

## Short Contents

---

	<i>Preface</i>	xvii
CHAPTER 1	Thinking Clearly in a Data-Driven Age	1
<b>PART I</b>	<b>ESTABLISHING A COMMON LANGUAGE</b>	11
CHAPTER 2	Correlation: What Is It and What Is It Good For?	13
CHAPTER 3	Causation: What Is It and What Is It Good For?	37
<b>PART II</b>	<b>DOES A RELATIONSHIP EXIST?</b>	53
CHAPTER 4	Correlation Requires Variation	55
CHAPTER 5	Regression for Describing and Forecasting	74
CHAPTER 6	Samples, Uncertainty, and Statistical Inference	94
CHAPTER 7	Over-Comparing, Under-Reporting	113
CHAPTER 8	Reversion to the Mean	138
<b>PART III</b>	<b>IS THE RELATIONSHIP CAUSAL?</b>	157
CHAPTER 9	Why Correlation Doesn't Imply Causation	159
CHAPTER 10	Controlling for Confounders	193
CHAPTER 11	Randomized Experiments	218
CHAPTER 12	Regression Discontinuity Designs	243

CHAPTER 13	Difference-in-Differences Designs	266
CHAPTER 14	Assessing Mechanisms	290
<b>PART IV</b>	<b>FROM INFORMATION TO DECISIONS</b>	303
CHAPTER 15	Turn Statistics into Substance	305
CHAPTER 16	Measure Your Mission	336
CHAPTER 17	On the Limits of Quantification	357
	<i>Index</i>	371

## Contents

---

	<i>Preface</i>	xvii
	Organization	xviii
	Who Is This Book For?	xix
	Acknowledgments	xx
CHAPTER 1	Thinking Clearly in a Data-Driven Age	1
	What You'll Learn	1
	Introduction	1
	Cautionary Tales	2
	Abe's hasty diagnosis	2
	Civil resistance	3
	Broken-windows policing	5
	Thinking and Data Are Complements, Not Substitutes	7
	Readings and References	9
<b>PART I</b>	<b>ESTABLISHING A COMMON LANGUAGE</b>	<b>11</b>
CHAPTER 2	Correlation: What Is It and What Is It Good For?	13
	What You'll Learn	13
	Introduction	13
	What Is a Correlation?	13
	Fact or correlation?	17
	What Is a Correlation Good For?	19
	Description	19
	Forecasting	20
	Causal inference	23
	Measuring Correlations	24
	Mean, variance, and standard deviation	24
	Covariance	27
	Correlation coefficient	28
	Slope of the regression line	29
	Populations and samples	29
	Straight Talk about Linearity	30
	Wrapping Up	33
	Key Terms	33

	Exercises	34
	Readings and References	36
<b>CHAPTER 3</b>	<b>Causation: What Is It and What Is It Good For?</b>	<b>37</b>
	What You'll Learn	37
	Introduction	37
	What Is Causation?	38
	Potential Outcomes and Counterfactuals	39
	What Is Causation Good For?	40
	The Fundamental Problem of Causal Inference	41
	Conceptual Issues	42
	What is the cause?	42
	Causality and counterexamples	44
	Causality and the law	47
	Can causality run backward in time?	47
	Does causality require a physical connection?	48
	Causation need not imply correlation	49
	Wrapping Up	49
	Key Terms	50
	Exercises	50
	Readings and References	52
<b>PART II</b>	<b>DOES A RELATIONSHIP EXIST?</b>	<b>53</b>
<b>CHAPTER 4</b>	<b>Correlation Requires Variation</b>	<b>55</b>
	What You'll Learn	55
	Introduction	55
	Selecting on the Dependent Variable	56
	The 10,000-hour rule	57
	Corrupting the youth	59
	High school dropouts	62
	Suicide attacks	63
	The World Is Organized to Make Us Select on the Dependent Variable	64
	Doctors mostly see sick people	65
	Post-mortems	65
	The <i>Challenger</i> disaster	67
	The financial crisis of 2008	69
	Life advice	69
	Wrapping Up	70
	Key Term	70
	Exercises	70
	Readings and References	72
<b>CHAPTER 5</b>	<b>Regression for Describing and Forecasting</b>	<b>74</b>
	What You'll Learn	74
	Introduction	74
	Regression Basics	74
	Linear Regression, Non-Linear Data	79

	The Problem of Overfitting	87
	Forecasting presidential elections	87
	How Regression Is Presented	89
	A Brief Intellectual History of Regression	89
	Wrapping Up	91
	Key Terms	91
	Exercises	92
	Readings and References	93
CHAPTER 6	Samples, Uncertainty, and Statistical Inference	94
	What You'll Learn	94
	Introduction	94
	Estimation	94
	Why Do Estimates Differ from Estimands?	96
	Bias	96
	Noise	97
	What Makes for a Good Estimator?	98
	Quantifying Precision	99
	Standard errors	99
	Small samples and extreme observations	101
	Confidence intervals	102
	Statistical Inference and Hypothesis Testing	103
	Hypothesis testing	103
	Statistical significance	104
	Statistical Inference about Relationships	105
	What If We Have Data for the Whole Population?	106
	Substantive versus Statistical Significance	107
	Social media and voting	107
	The Second Reform Act	108
	Wrapping Up	109
	Key Terms	109
	Exercises	110
	Readings and References	111
CHAPTER 7	Over-Comparing, Under-Reporting	113
	What You'll Learn	113
	Introduction	113
	Can an octopus be a soccer expert?	113
	Publication Bias	118
	<i>p</i> -hacking	119
	<i>p</i> -screening	120
	Are Most Scientific "Facts" False?	122
	ESP	122
	Get out the vote	123
	<i>p</i> -hacking forensics	124
	Potential Solutions	126
	Reduce the significance threshold	126
	Adjust <i>p</i> -values for multiple testing	127
	Don't obsess over statistical significance	127

	Pre-registration	127
	Requiring pre-registration in drug trials	128
	Replication	128
	Football and elections	129
	Test important and plausible hypotheses	130
	The power pose	131
	Beyond Science	131
	Superstars	132
	Wrapping Up	134
	Key Terms	134
	Exercises	134
	Readings and References	136
<b>CHAPTER 8</b>	<b>Reversion to the Mean</b>	<b>138</b>
	What You'll Learn	138
	Introduction	138
	Does the truth wear off?	138
	Francis Galton and Regression to Mediocrity	139
	Reversion to the Mean Is Not a Gravitational Force	142
	Seeking Help	145
	Does knee surgery work?	146
	Reversion to the Mean, the Placebo Effect, and Cosmic Habituation	147
	The placebo effect	147
	Cosmic habituation explained	148
	Cosmic habituation and genetics	150
	Beliefs Don't Revert to the Mean	150
	Wrapping Up	152
	Key Words	152
	Exercises	152
	Readings and References	155
<b>PART III</b>	<b>IS THE RELATIONSHIP CAUSAL?</b>	<b>157</b>
<b>CHAPTER 9</b>	<b>Why Correlation Doesn't Imply Causation</b>	<b>159</b>
	What You'll Learn	159
	Introduction	159
	Charter schools	160
	Thinking Clearly about Potential Outcomes	163
	Sources of Bias	168
	Confounders	168
	Reverse causality	169
	The 10,000-hour rule, revisited	170
	Diet soda	173
	How Different Are Confounders and Reverse Causality?	174
	Campaign spending	174
	Signing the Bias	176
	Contraception and HIV	179
	Mechanisms versus Confounders	181
	Thinking Clearly about Bias and Noise	183

	Wrapping Up	186
	Key Terms	187
	Exercises	187
	Readings and References	191
CHAPTER 10	Controlling for Confounders	193
	What You'll Learn	193
	Introduction	193
	Party whipping in Congress	193
	A note on heterogeneous treatment effects	197
	The Anatomy of a Regression	198
	How Does Regression Control?	201
	Controlling and Causation	209
	Is social media bad for you?	210
	Reading a Regression Table	211
	Controlling for Confounders versus Mechanisms	213
	There Is No Magic	214
	Wrapping Up	215
	Key Terms	215
	Exercises	216
	Readings and References	217
CHAPTER 11	Randomized Experiments	218
	What You'll Learn	218
	Introduction	218
	Breastfeeding	219
	Randomization and Causal Inference	221
	Estimation and Inference in Experiments	224
	Standard errors	224
	Hypothesis testing	225
	Problems That Can Arise with Experiments	225
	Noncompliance and instrumental variables	226
	Chance imbalance	232
	Lack of statistical power	234
	Attrition	235
	Interference	236
	Natural Experiments	237
	Military service and future earnings	238
	Wrapping Up	239
	Key Terms	239
	Exercises	240
	Readings and References	242
CHAPTER 12	Regression Discontinuity Designs	243
	What You'll Learn	243
	Introduction	243
	How to Implement an RD Design	247
	Are extremists or moderates more electable?	249
	Continuity at the Threshold	251
	Does continuity hold in election RD designs?	255

