STA130H1S: An Introduction to Statistical Reasoning and Data Science

Syllabus - Draft (subject to change before the start of the course)

Fall 2018

Teaching Team

- Prof. N. Taback,nathan.taback@utoronto.ca, SS6027C.
- Prof. N. Moon, nathalie.moon@utoronto.ca, SS6024A.
- A large team of teaching assistants

Course Office Hours

- Prof. Taback, Monday 12:30 -13:30.
- Prof. Moon, Monday 16:30 - 17:30.
- The schedule of the TA office hours will be posted as soon as it is available.

Course Schedule

Lectures

- L0101: Monday, 10:00-12:00 in PB B150.
- L0201: Monday, 14:00-16:00 in WB 116.
- There are no lectures on Monday, October 8 (Thanksgiving).
- On Thursday, December, 6 your lecture section will meet in the Bahen Atrium for project presentations.

Tutorials

- **Fridays:** Tutorial either 10:00-12:00 or 14:00-16:00, corresponding to your lecture time. Your tutorial assignment and location will be posted on the the UofT Portal page for the course https://portal.utoronto.ca/. To receive course credit for tutorials you should only attend the tutorial in which you are enrolled.
- There are no tutorials on the following Fridays:
    - September 7. You should complete the online orientation to R and R Studio in lieu of this tutorial.
    - October 22 (Term Test)
    - November 5 (Reading Week)

Preparation

- **In between Mondays and Fridays:** On Mondays, you will be provided with work to do in preparation for that Friday's tutorial. This may include computational work, readings, and/or practice problems. Often you will be required to submit part of this work before tutorial and to bring part of this completed work to the tutorial.

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 tutorials will introduce students to communicating statistical knowledge.

Learning objectives

By the end of this course, you 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

Course Website

Class slides, notes, and other important information can be found on the course website.

Optional Course Textbooks

Baumer, B.S., Kaplan, D.T., and Horton, N.J. Modern Data Science with R. 2017. CRC Press.

Diez, D. M., Barr, C.D., and Cetinkaya-Rundel, M. Introdcutory Statistics with Randomization and Simulation First Edition. This is a free book.

Hands-On Programming with R, by Garrett Grolemund. This is a free book.

R for Data Science, by Hadley Wickham and Garrett Grolemund is a wonderful resource. This is a free book.

Computing

See the R Resources section of the course website.

Discussion Forum

We will be using Piazza as a platform for discussions. You can find our class page at: https://piazza.com/utoronto.ca/fall2018/sta130h1/home. Students will be able to post anonymously to classmates, but not instructors.

Be sure to read Piazza's Privacy Policy and Terms of Service 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.

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, September 21)	NA	NA
Tutorial	20%	Weekly assignments due in tutorial	10-12, 2-4	TBA
Term Test	20%	Friday, October 26	10-12, 2-4	TBA
Final Project	20%	Thursday, December 6	10-12, 2-4	TBA
Final Exam	35%	TBA	TBA	Scheduled by Faculty Arts & Science

Important notes about tutorials

1. Your tutorial grade includes any assigned work that is due before tutorial and any work done during tutorial. Note that if you miss a tutorial, there is no way to make up this component of your mark.
2. If you are enrolled in the 10-12 lecture section then your tutorial will be on Friday 10-12, and if you are enrolled in the 2-4 lecture section then your tutorial will be on Friday 2-4.

Missed Tests or Assignments

- If a 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 within one week of the test. Submit the form to your TA if it is related to missed tutorial work. Submit the form to your instructor if it is related to the test.
- 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. 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.
- If a tutorial is missed for a valid reason then the weight for that tutorial work will be shifted to the remaining tutorials.
- If the term test is missed for a valid reason then the weight for the test will be shifted to the final exam. In other words, the final exam will be worth 57% of your final grade.
- Other reasons for missing an assignment, tutorial work, or the term 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 assignment, tutorial work, or test.

Marking Concerns

Any requests to have your test remarked must contain a written justification for consideration. Marking requests should be made within one week of receiving your test. Please note that we reserve the right to look at all questions on your test when you re-submit an assessment for reconsideration.

How to Communicate with Your Instructors

Questions about course material or organization, such as,

- What do I change the colour of my plotting symbol?
- How do I do question 3.7 in the textbook?
- 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. Questions can be posted anonymously (so that the author is anonymous to other students but not to the instructors), if desired.

If your communication is private, such as, I missed the test because I was ill, then e-mail your instructor. If you missed a tutorial then e-mail your TA. Use your utoronto.ca e-mail account to ensure that your message doesn't automatically go to a Junk folder and 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 not permitted to share complete solutions or to directly share code for anything that is to be handed in. When an assignment is required to be completed as a team, you may share solutions and code with other members of your team, but not with another team in the class. For example, "For question 2.1 what R function did you use?" is a fair question; "Please show me your R code for question 2.1" is not.

If you have any questions about what is or is not permitted in this course, please do not hesitate to contact your instructor (Prof. Gibbs if your last (family) name starts with A-L and Prof. Taback if your last (family) name starts with M-Z).

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 class room, or course materials, please contact Accessibility Services as soon as possible: accessibility.services@utoronto.ca or http://accessibility.utoronto.ca

Your Responsibilities

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 classroom sessions and tutorials to be effective, prepare by learning about the week's concepts through completing the recommended problems and readings.