Experimental Methods
Department of Political Science and International Relations
Trinity Term 2020

Lecture: Ray Duch raymond.duch@nuffield.ox.ac.uk
Lab: Mats Ahrenshop mats.ahrenshop@politics.ox.ac.uk

Course Objectives

The course covers the design, implementation, and analytic tools necessary for conducting social science experiments and analysing experimental data. The course consists of two modules; module 1 runs from week 1-4 and an entirely optional module 2 from week 5-8.

Module 1 begins in Week 1 with a brief review of causal inference and potential outcomes as they relate to experimental design; estimating Average Treatment Effects, regression in analyzing experimental results, and alternative designs to simple binary treatment. Week 2 will focus on Randomization Inference (RI) including hypothesis testing, p-values, sharp nulls and RI regression, as well as covariates and block randomized designs. Week 3 will concern contingencies such as noncompliance and interference. Week 4 covers topics in external validity, such as subject pools, design strategies, and effect heterogeneity. Lectures will be accompanied by lab sessions to verify recent experimental studies using R.

Module 2 is entirely optional and students will not be examined on this material. Participants will learn how to program experiments with o-Tree, work with the subject recruitment software, analyse experimental data and run a lab experiment at the Nuffield Centre for Experimental Social Sciences (CESS). In addition, participants will learn how to program online experiments in Qualtrics and how to work with crowd-sourced subject pools. Module 2 will also have a CTSE (Comparative Time Sharing Experiments) component. Students will have an opportunity to design a short experiment that would be competitively evaluated for inclusion in large scale cross-national CESS Online omnibus experimental study. The subject payment costs and fielding costs would be covered by CESS.

Participants will have the opportunity to present their own experimental research and receive feedback from an experienced team of instructors. Upon completion of the course participants should be able to (1) formulate research questions that can be addressed using experiments, (2) design and carry out experiments, and (3) analyse and interpret results from social sciences experiments.

The course is appropriate for participants from any discipline who expect to include experimental social research as part of their research agenda. It is also appropriate for participants who want to become informed consumers of experimental research scholarship.

Course Prerequisites

Participants should have a basic background in research design and statistics. For example, with respect to research design, they should understand basic concepts such as causal inference, exogeneity, control group, and confounding effects. With respect to basic statistics, they should understand the principals of ordinary least squares regression; how to calculate simple measures of association; and have some familiarity with a statistical software package. The hands-on experimental data analysis lab sessions will use R.

Course Materials and Logistics

Lectures and lab sessions take place on Zoom with url provided. All materials are available on `https://github.com/mahrenshop/experimental_methods2020`. Core readings and the reading for verificiation exercise are mandatory. Students are requested to have R up and running on their machines for the lab sessions, and students should familiarize themselves with R prior to the first lab session if they have not already done so.

Students will need to have installed:

- Zoom (free account sufficient!)
- R and RStudio

Course Assignments

In order to receive credit students submit a pre-analysis plan at the end of term. The proposal outlines a research question and an experimental design to address the question. Furthermore, each lab session will assign short R programming exercises that are evaluated. The evaluation is as follows:

- PAP: 80% (due 4 June 1500 UTC+1)
- R exercises: 20%
  - G&G ex. 2.7 and 2.8 (due Friday 8 May 1500 UTC+1)
  - G&G ex. 3.5 and 3.6 (due Friday 15 May 1500 UTC+1)
  - G&G ex. 5.10 (due Friday 22 May 1500 UTC+1)

