

# The Theory and Practice of Field Experiments: An Introduction from the EGAP Learning Days

Jake Bowers,<sup>1</sup> Maarten Voors,<sup>2</sup> and Nahomi Ichino<sup>3</sup>

March 3, 2021

<sup>1</sup>Author order randomized. <https://jakebowers.org>

<sup>2</sup><https://sites.google.com/site/maartenvoors/>

<sup>3</sup><https://www.nahomiichino.com/>



# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                             | <b>5</b>  |
| 1.1      | How to use the book . . . . .                   | 6         |
| 1.2      | We would love to hear from you! . . . . .       | 6         |
| 1.3      | Acknowledgments . . . . .                       | 6         |
| <b>2</b> | <b>The Research Design Process</b>              | <b>7</b>  |
| 2.1      | Core Content . . . . .                          | 7         |
| 2.2      | Slides . . . . .                                | 8         |
| 2.3      | Design Form and Pre-registration . . . . .      | 9         |
| 2.4      | Resources . . . . .                             | 9         |
| <b>3</b> | <b>Causal Inference</b>                         | <b>11</b> |
| 3.1      | Core Content . . . . .                          | 11        |
| 3.2      | Slides . . . . .                                | 12        |
| 3.3      | Resources . . . . .                             | 12        |
| <b>4</b> | <b>Randomization</b>                            | <b>15</b> |
| 4.1      | Core Content . . . . .                          | 15        |
| 4.2      | Slides . . . . .                                | 16        |
| 4.3      | Resources . . . . .                             | 16        |
| <b>5</b> | <b>Hypothesis Testing</b>                       | <b>19</b> |
| 5.1      | Core Content . . . . .                          | 19        |
| 5.2      | Slides . . . . .                                | 20        |
| 5.3      | Resources . . . . .                             | 20        |
| <b>6</b> | <b>Estimands and Estimators</b>                 | <b>23</b> |
| 6.1      | Core Content . . . . .                          | 23        |
| 6.2      | Slides . . . . .                                | 24        |
| 6.3      | Resources . . . . .                             | 25        |
| <b>7</b> | <b>Statistical Power and Design Diagnosands</b> | <b>27</b> |
| 7.1      | Core Content . . . . .                          | 27        |
| 7.2      | Slides . . . . .                                | 28        |
| 7.3      | Resources . . . . .                             | 28        |

---

|           |   |           |
|-----------|---|-----------|
| <b>8</b>  | <b>Measurement</b>  | <b>31</b> |
| 8.1       | Core Content . . . . .  | 31        |
| 8.2       | Slides . . . . .  | 32        |
| 8.3       | Resources . . . . .   | 32        |
| <b>9</b>  | <b>Threats to the Internal Validity of Randomized Experiments</b> | <b>35</b> |
| 9.1       | Core Content . . . . .  | 35        |
| 9.2       | Slides . . . . .  | 36        |
| 9.3       | Resources . . . . .   | 36        |
| <b>10</b> | <b>Ethical Considerations</b>                                     | <b>39</b> |
| 10.1      | Core content . . . . .  | 39        |
| 10.2      | Slides . . . . .  | 40        |
| 10.3      | Resources . . . . .   | 40        |
| <b>A</b>  | <b>Glossary of Terms</b>  | <b>41</b> |
| A.1       | Key Concepts . . . . .  | 41        |
| A.2       | Statistical Inference . . . . .                                   | 42        |
| A.3       | Randomization Strategies . . . . .                                | 43        |
| A.4       | Factorial Designs . . . . .                                       | 44        |
| A.5       | Threats . . . . .   | 44        |
| <b>B</b>  | <b>Introduction to R and RStudio</b>                              | <b>47</b> |
| B.1       | R and RStudio . . . . .   | 47        |
| B.2       | Downloading R and RStudio . . . . .                               | 47        |
| B.3       | RStudio Interface . . . . .                                       | 49        |
| B.4       | Learning to Use R . . . . .                                       | 53        |

# Chapter 1

## Introduction

Over the past decade, [Evidence in Governance and Politics \(EGAP\)](#) has organized [Learning Days](#) workshops with the aim of building experimental social-science research capacity among principal investigators (PIs) – both researchers and practitioners – in Africa and Latin America. By sharing the practical and statistical methods of randomized field experiments with workshop participants, the Learning Days effort hopes to identify and nurture researcher networks around the world and to create strong, productive connections between these researchers and EGAP members.

The Learning Days workshops are a combination of design clinics, research presentations, guided work with statistical software, and topical lectures by a small group of instructors, largely professors and PhD students from the EGAP network. The workshops focus on methods for the design and analysis of randomized experiments in the field rather than on randomized experiments in the lab or non-randomized studies.

**This book** grew out of a desire to share the materials we developed for the Learning Days. The current version is written primarily for **instructors and organizers** of similar workshops and courses aimed at PIs — i.e., professors, post-doctoral fellows, PhD students, and NGO/government agency evaluators — who will implement randomized studies of programs related to institutions, governance, and development. Much of the material will also be useful as a refresher for past participants of the Learning Days workshops.

This book is a comprehensive overview of causal inference methods for researchers developing an experimental research design. It is organized in **modules** and covers topics such as causal inference, randomization, hypothesis testing, estimands, estimators, statistical power, measurement, threats to internal validity, and the ethics of experimentation. The modules appear in the order the Learning Days instructors have found most useful. However, the modules are linked to one another and can be reordered to suit your needs as an instructor. In the appendix, we include some course preliminaries including a [glossary of terms](#) and an [introduction to R and RStudio](#).

The book includes **slides** on the core content, the [EGAP Research Design Form](#), and **references** to research examples and slides used in previous Learning Days. This material builds

---

significantly on and links to EGAP’s work on methodology, summarized in the [EGAP Methods Guides](#). We have made significant extensions to the past Learning Days’ materials on hypothesis testing, estimation, and statistical power and added new modules on the research design process, measurement, and ethical considerations. The slides and modules presented here contain too much information to be covered in a single week, the usual length of a Learning Days workshop. We have chosen to present more rather than less information, however, to help instructors tailor their courses to their specific audiences.

## 1.1 How to use the book

To gain the most benefit from this book, please have [R](#) and [RStudio](#) installed on your machine. In fact, the slides assume that you will use Rmarkdown to adapt them for your own purposes.

To get going with R, see the module [Introduction to R and RStudio](#).

You can copy this book or parts thereof (e.g., slides, etc.) either by using the Download button on the front page of [http://github.com/egap/theory\\_and\\_practice\\_of\\_field\\_experiments](http://github.com/egap/theory_and_practice_of_field_experiments) or by using github directly (by forking this repository).

We are happy for anyone to use the materials as long as EGAP is attributed. See Creative Commons Attribution-ShareAlike 4.0 International License for the precise terms.

## 1.2 We would love to hear from you!

If you have any questions, feedback or have organized your own event, please get in touch! Simply post an issue on [Github](#) or contact us via email, [admin@egap.org](mailto:admin@egap.org).

## 1.3 Acknowledgments

The materials included in this book have been developed over the past years by various Learning Days instructors. These include (in alphabetical order) Jake Bowers, Jasper Cooper, Ana De la O, Lindsay Dolan, Natalia Garbiras Díaz, Macartan Humphreys, Nahomi Ichino, Salif Jaiteh, Gareth Nellis, Dan Nielson, Rafael Piñeiro, Fernando Rosenblatt, Tara Slough, Peter van der Windt and Maarten Voors. We thank Natalia Garbiras Díaz, Macartan Humphreys, Anghella Brigeth Rosero Rodriguez, and Tara Slough in particular for comments on an early draft of the book.

At EGAP, Matt Lisiecki, Ingrid Lee, Goldie Negelev, Max Mendez-Back and others have provided wonderful support. Learning Days have been generously funded by the Hewlett Foundation and supported by institutions around the world including the African School of Economics (Benin), Universidad Diego Portales (Chile), Universidad de los Andes (Colombia), Ghana Center for Democratic Development (Ghana), Mercy Corps (Guatemala), Invest in Knowledge (Malawi), NYU Abu Dhabi (UAE), Universidad Católica del Uruguay (Uruguay).

# Chapter 2

## The Research Design Process

This book aims to help you understand and design randomized field experiments. But before we dive into the details of research design, we need a good research question – a question that will advance knowledge or help make a policy decision or both. There is no simple recipe for finding or developing a good scientific or policy question, but our theories are important for articulating good questions that underlie impactful research. After formulating our question, we develop the best design possible within our resource constraints, using our knowledge of causal inference and statistics from the modules that follow.

This module introduces the [EGAP Research Design Form](#), a checklist to guide you through the many stages of the research process. The Learning Days workshops are organized around the Research Design Form. We also point you towards the [DeclareDesign](#) software package to explore the implications of different choices we could make for our research designs. Finally, this module discusses pre-analysis plans and pre-registration. When plan our analyses and make these plans public, we improve our chances of persuading others with our results.

### 2.1 Core Content

- A **good research question** advances science and/or is a question the answer to which will inform a policy decision.
- Certain **research designs** are better able to address certain questions. We want to choose the design that best answers our key questions within our constraints.
- The questions we ask arise — often implicitly — from our values and from our understanding about how the world works. These **theories** make our questions relevant. And the experiments that we execute teach us about the theory. That is, we hope that the evidence and data arising from these research designs improve our understanding.
- Core components of a research design
- Introduce core components of the [EGAP Research Design Form](#).
- Introduce a research design software package, [DeclareDesign](#).

