Experimental Methods
University of Oxford
Trinity Term 2021

Lecture: Ray Duch raymond.duch@nuffield.ox.ac.uk

Wednesdays, 1600-1800 UTC+1

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 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. The theme for this module is around computational social science and experiments. Participants will learn how to program experiments with o-Tree and Qualtrics. Substantively, students will be introduced to topics in machine learning and experimental work as well as to applications in social media experimentation.

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 practice sessions take place on Zoom with url provided. All materials are available on RayDuch Github. 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

Module 1			
Day	Time ($UTC+1$)	Place	Topic
28 April	16:00 - 18:00	Zoom	Rubin Causal Model, ATE, alternative designs
5 May	16:00 - 18:00	Zoom	Randomization Inference (RI), Covariates
12 May	16:00 - 18:00	Zoom	Noncompliance, Interference, Mediation
19 May	16:00 - 18:00	Zoom	External validity: effect heterogeneity, subject pools, design strategies
Module 2			
26 May	15:30 - 16:30	Zoom	Machine Learning and Experimental Work
2 June	15:30 - 16:30	Zoom	Social Media Experimentation
9 June	15:30 - 16:30	Zoom	Programming Online Survey Experiments in Qualtrics
16 June	15:30 - 16:30	Zoom	Programming Experiments in o-Tree

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

Additional readings

- Gary Charness, Uri Gneezy, and Michael A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1 – 8, 2012
- Charles Bellemare, Luc Bissonnette, and Sabine Kröger. Simulating power of economic experiments: the powerbbk package. *Journal of the Economic Science Association*, 2(2):157–168, 2016
- Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014

– Molly Offer-Westort, Alexander Coppock, and Donald P. Green. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, n/a(n/a), 2021

Video Applications

– `https://www.youtube.com/watch?v=K9G3qgunYrI`

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

Week 3: Noncompliance, Interference, Mediation

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, 10

Additional readings

– Chapters 14, 15, 16 in Druckman and Green (2021) Advances in Experimental Political Science (CUP) - available via SOLO

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.

– Chapters 9 and 21 in Druckman and Green (2021) Advances in Experimental Political Science (CUP) - available via SOLO

Note that for Module 2, it will be good preparation for students to read Ch. 10 in Druckman and Green (2021) Advances in Experimental Political Science (CUP) - available via SOLO

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 con-

stituent 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*, 22(1):1–30, 2014

Sensitive Questions: Experiments

- 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

Adaptive Design

- Offer-Westort, Molly and Vitor Hadad and Susan Athey. Adaptive experimentation tutorial, 2021b. `https://mollyow.shinyapps.io/adaptive/`
- Molly Offer-Westort, Alexander Coppock, and Donald P. Green. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, n/a(n/a), 2021
- Molly Offer-Westort, Leah R Rosenzweig, and Susan Athey. Optimal policies to battle the coronavirus "infodemic" among social media users in sub-saharan africa, Feb 2021

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] Charles Bellemare, Luc Bissonnette, and Sabine Kröger. Simulating power of economic experiments: the powerbbk package. *Journal of the Economic Science Association*, 2(2):157–168, 2016.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[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.

[10] Gary Charness, Uri Gneezy, and Michael A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1 – 8, 2012.

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

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014.

[22] 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.

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

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

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

[26] 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.

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

[28] Molly Offer-Westort, Alexander Coppock, and Donald P. Green. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, n/a(n/a), 2021.

[29] Molly Offer-Westort, Leah R Rosenzweig, and Susan Athey. Optimal policies to battle the coronavirus "infodemic" among social media users in sub-saharan africa, Feb 2021.

[30] Offer-Westort, Molly and Vitor Hadad and Susan Athey. Adaptive experimentation tutorial, 2021b. `https://mollyow.shinyapps.io/adaptive/`.