	Noncompliance and the Fuzzy RD	256
	Bombing in Vietnam	257
	Motivation and Success	261
	Wrapping Up	262
	Key Terms	262
	Exercises	262
	Readings and References	264
<b>CHAPTER 13</b>	<b>Difference-in-Differences Designs</b>	<b>266</b>
	What You'll Learn	266
	Introduction	266
	Parallel Trends	267
	Two Units and Two Periods	269
	Unemployment and the minimum wage	269
	$N$ Units and Two Periods	272
	Is watching TV bad for kids?	273
	$N$ Units and $N$ Periods	275
	Contraception and the gender-wage gap	276
	Useful Diagnostics	278
	Do newspaper endorsements affect voting decisions?	278
	Is obesity contagious?	279
	Difference-in-Differences as Gut Check	282
	The democratic peace	282
	Wrapping Up	285
	Key Terms	285
	Exercises	286
	Readings and References	288
<b>CHAPTER 14</b>	<b>Assessing Mechanisms</b>	<b>290</b>
	What You'll Learn	290
	Introduction	290
	Causal Mediation Analysis	291
	Intermediate Outcomes	292
	Cognitive behavioral therapy and at-risk youths in Liberia	293
	Independent Theoretical Predictions	294
	Do voters discriminate against women?	294
	Testing Mechanisms by Design	295
	Social pressure and voting	295
	Disentangling Mechanisms	296
	Commodity price shocks and violent conflict	296
	Wrapping Up	298
	Key Terms	299
	Exercises	299
	Readings and References	300
<b>PART IV</b>	<b>FROM INFORMATION TO DECISIONS</b>	<b>303</b>
<b>CHAPTER 15</b>	<b>Turn Statistics into Substance</b>	<b>305</b>
	What You'll Learn	305



Introduction	305	
What's the Right Scale?	305	
Miles-per-gallon versus gallons-per-mile	306	
Percent versus percentage point	309	
Visual Presentations of Data	309	
Policy preferences and the Southern realignment	311	
Some rules of thumb for data visualization	314	
From Statistics to Beliefs: Bayes' Rule	314	
Bayes' rule	317	
Information, beliefs, priors, and posteriors	318	
Abe's celiac revisited	319	
Finding terrorists in an airport	322	
Bayes' rule and quantitative analysis	325	
Expected Costs and Benefits	328	
Screening frequently or accurately	329	
Wrapping Up	331	
Key Words	331	
Exercises	332	
Readings and References	334	
CHAPTER 16	Measure Your Mission	336
	What You'll Learn	336
	Introduction	336
	Measuring the Wrong Outcome or Treatment	337
	Partial measures	337
	Metal detectors in airports	337
	Intermediate outcomes	339
	Blood pressure and heart attacks	340
	Ill-defined missions	341
	Climate change and economic productivity	342
	Do You Have the Right Sample?	343
	External validity	343
	Malnutrition in India and Bangladesh	343
	Selected samples	344
	College admissions	345
	Why can't major league pitchers hit?	345
	Strategic Adaptation and Changing Relationships	349
	The duty on lights and windows	349
	The shift in baseball	350
	The war on drugs	351
	Wrapping Up	353
	Key Words	353
	Exercises	353
	Readings and References	355
CHAPTER 17	On the Limits of Quantification	357
	What You'll Learn	357
	Introduction	357
	Decisions When Evidence Is Limited	358
	Cost-benefit analysis and environmental regulation	358

Floss your teeth and wear a mask	359
Floss your teeth	359
Wear a mask	360
Quantification and Values	361
How quantitative tools sneak in values	361
Algorithms and racial bias in health care	361
How quantification shapes our values	363
Think Clearly and Help Others Do So Too	367
Exercises	367
Readings and References	368
<i>Index</i>	371

## CHAPTER 1

# Thinking Clearly in a Data-Driven Age

---

### What You'll Learn

- Learning to think clearly and conceptually about quantitative information is important for lots of reasons, even if you have no interest in a career as a data analyst.
- Even well-trained people often make crucial errors with data.
- Thinking and data are complements, not substitutes.
- The skills you learn in this book will help you use evidence to make better decisions in your personal and professional life and be a more thoughtful and well-informed citizen.

### Introduction

We live in a data-driven age. According to former Google CEO Eric Schmidt, the contemporary world creates as much new data every two days as had been created from the beginning of time through the year 2003. All this information is supposed to have the power to improve our lives, but to harness this power we must learn to think clearly about our data-driven world. Clear thinking is hard—especially when mixed up with all the technical details that typically surround data and data analysis.

Thinking clearly in a data-driven age is, first and foremost, about staying focused on ideas and questions. Technicality, though important, should serve those ideas and questions. Unfortunately, the statistics and quantitative reasoning classes in which most people learn about data do exactly the opposite—that is, they focus on technical details. Students learn mathematical formulas, memorize the names of statistical procedures, and start crunching numbers without ever having been asked to think clearly and conceptually about what they are doing or why they are doing it. Such an approach can work for people to whom thinking mathematically comes naturally. But we believe it is counterproductive for the vast majority of us. When technicality pushes students to stop thinking and start memorizing, they miss the forest for the trees. And it's also no fun.

Our focus, by contrast, is on conceptual understanding. What features of the world are you comparing when you analyze data? What questions can different kinds of comparisons answer? Do you have the right question and comparison for the problem you are trying to solve? Why might an answer that sounds convincing actually

be misleading? How might you use creative approaches to provide a more informative answer?

It isn't that we don't think the technical details are important. Rather, we believe that technique without conceptual understanding or clear thinking is a recipe for disaster. In our view, once you can think clearly about quantitative analysis, and once you understand why asking careful and precise questions is so important, technique will follow naturally. Moreover, this way is more fun.

In this spirit, we've written this book to require no prior exposure to data analysis, statistics, or quantitative methods. Because we believe conceptual thinking is more important, we've minimized (though certainly not eliminated) technical material in favor of plain-English explanations wherever possible. Our hope is that this book will be used as an introduction and a guide to how to think about and do quantitative analysis. We believe anyone can become a sophisticated consumer (and even producer) of quantitative information. It just takes some patience, perseverance, hard work, and a firm resolve to never allow technicality to be a substitute for clear thinking.

Most people don't become professional quantitative analysts. But whether you do or do not, we are confident you will use the skills you learn in this book in a variety of ways. Many of you will have quantitative analysts working for or with you. And all of you will read studies, news reports, and briefings in which someone tries to convince you of a conclusion using quantitative analyses. This book will equip you with the clear thinking skills necessary to ask the right questions, be skeptical when appropriate, and distinguish between useful and misleading evidence.

## Cautionary Tales

To whet your appetite for the hard work ahead, let's start with a few cautionary tales that highlight the importance of thinking clearly in a data-driven age.

### Abe's Hasty Diagnosis

Ethan's first child, Abe, was born in July 2006. As a baby, he screamed and cried almost non-stop at night for five months. Abe was otherwise happy and healthy, though a bit on the small side. When he was one year old the family moved to Chicago, without which move, you'd not be reading this book. (That last sentence contains a special kind of claim called a *counterfactual*. Counterfactuals are really important, and you are going to learn all about them in chapter 3.) After noticing that Abe was small for his age and growing more slowly than expected, his pediatrician decided to run some tests.

After some lab work, the doctors were pretty sure Abe had celiac disease—a digestive disease characterized by gluten intolerance. The good news: celiac disease is not life threatening or even terribly serious if properly managed through diet. The bad news: in 2007, the gluten-free dietary options for kids were pretty miserable.

It turns out that Abe actually had two celiac-related blood tests. One came back positive (indicating that he had the disease), the other negative (indicating that he did not have the disease). According to the doctors, the positive test was over 80 percent accurate. "This is a strong diagnosis," they said. The suggested course of action was to put Abe on a gluten-free diet for a couple of months to see if his weight increased. If it did, they could either do a more definitive biopsy or simply keep Abe gluten-free for the rest of his life.

