Stat 140: DESIGN OF EXPERIMENTS

Department of Statistics, Faculty of Arts and Sciences, Harvard University

Instructors: Marie-Abele Bind Office: Science Center 608 (ma.bind@mail.harvard.edu)

Office hours: Tuesdays 9:30-10:30am (Room 309) or by appointment

Course website: https://canvas.harvard.edu/courses/70155

Teaching fellow: Alice Sommer (Office hours: Thursday 5:30-6:30pm at Science Center Room 304 and Friday 10:30-11:30am at Science Center Room 400N)

References:

- On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9, Neyman J., *Statistical Science*, 1923, 1990
- On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Neyman J., *Journal of the Royal Statistical Society*, 1934
- Design of Experiments, Fisher R., *Olivier and Boyd*, 1935
- The Design and Analysis of Experiments, Kempthorne O., *Robert Krieger Publishing Company*, 1952
- Experimental Design, Cochran W., *Wiley Classics Library*, 1957
- The Theory of the Design of Experiments, Cox D. and Reid N., *Chapman & Hall/CRC*, 2000
- A Modern Theory of Factorial Designs, Mukerjee, R. and Wu J., *Springer Series in Statistics*, 2006
- Experimental Design for Biologists, Glass D. and Glass D., *Cold Spring Harbor Laboratory Press*, 2007
- Rerandomization to improve covariate balance in experiments, Morgan K. and Rubin D., *Annals of Statistics*, 2012
- Causal Inference for Statistics, Social, and Biomedical Sciences, Imbens G. and Rubin D., *Cambridge University Press*, 2015

Class meetings: Tuesday and Thursday 10:30-11:45am at Science Center Room 309

Sections: Thursday 4:30-5:30pm and Friday 9:30-10:30am at Science Center Room 304

Course objectives: The course covers topics on design of randomized experiments and causal inference at an undergraduate level. The objective is to equip students with classical and modern methodologies for designing and analyzing large and complex experiments in different scientific disciplines.

Intended audience: Stat 140 is intended for undergraduate and graduate students interested in designing and analyzing randomized experiments to estimate causal effects. Students from life sciences, biostatistics, epidemiology, economics, social science, political science, education, and any other field are warmly encouraged to attend.

Prerequisites: Both introductory courses in probability and inference (i.e., Stat 110 and Stat 111) are required. An introductory background in programming and linear models (e.g., Stat 139) is also preferred.

Assignments: There will be four assignments that give students an opportunity to master the methods discussed in class. There will be two in class midterms. These will generally be conceptual in nature.

Project and class presentation: The class projects will be completed in groups of size 1-3, depending on the actual enrollment. Students may choose topics of their own that are relevant to the course. Instructors will propose some projects too, and students (or groups) will be free to choose from them. Grading of projects will be based on (i) a written report and (ii) an in-class project presentation.

Grading: Assignments (40%, two days of grace period), class participation (5%), midterms (30%), project report (15%), project presentation (10%),

Topics:

1. History and basic principles of design of experiments
2. History of "big ideas" of causal inference (e.g., missing data problem and potential outcomes framework)
3. Assignment mechanism and essential assumptions
4. One-factor with two levels experiment
    - Fisherian inference
    - Neymanian inference
    - Bayesian inference
5. Matched-pair experiment
6. One-factor with more than two levels experiment
7. Randomized block
8. Negative and positive controls
9. Two-factor experiments
10. Full and fractional factorial experiment
11. Rerandomization
12. Complex randomization structure (e.g., crossover)
13. Complex settings and applications (e.g., high-dimensional, observational studies)

Tentative schedule:

Date	**Modules**	**Assignments**
Jan 28	Course overview and Module 1 (History and basic principles of design of experiments)	
Jan 30 - last class of shopping period	Module 2 (History of "big ideas" in causal inference)	
Feb 4	Guest lecture (Essential concepts): Donald Rubin	Assignment 1 and Project posted
Feb 6	Module 3 (Assignment mechanism)	
Feb 11	Module 4 (Fisherian inference)	**Assignment 1 due** and Assignment 2 posted
Feb 13	Guest lecture (P-value): Ronald Wasserstein	
Feb 18	Module 5 (Neymanian inference)	**Assignment 2 due**
Feb 20	**Midterm** 1	
Feb 25	Module 6 (Bayesian inference)	Assignment 3 posted
Feb 27	Module 7 (Matched-pair experiments)	

March 3	Class cancelled	**Assignment 3 due**
March 5	Module 8 (>2 levels)	
March 10	Class cancelled	
March 12	Guest lecture (Controls): Kristen Hunter	**Abstract of project due**
March 17	Spring break	
March 19	Spring break	
March 24	Module 9 (Rerandomization)	
March 26	Module 10 (Observational studies)	**Midterm 2 (online)**
March 31	Module 11 (Crossover experiment)	
April 2	Module 12 (Block randomized experiments)	**Updated abstract and one-page project due** and Assignment 4 posted
April 7	Module 13 (Two-factor)	
April 9	Module 14 (Full factorial)	**Assignment 4 due**
April 14	Guest lecture (Fractional factorial): Nicole Pashley	**Updated abstract and two-page project due**
April 16	Module 15 (Non-compliance)	**Midterm 3 (online)**
April 21	Module 16 (Missing data)	**Updated abstract and three-page project due**
April 23	Module 17 (High-dimensional)	
April 28	**Presentations**	**Full project report due**
April 30	Reading period	
May 5	Reading period	
May 7		
May 12		