# IPSA-USP Summer School 2020
# Causal Inference and Field Experiments

13-17 January 2020

Lectures: 9.00-12.00, Labs: 13.30-18.00



| | | | |
|---|---|---|---|
| Instructor: | Dr. Florian Foos | LSE | f.foos@lse.ac.uk |
| Teaching Assistant: | Andrea Junqueira | Texas A&M | andrea_junqueira@tamu.edu |



**Information:** "Correlation is not causation". You must have heard this warning many times. But, what then is causation, and how can we test causal hypotheses, and identify the effects of policies and programs? Starting from the ideal of a randomized experiment, the module introduces participants to the design-based approach to causal inference, based on the potential outcomes framework. Instead of using fancy modelling techniques to correct post-hoc for potential biases, the module encourages students to think about challenges to causal inference at the design stage of a study. Published work will be evaluated based on how it addresses three key assumptions underlying causal inference: independence, excludability, and non-interference.

After introducing students to the design-based approach to causal inference in general, the module will cover the design, conduct and analysis of randomized field experiments, in particular. The goal of the course is to provide participants with the methodological knowledge and the practical skills to design, analyse, and eventually conduct their own field experiments. The module is taught as a combination of lectures and applied computer labs.

**Prerequisites:** The only pre-requisite is any course covering (at any level of detail) linear regression. There is relatively little assumed knowledge, and the aim is to build the statistical foundations from the ground up. If you have conducted a hypothesis test of any kind, you probably have the requisite skills.

**Learning Outcomes:** Students will understand the potential outcomes framework, and the key assumptions underlying causal inference. They will also understand how to design, analyse and interpret randomized field experiments, and they will be aware of the specific challenges that field experimentalists face. Moreover, they will gain the practical skills of applying these insights about experimental design and statistical knowledge to experimental data.

**Problem Sets:** Students will complete two short practice problem sets in the evenings of days 2 and 4. The submission deadline is 9am the next morning, and we will provide feedback in the afternoon.

**Experimental Design:** For the research design session on day 5, students will be asked to come prepared with a 2-pager outlining an experimental design including 1) causal research question 2) hypotheses 3) experimental design (sample, experimental conditions, outcome variables, type of random assignment).

**Course Page:** There is a course dropbox to which all students will have access to.

**Required Textbook:**
Gerber, Alan and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*, New York: W.W. Norton, 2012.

**Recommended Texts on Field Experiments:**
John, Peter. *Field Experiments in Political Science and Public Policy: Practical Lessons in Design and Delivery*, Routledge, 2017.

Glennerster, Rachel and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*, Princeton University Press. 2013.

Karlan, Dean and Jacob Appel. *Failing in the Field*, Princeton University Press, 2016.

**Software:** Students will have a choice between using R or Stata. If you are unfamiliar with both languages, I would suggest using R, as it is free and open source. All analyses that we will conduct are easily done in either language, and code will be provided in both languages. However, most demonstrations will focus on R.

R Intro:

Imai, Kosuke. *Quantitative Social Science. An Introduction*. Princeton: Princeton University Press, 2017.

Grolemund, Garrett and Hadley Wickham. Learn R Online with R for Data Science: `https://r4ds.had.co.nz`

Phillips, Nathaniel D. The Pirate's Guide to R: `https://bookdown.org/ndphillips/YaRrr/`

**Lecture Outline**

1. Introduction to the Potential Outcomes Framework
2. Sampling Variability and Randomization Inference
3. Analyze as you Randomize: Blocking, clustering and covariates
4. Dealing with complications (non-compliance and attrition)
5. Designing and Executing Field Experiments

## 13 January

### Lecture: Introduction to the Potential Outcomes Framework

Gerber and Green: Chapters 1 and 2

Holland, Fundamental Problem of Causal Inference

Gelman, Forward Causal Inference

READ BEFORE DAY 1: Page, Stewart, 1998: Accepting the Gay Person, *Journal of Homosexuality* , 36: 2, 31-39.

- What is causal inference
- What is a field experiment
- Unobserved heterogeneity
- Potential outcomes
- Independence assumption
- Random assignment

- Potential and realised outcomes
- Difference-in-means estimator of the ATE
- Unbiasedness of ATE estimator
- Excludability and non-interference

### Lab: Basics of experimental data analysis

- Conduct a random assignment
- Load in a simulated dataset that includes $Y(0)$ and $Y(1)$.
- Estimate the ATE using simulated and real data.

## 14 January

### Lecture: Sampling variability and randomization inference

Gerber and Green: Chapter 3.

Coppock, Alexander: Randomization Inference Procedures with ri2 https://egap.org/methods-guides/10-procedures-ri2

Gomila, Robin: Logistic or Linear? Estimating Causal Effects of Treatments on Binary Outcomes Using Regression Analysis https://psyarxiv.com/4gmbv

- Sampling distribution of the ATE
- The true standard error
- Estimating the standard error
- Randomization inference
- Comparing linear regression and randomization inference

### Lab: Randomisation inference

- Generate a sampling distribution of the ATE
- Estimate the standard error
- Learn to obtain p-values via randomization inference

## 15 January

### Lecture: Analyze as you Randomize

Gerber and Green: Chapters 3 and 4

- Research design and experimental data analysis

1. Blocks
2. Clusters
3. Covariate-adjustment

### Lab: Analysing block and cluster randomised experiments

- Perform block and cluster random assignment
- Learn how clustering can widen a sampling distribution
- Learn how blocking can tighten a sampling distribution
- Learn how covariate adjustment can tighten a sampling distribution

## 16 January

### Lecture: Non-compliance and attrition

Gerber and Green: Chapter 5, 6 and 7

- One-sided noncompliance
- Two-sided noncompliance
- The Intent-to-Treat Effect (ITT)
- The Complier Average Causal Effect (CACE)
- Two-stage least squares estimation of the CACE
- Design-based solutions (placebo-designs)
- Attrition as a potential outcome: MIPO, MIPO $| X$, MITA
- How to avoid attrition
- What to do and not to do about attrition (Manski- and Lee bounds)

### Lab: CACE and attrition checks

- How to estimate the CACE using 2sls
- Conducting attrition checks
- Lee bound estimation

## 17 January

### Lecture: Designing and Executing Field Experiments

Gerber and Green: Chapters 12 and 13.

John: Chapters 1, 2 and 3.

Coppock, Alexander: 10 Things to Know About Statistical Power: https://egap.org/methods-guides/10-things-you-need-know-about-statistical-power

Grady Chris and Nuole Chen: 10 Things to Know About Pre-Analysis Plans: https://egap.org/methods-guides/10-things-pre-analysis-plans

Humphreys, Macartan. "Reflections on the ethics of social experimentation." *Journal of Globalization and Development* 6.1 (2015): 87-112.

- Ethics of field experimentation
- Working with partner organisations
- Power analysis
- Pre-Analysis Plans (PAPs)
- Pre-registration

### Lab: Workshop your designs

- Work on your own experimental designs
- Write a PAP