Ethan asked for a look at the report on Abe's bloodwork. The doctors indicated they didn't think that would be useful since Ethan isn't a doctor. This response was neither

surprising nor hard to understand. People, especially experts and authority figures, often don't like acknowledging the limits of their knowledge. But Ethan wanted to make the right decision for his son, so he pushed hard for the information. One of the goals of this book is to give you some of the skills and confidence to be your own advocate in this way when using information to make decisions in your life.

Two numbers characterize the effectiveness of any diagnostic test. The first is its false negative rate, which is how frequently the test says a sick person is healthy. The second is its false positive rate, which is how frequently the test says a healthy person is sick. You need to know *both* the false positive rate and the false negative rate to interpret a diagnostic test's results. So Abe's doctors' statement that the positive blood test was 80 percent accurate wasn't very informative. Did that mean it had a 20 percent false negative rate? A 20 percent false positive rate? Do 80 percent of people who test positive have celiac disease?

Fortunately, a quick Google search turned up both the false positive and false negative rates for both of Abe's tests. Here's what Ethan learned. The test on which Abe came up positive for celiac disease has a false negative rate of about 20 percent. That is, if 100 people with celiac disease took the test, about 80 of them would correctly test positive and the other 20 would incorrectly test negative. This fact, we assume, is where the claim of 80 percent accuracy came from. The test, however, has a false positive rate of 50 percent! People who don't have celiac disease are just as likely to test positive as they are to test negative. (This test, it is worth noting, is no longer recommended for diagnosing celiac disease.) In contrast, the test on which Abe came up negative for celiac disease had much lower false negative and false positive rates.

Before getting the test results, a reasonable estimate of the probability of Abe having celiac disease, given his small size, was around 1 in 100. That is, about 1 out of every 100 small kids has celiac disease. Armed with the lab reports and the false positive and false negative rates, Ethan was able to calculate how likely Abe was to have celiac disease given his small size and the test results. Amazingly, the combination of testing positive on an inaccurate test and testing negative on an accurate test actually meant that the evidence suggested that Abe was much *less* likely than 1 in 100 to have celiac disease. In fact, as we will show you in chapter 15, the best estimate of the likelihood of Abe having celiac, given the test results, was about 1 in 1,000. The blood tests that Abe's doctors were sure supported the celiac diagnosis actually strongly supported the opposite conclusion. Abe was almost certain not to have celiac disease.

Ethan called the doctors to explain what he'd learned and to suggest that moving his pasta-obsessed son to a gluten-free diet, perhaps for life, was not the prudent next step. Their response: "A diagnosis like this can be hard to hear." Ethan found a new pediatrician.

Here's the upshot. Abe did not have celiac disease. The kid was just a bit small. Today he is a normal-sized kid with a ravenous appetite. But if his father didn't know how to think about quantitative evidence or lacked the confidence to challenge a mistaken expert, he'd have spent his childhood eating rice cakes. Rice cakes are gross, so he might still be small.

## Civil Resistance

As many around the world have experienced, citizens often find themselves in deep disagreement with their government. When things get bad enough, they sometimes decide to organize protests. If you ever find yourself doing such organizing, you will face many important decisions. Perhaps none is more important than whether to build

a movement with a non-violent strategy or one open to a strategy involving more violent forms of confrontation. In thinking through this quandry, you will surely want to consult your personal ethics. But you might also want to know what the evidence says about the costs and benefits of each approach. Which kind of organization is most likely to succeed in changing government behavior? Is one or the other approach more likely to land you in prison, the hospital, or the morgue?

There is some quantitative evidence that you might use to inform your decisions. First, comparing anti-government movements across the globe and over time, governments more often make concessions to fully non-violent groups than to groups that use violence. And even comparing across groups that do use violence, governments more frequently make concessions to those groups that engage in violence against military and government targets rather than against civilians. Second, the personal risks associated with violent protest are greater than those associated with non-violent protest. Governments repress violent uprisings more often than they do non-violent protests, making concerns about prison, the hospital, and the morgue more acute.

This evidence sounds quite convincing. A non-violent strategy seems the obvious choice. It is apparently both more effective and less risky. And, indeed, on the basis of this kind of data, political scientists Erica Chenoweth and Evan Perkoski conclude that “planning, training, and preparation to maintain nonviolent discipline is key—especially (and paradoxically) when confronting brutal regimes.”

But let’s reconsider the evidence. Start by asking yourself, In what kind of a setting is a group likely to engage in non-violent rather than violent protest? A few thoughts occur to us. Perhaps people are more likely to engage in non-violent protest when they face a government that they think is particularly likely to heed the demands of its citizens. Or perhaps people are more likely to engage in non-violent protest when they have broad-based support among their fellow citizens, represent a group in society that can attract media attention, or face a less brutal government.

If any of these things are true, we should worry about the claim that maintaining non-violent discipline is key to building a successful anti-government movement. (Which isn’t to say that we are advocating violence.) Let’s see why.

Empirical studies find that, on average, governments more frequently make concessions in places that had non-violent, rather than violent, protests. The claimed implication rests on a particular interpretation of that difference—namely, that the higher frequency of government concessions in non-violent places is *caused* by the use of non-violent tactics. Put differently, all else held equal, if a given movement using violent methods had switched to using non-violent methods, the government would have been more likely to grant concessions. But is this causal interpretation really justified by the evidence?

Suppose it’s the case that protest movements are more likely to turn to violence when they do not have broad-based support among their fellow citizens. Then, when we compare places that had violent protests to places that had non-violent protests, all else (other than protest tactics) is not held equal. Those places differ in at least two ways. First, they differ in terms of whether they had violent or non-violent protests. Second, they differ in terms of how supportive the public was of the protest movement.

This second difference is a problem for the causal interpretation. You might imagine that public opinion has an independent effect on the government’s willingness to grant concessions. That is, all else held equal (including protest tactics), governments might be more willing to grant concessions to protest movements with broad-based public support. If this is the case, then we can’t really know whether the fact that governments

grant concessions more often to non-violent protest movements than to violent protest movements is because of the difference in protest tactics or because the non-violent movements also happen to be the movements with broad-based public support. This is the classic problem of mistaking correlation for causation.

It is worth noting a few things. First, if government concessions are in fact due to public opinion, then it could be the case that, were we actually able to hold all else equal in our comparison of violent and non-violent protests, we would find the opposite relationship—that is, that non-violence is not more effective than violence (it could even be less effective). Given this kind of evidence, we just can't know.

Second, in this example, the conclusion that appears to follow if you don't force yourself to think clearly is one we would all like to be true. Who among us would not like to live in a world where non-violence is always preferred to violence? But the whole point of using evidence to help us make decisions is to force us to confront the possibility that the world may not be as we believe or hope it is. Indeed, it is in precisely those situations where the evidence seems to say exactly what you would like it to say that it is particularly important to force yourself to think clearly.

Third, we've pointed to one challenge in assessing the effects of peaceful versus violent protest, but there are others. For instance, think about the other empirical claim we discussed: that violent protests are more likely to provoke the government into repressive crack-downs than are non-violent protests. Recall, we suggested that people might be more likely to engage in non-violent protest when they are less angry at their government, perhaps because the government is less brutal. Ask yourself why, if this is true, we have a similar problem of interpretation. Why might the fact that there are more government crack-downs following violent protests than non-violent protests *not* mean that switching from violence to non-violence will reduce the risk of crack-downs? The argument follows a similar logic to the one we just made regarding concessions. If you don't see how the argument works yet, that's okay. You will by the end of chapter 9.

## Broken-Windows Policing

In 1982, the criminologist George L. Kelling and the sociologist James Q. Wilson published an article in *The Atlantic* proposing a new theory of crime and policing that had enormous and long-lasting effects on crime policy in the United States and beyond.

Kelling and Wilson's theory is called *broken windows*. It was inspired by a program in Newark, New Jersey, that got police out of their cars and walking a beat. According to Kelling and Wilson, the program reduced crime by elevating "the level of public order." Public order is important, they argue, because its absence sets in motion a vicious cycle:

A piece of property is abandoned, weeds grow up, a window is smashed. Adults stop scolding rowdy children... Families move out, unattached adults move in. Teenagers gather in front of the corner store. The merchant asks them to move; they refuse. Fights occur. Litter accumulates. People start drinking in front of the grocery...