Module 1			
Day	Time ($UTC+1$)	Place	Topic
30 April	13:00 - 15:00	Zoom	Rubin Causal Model, ATE, alternative designs
	15:00 - 15:30	Zoom	Verification exercise: Study 1
7 May	13:00 - 15:00	Zoom	Randomization Inference (RI), Covariates
	15:00 - 15:30	Zoom	Verification exercise: Study 2
14 May	13:00 - 15:00	Zoom	Noncompliance, Interference
	15:00 - 15:30	Zoom	Verification exercise: Study 3
21 May	13:00 - 15:00	Zoom	External validity: effect heterogeneity, subject pools, design strategies
	15:00 - 15:30	Zoom	Topics in reproducible social science
Module 2			
26 May	15:30 - 16:30	Zoom	o-Tree I
2 June	15:30 - 16:30	Zoom	o-Tree II
9 June	15:30 - 16:30	Zoom	Programming Online Survey Experiments in Qualtrics
16 June	15:30 - 16:30	Zoom	CTSE

Week 1: Rubin Causal Model, ATE, alternative designs

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 2
- Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principals for Social Research*. Cambridge University Press, 2007
  - Ch. 1-2

Verification exercise

- Lauren Young (2019) The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe. *American Political Science Review* 113(1):140–155.

Week 2: Randomization Inference, Covariates

Core readings

- Susan Athey and Guido Imbens. The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, 1:73–140, 2017
- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 3-4

Verification exercise

- Alicia Dailey Cooperman. Randomization inference with rainfall data: Using historical weather patterns for variance estimation. *Political Analysis*, 25(3):277–288, 2017

Week 3: Noncompliance, Interference

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 5, 6, 8

Verification exercise

- Kalla, Joshua L., and David E. Brockman (2020) Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. *American Political Science Review*: 1–16.

Week 4: External validity: Effect heterogeneity, subject pools, design strategies

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 9
- David E. Broockman, Joshua L. Kalla, and Jasjeet S. Sekhon. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis*, 25(4):435?464, 2017.
- Duch R., Laroze D., Robinson T. and Beramendi P. 2020. Multi-modes for Detecting Experimental Measurement Error in *Political Analysis* 28(2):263-283.
- Alexander Coppock. Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, pages 1–16, 2018.

Topics in reproducible social science

- Garrett Christensen. Manual of Best Practices in Transparent Social Science Research (BITSS).
- Chris Grady and Nuole (Lula) Chen. 10 Things to Know About Pre-Analysis Plans. EGAP Methods Guides
- Duflo, Banerjee, Finkelstein, Katz, Olken, Sautman (2020) In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plan for RCTs in Economics. *NBER Working Paper No. 26993*

Additional readings

- Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016
- Justin Grimmer, Solomon Messing, and Sean J. Westwood. Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis*, 25(4):413?434, 2017
- Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized programme evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013
- Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295 EP –, 09 2012
- Donald P. Green and Holger L. Kern. Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public Opinion Quarterly*, 76(3):491–511, 2012
- Raymond M. Duch, Denise Laroze, Thomas S. Robinson, and Pablo Beramendi. Multi-modes for detecting experimental measurement error. Nuffield College Centre for Experimental Social Sciences Working Paper Series, 2018.

Online

- Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351?368, 2012
- Daniel G. Goldstein, Siddharth Suri, R. Preston McAfee, Matthew Ekstrand-Abueg, and Fernando Diaz. The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research*, 51(6):742–752, 2014
- Ryan T. Moore and Sally A. Moore. Blocking for sequential political experiments. *Political Analysis*, 21(4):507?523, 2013
- Connor Huff and Dustin Tingley. "who are these people?" evaluating the demographic characteristics and political preferences of mturk survey respondents. *Research & Politics*, 2(3):2053168015604648, 2015
- Kevin J. Mullinix, Thomas J. Leeper, James N. Druckman, and Jeremy Freese. The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2):109–138, 2015
- Alexander Coppock and Donald P. Green. Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science research. *Political Science Research and Methods*, 3(1):113–131, 2015

Social Media

- Justin Grimmer, Solomon Messing, and Sean J. Westwood. How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation. *American Political Science Review*, 106(4):703?719, 2012
- Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295 EP –, 09 2012

