POLI110: The Basic Theory of Experiments

Zed Hoffman-Weldon

March 2025

Abstract

These notes are an introduction to the theory of experiments in social science. A well-run experiment can guarantee that any effect you find is an unbiased estimate of the true causal effect, and not a result of spurious correlation or bias. These notes explain why. I introduce a basic model of causality, explain the conditions under which experiments lead to unbiased estimates of causal effects, and discuss common violations of those conditions. Finally, I provide some suggestions for further reading.

1 The Potential Outcomes Model

To understand why experiments are so good at getting at causal effects, we need a formal way of thinking about causation. The way I will introduce is called the potential outcomes model, which was figured out by Jerzy Neyman and Donald Rubin.[1] It is not the only way of thinking about causality, but it is useful for analyzing experiments.

Let's say that Y is the effect we care about, and we want to estimate the effect, if any, of X on Y. For example, X could be direct cash transfers, and Y could be long-run income. Does receiving a direct cash transfer now improve one's income in the long run?

To answer this question, we have to think about two possible worlds. The first possible world is the world where someone receives a direct cash transfer. In the other possible world, they don't receive that cash transfer. Suppose that the two possible worlds are identical, except for that change and its potential consequences.[2] X can take two values: In the first world, $X = 1$, because the cash transfer was received. In the second world, $X = 0$, because there was no cash transfer. Here's the crucial step. Let's say that your long-run income in World 1 is $Y(1)$, because it's your income in the case you get a cash transfer. In World 2, let's say that it's $Y(0)$, because you did not receive a cash transfer. Economists call these **potential outcomes.** There's some randomness in the possible worlds: It's possible you get a cash transfer that would have helped you, but you contract a rare and deadly disease that has negative effects on your income. But that event is unlikely, so what we really care about is the **average causal effect** of cash transfers on long-run income. Because the possible worlds are otherwise identical,[3] this average causal effect is just

$$\tau = \mathbb{E}\left(Y(1) - Y(0)\right). \qquad \text{(Average Causal Effect)}$$

1. Guido W. Imbens and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge, UK: Cambridge University Press, 2015), 23.

2. Actually, we only need them to be "sufficiently close." See David Lewis, "Causation," in *Causation*, ed. Ernest Sosa and Michael Tooley, Oxford Readings in Philosophy (Oxford, UK: Oxford University Press, 1993), 196.

3. I am tacitly making a standard assumption. It is called the "stable unit treatment value" assumption. Basically, it means that there cannot be spillover effects; the outcome in one possible world cannot affect any other possible world. This is reasonable when we are thinking of an average over possible worlds, but more dubious when we are thinking about natural or field experiments.

If you don't know what the $\mathbb{E}$ represents, think of it as an "averaging function" which outputs the average of its input, so $\mathbb{E}(X)$ is the average of X.[4] The greek letter τ is called "tau," pronounced like "cow" but with a "t," and it's the traditional symbol used for representing a caual effect.

Unfortunately, there's a problem: We only ever live in one possible world or another, and we can't observe both. If we measure $Y(1)$, we can't also measure $Y(0)$. Likewise, if we measure $Y(0)$, we can't also measure $Y(1)$. This is sometimes called the **fundamental problem of causal inference.** We can get around this by using experiments.

2 Experiments and Causality

The core idea is that experiments eliminate spurious correlation and bias by randomizing assignment to treatment and control. Remember that spurious correlation arises from a directed acyclic graph that looks like Figure 1.[5] If there is a causal arrow connecting X and Y, we have measurement bias instead. If X is randomly assigned, then by definition it cannot be caused by anything. So there can be no arrows leading into X, meaning we can't have spurious correlation or measurement bias.

