STAT 547: Topics on Replicability in Statistics

Teaching team

Instructor: Lucy Gao

- Email: lucygao at ubc dot ca

- Office hours: By appointment, in person or on Zoom. I will be typically be in-office on Tuesdays and Thursdays.

Course description

Being able to replicate an experimental result has long been an important gold standard in science. This idea is foundational for the widespread usage of statistical methodology (especially hypothesis tests and confidence intervals) in science. However, there are many ways to misapply statistical methods in scientific studies, which can lead to irreplicable results. In this course, we will focus on two particular settings that are particularly likely to cause replicability challenges:

1. *Multiple testing*, i.e. performing multiple hypothesis tests and confidence tests in a single study
2. *Selective inference*, i.e. letting the study data guide your choice of hypothesis or parameter of interest.

My goals are to:

- Expose you to scientific problems where multiple testing and selective inference are needed
- Explore statistical consequences of applying "classical"/"textbook" statistical methodology to multiple testing and selective inference
- Discuss the downstream adverse scientific consequences of the aforementioned statistical consequences
- Introduce "modern" statistical practices that can help avoid adverse scientific consequences

Pre-requisites

You must be comfortable with (frequentist) inference at the level of a graduate mathematical statistics class using a textbook at the level of Casella and Berger or beyond. This might come from STAT 560/561 at UBC.

Communications

- Please post non-private questions or discussion about the course in the #stat547T channel of the UBC Statistics Slack Workspace.
- Course announcements will also be posted to this channel.
- Please email or DM me for any private questions or concerns.
- Note that I can't guarantee that I will be able to respond to messages outside of typical workday hours (0900-1700 PT M-F except on holidays).

Class meetings

- Class meets in person on Tuesdays and Thursdays [redacted] from 12:30pm-2pm for 6 weeks: Jan 9 - Feb 15, 2024.
- On rare occasions, class may need to be rescheduled, cancelled or held virtually on Zoom.
- Class meetings will be a mix of lecture and computer labs.

Regular attendance and participation are critical to your learning in the course. But if you need to miss class (e.g. due to illness or conference travel):

- The lecture notes and lab outlines will be posted online.
- Make a connection early in the term to another student in the class. You can help each other by answering each others' questions about what happened in class.
- You can ask follow-up questions on Slack, or Zoom office hours, or in person office hours after you're well.

Note that if you're sick, it's important that you stay home, no matter what you think you may be sick with (e.g., cold, flu, other). This precaution will help reduce risk and keep everyone safer.

Assessments

I would love for the course to have grade-free formative assessments only. However, the university requires a final grade. It will be calculated as follows:

Category	Contribution	Notes
Homework	40%	2 total: one around Week 3, and one around Week 5
Lab Participation	20%	4 total: one in each of Weeks 1-4
Project and Presentation	40%	Week 6

Homework

These will be a handful of exercises in the style of mathematical statistics. Most of the grading will be binary (you make a good effort at the problem or not). Your solutions must be LaTeXed and submitted as a **PDF** via Canvas. My solutions will be posted to Canvas no more than 3 business days after the homework deadline.

Lab Participation

Labs will mostly on exploring statistical consequences of applying classical statistical methodology for multiple testing and selective inference through computer simulation. I strongly believe that these will enhance and/or cement your understanding of the lecture material. Grading will be **effort-based**, i.e. did you show up and make a good-faith effort.

Project and Presentation

You will watch a 1 hour talk and discussion from the online International Seminar on Selective Inference, digest it, and give your peers a 5 minute summary of the talk in the last week of class. This can (and probably should) be an informal and non-technical summary: think of this like your friend in the class asking "Hey, you went to that talk, right? What was it about? Was it cool?" This will be followed by a 5-10 minute question and discussion period with me and your peers. Grading will be 75% effort-based (did you show up and make a good-faith effort) and 25% based on the quality of the summary and the discussion.

Concession and Late Work

Late Work

You have three "late days" to be used at your discretion during the term. When you have run out of late days, any further late days will result in the grade of the late work to be multiplied by 0.8 each day that it is late. Turning in work after I have posted grades to Canvas will result in a grade of 0. I will start grading no more than 3 business days after the assessment deadline, and I will not provide "real-time updates" of how close I am to posting grades.

Academic Concession

UBC no longer requires a doctor's note (or supporting documentation) for academic concession. A self-declaration will suffice – here is a template you can use. Please submit this to the instructor **as soon as possible**. If you have grounds for academic concession, then we will work together to find a fair solution.