Residents will think that crime, especially violent crime, is on the rise... They will use the streets less often... Such an area is vulnerable to criminal invasion.

This idea that policing focused on minimizing disorder can reduce violent crime had a big impact on police tactics. Most prominently, the broken-windows theory was the

guiding philosophy in New York City in the 1990s. In a 1998 speech, then New York mayor Rudy Giuliani said,

We have made the “Broken Windows” theory an integral part of our law enforcement strategy...

You concentrate on the little things, and send the clear message that this City cares about maintaining a sense of law and order... then the City as a whole will begin to become safer.

And, indeed, crime in New York city did decline when the police started focusing “on the little things.” According to a study by Hope Corman and H. Naci Mocan, misdemeanor arrests increased 70 percent during the 1990s and violent crime decreased by more than 56 percent, double the national average.

To assess the extent to which broken-windows policing was responsible for this fall in crime, Kelling and William Sousa studied the relationship between violent crime and broken-windows approaches across New York City’s precincts. If minimizing disorder causes a reduction in violent crime, they argued, then we should expect the largest reductions in crime to have occurred in neighborhoods where the police were most focused on the broken-windows approach. And this is just what they found. In precincts where misdemeanor arrests (the “little things”) were higher, violent crime decreased more. They calculated that “the average NYPD precinct... could expect to suffer one less violent crime for approximately every 28 additional misdemeanor arrests.”

This sounds pretty convincing. But let’s not be too quick to conclude that arresting people for misdemeanors is the answer to ending violent crime. Two other scholars, Bernard Harcourt and Jens Ludwig, encourage us to think a little more clearly about what might be going on in the data.

The issue that Harcourt and Ludwig point out is something called *reversion to the mean* (which we’ll talk about a lot more in chapter 8). Here’s the basic concern. In any given year, the amount of crime in a precinct is determined by lots of factors, including policing, drugs, the economy, the weather, and so on. Many of those factors are unknown to us. Some of them are fleeting; they come and go across precincts from year to year. As such, in any given precinct, we can think of there being some “baseline” level of crime, with some years randomly having more crime and some years randomly having less (relative to that precinct-specific baseline).

In any given year, if a precinct had a high level of crime (relative to its baseline), then it had bad luck on the unknown and fleeting factors that help cause crime. Probably next year its luck won’t be as bad (that’s what *fleeting* means), so that precinct will likely have less crime. And if a precinct had a low level of crime (relative to its baseline) this year, then it had good luck on the unknown and fleeting factors, and it will probably have worse luck next year (crime will go back up). Thus, year to year, the crime level in a precinct tends to revert toward the *mean* (i.e., the precinct’s baseline level of crime).

Now, imagine a precinct that had a really high level of violent crime in the late 1980s. Two things are likely to be true of that precinct. First, it is probably a precinct with a high baseline of violent crime. Second, it is also probably a precinct that had a bad year or two—that is, for idiosyncratic and fleeting reasons, the level of crime in the late 1980s was high relative to that precinct’s baseline. The same, of course, is true in reverse for precincts that had a low level of crime in the late 1980s. They probably have a low baseline of crime, and they also probably had a particularly good couple of years.



Why is this a problem for Kelling and Sousa's conclusions? Because of reversion to the mean, we would expect the most violent precincts in the late 1980s to show a reduction in violent crime on average, even with no change in policing. And unsurprisingly, given the police's objectives, but unfortunately for the study, it was precisely those high-crime precincts in the 1980s that were most likely to get broken-windows policing in the early 1990s. So, when we see a reduction in violent crime in the precincts that had the most broken-windows policing, we don't know if it's the policing strategy or reversion to the mean that's at work.

Harcourt and Ludwig go a step further to try to find more compelling evidence. Roughly speaking, they look at how changes in misdemeanor arrests relate to changes in violent crime in precincts that had similar levels of violent crime in the late 1980s. By comparing precincts with similar starting levels of violent crime, they go some way toward eliminating the problem of reversion to the mean. Surprisingly, this simple change actually flips the relationship! Rather than confirming Kelling and Sousa's finding that misdemeanor arrests are associated with a reduction in violent crime, Harcourt and Ludwig find that precincts that focused more on misdemeanor arrests actually appear to have experienced an *increase* in violent crime. Exactly the opposite of what we would expect if the broken-windows theory is correct.

Now, this reversal doesn't settle the matter on the efficacy of broken-windows policing. The relationship between misdemeanor arrests and violent crime that Harcourt and Ludwig find could be there for lots of reasons other than misdemeanor arrests causing an increase in violent crime. For instance, perhaps the neighborhoods with increasing misdemeanors are becoming less safe in general and would have experienced more violent crime regardless of policing strategies. What these results do show is that the data, properly considered, certainly don't offer the kind of unequivocal confirmation of the broken-windows ideas that you might have thought from Kelling and Sousa's finding. And you can only see this if you have the ability to think clearly about some subtle issues.

This flawed thinking was important. Evidence-based arguments like Kelling and Sousa's played a role in convincing politicians and policy makers that broken-windows policing was the right path forward when, in fact, it might have diverted resources away from preventing and investigating violent crime and may have created a more adversarial and unjust relationship between the police and the disproportionately poor and minority populations who were frequently cited for "the small stuff"

## Thinking and Data Are Complements, Not Substitutes

Our quantitative world is full of lots of exciting new data and analytic tools to analyze that data with fancy names like machine learning algorithms, artificial intelligence, random forests, and neural networks. Increasingly, we are even told that this new technology will make it possible for the machines to do the thinking for us. But that isn't right. As our cautionary tales highlight, no data analysis, no matter how futuristic its name, will work if we aren't asking the right questions, if we aren't making the right comparisons, if the underlying assumptions aren't sound, or if the data used aren't appropriate. Just because an argument contains seemingly sophisticated quantitative data analysis, that doesn't mean the argument is rigorous or right. To harness the power of data to make better decisions, we must combine quantitative analysis with clear thinking.

Our stories also illustrate how our intuitions can lead us astray. It takes lots of care and practice to train ourselves to think clearly about evidence. The doctors'

intuition that Abe had celiac disease because of a test with 80 percent accuracy and the researchers' intuition that broken-windows policing works because crime decreased in places where it was deployed seem sensible. But both intuitions were wrong, suggesting that we should be skeptical of our initial hunches. The good news is that clear thinking can become intuitive if you work at it.

Data and quantitative tools are not a substitute for clear thinking. In fact, quantitative skills without clear thinking are quite dangerous. We suspect, as you read the coming chapters, you will be jarred by the extent to which unclear thinking affects even the most important decisions people make. Through the course of this book, we will see how misinterpreted information distorts life-and-death medical choices, national and international counterterrorism policies, business and philanthropic decisions made by some of the world's wealthiest people, how we set priorities for our children's education, and a host of other issues from the banal to the profound. Essentially, no aspect of life is immune from critical mistakes in understanding and interpreting quantitative information.

In our experience, this is because unclear thinking about evidence is deeply ingrained in human psychology. Certainly our own intuitions, left unchecked, are frequently subject to basic errors. Our guess is that yours are too. Most disturbingly, the experts on whose advice you depend—be they doctors, business consultants, journalists, teachers, financial advisors, scientists, or what have you—are often just as prone to making such errors as the rest of us. All too often, because they are experts, we trust their judgment without question, and so do they. That is why it is so important to learn to think clearly about quantitative evidence for yourself. That is the only way to know how to ask the right questions that lead you, and those on whose advice you depend, to the most reliable and productive conclusions possible.

How could experts in so many fields make important errors so often? Expertise, in any area, comes from training, practice, and experience. No one expects to become an expert in engineering, finance, plumbing, or medicine without instruction and years of work. But, despite its fundamental and increasing importance for so much of life in our quantitative age, almost no one invests this kind of effort into learning to think clearly with data. And, as we've said, even when they do, they tend to be taught in a way that over-emphasizes the technical and under-emphasizes the conceptual, even though the fundamental problems are almost always about conceptual mistakes in thinking rather than technical mistakes in calculation.

The lack of expertise in thinking presents us with two challenges. First, if so much expert advice and analysis is unreliable, how do you know what to believe? Second, how can you identify those expert opinions that do in fact reflect clear thinking?