Field Experiments

- Esther Duflo, Rachel Glennerster, and Michael Kremer. Chapter 61 using randomization in development economics research: A toolkit. volume 4 of *Handbook of Development Economics*, pages 3895 – 3962. Elsevier, 2007
- Marianne Bertrand, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. Obtaining a driver's license in india: An experimental approach to studying corruption. *Quarterly Journal of Economics*, 122(4):1639–76, 2007
- Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? *American Economics Review*, 94:991, 2004
- Karthik Muralidharan and Venkatesh Sundararaman. Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, 119(1):39–77, 2011
- R. Glennerster. Chapter 5 - the practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In Abhijit Vinayak Banerjee and Esther Duflo, editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 175 – 243. North-Holland, 2017

Conjoint

- Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 531:1–30, 2013

List and Endorsement

- Graeme Blair, Kosuke Imai, and Jason Lyall. Comparing and combining list and endorsement experiments: Evidence from afghanistan. *American Journal of Political Science*, 58(4):1043–1063, 2014

References

[1] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

[2] Susan Athey and Guido Imbens. The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, 1:73–140, 2017.

[3] Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351?368, 2012.

[4] Marianne Bertrand, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. Obtaining a driver's license in india: An experimental approach to studying corruption. *Quarterly Journal of Economics*, 122(4):1639–76, 2007.

[5] Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? *American Economics Review*, 94:991, 2004.

[6] Graeme Blair, Kosuke Imai, and Jason Lyall. Comparing and combining list and endorsement experiments: Evidence from afghanistan. *American Journal of Political Science*, 58(4):1043–1063, 2014.

[7] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295 EP –, 09 2012.

[8] David E. Broockman, Joshua L. Kalla, and Jasjeet S. Sekhon. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis*, 25(4):435?464, 2017.

[9] Alicia Dailey Cooperman. Randomization inference with rainfall data: Using historical weather patterns for variance estimation. *Political Analysis*, 25(3):277–288, 2017.

[10] Alexander Coppock. Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, pages 1–16, 2018.

[11] Alexander Coppock and Donald P. Green. Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science research. *Political Science Research and Methods*, 3(1):113–131, 2015.

[12] Raymond M. Duch, Denise Laroze, Thomas S. Robinson, and Pablo Beramendi. Multimodes for detecting experimental measurement error. Nuffield College Centre for Experimental Social Sciences Working Paper Series, 2018.

[13] Esther Duflo, Rachel Glennerster, and Michael Kremer. Chapter 61 using randomization in development economics research: A toolkit. volume 4 of *Handbook of Development Economics*, pages 3895 – 3962. Elsevier, 2007.

[14] Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.

[15] R. Glennerster. Chapter 5 - the practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In Abhijit Vinayak Banerjee and Esther Duflo, editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 175 – 243. North-Holland, 2017.

[16] Daniel G. Goldstein, Siddharth Suri, R. Preston McAfee, Matthew Ekstrand-Abueg, and Fernando Diaz. The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research*, 51(6):742–752, 2014.

[17] Donald P. Green and Holger L. Kern. Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public Opinion Quarterly*, 76(3):491–511, 2012.

[18] Justin Grimmer, Solomon Messing, and Sean J. Westwood. How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation. *American Political Science Review*, 106(4):703?719, 2012.

[19] Justin Grimmer, Solomon Messing, and Sean J. Westwood. Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis*, 25(4):413?434, 2017.

[20] Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 531:1–30, 2013.

[21] Connor Huff and Dustin Tingley. "who are these people?" evaluating the demographic characteristics and political preferences of mturk survey respondents. *Research & Politics*, 2(3):2053168015604648, 2015.

[22] Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized programme evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

[23] Ryan T. Moore and Sally A. Moore. Blocking for sequential political experiments. *Political Analysis*, 21(4):507?523, 2013.

[24] Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principals for Social Research*. Cambridge University Press, 2007.

[25] Kevin J. Mullinix, Thomas J. Leeper, James N. Druckman, and Jeremy Freese. The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2):109–138, 2015.

[26] Karthik Muralidharan and Venkatesh Sundararaman. Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, 119(1):39–77, 2011.