- 
- The move in social science towards the **review of designs, rather than outcomes**.
  - **Pre-registration** – what it is, and why and how we should do it.

## 2.2 Slides

Below are slides with the core content that we cover in our lecture on research design. You can directly use these slides or make your own local copy and edit.

- [R Markdown Source](#)
- [PDF Version](#)
- [HTML Version](#)

You can also see the slides used in previous EGAP Learning Days:

- [The DeclareDesign presentation from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [The DeclareDesign presentation from EGAP Learning Days in Salima, Malawi, February 2017](#)
- [The DeclareDesign presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)

You can also see slides for Design Talks in previous EGAP Learning Days, where presenters focus on issues that come up in designing the research, rather than the results:

- [Design Talk from EGAP Learning Days at the African School of Economics, Benin, March 2018](#)
- [Design Talk 1 from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [Design Talk 2 from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [Design Talk 3 from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [Design Talk 1 from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)
- [Design Talk 2 from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)
- [Design Talk 3 from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)
- [Design Talk from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)



---

## 2.3 Design Form and Pre-registration

- [EGAP Research Design Form](#). A checklist we created for the Learning Days to guide you through the stages of the research process.
  - [Docx Version](#)
  - [PDF Version](#)
  - [HTML Version](#)
- Links to repositories for pre-registration/pre-analysis plans:
  - EGAP registry, hosted by OSF (<https://egap.org/registry/>)
  - AEA RCT registry (<https://www.socialscienceregistry.org/>)
  - OSF (<https://osf.io/registries>)
- Examples of other pre-registrations/pre-analysis plans:
  - SMS Messages in Mozambique from the [US Federal Government](#)
  - Police Body-Cameras from the [Lab @ DC](#)

## 2.4 Resources

### 2.4.1 EGAP Methods Guides

- EGAP Methods Guide [10 Things to Know about Pre-Analysis Plans](#)
- EGAP Methods Guide [10 Things to Know about Measurement in Experiments](#)

### 2.4.2 Books, Chapters, and Articles

- [Preregistration as a Tool for Strengthening Federal Evaluation](#). A white paper from the US Government's Office of Evaluation Sciences. You can also see examples of their pre-analysis plans on all of [their field experiment pages](#).
- Garret S. Christensen, Jeremy Freese, and Edward Miguel. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Oakland, California: University of California Press, 2019. The book summarizes new approaches in social science research on transparency and reproducibility.
- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton & Company, 2012. Chapter 12 includes some examples of experimental research designs.

---

### 2.4.3 Tools

- [DeclareDesign](#), an exciting and comprehensive set of software tools for describing, assessing, and conducting empirical research.

# Chapter 3

## Causal Inference

Much of social science is about causality. We might ask questions like whether voter registration increases political participation, whether bottom-up accountability can improve health outcomes, or whether personal narratives of immigrants help reduce prejudicial attitudes towards them.

Over the past decade, social science has become much more serious about how causal claims are made, building on a long history of work on causality dating back to the classic writings of Fisher and Rubin. We make greater use of experiments, and randomization has become the gold standard for addressing causal questions.

In this module, we introduce the counterfactual approach to causal inference and how statements with causal claims can be interpreted. We introduce the potential outcome framework and how random assignment helps us make claims about what would have happened in the absence of the policy, action or program we study. We discuss the three core assumptions for causal inference: random assignment of subjects to treatment, non-interference, and excludability.

### 3.1 Core Content

- What do we mean when we say “**cause**”? And why does it matter to be clear about the meaning of causal claims?
- An introduction to **potential outcomes** as a way to think about alternative states of the world.
- **Randomization** helps us learn about counterfactual causal claims in a particularly useful way.
- The three key **core assumptions** for causal inference: random assignment of subjects to treatment, non-interference, and excludability.
- Comparison of randomized studies with observational studies.

- 
- Randomization brings high internal validity, but it can't promise **external validity**.
  - Your causal question is closely linked so your [research design](#).

## 3.2 Slides

Below are slides with the core content that we cover in our lecture on causality. You can directly use these slides or make your own local copy and edit.

- [R Markdown Source](#)
- [PDF Version](#)
- [HTML Version](#)

You can also see the slides used in previous EGAP Learning Days:

- [The causal inference presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019](#)
- [The causal inference presentation from EGAP Learning Days at Universidad de Los Andes, Bogotá, April 2019](#)
- [The causal inference presentation from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [The causal inference presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The introduction to experiments presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The causal inference presentation from EGAP Learning Days in Salima, Malawi, February 2017](#)
- [The causal inference presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)

## 3.3 Resources

### 3.3.1 EGAP Methods Guides

- [EGAP Methods Guide 10 Things You Need to Know about Causal Inference](#)
- [EGAP Methods Guide 10 Strategies for Figuring Out If X Caused Y](#)
- [EGAP Methods Guide 10 Things You Need to Know about Mechanisms](#)
- [EGAP Methods Guide 10 Things to Know About External Validity](#)

---

## 3.3.2 Books, Chapters, and Articles

### 3.3.2.1 Classics

- Ronald A. Fisher. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935. Fisher introduces the idea of randomization and hypothesis testing as a way to learn about causal inference.
- Donald B. Rubin. “Estimating the Causal Effects of Treatments in Randomized and Nonrandomized Studies”. In: *J. Educ. Psych.* 66 (1974), pp. 688–701. Rubin introduces the idea of potential outcomes and links counterfactual conceptualizations of causality to statistical inference.

### 3.3.2.2 Contemporary Overviews

- Henry E Brady. “Causation and explanation in social science”. In: *The Oxford Handbook of Political Science*. 2008. URL: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-10>.
- Gerber and Green, *Field Experiments: Design, Analysis, and Interpretation*, Chapter 1. This book is a great resource for many topics in experimental design.
- Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2007, Chapter 1. This book includes nice examples of thinking through making causal claims from observational data.
- Rachel Glennerster and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press, 2013. This is a great introduction to running field experiments and discusses many examples.

## 3.3.3 EGAP Policy Briefs

Some examples of causal questions:

- EGAP Policy Brief 38: Are radio voter education campaigns effective in discouraging voters from voting for parties/candidates that engage in vote-buying?
- EGAP Policy Brief 51: Can free and anonymous information communication technology strengthen local accountability and improve the delivery of public services?
- EGAP Policy Brief 58: Can bottom-up accountability generate improvements in health outcomes?
- EGAP Policy Brief 69: Does bottom-up, citizen-based monitoring improve public service delivery?



# Chapter 4

## Randomization

The module on [causal inference](#) discussed the crucial role of randomization for drawing valid inferences from a comparison of treated and untreated groups. In this module, we move from theory to the first of many concrete choices for your research design.

We introduce four common ways to randomize treatment – simple, complete, block, and clustered – and when these different types of randomization may be available and appropriate. We also cover several popular designs including factorial designs and encouragement designs. The module provides some guidance on implementation, including best practices for checking for balance and ensuring replicability.

### 4.1 Core Content

- What is **randomization**? Random assignment is **not** the same as random sampling.
- Four common ways to randomize treatment:
  - **Simple**: randomly assign units to treatment (like a coin flip).
  - **Complete**: within a list of eligible units, assign a fixed number to receive a treatment (like drawing from a urn).
  - **Block (or stratified)**: assign treatment within specific strata or blocks, as if you are running an experiment within each block.
  - **Cluster**: assign groups (clusters) of observations to the same treatment condition.
- Some popular designs:
  - Randomized **access**: randomization to availability of a treatment.
  - Randomized **delayed access**: randomize the timing of access.
  - **Factorial**: randomize units to combinations of treatment arms.

- 
- **Encouragement**: randomize the invitation to receive treatment.
  - How do you check whether your randomization produced balance on observables? Typically we conduct randomization tests also known as balance tests using the  $d^2$  omnibus test from **xBalance** in the **RIttools** package (because it is randomization inference) or approximate this result with an  $F$ -test.
  - There are, of course, **limits to randomization**. We discuss some here and direct you to the module on [threats](#) for more.

## 4.2 Slides

Below are slides with the core content that we cover in our lecture on randomization. You can directly use these slides or make your own local copy and edit.

- [R Markdown Source](#)
- [PDF version](#)
- [HTML version](#)

The linked files shows how to [do replicable randomization in R](#). You can also see more examples of randomization in R at [10 Things You Need to Know About Randomization](#).

