anything that is to be handed in (e.g., weekly problems and tutorial assignments).** 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. 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

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 instructors 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.

No videotaping of lectures will be permitted under any circumstances. If you would like to make an audio recording of the lectures 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

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 and tutorials:

- Attending and participating in lectures and tutorials,
- Completing the weekly problems and readings,
- Keeping up-to-date in the course – do not leave weekly problems, studying or the project to the last minute, and
- **Asking questions!** Post/watch the course discussion forum on Piazza and visit instructor and/or TA office hours (TA Office hours will be posted on Quercus).