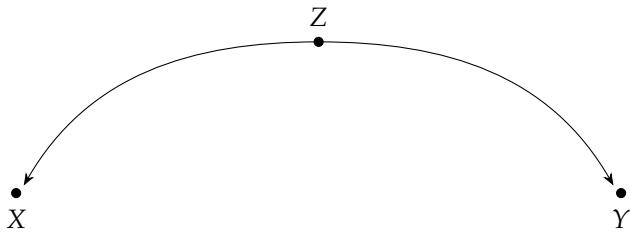

Figure 1: Spurious Correlation Depicted in a Directed Acylic Graph

Let's work through this in more detail. In the last section, we had only one person. The problem was that we couldn't observe both their treated outcome $Y(1)$; and their untreated outcome, $Y(0)$. If we have a lot of people, however, we can split them in half and observe $Y(1)$ for half and $Y(0)$ for the other half. Then, we can subtract the averages of the two groups to get the Average Causal Effect. Here's how this works. Suppose we have a population of people, who we number $1, 2, \ldots, n-1, n$. Suppose that i is the number of one such person. We'll write person i's potential outcomes $Y_i(1)$ and $Y_i(0)$. We want to know the Average Causal Effect $\mathbb{E}\tau_i$, but we can't measure both potential outcomes. Instead, we randomly split the group in half. Half of the people are randomly assigned to receive treatment: they get a cash transfer. Half of the people are randomly assigned not to receive treatment: they don't get a cash transfer. These groups are called the **treatment group** and **control group**, respectively. For every person in the treatment group, we observe $Y_i(1)$. For every person in the control group, we observe $Y_i(0)$. If we take the average of these groups, and subtract it, that will be an unbiased estimate of $\mathbb{E}(\tau_i)$.

To state this formally, let Z_i be a random variable which is 1 if person i receives treatment, and 0 if they do not. We need to make a distinction between Z and X, because one's potential outcomes can affect receipt of treatment. For example, suppose you work for a health insurance authority, and are interested in if early cancer screening reduces cancer mortality.

4. Formally, $\mathbb{E}$ is the "expectation operator." See e.g. Rick Durrett, *Probability: Theory and Examples*, https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf.

5. Judea Pearl, *Causality*, 2nd ed. (2000; Cambridge, UK: Cambridge University Press, 2009), 79.

You invite people to get screened for cancer: the treatment group is people who show up, the control group are people who don't. You work for the insurer, so you know what happens to people who show up and get screened, and those who don't. Surprisingly, you find that cancer screenings make people *more* likely to get cancer! What happened? The answer is that people are busy, and so those who show up will be unemployed, or especially worried about cancer, and so forth. They will have attributes—like a family history of cancer—which affect both receipt of treatment and their potential outcomes. This creates measurement bias! We express the dependence of $Y(X)$ on Z by writing $Y(X \mid Z)$. The symbol $\mid$ means "given," or "in the case that." You should read $Y_i(1 \mid Z = 1)$ as "person i's treated outcome, in the case that i actually received treatment."

Theorem. *In a random experiment, the difference in means estimator*

$$\hat{\tau} = \overline{Y(1 \mid Z = 1)} - \overline{Y(0 \mid Z = 0)}$$

is an unbiased estimator of the Average Causal Effect. The overbar denotes sample average.

Proof. An estimator is unbiased if its expectation equals the true value. I will prove that $\mathbb{E}\hat{\tau} = \tau$. Consider this expression directly:

$$\mathbb{E}\hat{\tau} = \mathbb{E}(\overline{Y(1 \mid Z = 1)} - \overline{Y(0 \mid Z = 0)}).$$

The average of a difference is the difference of the averages,[6] so we have

$$\mathbb{E}\hat{\tau} = \mathbb{E}(\overline{Y(1 \mid Z = 1)}) - \mathbb{E}(\overline{Y(0 \mid Z = 0)}).$$

Because assignment to treatment is random, we know that Y and X are conditionally independent on Z. Heuristically, this means that if we know whether or not a unit received treatment—its value of Z—we don't learn anything else about Y if someone told us X. See the mammogram example above for how this can fail. Mathematically, this implies

$$\mathbb{E}\overline{Y(1 \mid Z = 1)} = \mathbb{E}\overline{Y(1)}.$$

The same holds for $\overline{Y(0 \mid Z = 0)}$. Thus, we have

$$\mathbb{E}\hat{\tau} = \mathbb{E}\overline{Y(1)} - \mathbb{E}\overline{Y(0)}.$$

On average, the sample mean is the population average,[7] so the proof is complete. ■

Even better, the weak law of large numbers says approximately that as the sample size gets very large, the variance of the sample mean goes to zero.[8] So we can be sure that as we replicate the experiment or increase the sample size, it's likely that we will measure something very close to the true value. This is excellent news.

6. Formally, this is a special case of the linearity of expectation. See any probability text, like Durrett, *Probability*.

7. Again, I have assumed a stable unit treatment value here.

8. For a rigorous statement, see Adrianus Willem van der Vaart, *Asymptotic Statistics* (Cambridge, UK: Cambridge University Press, 1998), 15.