You can also see the slides used in previous EGAP Learning Days:

- [The design issues presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019 \(first section reviews randomization designs\)](#)
- [The randomization presentation from EGAP Learning Days at Universidad de Los Andes, Bogotá, April 2019](#)
- [The randomization presentation from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [The randomization presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The randomization presentation from EGAP Learning Days in Salima, Malawi, February 2017](#)
- [The randomization presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)

## 4.3 Resources

### 4.3.1 EGAP Methods Guides

- EGAP Methods Guide [10 Things You Need to Know About Randomization](#)
- EGAP Methods Guide [10 Things You Need to Know About Cluster Randomization](#)



---

### 4.3.2 Books, Chapters, and Articles

- [Standard operating procedures for Don Green’s lab at Columbia University](#). A comprehensive set of procedures and rules of thumb for conducting experimental studies.
- Glennerster and Takavarasha, *Running Randomized Evaluations: A Practical Guide*. Chapter 2 on randomization.
- Gerber and Green, *Field Experiments: Design, Analysis, and Interpretation*. Chapter 2: Causal Inference and Experimentation

### 4.3.3 EGAP Policy Briefs

#### *Factorial designs*

- [EGAP Policy Brief 57: How Media Influence Social Norms: Evidence from Mexico](#)
- [EGAP Policy Brief 58: Does Bottom-Up Accountability Work?](#)

#### *Randomizing access*

- [EGAP Policy Brief 24: Reducing Elite Capture in the Solomon Islands](#)

#### *Randomizing delayed access*

- [EGAP Policy Brief 35: Reducing Reconvictions Among Released Prisoners](#)
- [EGAP Policy Brief 60: Reducing Youth Support for Violence through Training and Cash Transfers in Afghanistan](#)

#### *Cluster randomization*

- [EGAP Policy Brief 22: Getting Out the Vote](#)

#### *Blocked cluster randomization*

- [EGAP Policy Brief 54: Incumbent Malfeasance Revelations](#)
- [EGAP Policy Brief 56: Reporting Corruption](#)

### 4.3.4 Tools

- [RIttools](#), a set of tools for randomization-based inference including balance testing.



# Chapter 5

## Hypothesis Testing

We cannot directly observe causal effects because of the *fundamental problem of counterfactual causal inference* ([causal inference module](#)). So how can we learn about these *unobserved causal effects* using what we do observe? In a randomized experiment, we can assess *guesses* or *hypotheses* about the unobserved causal effects by comparing what we observe in a given experiment to what we would observe if we were able to repeat the experimental manipulation and the guess or hypothesis were true.

In this module we introduce hypothesis testing, how it relates to causal inference,  $p$ -values, and what to do when we have multiple hypotheses to test.

### 5.1 Core Content

- What is a **good hypothesis**?
- The relationship between hypothesis testing and causal inference.
- **Hypothesis tests.**
  - Null hypotheses.
  - Estimators versus test statistics.
  - In an experiment, a reference distribution for a hypothesis test comes from the experimental design and the randomization.
  - $p$ -values and how to interpret the results of hypothesis tests.
- A good hypothesis test should (1) cast doubt on the truth rarely (i.e., have a controlled and low false positive rate), and (2) easily distinguish signal from noise (i.e., cast doubt on falsehoods often; have high statistical power).
- How would we know when our hypothesis test is doing a good job? ([Power analysis](#) is its own module).
  - False positive rates.

- 
- Correct coverage of a confidence interval.
  - Assessing the false positive rate of a hypothesis test for a given design and choice of test statistic; the case of cluster-randomized trials and robust cluster standard errors.
  - Be careful when testing **many hypotheses**, such as when you have more than two treatment arms or you are assessing the effects of a treatment on multiple outcomes. We should be careful to **adjust the  $p$ -values or confidence intervals** to reflect the number of tests/intervals produced.

## 5.2 Slides

Below are slides with the core content that we cover in our lecture on hypothesis testing. You can directly use these slides or make your own local copy and edit.

- [R Markdown Source](#)
- [PDF Version](#)
- [HTML Version](#)

You can also see the slides used in previous EGAP Learning Days:

- [The hypothesis testing presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019](#)
- [The hypothesis testing presentation from EGAP Learning Days at Universidad de Los Andes, Bogotá, April 2019](#)
- [The hypothesis testing presentation from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [The hypothesis testing presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The hypothesis testing presentation from EGAP Learning Days in Salima, Malawi, February 2017](#)
- [The hypothesis testing presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)

## 5.3 Resources

### 5.3.1 EGAP Methods Guides

- [EGAP Methods Guide 10 Things to Know about Hypothesis Testing](#)
- [EGAP Methods Guide 10 Things You Need to Know about Multiple Comparisons](#)

---

### 5.3.2 Books, Chapters, and Articles

- Gerber and Green, *Field Experiments: Design, Analysis, and Interpretation*. Chapter 3: Sampling Distributions, Statistical Inference, and Hypothesis Testing.
- Paul R. Rosenbaum. “Design of observational studies”. In: *Springer series in statistics* (2010). Chapter 2: Causal Inference in Randomized Experiments.
- Paul R. Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, 2017. Part I: Randomized Experiments.



# Chapter 6

## Estimands and Estimators

Randomized experiments generate good guesses about the average outcome under treatment and the average outcome under control. This allows us to write down unbiased estimators of average treatment effects. We can also use the randomization to describe how estimates generated by an estimator can vary from experiment to experiment in the form of standard errors and confidence intervals.

In this module, we introduce several types of estimands, the target quantity to be estimated. The choice of estimand is a scientific and policy-informed decision – what quantity is useful for us to learn about? In addition, we want to select an appropriate estimator for this estimand as part of the research design. We discuss how estimators are applied to data to generate an estimate of our estimand and how to characterize the variability of this estimate.

### 6.1 Core Content

- A **causal effect**,  $\tau_i$ , is a comparison of unobserved potential outcomes for each unit  $i$ . For example, this can be a difference or a ratio of unobserved potential outcomes.
- To learn about  $\tau_i$ , we can treat  $\tau_i$  as an **estimand** or target quantity to be estimated (this module) or as a target quantity to be hypothesized about ([hypothesis testing module](#)).
- Many focus on the **average treatment effect (ATE)**,  $\bar{\tau} = \sum_{i=1}^n \tau_i$ , in part, because it allows for easy **estimation**.
- An **estimator** is a recipe for calculating a guess about the value of an estimand. For example, the difference between the mean of observed outcomes for  $m$  treated units and the mean of observed outcomes for  $N - m$  untreated units is one estimator of  $\bar{\tau}$ .
- Different randomizations will produce different values of the same estimator targeting the same estimand. A **standard error** summarizes this variability in an estimator.
- A  $100(1 - \alpha)\%$  **confidence interval** is a collection of hypotheses that cannot be rejected at the  $\alpha$  level. We tend to report confidence intervals containing hypotheses

---

about values of our estimand and use our estimator as a test statistic.

- Estimators should (1) avoid systematic error in their guessing of the estimand (be unbiased); (2) vary little in their guesses from experiment to experiment (be precise or efficient); and perhaps ideally (3) converge to the estimand as they use more and more information (be consistent).
- **Analyze as you randomize** in the context of estimation means that (1) our standard errors should measure the variability from randomization and (2) our estimators should target estimands defined in terms of potential outcomes.
- We do not **control for** background covariates when we analyze data from randomized experiments. But covariates can make our estimation more **precise**. This is called **covariance adjustment**. Covariance adjustment in randomized experiments differs from controlling for variables in observational studies.
- A policy intervention (like a letter that encourages exercise) may *intend* to change behavior via an **active dose** (actual exercise). We can learn about the causal effect of the intention by randomly assigning letters; this is the **intent to treat effect, ITT**.
- We can learn about the causal effect of actual exercise by using the random assignment of letters as an **instrument** for the active dose (exercise itself) in order to learn about the causal effect of exercise **among those who would change their behavior after receiving the letter**. The average causal effect versions of these effects are often known as the **complier average causal effect** or the **local average treatment effect**.

## 6.2 Slides

Below are slides with the core content that we cover in this session.

- [R Markdown Source](#)
- [PDF Version](#)
- [HTML Version](#)

You can also see the slides used in previous EGAP Learning Days:

- The estimation presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019
- The estimation presentation from EGAP Learning Days at Universidad de Los Andes, Bogotá, April 2019
- The estimation presentation from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018
- The estimation presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016



---

You can also see discussion of the problems of estimating the effect of the active dose of a treatment in these slides (as well as discussion of the problems that missing data on outcomes cause for estimation of average causal effects):

- [The design issues presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019 \(first section reviews randomization designs\)](#)
- [The spillovers and attrition presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The threats presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The complications presentation from EGAP Learning Days in Salima, Malawi, February 2017](#)
- [The threats presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016 \(the middle section reviews ITT and non-compliance \)](#)

## 6.3 Resources

### 6.3.1 EGAP Methods Guides

- [EGAP Methods Guide 10 Types of Treatment Effect You Should Know About](#)
- [EGAP Methods Guide 10 Things to Know about Covariate Adjustment](#)
- [EGAP Methods Guide 10 Things to Know about Missing Data](#)
- [EGAP Methods Guide 10 Things to Know about the Local Average Treatment Effect](#)
- [EGAP Methods Guide 10 Things to Know about Spillovers](#)

### 6.3.2 Books, Chapters, and Articles

- Gerber and Green, *Field Experiments: Design, Analysis, and Interpretation*. Chapter 2.7 on excludability and non-interference, Chapter 3, Chapter 5 on one-sided noncompliance, Chapter 6 on two-sided noncompliance, Chapter 7 on attrition, Chapter 8 on interference between experimental units.
- Jake Bowers and Thomas Leavitt. “Causality & Design-Based Inference”. In: *The SAGE Handbook of Research Methods in Political Science and International Relations*. Ed. by Luigi Curini and Robert Franzese. Sage Publications Ltd, 2020.