This book provides a framework for addressing these challenges. Each of the coming chapters explains and illustrates, through a variety of examples, fundamental principles of clear thinking in a data-driven world. Part 1 establishes some shared language—clarifying what we mean by correlation and causation and what each is useful for. Part 2 discusses how we can tell whether a statistical relationship is genuine. Part 3 discusses how we can tell if that relationship reflects a causal phenomenon or not. And part 4 discusses how we should and shouldn't incorporate quantitative information into our decision making.

Our hope is that reading this book will help you internalize the principles of clear thinking in a deep enough way that they start to become second nature. You will know you are on the right path when you find yourself noticing basic mistakes in how people think and talk about the meaning of evidence everywhere you turn—as you watch

the news, peruse magazines, talk to business associates, visit the doctor, listen to the color commentary during athletic competitions, read scientific studies, or participate in school, church, or other communal activities. You will, we suspect, find it difficult to believe how much nonsense you're regularly told by all kinds of experts. When this starts to happen, try to remain humble and constructive in your criticisms. But do feel free to share your copy of this book with those whose arguments you find are in particular need of it. Or better yet, encourage them to buy their own copy!

## Readings and References

The essay on non-violent protest by Erica Chenoweth and Evan Perkoski that we quote can be found at <https://politicalviolenceataglance.org/2018/05/08/states-are-far-less-likely-to-engage-in-mass-violence-against-nonviolent-uprisings-than-violent-uprisings/>.

The following book contains more research on the relationship between non-violence and efficacy:

Erica Chenoweth and Maria J. Stephan. 2011. *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict*. Columbia University Press.

The following articles were discussed in this order on the topic of broken windows policing:

George L. Kelling and James Q. Wilson. 1982. "Broken Windows: The Police and Neighborhood Safety." *The Atlantic*. March <https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/>.

Archives of Rudolph W. Giuliani. 1998. "The Next Phase of Quality of Life: Creating a More Civil City." February 24. <http://www.nyc.gov/html/rwg/html/98a/quality.html>.

Hope Corman and H. Naci Mocan. 2005. "Carrots, Sticks, and Broken Windows." *Journal of Law and Economics* 48(1):235–66.

George L. Kelling and William H. Sousa, Jr. 2001. Do Police Matter? An Analysis of the Impact of New York City's Police Reforms. Civic Report for the Center for Civic Innovation at the Manhattan Institute.

Bernard E. Harcourt and Jens Ludwig. 2006. "Broken Windows: New Evidence from New York City and a Five-City Social Experiment." *University of Chicago Law Review* 73:271–320. *The published version has a misprinted sign in the key table. For the correction, see Errata, 74 U. Chi. L. Rev. 407 (2007).*

## Index

---

Note: Page numbers in italic type indicate figures or tables.

- abused children, 364–65
- academic performance: charter schools' effect on, 160–67, *161*, *163*, 177–78, 237–38, 290–93; standardized test scores in relation to, 345; television's effect on, 273–75, *274*, *275*
- Achen, Christopher, 311
- ACU. *See* American Conservative Union
- adaptation, 22, 337–38, 349–52
- airport security: screening for terrorists with behavioral observation, 322–24; screening for terrorists with metal detectors, 337–38, 338, 339
- algorithms, and racial bias, 361–63, 363
- Altenburger, Kristen, 22
- always-takers, 227, 239
- American Conservative Union (ACU), 193–94, 196–97, 206–9, 212–13, 214
- Anzia, Sarah, 294–95
- Aronow, Peter, 123–24
- attrition, 235–36, 240
- average treatment effect (ATE), 164–67, 187, 198
- average treatment effect on the treated (ATT), 165–67, 187, 268
- average treatment effect on the untreated (ATU), 165–67, 187
- back pain, 65
- Bailey, Martha, 277
- bandwidth, 248, *249*
- Bangladesh Integrated Nutrition Project (BINP), 344
- Bartels, Larry, 311
- baseball: defensive shifts in, 350–51; pitchers' batting performance in, 345–49, *347*, *348*
- baseline differences, 166–68, 187
- Basinger, Scott, 18
- Bayes, Thomas, 317
- Bayesian statistics, 328
- Bayes' rule, 317–28, 332
- beliefs: Bayes' rule and, 317–28; and reversion to the mean, 151–52; and quantitative evidence, 314–28
- Bem, Daniel, 122–23
- Berger, Jonah, 261
- Berlinski, Samuel, 108
- Berry, Chris, 294–95
- best fit, 16, 34, 76. *See also* lines of best fit
- best linear approximation to the conditional expectation function (BLACEF), 199
- bias: baseline differences as, 166–68; in correlation and causation questions, 183–86; defined, 95, 109; estimator quality and, 98–99; over- and underestimating, 176–81, *178*; overview of, 96–97; precision and, 98, 99; randomized experiments as safeguard against, 210–11, 221–23; sources of, 168–74; as threat to clear thinking, 5. *See also* publication bias
- Biden, Joe, 249, 251
- Bill and Melinda Gates Foundation, 62, 101–2
- binary variables, 14
- Bin Laden, Osama, 63
- bins, 247–48
- Blattman, Chris, 236–37, 293
- blocking, 223, 234, 239
- Body Vibes, 38–41, 50, 221–22, 226–32, 235–36
- breastfeeding, 219–21
- broken-windows theory, 5–7
- Brooks, Juanita, 315
- Budd, Chris, 116, 117
- Burke, Marshall, 342–43
- Bush, George W., 160
- but-for test, 42, 47
- Cage, Nicolas, 186
- campaign spending, 174–76
- Card, David, 269
- Carney, Dana, 131
- CATE. *See* complier average treatment effect
- Caughy, Devin, 256
- causal effects, 50, 159, 187
- causal mediation analysis, 291–92, 299
- causation/causal inference, 37–50; average effects as key to, 41–42, 46, 49; conceptual issues in, 42–49; controlling and, 209–11; correlation mistaken for, 5, 23–24, 49, 159–74, 183–86; counterexamples and, 44–45; counterfactuals' role in, 39–40, 42–44, 49; defined, 38; explanation

- causation/causal inference (*continued*)  
of, 38–39; fundamental problem of, 41–42, 50; the law and, 42, 47; physical connection not required for, 48–49; problem of a single cause vs. multiple causes in, 43–44; proximate causes and, 43–44, 47; randomized experiments and, 221–25; significance of knowledge about, 37; time and, 47–48; uses of, 40–41
- cause-in-fact, 47
- celiac disease, 2–3, 319–21
- Central Limit Theorem, 102, 104
- Challenger* disaster, 67–68
- chance: prediction threatened by possibility of, 21.  
*See also* luck; noise
- chance imbalance, 232–34, 240
- changing relationship, 352
- charter schools, and academic performance, 160–67, 161, 163, 177–78, 237–38, 290–93
- Chenoweth, Erica, 4
- Chicago Cubs, 351
- Christakis, Nicholas, 280–81
- civil resistance, 3–5
- civil war–economy relationship, 168–69, 174, 182, 214
- clear thinking and conceptual understanding: about bias and noise, 183–86; bias as threat to, 5; data as complement to, 8, 108; errors in, 8; examples showing the need for, 2–8; improper training in, 8; moral responsibility for, 367; about potential outcomes, 163–67; technicality subordinate to, 1–2
- climate change, economic effects of, 342–43
- Clinton, Hillary, 249, 251
- cognitive behavioral therapy, 293–94
- Cohen-Cole, Ethan, 281
- coin clipping, 349–50, 349n
- coin flips
- Coleman Study, 273
- college admissions, 345
- Collins, Janet, 315–19
- Collins, Malcolm, 315–19, 317n
- Collins, Nick, 113
- commodity price shocks–violent conflict relationship, 296–98. *See also* economy: civil war's effect on
- complier average treatment effect (CATE), 230, 240, 260
- compliers, 227, 239, 260. *See also* noncompliance
- conceptual understanding. *See* clear thinking and conceptual understanding
- conditional correlation: example of, 21–22
- conditional mean function, 78–79, 79, 92
- conditional probability, 316, 332
- confidence intervals, 102–3, 110
- confounders: in case examples, 172–81, 194; controlling for, 193–215; defined, 187; explanation of, 168, 169; mechanisms vs., 181–83; over- and under-estimating, 176–81; reverse causality in relation to, 174–76
- continuity at the threshold, 251–56, 262
- contraception, 179–80
- contraception–gender-wage gap relationship, 276–77
- controlling, 193, 193–215; causation and, 209–11; defined, 215; examples of, 194–97; heterogenous treatment effects and, 197–98; matching and, 214–15; for mechanisms, 213–14; regression and, 198–209
- control variables, 198, 215
- Corman, Hope, 6
- coronavirus pandemic, 329–31, 360–61
- correlation, 13–35; causation not implied by, 5, 23–24, 49, 159–74, 183–86; conditional, 21–22; defined, 13, 33; description as a useful function of, 19–20; ethical issues with, 22–23; facts vs., 17–19; measuring, 24–29; multivariable, 21–22; positive, negative, and absence of, 13, 33; prediction as a useful function of, 20–23; usefulness of, 19–24; variation in both variables required to determine, 17–18, 55–70
- correlation coefficient, 28–29, 34
- corruption of the youth, 59–62
- cosmic habituation, 139, 149–50, 151
- cost-benefit considerations, 329–31, 358–59, 365–66
- counterexamples, 44–45
- counterfactual comparisons, 39, 50
- counterfactuals: examples of, 2, 38–39; potential outcomes and, 39–40; role of, in causal reasoning, 39–40, 42–44, 49
- covariance, 27–28, 34
- crime: temperature in relation to, 15–17, 15–17, 24, 25, 26–27, 74–79, 75, 77–79
- crime policy: broken-windows theory and, 5–7
- Cuddy, Amy, 131
- Dal Bo, Ernesto, 297–98
- Dal Bo, Pedro, 297–98
- data: clear thinking as complement to, 8, 108; non-linear, 30–32, 83–86; visual presentation of, 309–14, 310, 312, 313. *See also* quantitative evidence
- defiers, 227, 240, 260
- Dell, Melissa, 257, 260
- demand effects, 139, 152
- Democratic Party, 249–51, 311–14
- democratic peace theory, 282–85, 283
- Denver, John, 60
- dependent (outcome) variables, 57, 75, 91, 215. *See also* selecting on the dependent variable
- description, as a useful function of correlation, 19–20
- deviation from the mean, 26, 34
- Dewan, Torun, 108
- diet soda, 173–74
- difference-in-differences design, 266–86; defined, 285; examples of, 269–77, 270–72, 274–75; overview of, 266–67; parallel trends and, 267–69, 271–72, 278, 280–81, 285–86; units and periods in, 269–77; useful diagnostics for, 278–81; usefulness of, 282–85
- difference in means, 187
- distribution: defined, 24
- Drug Enforcement Administration, 352

