Welcome to DSCI 554: Experimentation and Causal Inference

Contents

This frequentist course focuses on statistical evidence from randomized experiments versus observational studies along with applications of randomization, e.g., A/B testing for website optimization.

High-Level Goals

By the end of the course, students are expected to:

- Distinguish between experimentally-generated data and observational data, with particular reference to the strength of ensuing statistical conclusions regarding causality.
- Fit and interpret regression models for observational data, with particular reference to adjustment for potential confounding variables.
- Apply the principle of "block what you can, randomize what you cannot" in designing an A/B testing experiment.

Teaching Team

Position	Name	Slack Handle
Lecture/Lab Instructor	Payman Nickchi	`@Payman Nickchi`
Teaching Assistant	Anne-Sophie Fratzscher	`@Anne-Sophie Fratzscher (TA)`
Teaching Assistant	Haley Oleynik	`@Haley Oleynik (TA)`
Teaching Assistant	Mahdi Asmae	`@Mahdi (TA)`
Teaching Assistant	Tony Liang	`@Tony Liang (TA)`
Teaching Assistant	Jared Connoy	`@Jared Connoy (TA)`
Teaching Assistant	Ailar Mahdizadeh	`@Ailar Mahdizadeh`

Lecture Schedule

This course occurs during **Block 6** in the 2024/25 school year.

Course notes can be accessed **here**. **Typically, you should review these notes before each lecture.** Moreover, there is optional reading material.

Lecture	Topic	Optional Reading Material
1	**Multiple Comparisons**	• This chapter from Handbook of Biological Statistics by McDonald
2	**Confounding and Randomized versus Non-randomized Studies**	• section 1.3.5 - 1.5.2, inclusive (pages 19 - 26), from OpenIntro Statistics • Improving Library User Experience with A/B Testing: Principles and Process
3	**Randomization and Blocking**	• Refresher on ANOVA from Handbook of Biological Statistics • section 1.4 Experiments from OpenIntro Statistics
4	**More Blocking and Power**	• section 5.4 (pages 239 - 245), from OpenIntro Statistics • Stopping rules and regression to the mean
5	**More Power and Early Stopping in A/B Testing**	• Peeking in A/B testing at Etsy
6	**Observational Data: Stratifying and Modelling**	• Confounding in Observational Studies Explained • section 8.4 (pages 386 - 395) from OpenIntro Statistics
7	**Observational Data: Different Sampling Schemes**	• BU SPH on case-control sampling

Lecture	Topic	Optional Reading Material
		• Basics of observational study design
8	**Matched Case-Control Scheme, Ordinal Regressors, and Final Wrap-Up**	

See the **lecture learning objectives** for a detailed breakdown of lecture-by-lecture learning objectives.

Deliverables

This is an **assignment-based course**. The following deliverables will determine your course grade:

Assessment	Weight
Lab Assignment 1	12.5%
Lab Assignment 2	12.5%
Lab Assignment 3	12.5%
Lab Assignment 4	12.5%
Quiz 1	25%
Quiz 2	25%

> **Note:** A +1% final bonus mark will be granted to everybody if the class reaches a 60% response rate (or above) in the final teaching evaluations.

Lectures

Refer to the MDS calendar for lecture times and room numbers.

Lab Topics and Due Dates

	Lab Topic	Due Date
1	Simpson's Paradox, Multiple Testing, and A/B Testing (Lectures 1 and 2)	2025-03-29 06:00 p.m.
2	Statistical Questions, Experimental Terminology, A/B Testing Communication, and Three-Way ANOVA with Blocking (Lectures 3 and 4)	2025-04-05 06:00 p.m.
3	Power Analysis, Early Stopp ing in A/B Testing, and Causality Through Observational Studies (Lectures 5 and 6)	2025-04-13 06:00 p.m.
4	Practicum of Causality Through an Observational Study (Lectures 7 and 8)	2025-04-20 6:00 p.m.

Quizzes

Refer to the MDS calendar.

Office hours

Refer to the MDS calendar.

Communication

Slack Channel: https://ubc-mds.slack.com/messages/554_exper-causal-inf

- We will use Slack as the main communication channel.
- If you have any questions regarding the course content, lectures, labs, autograders, or any other course-related matters, we kindly request that you avoid direct messaging (DM) the instructor or TAs. Instead, please post your question on this Slack channel. This approach not only enables our TAs to respond promptly but also benefits other students who might have similar questions.
- **Response time:** We will try our best to reply to your inquiries as soon as possible during the normal working hours (9:00 a.m.- 5:00 p.m. Mon-Fri). If you send us a message outside of regular working hours, please expect a response on the next working day.

Reference Material

- Seltman HJ, Experimental Design and Analysis, 2015.
- Oehlert GW, A First Course in Design and Analysis of Experiments, 2010.
- O'Neil, Cathy and Schutt, Rachel. "Causality," Ch. 11 of Doing Data Science: Straight Talk from the Frontline, O'Reilly Media, 2013.
- Tang, Diane, et al. "Overlapping Experiment Infrastructure: More, Better, Faster Experimentation." Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010.

Further reading:

- Work by Judea Pearl, such as "The Book of Why".

Recommended Course Reviews

This course is taught in `R` (we will follow the `tidyverse` style guide) with a reasonable mathematical, statistical, and programming basis. We strongly recommend reviewing the following courses:

- *DSCI 551: Descriptive Statistics and Probability for Data Science*, for basic statistical and probabilistic concepts, and familiarity with the mathematical notation.
- *DSCI 552: Statistical Inference and Computation I*, for statistical inference concepts with a frequentist approach.
- *DSCI 561: Regression I*, for ordinary ordinary least-squares (OLS).

- *DSCI 562: Regression II*, for generalized linear models (GLMs).
- *DSCI 531: Data Visualization I*, for plotting tools using the package `ggplot2`.

Policies

See the general MDS policies.

Attribution

The course is built upon previous years' materials developed by previous instructors.

License