3 Natural and Field Experiments

There are many interesting questions, especially macrosocial questions, which cannot be measured in a laboratory. Famous examples include the effect of contact with outgroups on bigotry, and the causes of social revolutions. What's a girl to do?

Social scientists have found clever ways of exploiting "as-if" randomness in the world to approximate experimental conditions. Basically, we look for some feature in the real world which, is not directly causally related to the outcome, but assigns units to treatment or control. Then, within those treatment or control groups, we can estimate the causal effect as if it were a laboratory experiment.

Natural and field experiments are more complicated than laboratory experiments, because the way they assign treatment is imperfect. Rather than getting deep into the weeds here, I will discuss one common problem with natural experiments and defer the rest to further reading.

A major problem in field experiments which does not usually happen in the lab is "non-compliance."[9] Unlike in the lab, in a field experiment, not everyone assigned to treatment takes treatment. If you send canvassers to households with instructions to talk about trans rights, some people will not be home.[10] Worse, the people who were home probably differ systematically from those who were not. We cannot be sure if that difference is salient for the outcome variable, so we have reintroduced the possibility of measurement bias and spurious correlation.

There are two possible approaches to handling this. The first is focusing on the **average intent-to-treat effect**. An experimenter do not know if who was home is random, but they can be sure that the doors your canvassers knock on is random, because the experimenter did the random assignment. In that case, if one can measure outcomes, even for people who were assigned to treatment but didn't take it, one can estimate the effect of being assigned to treatment on the outcome. Be careful: This is not the same as the average causal effect of X on Y, because the treatment is no longer $X = 1$, it is *assignment to treatment* $(Z = 1)$. In the example, we are measuring the causal effect of having one's door knocked on by a canvasser, not the conversation itself.

The second approach is the **average treatment effect on the treated.** In this approach, we break people up into four categories. People who take treatment when assigned to treatment and don't when assigned to control are called **compliers.** People who do the opposite of what they're told are called **deniers.** Finally, there are **always-takers** and **never-takers.** If we assume there are no deniers, then the only people in the control group who take treatment are the always-takers, so we can identify the proportion of compliers. Using this knowledge, we can modify the average intent-to-treat effect to get at the average treatment effect on the treated. Be careful: This is not the same as the average causal effect, because we are only looking at the effect on people who took treatment. Furthermore, if we believe the proportion of deniers is significant, this method does not work.

A final caution about experiments: Not every interesting question can be answered using an experiment of every type. There is a tendency in social science to answer questions which work well with our methods, rather than answering interesting questions.[11] Experiments of any type are one tool for answering interesting question: be sure you have other tools in your box.

9. In natural experiments, the problem of non-compliance takes the form of instrumental variable designs.

10. This was the procedure in David Broockman and Joshua Kalla, "Durably reducing transphobia: A field experiment on door-to-door canvassing," *Science* 352, no. 6282 (2016): 220–224.

11. This was one of Joan Robinson's criticisms of American economics. I consider it correct. Joan Robinson, *The Accumulation of Capital*, 3rd ed., Palgrave Classics in Economics (1956; Houndmills, UK: Palgrave MacMillan, 2013), xxxi.

4 Further Reading

(a) For a gentle introduction to the basics of probability, see Hacking.[12] For a more rigorous overview, Durrett[13] is a classic, but I find it overly didactic. Bauer[14] is pure measure theory, but it suits my tastes better, and it has a sequel on probability theory.

(b) For more about the potential outcomes model, see Angrist and Pischke.[15] A book by one of the model's originators is Imbens and Rubin.[16] There is also a relatively clear but very basic chapter on this model in Wasserman.[17] You should also read Pearl,[18] especially Chapter 1.

(c) For more about natural experiments in the social sciences, see Dunning.[19] The book contains some notable errors, so be careful. Some specific techniques are also covered in Angrist and Pischke.[20] One particularly important kind of natural experiment in social science is the difference-in-difference estimator. For an overview of this, see Cattaneo, Idrobo, and Titiunik;[21] and for a discussion of how it is usually applied in political science see de la Cuesta and Imai.[22] If you find the idea of as-if randomization strange, you should know that actually as-if randomness is at work *even in laboratory experiments*. For a defence of this perhaps surprising idea, see Jaynes.[23] Theorists in natural science typically call this kind of randomness "phenomenological," in the same way that a physical theorist might be a phenomenologist.