### 6.3.3 Tools

- [DeclareDesign](#)
- [estimatr](#) package for R



# Chapter 7

## Statistical Power and Design Diagnosands

Before we run a study, we would like to know whether a particular design has the statistical power to detect an effect if it exists. It is difficult to learn from an under-powered study, since it would be unclear whether a null result indicates that there was no effect or just that we failed to detect a non-zero effect that exists. A power analysis can help you improve your design and allocate your resources better; it may even help you decide against conducting the study.

In this module, we introduce statistical power, core approaches to calculating power through analytical calculations and through simulation, and how design features such as blocking, covariate adjustment, and clustering impact power.

### 7.1 Core Content

- **Statistical power** is the ability of a study to detect an effect given that it exists.
- **Power analysis** is something we do before a study. It helps you figure out the sample you need or what effects you can detect. It is an essential step in research design and helps you communicate your design.
- Common approaches to power calculation:
  - **Analytical** power calculations (using a formula)
  - **Simulations** (for example, using `DeclareDesign`)
- **Covariate adjustment** and **blocking** can increase power.
- For **clustered designs** you need to take account of the intra-cluster correlation (the within-cluster variance relative to the overall variance).
- Power is closely linked to [study design](#), [hypothesis testing](#) and [estimation](#).

---

## 7.2 Slides

Below are slides with the core content that we cover in our lecture on power. You can directly use these slides or make your local copy and edit.

- [R Markdown Source](#)
- [PDF version](#)
- [HTML version](#)

You can also see the slides used in previous EGAP Learning Days:

- [The power presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019](#)
- [The power presentation from EGAP Learning Days at Universidad de Los Andes, Bogotá, April 2019](#)
- [The power presentation from EGAP Learning Days at Universidad Católica del Uruguay, Montevideo, March 2018](#)
- [The power presentation from EGAP Learning Days in Guatemala City, Guatemala, August 2017](#)
- [The power presentation from EGAP Learning Days in Salima, Malawi, February 2017](#)
- [The power presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)

## 7.3 Resources

### 7.3.1 EGAP Methods Guides

- [EGAP Methods Guide 10 Things to Know about Statistical Power](#)
- [EGAP Methods Guide 10 Things to Know about Covariate Adjustment](#)
- [EGAP Methods Guide 10 Things Your Null Results Might Mean](#)

### 7.3.2 EGAP Policy Briefs and PAPs

Some examples of power analysis in designs:

- [Pre-Analysis Plan. Accountability Can Transform \(ACT\) Health: A Replication and Extension of Bjorkman and Svensson \(2009\)](#)
- [EGAP Policy Brief 58: Can bottom-up accountability generate improvements in health outcomes?](#)

---

### 7.3.3 Tools

- Interactive power analysis
  - [EGAP Power Calculator](#)
  - [rpsychologist](#)
- R packages for power analysis
  - [pwr](#)
  - [DeclareDesign](#), see also <https://declaredesign.org/>



# Chapter 8

## Measurement

To estimate effects and test hypotheses, we often use an outcome of interest measured with quantitative data from surveys, behavioral games, or administrative records. For causal questions, we typically use data on immediate and final outcomes and core mechanisms. We use baseline data to identify relevant subgroups, adjust our estimates, or help block-randomize our treatment. Measurements should be valid and reliable. Be aware that data can be noisy (random error) and/or biased (systematic error).

This module discusses *what to measure* and *how to measure*. It shows how good measurement is closely linked to your [research design](#) and [statistical power](#).

### 8.1 Core Content

- When we represent some attribute of a unit by some number, letter, word, or symbol in some systematic way (perhaps in a cell in a simple dataset), we are **measuring**.
- A **valid** measure of a concept or phenomenon of interest should clearly represent that underlying and often abstract entity.
- A **reliable** measure of a concept would provide the same score for the unit of measurement (for example, a person or a village) if conditions were not changed.
- We can assess our theories of measurement using multiple approaches to measuring outcomes, covariates, or differences between units implied by different accounts of causal mechanisms.
- Invalid measurement can make it hard for your [research design](#) to effectively distinguish between alternative explanations for the relationship between treatment and outcome.
- Unreliable measurement can diminish [statistical power](#).
- Difficult measurement may call for a pilot study focused on measurement itself.

---

## 8.2 Slides

Below are slides with the core content that we cover in our lecture on measurement. You can directly use these slides or make your own local copy and edit.

- [R Markdown Source](#)
- [PDF Version](#)
- [HTML Version](#)

## 8.3 Resources

### 8.3.1 EGAP Methods Guides

- EGAP Methods Guide [10 Things to Know about Measurement in Experiments](#)
- EGAP Methods Guide [10 Things to Know about Survey Design](#)
- EGAP Methods Guide [10 Things to Know about Survey Implementation](#)

### 8.3.2 Books, Chapters, and Articles

- Robert Adcock and David Collier. “Measurement Validity: A shared Standard for Qualitative and Measurement Validity: A shared Standard for Qualitative and Quantitative Research.” In: *American Political Science Review* 95.3 (2001), pp. 529–546
- Alexandra Scacco and Shana S. Warren. “Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria”. In: *American Political Science Review* 112.3 (2018), pp. 654–677
- William R. Shadish, Thomas D. Cook, Donald Thomas Campbell, et al. *Experimental and quasi-experimental designs for generalized causal inference/William R. Shadish, Thomas D. Cook, Donald T. Campbell*. Boston: Houghton Mifflin, 2002
- Pedro C. Vicente. “Is Vote Buying Effective? Evidence from a Field Experiment in West Africa”. In: *Economic Journal* 124.574 (2014), F356–87

### 8.3.3 EGAP Policy Briefs

*Using survey data at multiple levels*

- [EGAP Policy Brief 58: Does Bottom-Up Accountability Work?](#)

*Using text messages*

- [EGAP Policy Brief 27: ICT and Politicians in Uganda](#)
- [EGAP Policy Brief 56: Reporting Corruption in Nigeria](#)

*Using administrative data*



- 
- [EGAP Policy Brief 16: Spillover Effects of Observers in Ghana](#)
  - [EGAP Policy Brief 67: Electoral Administration in Kenya](#)



# Chapter 9

## Threats to the Internal Validity of Randomized Experiments

Randomized experiments can run into issues that undermine their ability to demonstrate causal effects – that is, threaten the internal validity of randomized experiments. Some units might be missing outcome data and that missingness may be due to the treatment. They may not take the treatment status assigned to them or be subject to spillover effects from a treated neighbor.

In this module, we cover some common threats and some best practices to avoid or work around them.

### 9.1 Core Content

- Review the three core assumptions discussed in the [causal inference](#) module.
- We have said “Analyze as you randomize” in the module on [estimands and estimators](#). Remember that you randomized treatment **assignment**, not whether the treatment is received or whether a unit participates in data collection.
- **Missing data on the outcome (attrition)** is especially a problem if the patterns of missingness are caused by the treatment itself. This is a very common problem.
  - Do not drop observations that are missing outcome data from your analysis.
  - You may be able to **bound** estimates of treatment effects.
- **Non-compliance.** The effect of treatment assignment is not the same as the effect of receiving the treatment. Sometimes units will not comply with their assigned treatment status.
  - One-sided compliance occurs when some units assigned to treatment fail to take the treatment, but all units assigned to control do not take the treatment.

- 
- The local average treatment effect (LATE, also known as the complier average causal effect, CACE) is the average effect for the units that take the treatment when assigned, but not otherwise. If the monotonicity assumption and the exclusion restriction hold, we may be able to estimate LATE when we have non-compliance.
  - **“Spillover effects” or interference between units** is a violation of one of the core assumptions for causal inference ([causal inference](#)).
    - However, this may not be a problem if you are interested in spillover effects and/or have designed your research to account for it.
  - **Hawthorne effects** are when subjects behave differently because they are being observed.
  - **Non-excludability**. Treating treatment and control units differently, such as with different data collection processes or extra attention to the treated units, can confuse interpretation of experimental results.
    - If Hawthorne effects are present for treated units but not control units, then we have a violation of the excludability assumption.

## 9.2 Slides

Below are slides with the core content that we cover in our lecture on threats to the internal validity of randomized experiments. You can directly use these slides or make your local copy and edit.

- [R Markdown Source](#)
- [PDF version](#)
- [HTML version](#)

You can also see the slides used in previous EGAP Learning Days:

- [The threats presentation from EGAP Learning Days at the African School of Economics, Abomey-Calavi, June 2019](#) (first section reviews randomization designs)
- [The attrition and missing data presentation from EGAP Learning Days at Universidad Diego Portales in Santiago, Chile, May 2016](#)

## 9.3 Resources

### 9.3.1 EGAP Methods Guide

- EGAP Methods Guide [10 Things to Know about Missing Data](#)
- EGAP Methods Guide [10 Types of Treatment Effect You Should Know About](#)

- 
- [EGAP Methods Guide 10 Things to Know about the Local Average Treatment Effect](#)

### 9.3.2 Books, Chapters, and Articles

- [Standard operating procedures for Don Green’s lab at Columbia University](#). A comprehensive set of procedures and rules of thumb for conducting experimental studies.
- Gerber and Green, *Field Experiments: Design, Analysis, and Interpretation*. Chapters 5–8 address non-compliance, attrition, and interference.