- Dube, Oeindrila, 298  
dummy variables, 196, 215
- earnings. *See* income
- economy: civil war's effect on, 168–69, 174, 182, 214; climate change's effect on, 342–43. *See also* commodity price shocks–violent conflict relationship
- efficient-market hypothesis, 132, 151
- elections: age in relation to voter turnout in, 19–21, 20, 79–86, 80–86, 89; campaign spending's effect on, 174–76; candidates' ideology as influence on, 193–97, 194, 206–9, 207–9, 212–13, 213; effect of expanding the franchise on, 108; football results in relation to, 129–30; forecasting of presidential, 87–89; gender discrimination and, 294–95; get-out-the-vote campaigns and, 123–24; incumbency and, 256; of moderates vs. extremists, 249–51, 252, 255–56; newspaper endorsements' effect on, 278–79, 279; RD design's appropriateness for, 255–56; scandals' influence on, 18–19, 18; social pressure's influence on voting in, 107–8, 295–96
- elections, in relation to football results, 129–30
- Enders, Walter, 338
- Environmental Protection Agency (EPA), 358–59
- Ericsson, K. Anders, 171–72
- errors: defined, 76, 92; examples of, 77, 78; standard, 99–101, 110, 224–25
- estimands, 95–96, 106–7, 109
- estimates, 95–96, 106–7, 109
- estimation: overview of, 94–95; over- vs. under-, 176–81; in randomized experiments, 224–25
- estimator, 95, 98–99, 98, 109
- ethics, 22–23. *See also* values
- evidence. *See also* quantitative evidence
- exclusion restriction, 231, 238, 239, 240, 259
- exogeneity, 231, 238, 240, 259, 260
- expectations (expected values), 97, 109, 221–23, 232
- experiments. *See* randomized experiments
- experts and authorities, limits of reasoning and judgment of, 3, 8–9
- explanatory variables. *See* independent (explanatory) variables
- external validity, 343–44, 353
- extrapolation: misuses of, 32
- extrasensory perception (ESP), 122–23, 326
- extreme observations, 101–2
- Facebook, 107–8, 210–11
- facts: correlation vs., 17–19; publication bias as cause for skepticism about, 122–26, 138–39
- false negative rates, 3
- false positive rates, 3
- Farber, Henry, 284
- favorite equation: bias and, 152, 154, 176–77, 222–23, 270, 274; and causal inference, 166, 183–84; explanation of, 94–95; noise and, 118, 138, 232; parallel trends and, 268; publication bias and, 118; randomization and, 222–23; reversion to the mean and, 148–49; statistical inference and, 105
- Fearon, James, 43
- features of the world. *See* variables
- Ferdinand, Archduke, 42–44, 47
- Feynman, Richard, 68
- file drawer problem, 121, 235
- first differences, 272, 275–76, 286
- first-stage effect, 231, 240
- Fisher, R. A., 223
- fixed effects, 272, 275–76, 277, 286
- Fleming, Thomas, 340
- Fletcher, Jason, 281
- flossing, 359–60
- Food and Drug Administration (FDA), 41–42, 329
- football results, in relation to elections, 129–30
- forecasting. *See* prediction
- Fowler, James, 280–81
- France, mayors' salaries in, 254–55
- frequentist statistics, 328
- fundamental problem of causal inference, 41–42, 50
- fuzzy regression discontinuity design, 257–61, 262
- gallons-per-mile vs. miles-per-gallon, 306–8, 307
- Galton, Francis, 90, 139–42
- Garro, Haritz, 248
- Gauss, Carl Friedrich, *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, 90
- Gelman, Andrew, 137
- gender: height-income relationship and, 202–6, 204–5; voter discrimination based on, 294–95; wage gap in relationship to contraception, 276–77
- gene-disease linkages, 150, 151
- General Accountability Office (GAO), 322, 324
- Gentzkow, Matthew, 273, 275
- Gerber, Alan, 295
- get-out-the-vote campaigns, 123–24
- Giuliani, Rudy, 6
- Gladwell, Malcolm, 57–59, 171
- Goldin, Claudia, 276
- golf scores, 143–45, 144, 145
- Goop, 38
- Gore, Tipper, 59–60
- government programs, RD designs for evaluating, 246–47
- government type–oil production correlation, 14–15, 14, 55–56, 56
- Gowa, Joanne, 284
- Green, Donald (Don), 123–24, 284, 295
- Greenstone, Michael, 365–66
- Guggenheim, David, *Waiting for Superman*, 160–61
- Hall, Andrew, 251, 256
- Hamlet Evaluation System, 257–58, 260
- Harcourt, Bernard, 6–7
- Hawthorne effect, 139, 139n, 152

