INF2178 - Experimental Design

Rohan Alexander

11 January 2021

Key details

- Instructor: Rohan Alexander - rohan.alexander@utoronto.ca - (you are welcome to call me 'Rohan' - 'row'-'hun').
- TA: Dan Xu.
- Time and place:
    - Lecture: Tuesdays 1-3pm, starting 12 January, online synchronous via Zoom or BBCollaborate, with lectures recorded.
    - Lab: Tuesdays 3-4pm, starting 12 January, online synchronous via Zoom or BBCollaborate.
- Course website: https://rohanalexander.com/inf2178.html.
- Office hours: Tuesdays noon - 1pm via Zoom.
- Course content: https://rohanalexander.com/inf2178.html#content
- Course discussion forums: Slack.
- Communication: If you have a question, there is a decent chance that others have the same question or, at least, will benefit from the answer. Please post all questions to the course Slack so that everyone in the course can benefit from your questions and our answers. You are encouraged to post answers to the questions of other students, where appropriate. Of course, if you have a concern of a personal nature then please email me - rohan.alexander@utoronto.ca. All emails must be professional and include a descriptive subject line. Please be polite. I do not check Slack or email after 5pm or on weekends.
- Text: There is no one required textbook for this course. The course notes - https://www.tellingstorieswithdata.com - contain extensive references to helpful texts that we will refer to. The following may be especially useful:
    - Bryan, Jenny, 2020, *Happy Git and GitHub for the useR*, https://happygitwithr.com.
    - Garrett Grolemund and Hadley Wickham, 2017, *R for Data Science*, https://r4ds.had.co.nz/.
    - Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Christel Vermeersch, *Impact Evaluation in Practice*, http://hdl.handle.net/10986/25030.
    - Healy, Kieran, 2019, *Data Visualization: A Practical Introduction*, Princeton University Press, https://socviz.co/.
    - Hernán MA, Robins JM, 2020, *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC, https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
    - James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani, 2017 *An Introduction to Statistical Learning with Applications in R*, Corrected 7th printing, Springer, https://statlearning.com.
    - Kohavi, Ron, Diane Tang, and Ya Xu, 2020, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press (freely available through the library).
- Computing: Throughout the course we will be using R. In particular, you will be learning and expected to code in the tidyverse style, and all papers will be written using R Markdown. You will need to have R and RStudio installed on your computer: Download R here: https://www.r-project.org/. Download RStudio (free version) here: https://www.rstudio.com/products/rstudio/download/. You will also need to have a GitHub account: https://github.com/

Course objectives

Experimental design has a long and robust tradition within traditional applications such as agriculture, medicine, physics, and chemistry. It allows us to speak of causality with confidence. Typically, these are situations in which control groups can be established, randomization is appropriate, and ethical concerns can be assuaged. Unfortunately, such a set-up is rarely possible in the full extent of the modern applications where we want to understand causality.

This course covers the traditional approaches and statistical methods, but focuses on what to do when traditional experimental design methods cannot be implemented or are not appropriate (i.e. what feels like most of the time these days). We cover experiments in their modern guise especially the concerns that we might have when we can run them; but also methods that can provide some causal understanding even when we cannot conduct traditional experiments. Importantly, these approaches do not rely on 'big data' or fancy statistics, but instead on thoroughly interrogating the data that are available to get understanding through as simple means as possible.

This is a hands-on course in which you will conduct research projects using real-world data. This means that you will: obtain and clean relevant datasets; develop your own research questions; use the statistical techniques that you are introduced to in class to answer those questions; and finally communicate your results in a meaningful way. This course is designed around approaches that are used extensively in academia, government, and industry. Furthermore, it includes many aspects, such as data cleaning and preparation, that are critical, but rarely taught.

This course is different to many other courses at the University of Toronto. At the end of this course, you will have a portfolio of work that you could show off to a potential employer. You will have developed the skills to work successfully as an applied statistician or data scientist. And you will know how to fill gaps in your knowledge yourself. A lot of scholarships and jobs these days ask for GitHub and blog links etc to show off a portfolio of your work. This is the class that gives you a chance to develop these.

Course description

At the heart of every Data Science project exists the planning, design and execution of experiments. Such experiments aim at understanding the data, potentially cleaning it and performing the necessary data analysis for knowledge discovery and decision-making. Without knowing the experimental design processes that are used in practice, researchers may not be able to discover what is really hidden in their data. The first aim of this course is to look at existing experimental designs that take into account the questions that need to be answered as well as the nature of the data and the different parameters used by algorithms.

Subsequently, the course will introduce different qualitative and quantitative methods to assess the quality of the results. All concepts will be accompanied by examples and the students will have practical exercises and a project in which they will demonstrate their knowledge.