### 9.3.3 EGAP Policy Briefs

- [EGAP Policy Brief 11: Election Observers and Fraud in Ghana](#)
- [EGAP Policy Brief 16: Spillover Effects of Observers in Ghana](#)



# Chapter 10

## Ethical Considerations

*A randomized experiment involves one group of humans changing the lives of another group humans.* Those working in government do this as a matter of course — their very job is to provide food, shelter, safety, justice, etc., to their people. Academics, whose work does not usually have immediate impacts on the public, must remember to also carefully consider how their research might change the lives of those exposed to the intervention, as well as those not exposed. When one person influences the life of another, the influencer has responsibilities not to harm the person being influenced.

This module discusses the core topics on research ethics, such as privacy and autonomy; the basic principles relating to respect for persons, beneficence, and justice; and how informed consent helps communicate these principles to study participants.

### 10.1 Core content

- Research must weigh the **potential benefits** of the knowledge to be gained from the research against the **potential harms** it may do to human subjects.
- How would you feel if you were a research subject in your study? In the control group? In the treatment group? A relatively high-status member of the community? A relatively low-status member of the community?
- Key tenets: **privacy** and **autonomy**.
- Basic principles in the Belmont Report: **respect for persons, beneficence, justice**.
- **Informed consent**: Can you ensure that research subjects have the freedom to refuse to participate and/or drop out of the study if they want to? Can you ensure that research subjects can report problems that might arise?
- Challenges for social science experimental research in general:
  - Many more people may benefit (or suffer from) your intervention than directly participate in your study.

- 
- Changing election results or corruption can produce large societal changes. Is this beyond the remit of research?

## 10.2 Slides

Below are slides with the core content that we cover in this session.

- [R Markdown Source](#)
- [PDF Version](#)
- [HTML Version](#)

## 10.3 Resources

- [EGAP Research Principles and Work on Ethics](#)
- [Belmont Report](#)
- [Institutional Review Boards in the US](#)
- [Example: Research Ethics at Oxford University in the UK](#)
- [Example: Ethics for Researchers in the EU](#)
- [Example: Research Ethics at the Universidad Catolica de Chile](#)

### 10.3.1 Books, Chapters, and Articles

- Edward Asiedu et al. *A Call for Structured Ethics Appendices in Social Science Papers*. Working Paper 28393. National Bureau of Economic Research, Jan. 2021. DOI: [10.3386/w28393](https://doi.org/10.3386/w28393). URL: <http://www.nber.org/papers/w28393>
- David K. Evans. *Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science*. Working Paper 565. Center for Global Development, Feb. 2021. URL: <https://www.cgdev.org/sites/default/files/WP565-Evans-Ethical-issues-and-RCTs.pdf> pdf

### 10.3.2 Papers and Pre-Analysis Plans

Examples of PAPs and papers that discuss ethical issues:

- [Pre Analysis Plan: The Effects of Non-Food Item Vouchers in a Humanitarian Context The Case of the Rapid Response to Movements of Population Program in Congo](#)
- [Paper: Appendix E.1 in Countering violence against women by encouraging disclosure: A mass media experiment in rural Uganda](#)



# Appendix A

## Glossary of Terms

Below are some core terms frequently used throughout the book and more broadly in discussions of randomized field experiments.

### A.1 Key Concepts

See the module on [causal inference](#), [estimands and estimators](#).

- **Potential outcome**  $Y_i(T)$  The outcome  $Y$  that unit  $i$  *would* have under treatment condition  $T$ . We think of these as fixed quantities for a specific point in time.  $T$  can be 0 for control or 1 for treatment if there is only one type of treatment. See the module on [causal inference](#).
- **Treatment effect**  $\tau_i$  **for unit**  $i$  The contrast between potential outcomes under two treatment conditions for unit  $i$ . We typically define the treatment effect as the difference in potential outcomes under treatment and control,  $Y_i(1) - Y_i(0)$ . See the module on [causal inference](#).
- **Fundamental problem of causal inference** in the counterfactual framework. We can't observe both  $Y_i(1)$  and  $Y_i(0)$  for a given unit, so we can't get  $\tau_i$  directly. See the module on [causal inference](#).
- **Estimand** The thing you want to estimate. An example of an estimand is the average treatment effect. In counterfactual causal inference, this is a function of potential outcomes, not fully observed outcomes. See the module on [estimands and estimators](#).
- **Estimator** How you make a guess about the value of your estimand from the data you have (i.e., observed). An example of an estimator is the difference-in-means. See the module on [estimands and estimators](#).
  - **Average treatment effect, ATE** The average of the treatment effect for all individuals in your subject pool. This is a type of **estimand**. If we define  $\tau_i$  to be  $Y_i(1) - Y_i(0)$ , then the ATE is  $\overline{Y_i(1) - Y_i(0)}$ , which is also equivalent to  $\overline{Y_i(1)} - \overline{Y_i(0)}$ . Notice that we do not use the  $E[Y_i(1)]$  style of notation here because  $E[\cdot]$  means “average over repeated operations,” but  $\overline{Y}$  means “average over a set of observations”. See the module on [causal inference](#) and the module

---

on [estimands and estimators](#).

- **Random sampling** Selecting subjects from a population with known probabilities strictly between 0 and 1.
- **$k$ -arm experiment** An experiment that has  $k$  treatment conditions (including control). See the module on [randomization](#).
- **Random assignment** Assigning subjects to experimental conditions with known probabilities strictly between 0 and 1. This is equivalent to random sampling without replacement from the potential outcomes. There are several strategies for random assignment: simple, complete, cluster, block, blocked-cluster. See the module on [randomization](#).
- **External validity** Findings from your study teach you about contexts outside of your sample — in other locations or for other interventions.

## A.2 Statistical Inference

See modules on [hypothesis testing](#) and [statistical power](#).

- **Hypothesis** A simple, clear, falsifiable claim about the world. In counterfactual causal inference, this is a statement about a relationship among potential outcomes, like  $H_0 : Y_i(T_i = 0) = Y_i(T_i = 1) + \tau_i$  for the hypothesis that the potential outcome under treatment is the potential outcome under control plus some effect for each unit  $i$ . See the module on [hypothesis testing](#).
- **Null hypothesis** A conjecture about the world that you may reject after seeing the data. See the module on [hypothesis testing](#).
- **Sharp null hypothesis of no effect** The null hypothesis that there is no treatment effect for any subject. This means  $Y_i(1) = Y_i(0)$  for all  $i$ . We might write this as  $H_0 : Y_i(T_i = 0) = Y_i(T_i = 1)$ . See the module on [hypothesis testing](#).
- **$p$ -value** The probability of seeing a test statistic as large (in absolute value) as or larger than the test statistic calculated from observed data. See the module on [hypothesis testing](#).
- **One-sided vs. two-sided test** When you have a strong expectation that the effect is either positive or negative, you can conduct a one-sided test. When you do not have such a strong expectation, conduct a two-sided test. A one-sided test has more power than a two-sided test for the same experiment. See the module on [hypothesis testing](#).
- **Standard deviation** Square root of the mean-square deviation from the average of a variable. It is a measure of the dispersion or spread of a statistic.  $SD_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
- **False Positive Rate/Type I Error of a Test** A well-operating hypothesis test rejects a hypothesis about a true causal effect no more than  $\alpha$  % of the time. The false positive rate is the rate at which a test will cast doubt on a true hypothesis. It is the rate at which the test will encourage the analyst to say “statistically significant” when, in fact, there is no causal relationship. See the module on [hypothesis testing](#).
- **Sampling distribution** The distribution of estimates (e.g., estimates of the ATE) for all possible treatment assignments. In design-based statistical inference for random-

---

ized experiments, the distribution of estimates from an estimator is generated from randomizations. Many call this a “sampling distribution” because textbooks often use the idea of repeated samples from a population rather than repeated randomizations to describe this kind of variation.

- **Standard error** The standard deviation of the sampling distribution. A bigger standard error means that our estimates are more susceptible to sampling variation. See the module on [estimands and estimators](#).
- **Coverage of a confidence interval** A well-operating confidence interval contains the true causal effect  $100(1 - \alpha)$  % of the time. A confidence interval has *incorrect coverage* when it excludes the true parameter less than  $100(1 - \alpha)$ % of the time. For example, a 95% confidence interval is supposed to only exclude the true parameter less than 5% of the time.
- **Statistical power of a test** Probability that a test of causal effects will detect a statistically significant treatment effect if the effect exists. See the module on [statistical power](#). This depends on:
  - The number of observations in each arm of the experiment
  - Effect size (usually measured in standardized units)
  - Noisiness of the outcome variable
  - Significance level ( $\alpha$ , which is fixed by convention)
  - Other factors including what proportion of your units are assigned to different treatment conditions.
- **Intra-cluster correlation** How correlated the potential outcomes of units are within clusters compared to across clusters. Higher intra-cluster correlation hurts power.
- **Unbiased** An estimator is unbiased if you *expect* that it will return the right outcome. That means that if you were to run the experiment many times, the estimate might be too high or too low sometimes but it will be right on average. See the module on [estimands and estimators](#).
- **Bias** Bias is the difference between the average value of the estimator across its sampling distribution and the single, fixed value of the estimand. See the module on [estimands and estimators](#).
- **Consistency of an estimator** An estimator that produces answers that become ever nearer to the true value of the estimand as the sample size increases is a *consistent estimator* of that estimand. A consistent estimator may or may not be unbiased. See the module on [estimands and estimators](#).
- **Precision/Efficiency of an estimator** The variation in or width of the sampling distribution of an estimator. See the module on [estimands and estimators](#).