- health: causation in matters of, 38–42, 45–46; confounders and mechanisms in, 181–82; diagnoses influenced by limited sample of patients, 65; heart health, 181–82; help-seeking and reversion to the mean in matters of, 145–47; HIV and contraception, 179–80; intermediate outcomes' value in studies related to, 340; predictions concerning violations of codes of public, 22; quantitative evidence interpretation in matters of, 2–3, 319–21; racial bias in algorithms related to, 361–63, 363; of skin, 38–40, 221–22, 226–32, 235–36; social networks' effect on behavior associated with, 279–81
- Healy, Andrew, 129
- height of individuals: generational variation in, 90, 91, 139–43; and income, 201–6, 202–5
- Heinzerling, Lisa, 358
- help-seeking, and reversion to the mean, 145–47
- heterogenous treatment effects, 46, 50, 197–98
- high school dropouts, 62–63
- HIV, 179–80
- Ho, Daniel, 22
- homogeneous treatment effects, 230
- homophily, 280–81
- Hsiang, Solomon, 342–43
- hypothesis testing, 103–4, 110, 114, 225
- Imbens, Guido, 198
- income: height in relation to, 201–6, 202–5; merit scholarships' effect on, 244–46, 254–55; military service's effect on, 238–39
- incumbency advantage, 256
- independent (explanatory) variables, 75, 83–86, 92
- India, malnutrition in, 343–44
- in expectation. *See* expectations (expected values)
- instrumental variables (IV) analysis, 231–32, 238, 240, 260
- intent-to-treat (ITT) effect, 226–31, 238, 240
- intercept, 75, 92
- interference, 236–37, 240
- intermediate outcomes, 292–94, 339–41
- internal validity, 343, 353
- International Journal of Epidemiology*, 219
- Irvin, Veronica, 128
- Jacobson, Gary, 175
- Jamison, Julian, 293
- Journal of Personality and Social Psychology*, 122–23
- Kant, Immanuel, 282
- Kaplan, Robert, 128
- Katz, Lawrence, 276
- Kelling, George L., 5–7
- Kim, Soo Yeon, 284
- knee surgery, 146–47
- Knowledge is Power Program, 163
- Koehler, Jonathan, 315
- Krueger, Alan, 269
- Ladd, Jonathan, 278
- The Lancet Infectious Diseases* (journal), 179–80
- Larimer, Christopher, 295
- Larrick, Richard, 307–8
- LATE. *See* local average treatment effect
- law: causality and, 42, 47
- Law of Large Numbers, 102
- lead treatment variable, 278, 286
- Legendre, Adrien-Marie, *New Methods for the Determination of the Orbits of Comets*, 89–90
- Lenz, Gabriel, 278
- Liberian youth, interventions for at-risk, 236–37, 293–94
- life advice, 70
- linearity: importance of determining, 21; uses for describing non-linearity, 30–32
- lines of best fit, 16, 21, 29, 33, 74–76
- Ling, Jeff, 60
- local average treatment effect (LATE), 198, 216, 246, 247, 252, 260
- local linear approach, 248, 249
- long format, 272, 286
- long regression, 199
- luck, 113–18, 132–33, 143–44, 151. *See also* chance
- Ludwig, Jens, 6–7
- Maddon, Joe, 351
- Malhotra, Neil, 129
- malnutrition, in India and Bangladesh, 343–44
- margin of error, 102, 110
- mask-wearing, during coronavirus pandemic, 360–61
- matched-pair design, 223
- matching, 214–15
- McChrystal, Stanley, 63–64
- McGlaughlin, Dan, 170–73
- McGrath, Mary, 123–24
- McNamara, Robert, 257
- mean ( $\mu$ ), 24, 26, 34. *See also* reversion to the mean
- measurement: intermediate outcomes and, 339–41; mission in relation to, 336–43; partial, 337–38; and sample appropriateness, 345–49; strategic adaptation and, 349–52; of wrong outcome or treatment, 337–43
- mechanisms: assessment of, 290–99; causal mediation analysis and, 291–92; confounders vs., 181–83; controlling for, 213–14; defined, 187; designing studies to test, 295–96; disentangling, 296–98; independent theoretical predictions about, 294–95; intermediate outcomes and, 292–94; overview of, 290–91
- mediators. *See* mechanisms
- metal detectors in airports, 337–38
- Miguel, Edward, 342–43
- miles-per-gallon vs. gallons-per-mile, 306–8, 307
- military service–earnings relationship, 238–39
- Miller, Bill, 132–34
- minimum wage–unemployment relationship, 269–72, 270–72

- missions: aligning measurement with, 336–43; external validity and, 343–44; ill-defined, 341–43; sample appropriateness for, 343–49; strategic adaptation and, 349–52
- Mo, Cecelia, 129
- Mocan, H. Naci, 6
- Montagnes, Pablo, 129–30
- motivation in sports, 261
- multivariable correlation, 21–22
- National Aeronautics and Space Administration (NASA), 68
- National Heart, Lung, and Blood Institute, 128
- National Political Awareness Test (NPAT), 194–97, 206–9, 212, 214
- natural experiments, 237–39, 240
- Nature* (journal), 107–8
- Nature Human Behaviour* (journal), 127
- nearest neighbor matching, 214
- negative correlation: defined, 13, 33; examples of, 16
- Nelson, Leif D., 124, 126
- never-takers, 227, 240
- New England Journal of Medicine*, 148, 280
- newspaper endorsements, voter behavior affected by, 278–79, 279
- Newton, Isaac, 349n
- New Yorker* (magazine), 139
- n factorial, 115
- 95% confidence interval, 102–3, 110
- noise: in correlation and causation questions, 183–86; defined, 95, 97, 109, 152; effects of, 109; hypothesis testing for, 104, 114; in randomized experiments, 222, 223; reversion to the mean linked to, 138, 142. *See also* chance
- noncompliance, 226–32, 239, 256–61
- non-linearity, 30–32, 83–86
- NPAT. *See* National Political Awareness Test
- null hypothesis, 104–5, 110
- Obama, Barack, 160
- obesity, contagiousness of, 279–80
- observations: defined, 15
- Office of Information and Regulatory Affairs (OIRA), 358–59
- oil production–government type correlation, 14–15, 14, 55–56, 56
- omitted variable bias formula, 200–201, 200, 201, 216
- one-sided z-test, 104
- ordinary least squares (OLS) regression, 77, 92, 196
- ordinary least squares (OLS) regression coefficients, 77
- ordinary least squares (OLS) regression line, 29, 34, 77–81
- outcome variables. *See* dependent (outcome) variables
- out-of-sample predictions, 86, 86, 92
- over-comparing, 118–21, 119, 127–28, 131
- over-estimates, 176–78, 180, 182, 187
- overfitting, 86–89, 92
- Paltrow, Gwyneth, 38
- Pape, Robert, 63–64
- paradox of plenty, 14
- parallel trends, 267–69, 271–72, 278, 280–81, 285–86
- Parents Music Resource Center (PMRC), 59–62
- partial measures, 337–38, 352
- Pascal, Blaise, 43
- Pauling, Linus, 148
- Paul the Octopus, 113–18
- PCR tests. *See* polymerase chain reaction (PCR) tests
- Pence, Mike, 360
- percentage point change, 309, 331
- percent change, 309, 331–32
- Perkoski, Evan, 4
- p-hacking, 119–20, 120n, 122–26, 134, 206
- physicalism, 48–49
- pitchers, batting performance of, 345–49, 347, 348
- placebo effect, 147–48
- politics: campaign spending, 174–76; French mayors' salary-performance relationship, 254–55; Southern realignment in, 311–14, 312, 313; stock prices, 248–49. *See also* elections; U.S. Congress; voter behavior; voter turnout
- polling, 95, 97–101, 103–4
- polymerase chain reaction (PCR) tests, 329–31
- Pope, Devin, 261
- population difference in means, 165
- populations: defined, 109; samples in relation to, 29–30, 94
- positive correlation: defined, 13, 33; examples of, 16
- posterior beliefs, 318, 326–28, 327, 328, 332
- post-mortem analyses, 65–69
- post-treatment covariates, 181, 187
- potential outcomes, 163–67; defined, 50; explanation of, 39–40
- potential outcomes framework, 39, 50
- power pose, 131
- precision: bias and, 98, 99; defined, 109; estimator quality and, 97, 98–99
- prediction: considerations involved in, 21; correlation as useful for, 20–23; ethical issues in, 22–23; example of unreliable, 113–18; out-of-sample, 86, 86, 92; of presidential elections, 87–89; quantifying, 99–103
- pre-registration, of research studies, 127–28
- presidential elections, 87–89
- pre-treatment covariates, 181, 187
- pre-trends, 278, 279, 279, 286
- Preuss School, 160–63, 161, 163, 167–68
- prior beliefs, 318, 326–28, 327, 328, 332
- Proceedings of the National Academy of Sciences* (journal), 129
- product of the deviations, 28
- Project Vote Smart, 194
- proximate causality, 47
- proximate causes, 43–44
- p-screening, 120–21, 123–24, 134
- PS: *Political Science & Politics* (journal), 87