Assessment

Broadly, there are four components to the assessment for this course:

- Weekly quiz, 10 per cent, due immediately before the lecture (only the best eight count, and there is no quiz due in Week 1 or 10).
- Professional conduct, 1 per cent, due anytime during the teaching term.
- Three papers, each worth 25 per cent, due 29 January 2021, 12 February 2021, and 12 March 2021 (only the best two count).
- Test, 10 per cent, 23 March 2021.
- A final paper, worth 25 per cent, due 19 April 2021, with a peer review component worth 5 per cent due, 9 April 2021 and 12 April 2021.

Details of each are available here: https://rohanalexander.com/inf2178.html#assessment

Accommodations with regard to assessment

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 exist to provide for those situations.

Straight-forward (will automatically apply to all students - there's no need to ask for these):

- Quiz: Worst two quizzes are dropped.
- Papers: Worst paper is dropped.

So for those, if you have a situation, then just don't submit.

Slightly more involved:

- Test: There is a 24 hour period to take the test, but if you can't sit it then email me before the test opens and the weight will be shifted to the final paper.
- Peer review: No accommodation or late submission is possible for this because it would hold up the rest of the class. If you cannot submit then email me before the deadline and the weight will be shifted to the final paper.
- Final paper: The final paper is a critical piece of assessment. It's also up against deadlines for submission of grades. Extensions for valid reasons may be granted for a maximum of three days, however this isn't possible for all students (i.e. there are restrictions around graduating students). Hence, the exact extension will be at the discretion of the instructor. To be considered, an extension request must be sent to rohan.alexander@utoronto.ca by the business day before the due date so there is time to get advice from the Faculty about your particular circumstance.

Re-grading

Requests to have your work re-graded will not be accepted within 24 hours of the release of grades. This is to give you a chance to reflect. Similarly, requests to have your work re-graded more than seven days after the release of the grades will not be accepted. This is to ensure the course runs smoothly.

Inside that 1-7 day period if you would like to request a re-grade, please email rohan.alexander@utoronto.ca with a subject line that starts with INF2178. You must specify where the marking error was made in relation to the marking guide. Your entire assessment will be re-marked and it is possible that your grade could reduce.

Of course, genuine marking errors, such as a part of the rubric missed, should be brought to my attention as soon as possible.

Plagiarism and integrity

Please do not plagiarize. In particular, be careful to acknowledge the source of code - if it's extensive then through proper citation and if it's just a couple of lines from Stack Overflow then in a comment immediately next to the code.

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

I will not tolerate 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 seriously and dealt with accordingly. If you have any questions about what is or is not permitted in this course, please contact me.

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/ppjun011995.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.

Late policy

You are expected to manage your time effectively. If no extension has been granted and no accommodation applies, then the late submission of an assessment item carries a penalty of 10 percentage points per day to a maximum of one week after which it will no longer be accepted, e.g. a problem set submitted a day late that would have otherwise received 8/10 will receive 7/10, if that same problem set was submitted two days late then it would receive 6/10.

Writing

Papers and reports 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. Papers should demonstrate your understanding of the topics you are studying in the course and your confidence in using the terms, techniques and issues you have learned. As always, references must be properly included and cited. If you have concerns about your ability to do any of this then please make use of the writing support provided to graduate students by the SGS Graduate Centre for Academic Communication. The services are designed to target the needs of both native and non-native speakers and all programs are free.

Other

Children in the classroom

Although not overly relevant while we are remote, babies (bottle-feeding, nursing, etc) are welcome 'in' class as often as necessary. You'll likely hear my toddler and inevitably he'll run into the room at some point. It's fine if you're in the same position.

You are welcome to take breaks to feed your infant or express milk as needed, either in the classroom or elsewhere including (if relevant): https://familycare.utoronto.ca/childcare/breastfeeding-at-u-of-t/. A list of baby change stations is also available: https://familycare.utoronto.ca/childcare/baby-change-stations-at-u-of-t/. For older children, I understand that unexpected disruptions in childcare can happen. You are welcome to bring your child to 'class' in order to cover unforeseeable gaps in childcare.

Course Learning Outcomes and their Relationship with Assessment

Students who have successfully completed this course will:

1. Know how to how to clean and prepare a new dataset and quickly generate summary statistics, tables, and graphs. Demonstrated through completion of problem sets.
2. Be able to identify ethical considerations and adjust approaches accordingly. Demonstrated through all assessment items, but especially the papers and quizzes.
3. Be able to describe, identify, explain, and implement approaches to causal inference and the types of methods that could be used to speak to causality in various circumstances. Demonstrated through completion of the papers and quizzes.
4. Create graphs and tables that help to illustrate and support claims. Demonstrated through completion of the papers.
5. Conduct statistical analysis in a reproducible way and interpret results. Demonstrated through completion of the papers.
6. Communicate results in a way that is clear about their strengths, weaknesses, and the assumptions that underpin them. Demonstrated through completion of all aspects of assessment, especially the papers.