## A.3 Randomization Strategies

See the module on [randomization](#).

- **Simple** An independent coin flip for each unit. You are not guaranteed that your experiment will have a specific number of treated units.
- **Complete** Assign  $m$  out of  $N$  units to treatment. You know how many units will be treated in your experiment and each unit has a  $m/N$  probability of being treated.

---

The number of ways treatment can be assigned (number of permutations of treatment assignment) is  $\frac{N!}{m!(N-m)!}$ .

- **Block** First divide the sample into blocks, then do complete randomization separately in each block. A block is a set of units within which you conduct random assignment.
- **Cluster** Clusters of units are randomly assigned to treatment conditions. A cluster is a set of units that will always be assigned to the same treatment status.
- **Blocked-Cluster** First form blocks of clusters. Then in each block, randomly assign the clusters to treatment conditions using complete randomization.

## A.4 Factorial Designs

See the module on [randomization](#).

- **Factorial design** A design with more than one treatment, with each treatment assigned independently. The simplest factorial design is a 2 by 2.
- **Conditional marginal effect** The effect of one treatment, conditional on the other being held at a fixed value. For example:  $Y_i(T_1 = 1|T_2 = 0) - Y_i(T_1 = 0|T_2 = 0)$  is the marginal effect of  $T_1$  conditional on  $T_2 = 0$ .
- **Average marginal effect** Main effect of each treatment in a factorial design. It is the average of the conditional marginal effects for all the conditions of the other treatment, weighted by the proportion of the sample that was assigned to each condition.
- **Interaction effect** In a factorial design, we may also estimate interaction effects.
  - No interaction effect: one treatment does not amplify or reduce the effect of the other treatment.
  - Multiplicative interaction effect: the effect of one treatment depends on which condition a unit was assigned for the other treatment. This means one treatment *does* amplify or reduce the effect of the other. The effect of two treatments together is *not* the sum of the effect of each treatment.

## A.5 Threats

See the module on [threats](#).

- **Hawthorne effect** When a subject responds to being observed.
- **Spillover** When a subject responds to another subject's treatment status. Example: my health depends on whether my neighbor is vaccinated, as well as whether I am vaccinated.
- **Attrition** When outcomes for some subjects are not measured. This might be caused, for example, by people migrating, refusing to respond to endline surveys, or dying. This is especially problematic for inference when it is correlated with treatment status.
- **Compliance** A unit's treatment status matches its assigned treatment condition. Example of non-compliance: a unit assigned to treatment doesn't take it. Example of compliance: a unit assigned to control does not take treatment.
- **Compliance types** There are four types of units in terms of compliance:

- 
- **Compliers** Units that would take treatment if assigned to treatment and would be untreated if assigned to control.
  - **Always-takers** Units that would take treatment if assigned to treatment and if assigned to control.
  - **Never-takers** Units that would be untreated if assigned to treatment and if assigned to control.
  - **Defiers** Units that would be untreated if assigned to treatment and would take treatment if assigned to control.
  - **One-sided non-compliance** The experiment has only compliers and *either* always-takers or never-takers. Usually, we think of one-sided non-compliance as having only never-takers and compliers, meaning that that local average treatment effect is the effect of treatment on the treated.
  - **Two-sided non-compliance** The experiment may have all four latent groups.
  - **Encouragement design** An experiment that randomizes  $T$  (treatment assignment), and we measure  $D$  (whether the unit takes treatment) and  $Y$  (outcome). We can estimate the ITT and the LATE (local average treatment effect, aka CACE—complier average causal effect). It requires three assumptions.
    - **Monotonicity** Assumption of either no defiers or no compliers. Usually we assume no defiers which means that the effect of assignment on take up of treatment is either positive or zero but not negative.
    - **First stage** Assumption that there is an effect of  $T$  on  $D$ .
    - **Exclusion restriction** Assumption that  $T$  affects  $Y$  only through  $D$ . This is usually the most problematic assumption.
  - **Intention-to-treat effect (ITT)** The effect of  $T$  (treatment assignment) on  $Y$ .
  - **Local average treatment effect (LATE)** The effect of  $D$  (taking treatment) on  $Y$  for compliers. Also known as the complier average causal effect (CACE). Under the exclusion restriction and monotonicity, the LATE is equal to ITT divided by the proportion of your sample who are compliers.
  - **Downstream experiment** An encouragement design study that takes advantage of the randomization of  $T$  by a previous study. The outcome from that previous study is the  $D$  in the downstream experiment.



# Appendix B

## Introduction to R and RStudio

Throughout the book we include R code for estimation, simulation, and creating examples. We used RStudio to create the slides. To personalize them for your own purpose, we assume you will use R Markdown. Below, we include guides to setting up R and RStudio on your machine, as well as some basic commands that are frequently used.

### B.1 R and RStudio

R is a free software environment most commonly used for statistical analysis and computation. Because Learning Days participants arrive with different statistical backgrounds and preferred statistical software, we use R to ensure that everyone is on the same page. We advocate the use of R more generally for its flexibility, wealth of applications, and comprehensive support mostly through online fora.

RStudio is a free, open source integrated development environment with a user interface that makes R much more user-friendly. R Markdown, a feature of RStudio, enables the easy output of code, results, and text in a .pdf, .html, or .doc format.

### B.2 Downloading R and RStudio

#### B.2.1 Downloading R

R can be freely downloaded from CRAN at the link corresponding to your operating system:

- For **Windows**: <https://cran.r-project.org/bin/windows/base/>
- For **Mac OS X**: <https://cran.r-project.org/bin/macosx/>.
  - Select R-4.0.4.pkg for OS X 10.13 and higher.
  - Select R-3.6.3.nnpkg for OS X 10.11-10.12.
  - Select R-3.3.3.nnpkg for OS X 10.19-10.10.
  - Select R-3.2.1-snowleopard.pkg for OS X 10.6-10.8.

## B.2.2 Downloading RStudio

RStudio can be freely downloaded from the RStudio website, <https://www.rstudio.com/products/rstudio/download/>. In the table, click the blue **Download** button at the top of the left column, “RStudio Desktop Open Source License” as depicted below in Figure B.1. Once you select this button, the page will jump to a list of download options as depicted in Figure B.2.

- For **Windows**, select **Windows 10/8/7**.
- For **Mac OS X**, select **Mac OS X 10.13+**.

|                        | RStudio Desktop<br>Open Source License        | RStudio Desktop Pro<br>Commercial License | RStudio Server<br>Open Source License         | RStudio Server Pro<br>Commercial License              |
|------------------------|---|---|---|---|
|                        | <b>Free</b>                                   | <b>\$995</b><br>/year                     | <b>Free</b>                                   | <b>\$4,975</b><br>/year<br>(5 Named Users)            |
|                        | <b>DOWNLOAD</b><br><a href="#">Learn more</a> | <b>BUY</b><br><a href="#">Learn more</a>  | <b>DOWNLOAD</b><br><a href="#">Learn more</a> | <b>BUY</b><br><a href="#">Evaluation   Learn more</a> |
| Integrated Tools for R | ✓   | ✓   | ✓   | ✓   |
| Priority Support       |   | ✓   |   | ✓   |
| Access via Web Browser |   |   | ✓   | ✓   |

Figure B.1: Select Download in the “RStudio Desktop Open Source License” column.

### All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

| OS                  | Download                                    | Size      | SHA-256  |
|---------------------|---|-----------|----------|
| Windows 10/8/7      | <a href="#">RStudio-1.4.1106.exe</a>        | 155.97 MB | d2ff8453 |
| macOS 10.13+        | <a href="#">RStudio-1.4.1106.dmg</a>        | 153.35 MB | c64d2cda |
| Ubuntu 16           | <a href="#">rstudio-1.4.1106-amd64.deb</a>  | 118.45 MB | 1fc82387 |
| Ubuntu 18/Debian 10 | <a href="#">rstudio-1.4.1106-amd64.deb</a>  | 121.07 MB | 3b5d3835 |
| Fedora 19/Red Hat 7 | <a href="#">rstudio-1.4.1106-x86_64.rpm</a> | 138.18 MB | a9e6ddc4 |
| Fedora 28/Red Hat 8 | <a href="#">rstudio-1.4.1106-x86_64.rpm</a> | 138.16 MB | 35e57c1c |
| Debian 9            | <a href="#">rstudio-1.4.1106-amd64.deb</a>  | 121.33 MB | c7c9dd68 |
| OpenSUSE 15         | <a href="#">rstudio-1.4.1106-x86_64.rpm</a> | 123.57 MB | 3539d9c3 |

Figure B.2: Select the Windows 10/8/7 link for Windows or the Mac OS X 10.13+ link for Mac.



## B.3 RStudio Interface

When you open RStudio for the first time, there should be three panels visible, as depicted in Figure B.3 below.