- publication bias: defined, 134; in everyday life, 131–34; overview of, 118–21, 119; *p*-hacking, 119–20, 122–26; potential solutions to, 126–31; *p*-screening, 120–21, 123–24; reversion to the mean linked to, 149–50, 151; skepticism about facts because of, 122–26, 138–39
- p*-value, 104–5, 110, 114, 119–21, 127, 325
- quantitative evidence: acting on, 305, 328–31, 361–67; beliefs' interaction with, 314–28; cautionary tales about, 2–8; insufficiency of, 357–61; limits of, 357–67; and non-quantifiable properties, 358–59, 363–64, 366–67; scale for representing, 305–11; values' interaction with, 328–31, 357, 361–67; visual presentation of, 309–14. *See also* data
- Quarterly Journal of Political Science*, 108
- Querubin, Pablo, 257, 260
- racial bias, in health-related algorithms, 361–63, 363
- Radio Lab* (radio show), 139
- random assignment, 239
- randomized experiments, 218–39; attrition in, 235–36; causal inference and, 221–25; chance imbalance in, 232–34; estimation in, 224–25; examples of, 219–21; hypothesis testing and, 225; interference in, 236–37; natural experiments vs., 237–39; noise in, 222, 223; noncompliance in, 226–32; potential problems with, 225–37; surgical procedures as, 147; unbiasedness of, 106, 210–11, 221–23; underpowered, 234–35; value of, 218
- RD designs. *See* regression discontinuity (RD) designs
- reduced-form effect. *See* intent-to-treat (ITT) effect
- regression, 74–92; controlling and, 198–209; elements of, 198; intellectual history of, 89–91, 141; linear, 79–86; non-linear data and, 83–86; overfitting and, 87–89; overview of, 74–79; presentation of, 89, 89; reading regression tables, 211–13, 213; sharp vs. fuzzy, 256–57; usefulness of, 91
- regression coefficient (slope of regression line), 29, 34
- regression discontinuity (RD) designs, 243–62; applications of, 246–47; continuity at the threshold in, 251–56, 253; defined, 262; examples of, 244–46, 245–47, 248–51, 252, 257–61, 258, 259, 261; implementation of, 247–51; noncompliance and, 256–61; overview of, 243–46
- regression equation, 75, 92
- regression lines, 92. *See also* ordinary least squares (OLS) regression line
- regression parameters, 75–76, 92
- regression tables, 211–13, 213
- regression to the mean. *See* reversion to the mean
- regulations, government, 306–8, 358–59
- Reinhart, Carmen M., 69
- replication, 123, 128–30, 138–39
- replication crisis, 139
- representativeness: of samples, 21
- Republican Party, 311–14
- research design, 218, 239
- resource curse, 14
- reverse causality: in case examples, 174–76, 178–79; confounders in relation to, 174–76; controlling not effective for, 210; defined, 187; explanation of, 169–70, 170
- reversion to the mean, 138–52, 176–81; beliefs not susceptible to, 151–52; cosmic habituation and, 139, 149–50, 151; dangers of not recognizing, 150; defined, 152; discovery of, 139–41, 140; everyday life examples of, 145–47; examples of, 6–7, 90–91; explanation of, 141–45; instances not susceptible to, 151; noise linked to, 138, 142; placebo effect compared to, 147–48; signal linked to, 142
- Rogers Commission, 68
- Rogoff, Kenneth S., 69
- r*-squared statistic, 29, 89, 212–13
- running variable, 244–46, 262
- samples: appropriateness of, 343–49; defined, 109; populations in relation to, 29–30, 94; representativeness of, 21; selected, 344–49; small, 101–2
- sampling distribution, 99, 110
- Sanders, Bernie, 249–51
- San Diego City Schools, 160–63, 161, 163, 167
- Sandler, Todd, 338
- scale, for data representation, 305–11
- scatter plots, 15, 15
- Schmidt, Eric, 1
- scholarships–earnings relationship, 244–46, 254–55
- Schooler, Jonathan, 139, 149
- Science* (journal), 362–63
- Screening of Passengers by Observation Techniques (SPOT), 322–24
- Second Reform Act (United Kingdom, 1867), 108
- Sekhon, Jas, 256
- selected samples, 344–49, 353
- selecting on the dependent variable: defined, 56–57, 70; error of, 56–64; world seemingly organized to encourage, 64–70
- sensitivity analysis, 179
- Shapiro, Jesse, 273, 275
- sharp regression discontinuity design, 256, 262
- Sheridan, Margaret, 293
- short regression, 199
- signal (systematic factors), 142, 152
- significance, substantive vs. statistical, 107–8
- Silver, Nate, 87, 89
- Simmons, Uri, 124, 126
- Simonsohn, Joseph P., 124, 126
- Singer, David Andrew, 69
- skin health, 38–40, 221–22, 226–32, 235–36
- slope, 16, 34, 75. *See also* regression coefficient
- Snider, Dee, 60
- social media: voter behavior and, 107–8; well-being in relation to use of, 210–11, 211
- Soll, Jack, 307–8

- sorting, 254  
Sousa, William, 6–7  
Southern realignment in politics, 311–14, 312, 313  
Spenkuch, Jorg, 248  
Spiegelharter, David, 116, 117  
Springsteen, Bruce, 60  
spurious correlation, 184  
standard deviation ( $\sigma$ ), 27  
standard error, 99–101, 110, 224–25  
state fixed effects, 272  
statistical inference, 21, 105–7  
statistical power, 234–35, 240, 325–26, 332  
statistical significance, 104–5, 107–8, 110, 126–27  
statistics: beliefs and, 314–28; defined, 18; substantive use of, 305–31; values and, 328–31. *See also* quantitative evidence  
stock prices, 132–34, 151–52  
strategic adaptation, 22, 337–38, 349–52, 353  
stratification, 223, 234, 239  
substantive significance, 107–8  
suicide terrorism, 63–64  
summation ( $\Sigma$ ): defined, 24  
Summers, Larry, 365–66  
sum of squared errors, 29, 34, 76–79, 92  
superstars, 132–34  
  
Tamil Nadu Integrated Nutrition Project (TINP), 343–44  
Tampa Bay Rays, 351  
teeth flossing, 359–60  
television's effect on children's academic performance, 273–75  
temperature-crime relationship, 15–17, 15–17, 24, 25, 26–27, 74–79, 75, 77–79  
10,000-hour rule, 57–59, 170–73  
terrorists, identification of, 322–24, 337–38, 338, 339  
Thompson, Dan, 251  
time: causality and, 47–48, difference-in-differences design for studying changes over, 266–85  
toxic dumping, 365–67  
Transportation Security Administration (TSA), 322–24  
treatments, 39, 50, 164. *See also* heterogeneous treatment effects  
treatment variables, 198, 215  
Trump, Donald, 250, 360  
2008 financial crisis, 69  
  
unbiasedness, 109  
under-estimates, 176–78, 187  
under-reporting, 118–21, 119, 127–28, 131  
unemployment–minimum wage relationship, 269–72, 270–72  
U.S. Coast Guard, 352  
  
U.S. Congress, voting records of members of, 193–97, 194, 206–9, 207–9, 212–13, 213, 214  
U.S. Department of Education, 61  
U.S. House of Representatives, 63–64  
U.S. Secret Service, 61  
U.S. Senate, 60  
utilitarianism, 364  
  
values: and decision-making with quantitative evidence, 328–31, 357, 361–67; hidden in quantitative approach, 361–63; influenced by quantitative approaches, 364–67. *See also* ethics  
values, and quantitative evidence, 328–31  
Vargas, Juan, 298  
variables: control, 198, 215; defined, 13, 15; dependent (outcome), 57, 75, 91, 215; independent (explanatory), 75, 83–86, 92; instrumental, 231–32, 238, 240; running, 244–46; treatment, 198, 215  
variance ( $\sigma^2$ ), 26, 34  
Vietnam War, U.S. bombing strategy in, 257–61, 258, 259  
Vigen, Tyler, 184–85  
violent conflict–commodity price shocks relationship, 296–98. *See also* civil war–economy relationship  
vitamin C, 148  
voter behavior: and gender discrimination, 294–95; get-out-the-vote campaigns and, 123–24; policy preferences and, 311–14, 312, 313; social pressure and, 107–8, 295–96  
voter turnout: age in relation to, 19–21, 20, 79–86, 80–86, 89; newspaper endorsements' effect on, 278–79, 279  
  
Wainer, Howard, 101  
Wald, Abraham, 231  
Wald Estimator, 231–32, 240  
war. *See* civil war–economy relationship; democratic peace theory; violent conflict–commodity price shocks relationship  
war on drugs, 351–52  
welfarism, 364  
well-being, social media's effect on, 210–11, 211  
Wertham, Fredric, 60–61  
wide format, 272, 286  
William III, King, 349  
Wilson, James Q., 5  
window tax, 349–50  
World Bank, 343–44  
World War I, 42–44, 47  
  
Yap, Andy, 131  
Yoon, David, 284  
youth: corruption of, 59–62; as dropouts, 62–63  
  
z-test, 104