7. Discuss the importance of openness in science, of making their analysis and datasets public. Demonstrated through completion of the papers.

Relationship between Course Learning Outcomes and Program Learning Outcomes

The program-level student learning outcomes for the Master of Information program (https://ischool.utoronto.ca/areas-of-study/master-of-information/) are:

1. Students understand and are conversant with fundamental concepts, theories, practices, and the diverse horizons of information disciplines, and can respond to changing information practices and needs of society.
2. Students develop knowledge and values appropriate to their future exercise of economic, cultural, and/or social leadership, and thereby provide leadership in defining the social responsibility of information professionals to provide information services for all, regardless of age, educational level, or social, cultural, or ethnic background.
3. Students develop the ability to contribute through research and publication, to the continuous expansion and critical assessment of the body of knowledge underlying the information and archives sciences.
4. Students develop an understanding of the development of theory concerning information, where it is found, and how it is used.
5. Students develop an understanding of the application of new technological developments to the preservation and communication of information, and in the identification of the impact of such developments on society.
6. Students continue in life-long intellectual growth beyond graduation.

As a student in INF2178H you will become familiar with causal inference in a conceptual, theoretical and practical way (Outcome 1). Importantly, this understanding will be based on a solid ethical foundation that will allow students to conduct themselves appropriately as information professionals (Outcome 2). Students will be conducting original research on real-world datasets, thereby contributing to the body of knowledge underlying the information sciences (Outcome 3). Students will be expected to conduct their analysis in a reproducible manner and appreciate the importance of this (Outcome 5). Finally, the course will provide students with an opportunity to create a portfolio of their work and techniques that they can continue to build on beyond the course (Outcome 6).

Grading

Please consult the Faculty of Information's:

- Grade Interpretation Guidelines: http://ischool.utoronto.ca/wp-content/uploads/2020/08/grade_interpretation_revised_August2020.pdf
- The University Assessment and Grading Practices Policy: http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/grading.pdf
- The Guidelines on the Use of INC, SDF, & WDR: https://www.sgs.utoronto.ca/policies-guidelines/inc-sdf-wdr/

These documents will form the basis for grading in the course.

Accommodations

Students with diverse learning styles and needs are welcome in this course. If you have a disability or a health consideration that may require accommodations, please feel free to approach Student Services and/or the Accessibility Services Office http://www.studentlife.utoronto.ca/as as soon as possible. The Accessibility Services staff are available by appointment to assess needs, provide referrals and arrange appropriate accommodations. The sooner you let us know your needs, the quicker we can assist you in achieving your learning goals in this course.

Writing Support

As stated in the Faculty of Information's Grade Interpretation Guidelines, "work that is not well written and grammatically correct will not generally be considered eligible for a grade in the A range, regardless of its quality in other respects." With this in mind, please make use of the writing support provided to graduate students by the SGS Graduate Centre for Academic Communication. The services are designed to target the needs of both native and non-native speakers and all programs are free. Please consult the current workshop schedule http://www.sgs.utoronto.ca/currentstudents/Pages/Current-Years-Courses.aspx for more information.

Academic integrity

Please consult the University's site on Academic Integrity http://academicintegrity.utoronto.ca/. The Faculty of Information has a zero-tolerance policy on plagiarism as defined 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/ppjun011995.pdf.

You should acquaint yourself with the Code. Please review the material in 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.

Cite it Right covers relevant parts of the U of T Code of Behaviour on Academic Matters (1995). It is expected that all Faculty of Information students take the Cite it Right workshop and the online quiz. Completion of the online Cite it Right quiz should be made prior to the second week of classes as the workshop is now interactive. To review and complete the workshop, visit the Orientation Workshop portion of the Inforum site: https://inforum.library.utoronto.ca/workshops/orientation

Absence Declaration Tool

During the COVID-19 pandemic, the University is temporarily suspending the need for a doctor's note or medical certificate for absences from academic participation; students should use the Absence Declaration tool on ACORN to declare an absence if they require consideration for missed academic work; students are responsible for contacting instructors to request the academic consideration they are seeking; students should record each day of their absence as soon as it begins, up until the day before they return to classes or other academic activities.

Notice of video recording and sharing (Download and re-use prohibited)

This course, including your participation, will be recorded on video and will be available to students in the course for viewing remotely and after each session.

Course videos and materials belong to your instructor, the University, and/or other sources depending on the specific facts of each situation, and are protected by copyright. Do not download, copy, or share any course or student materials or videos without the explicit permission of the instructor.

For questions about recording and use of videos in which you appear please contact your instructor.

Minimum Technical Requirements

The University of Toronto has identified minimum technical requirements needed for students to access remote/online learning: https://www.viceprovoststudents.utoronto.ca/covid-19/tech-requirements-online-learning/