- Console (left panel)
- Accounting (upper right panel): includes Environment and History tabs
- Miscellaneous (lower right panel)

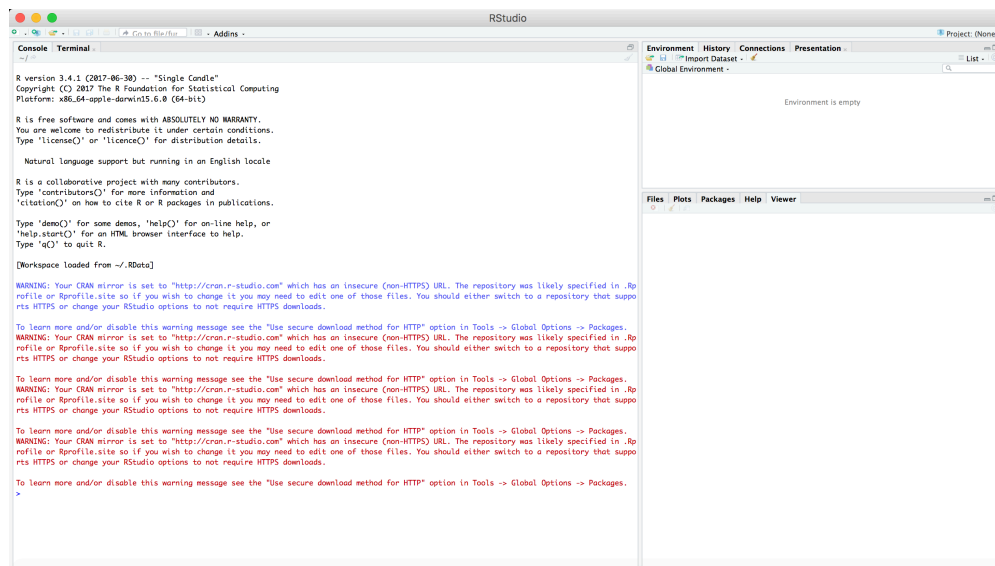


Figure B.3: When you open RStudio, there are three panels visible: the Console (left), Accounting (upper right), and Miscellaneous (lower right).

### B.3.1 Console

You can execute all operations in the console. For example if you enter `4 + 4` and hit the Enter/Return key, the Console will return `[1] 8`.

To make sure everyone is prepared to use R at Learning Days, we ask participants to run one line of code in the Console to download several R packages. Packages are fragments of reproducible code that allow for more efficient analysis in R. To run these lines, copy the following code into the Console and hit your Return/Enter key. You must be connected to the internet to download packages.

```
install.packages(c("ggplot2", "dplyr", "AER", "arm", "MASS", "sandwich",  
                  "lmtree", "estimatr", "coin", "randomizr", "DeclareDesign"))
```

If successfully downloaded, your Console will resemble Figure B.4, except that the urls will differ depending on your location.

[illegible]

Figure B.4: An image of the Console after executing the three lines of code listed above.

### B.3.2 Editor

In order to write and save reproducible code, we will open a fourth panel, the Editor, by clicking on the icon with a white page with a plus sign on the upper-left corner of the RStudio interface and selecting **R Script**, as depicted in Figure B.5.

Once the R script is opened, there should be four panels within the RStudio interface, now with the addition of the Editor panel. We can execute simple arithmetic by entering a formula in the editor and pressing **Control + Enter** (Windows) or **Command + Enter** (Mac). The formula and the “answer” will appear in the Console, as depicted in Figure B.6, with red boxes added for emphasis.

R can be used for any arithmetic operation including, but not limited to, addition (+), subtraction (-), scalar multiplication (\*), division (/), and exponentiation (^).

### B.3.3 Accounting

Beyond basic functions, we can also store values, data, and functions in the global environment. To assign a value to a variable, use the `<-` operator. All stored values, functions, and data will appear in the Environment tab in the Accounting panel. In Figure B.7, we define the variable `t` to take the value  $3 \times \frac{6}{14}$ , and can see that it is stored under Values.

We also load a dataset. Here, “ChickWeight” is a dataset built into R; most datasets will be loaded from the web or other files on your computer through an alternate method. We can see that ChickWeight contains 578 observations of 4 variables and is stored in the Environment. By clicking on the name ChickWeight a tab will enter with the dataset in your Editor window.

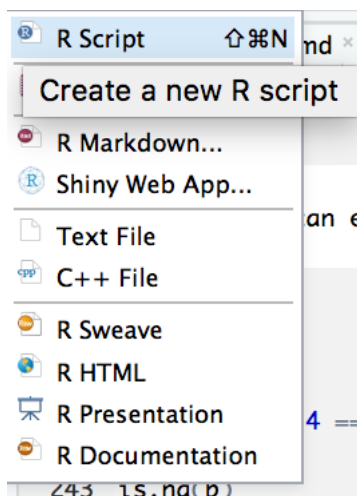


Figure B.5: Create a new R script and open the editor panel by selecting ‘R Script’ from the dropdown menu.

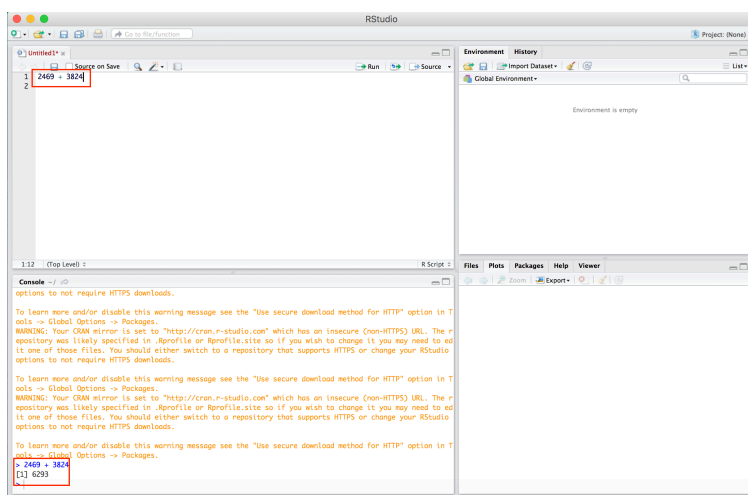


Figure B.6: An arithmetic expression is entered in the editor and evaluated in the console. The red boxes are added for emphasis.

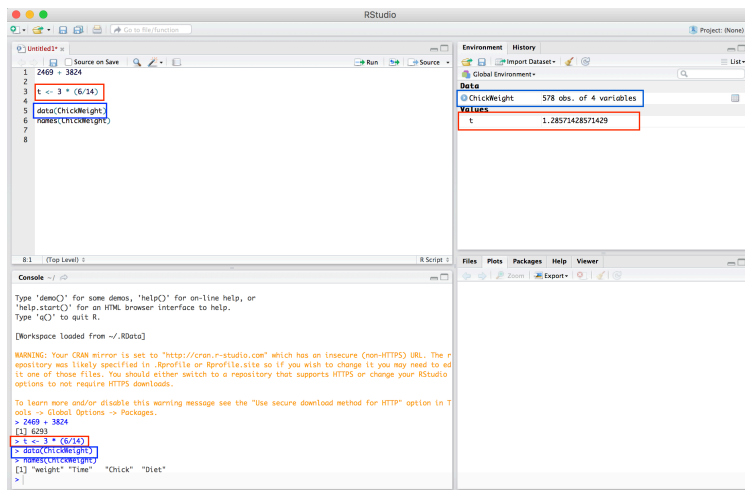


Figure B.7: The value  $3 * (6/14)$  is assigned to the variable `t` (red) and the dataset `ChickWeight` is added to the global environment (blue). The boxes are added for emphasis.

The Learning Days workshops use many tools in R to analyze and view data. For now, we can learn some basic tools to examine the data. The function `head()` allows us to see the first six rows of the dataset. `summary()` summarizes each of the columns of the dataset and `dim()` provides the dimensions of the dataset with first the number of rows and then columns.

```
head(ChickWeight) # First 6 observations in dataset
```

|   | weight | Time | Chick | Diet |
|---|--------|------|-------|------|
| 1 | 42     | 0    | 1     | 1    |
| 2 | 51     | 2    | 1     | 1    |
| 3 | 59     | 4    | 1     | 1    |
| 4 | 64     | 6    | 1     | 1    |
| 5 | 76     | 8    | 1     | 1    |
| 6 | 93     | 10   | 1     | 1    |

```
summary(ChickWeight) # Summary of all variables
```

|          | weight | Time         | Chick    | Diet |
|----------|--------|--------------|----------|------|
| Min.     | : 35   | Min. : 0.0   | 13       | : 12 |
| 1st Qu.: | 63     | 1st Qu.: 4.0 | 9        | : 12 |
| Median : | 103    | Median :10.0 | 20       | : 12 |
| Mean :   | 122    | Mean :10.7   | 10       | : 12 |
| 3rd Qu.: | 164    | 3rd Qu.:16.0 | 17       | : 12 |
| Max. :   | 373    | Max. :21.0   | 19       | : 12 |
|          |        |              | (Other): | 506  |

```
dim(ChickWeight) # Dimensions of the dataset in the order rows, columns
```

```
[1] 578 4
```

Unlike other statistical software, R allows users to store multiple datasets, possibly of different dimensions, simultaneously. This feature makes R quite flexible for analysis using multiple methods.