(d) For field experiments the canonical text is Gerber and Green.[24]

(e) If you are uncomfortable—as I am—with the Lewis-Neyman-Rubin picture of counterfactual causation, take a gander at Anscombe,[25] or anything else in that edited volume. For a defence of the idea that we can, in fact, directly observe causal effects and don't need to appeal to the notion of potential outcomes at all, see Fales.[26] I largely agree with Fales, but still think the Neyman-Rubin model is a useful way of thinking about experiments.

12. Ian Hacking, *An Introduction to Probability and Inductive Logic* (2001; Cambridge, UK: Cambridge University Press, 2009).
13. Durrett, *Probability*.
14. Heinz Bauer, *Measure and Integration Theory*, trans. Robert Burckel (Berlin, DE: de Gruyter, 2001).
15. Joshua Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton, NJ: Princeton University Press, 2008).
16. Imbens and Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences*.
17. Larry Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics (New York, NY: Springer, 2004).
18. Pearl, *Causality*.
19. Thad Dunning, *Natural Experiments in the Social Sciences* (Cambridge, UK: Cambridge University Press, 2012).
20. Angrist and Pischke, *Mostly Harmless Econometrics*.
21. Matias D. Cattaneo, Nicolás Idrobo, and Rocío Titiunik, *A Practical Introduction to Regression Discontinuity Design: Foundations*, Cambridge Elements (Cambridge, UK: Cambridge University Press, 2019).
22. Brandon de la Cuesta and Kosuke Imai, "Misunderstandings about the regression discontinuity design in the study of close elections," *Annual Review of Political Science* 19 (2016): 375–396.
23. Edwin T. Jaynes, "The Physics of Random Experiments," chap. 10 in *Probability Theory: The Logic of Science*, ed. G. Larry Bretthorst (Cambridge, UK: Cambridge University Press, 2003), 314–339.
24. Alan S. Gerber and Donald P. Green, *Field Experiments: Design, Analysis, and Interpretation* (New York, NY: W. W. Norton, 2012).
25. Gertrude Elizabeth Margaret Anscombe, "Causality and Determination," in *Causation*, ed. Ernest Sosa and Michael Tooley, Oxford Readings in Philosophy (Oxford, UK: Oxford University Press, 1993).
26. Evan Fales, *Causality and Universals* (New York, NY: Routledge, 1990).

References

Angrist, Joshua, and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2008.

Anscombe, Gertrude Elizabeth Margaret. "Causality and Determination." In *Causation*, edited by Ernest Sosa and Michael Tooley. Oxford Readings in Philosophy. Oxford, UK: Oxford University Press, 1993.

Bauer, Heinz. *Measure and Integration Theory*. Translated by Robert Burckel. Berlin, DE: de Gruyter, 2001.

Broockman, David, and Joshua Kalla. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352, no. 6282 (2016): 220–224.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. *A Practical Introduction to Regression Discontinuity Design: Foundations*. Cambridge Elements. Cambridge, UK: Cambridge University Press, 2019.

de la Cuesta, Brandon, and Kosuke Imai. "Misunderstandings about the regression discontinuity design in the study of close elections." *Annual Review of Political Science* 19 (2016): 375–396.

Dunning, Thad. *Natural Experiments in the Social Sciences*. Cambridge, UK: Cambridge University Press, 2012.

Durrett, Rick. *Probability: Theory and Examples*. https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf.

Fales, Evan. *Causality and Universals*. New York, NY: Routledge, 1990.

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

Hacking, Ian. *An Introduction to Probability and Inductive Logic*. 2001. Cambridge, UK: Cambridge University Press, 2009.

Imbens, Guido W., and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge, UK: Cambridge University Press, 2015.

Jaynes, Edwin T. "The Physics of Random Experiments." Chap. 10 in *Probability Theory: The Logic of Science*, edited by G. Larry Bretthorst, 314–339. Cambridge, UK: Cambridge University Press, 2003.

Lewis, David. "Causation." In *Causation*, edited by Ernest Sosa and Michael Tooley. Oxford Readings in Philosophy. Oxford, UK: Oxford University Press, 1993.

Pearl, Judea. *Causality*. 2nd ed. 2000. Cambridge, UK: Cambridge University Press, 2009.

Robinson, Joan. *The Accumulation of Capital*. 3rd ed. Palgrave Classics in Economics. 1956. Houndmills, UK: Palgrave MacMillan, 2013.

van der Vaart, Adrianus Willem. *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press, 1998.

Wasserman, Larry. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. New York, NY: Springer, 2004.