### B.3.4 Miscellaneous

R provides a suite of tools, ranging from built-in plot functions to packages to graph data, models, estimates, etc. The final Miscellaneous panel allows for the quick viewing of graphs in RStudio. Figure B.8 shows a plot in this panel. Learning Days will discuss how to plot data; for now, don't worry about the graphing the code in the Editor.

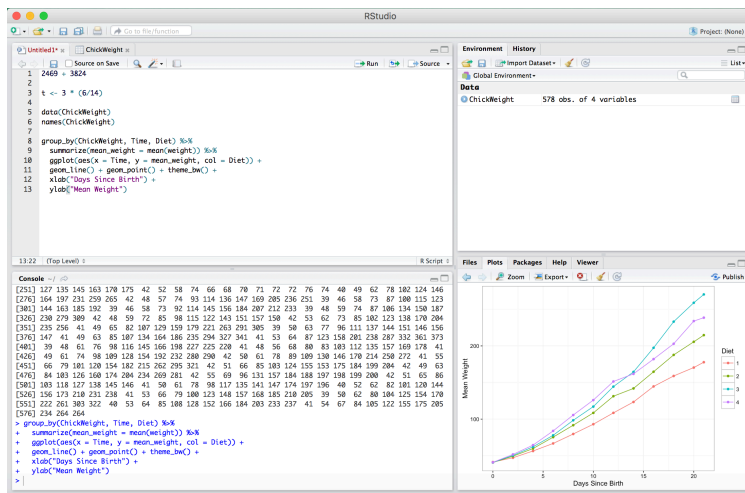


Figure B.8: An example plot of the ‘ChickWeight’ data made in R.

## B.4 Learning to Use R

### B.4.1 Online Resources

There are many helpful online resources to help you start learning R. We recommend two sources:

- Code School, which runs entirely through your browser <https://www.codeschool.com/courses/try-r>.
- Coursera, via an online R Programming course organized by Johns Hopkins University:
  - i. Go to <https://www.coursera.org>
  - ii. Create an account (this is free!)
  - iii. Sign up for R Programming at Johns Hopkins University (instructor: Roger Peng) under the “Courses” tab
  - iv. Read the materials and watch the videos from the first week. The videos from the first week are about 2.5 hours long total.

---

## B.4.2 Basic Practice

Here we provide some fragments of code to familiarize you with some basic practices in R. We recommend that you practice by typing the code fragments into your Editor and then evaluating them.

### B.4.2.1 Setting up an R Session

In general, we read other files such as data or functions into R and output results like graphs or tables into files not contained within an R session. To do this, we must give R an “address” at which it can locate such files. It may be most efficient to do this by setting a working directory, a file path at which relevant files are stored. We can identify the current working directory using `getwd()` and set a new one using `setwd()`. Note that the syntax of these filepaths varies by operating system.

```
getwd()
```

```
setwd("~/TaraLyn/EGAP Learning Days Admin/Workshop 2018_2 (Uruguay)/")
```

You may need to install packages beyond those listed above to execute certain functions. To install packages we use `install.packages("")`, filling in the package name between the "" marks, as follows. You need only install packages once.

```
install.packages("randomizr")
```

Once a package is installed, it can be loaded and accessed using `library()` where the package name is inserted between the parentheses (no "" marks).

```
library(randomizr)
```

To clear R’s memory of the stored data, functions, or values that appear in the accounting tab, use `rm(list = ls())`. It may be useful to set a random number seed to ensure that replication is possible in a different R session, particularly when we work with simulation-based methods.

```
rm(list = ls())
```

```
set.seed(2018) # Optional: Set a seed to make output replicable
```

### B.4.2.2 R Basics

We now explore some basic commands. In order to assign a scalar (single element) to a variable, we use the `<-` command as discussed previously:

```
# "<-" is the assignment command; it is used to define things. eg:  
(a <- 5)
```

```
[1] 5
```

We may also want to assign a vector of elements to a variable. Here we use the same `<-` command, but focus on how to create the vector.

---

```
(b <- 1:10)           # ":" is used to define a string of integers
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

```
(v <- c(1, 3, 2, 4, pi)) # use c() to make a vector with anything in it
```

```
[1] 1.000 3.000 2.000 4.000 3.142
```

We can then refer to elements of a vector by denoting their position in a vector inside hard brackets [].

```
# Extract elements of a vector:
```

```
b[1]           # Returns position 1
```

```
[1] 1
```

```
b[5:4]         # Returns positions 5 and 4, in that order
```

```
[1] 5 4
```

```
b[-1]          # Returns all but the first number
```

```
[1] 2 3 4 5 6 7 8 9 10
```

```
# Returns all numbers indicated as "TRUE"
```

```
b[c(TRUE, FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE, FALSE, FALSE)]
```

```
[1] 1 3 6 7
```

```
# Assign new values to particular elements of a vector
```

```
b[5] <- 0
```

There are a set of built-in functions that can be applied to vectors like `b`.

```
sum(b)         # Sum of all elements
```

```
[1] 50
```

```
mean(b)        # Mean of all elements
```

```
[1] 5
```

```
max(b)         # Maximum of all elements
```

```
[1] 10
```

```
min(b)         # Minimum of all elements
```

```
[1] 0
```

```
sd(b)          # Standard deviation of all elements
```

```
[1] 3.496
```

```
var(b)          # Variance of all elements
```

```
[1] 12.22
```

We can also apply arithmetic transformations to all elements of a vector:

```
b^2             # Square the variable
```

```
[1]  1  4  9 16  0 36 49 64 81 100
```

```
b^.5           # Square root of the variable
```

```
[1] 1.000 1.414 1.732 2.000 0.000 2.449 2.646 2.828 3.000 3.162
```

```
log(b)         # Log of variable
```

```
[1] 0.0000 0.6931 1.0986 1.3863 -Inf 1.7918 1.9459 2.0794 2.1972 2.3026
```

```
exp(b)         # e to the b
```

```
[1]  2.718  7.389 20.086 54.598  1.000 403.429 1096.633 2980.958 81
```

Finally, we can evaluate logical statements (i.e. “is condition X true?”) on all elements of a vector:

```
b == 2         # Is equal to
```

```
[1] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
b < 5         # Less than
```

```
[1] TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE
```

```
b >= 5        # Greater than or equal to
```

```
[1] FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE
```

```
b <= 5 | b / 4 == 2 # | means OR
```

```
[1] TRUE TRUE TRUE TRUE TRUE FALSE FALSE TRUE FALSE FALSE
```

```
b>2 & b<9     # & means AND
```

```
[1] FALSE FALSE TRUE TRUE FALSE TRUE TRUE TRUE FALSE FALSE
```

```
is.na(b)      # Indicates if data is missing
```

```
[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
which(b<5)    # Gives indices of values meeting logical requirement
```

```
[1] 1 2 3 4 5
```

The basic logic of these commands applies to data structures much more complex than scalars and vectors. Understanding of these basic features will help facilitate your understanding of



---

more advanced topics during Learning Days.



# Bibliography

- [1] Robert Adcock and David Collier. “Measurement Validity: A shared Standard for Qualitative and Measurement Validity: A shared Standard for Qualitative and Quantitative Research.” In: *American Political Science Review* 95.3 (2001), pp. 529–546.
- [2] Edward Asiedu et al. *A Call for Structured Ethics Appendices in Social Science Papers*. Working Paper 28393. National Bureau of Economic Research, Jan. 2021. DOI: [10.3386/w28393](https://doi.org/10.3386/w28393). URL: <http://www.nber.org/papers/w28393>.
- [3] Jake Bowers and Thomas Leavitt. “Causality & Design-Based Inference”. In: *The SAGE Handbook of Research Methods in Political Science and International Relations*. Ed. by Luigi Curini and Robert Franzese. Sage Publications Ltd, 2020.
- [4] Henry E Brady. “Causation and explanation in social science”. In: *The Oxford Handbook of Political Science*. 2008. URL: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-10>.
- [5] Garret S. Christensen, Jeremy Freese, and Edward Miguel. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Oakland, California: University of California Press, 2019.
- [6] David K. Evans. *Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science*. Working Paper 565. Center for Global Development, Feb. 2021. URL: <https://www.cgdev.org/sites/default/files/WP565-Evans-Ethical-issues-and-RCTs.pdf>.
- [7] Ronald A. Fisher. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935.
- [8] Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton & Company, 2012.
- [9] Rachel Glennerster and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press, 2013.
- [10] Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2007.
- [11] Paul R. Rosenbaum. “Design of observational studies”. In: *Springer series in statistics* (2010).
- [12] Paul R. Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, 2017.
- [13] Donald B. Rubin. “Estimating the Causal Effects of Treatments in Randomized and Nonrandomized Studies”. In: *J. Educ. Psych.* 66 (1974), pp. 688–701.

- 
- [14] Alexandra Scacco and Shana S. Warren. “Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria”. In: *American Political Science Review* 112.3 (2018), pp. 654–677.
- [15] William R. Shadish, Thomas D. Cook, Donald Thomas Campbell, et al. *Experimental and quasi-experimental designs for generalized causal inference*/William R. Shadish, Thomas D. Cook, Donald T. Campbell. Boston: Houghton Mifflin, 2002.
- [16] Pedro C. Vicente. “Is Vote Buying Effective? Evidence from a Field Experiment in West Africa”. In: *Economic Journal* 124.574 (2014), F356–87.