CHAPTER 1

# INTRODUCTION

## THE NEED FOR STATISTICS AND GRAPHICAL DISPLAY

Many of you are reading this book very likely because you're taking a course that some college administrator and/or faculty committee believes is important for you to suffer through. As in the other courses you're required to take, you're determined to grin and bear it for 12 to 15 weeks, have your passport stamped, and move on, perhaps to courses you actually *want* to take. Or you're an intrepid soul who wants to learn statistics on your own and apply best principles and practices in your job, generating and interpreting statistics or judging others' use of them.

Some of you are quite nervous about statistics. You've heard of or experienced statistics as formulas you loathed and soon forgot, as a mysterious language that you misapplied on more than one occasion, or as a pack of lies that others used to lead you away from the path of truth.

Some of you can expect in your current or future job to rely on others to complete the statistical analysis and graphical display. The wise among you will realize that you don't want to be captive to what these people tell you can and cannot be done. You've got to know enough to know what's possible even if you may not be able to or have the time to wield these tools yourself. Similarly, some of you want to learn enough about statistics and data collection strategies to know when those who criticize your efforts or your organization's actions are making a valid point and when they are inventing results behind a curtain of bewildering statistical language and gadgetry.

And then there are some of you (maybe even the same folks above) who believe that data and analysis can help humans make better and wiser decisions. They can. They have. We'll discuss these instances shortly and try to draw general guidelines for the conditions under which this is more or less likely to take place.

The policy analyst that this book hopes to help you become or to better communicate with is a broker between decision makers and social scientists and statisticians, whose language and methods are difficult for decision makers to fully understand. The policy analyst must integrate knowledge and insights from multiple disciplines and multiple methodologies. In seeking to cut through the clutter of messages that decision makers receive, in order to move bureaucracies, entrenched interests, and long-standing administrative cultures, the policy analyst, cum advocate, will be tempted to overstate his or her case and, in so doing, fuel the criticism of those who oppose those recommendations. This text is designed to help you spot when that enthusiasm gets the better of others and to check your own impulse to distort what the data say.

## ● THE POLITICS OF NUMBERS

That numbers matter is nowhere more tellingly validated than by the vast amount of money, time, and energy that go into collecting, analyzing, and presenting them and—on some occasions—by the conscious attempts to bend them to support someone's position (see, e.g., Prewitt, 1987). In these latter instances, data and statistics become the handmaidens of ideology or self-interest. It is as much the purpose of this book to enable you to spot these statistical charlatans as to create truthful, useful, and effective analyses and presentations yourself.

Solid and defensible statistical analysis can be important in moving policies in an effective direction. But you have to have your statistical house in order if you hope to promote positive change. If not, those who oppose the prescriptions that you draw from your evidence will attack your ideas through your methods. Don't give them the chance. Learn and apply statistics appropriately, lest disputes over them muddy the waters and immobilize the body politic. Learn to apply statistics—in Robert Abelson's (1995) terms—as principled argument.

Seek through sound analysis to mitigate disputes over methods. Such arguments—often noted more for their heat than for their light—simply befuddle the public, the media, and public officials. The exchanges between Paul Krueger and Paul Peterson and their colleagues (Howell & Peterson, 2004; Krueger & Zhu, 2004a, 2004b; Peterson & Howell, 2004) over the effects of school vouchers on inner-city minorities is a case in point. Arguments over how to handle missing values and how to code the race of a child have left the public, journalists, and many social scientists unsure of the actual effects of school vouchers and, thus, whether their implementation is a worthy goal. In so doing, they serve, according to the

economist Henry Aaron (1978), the profoundly conservative purpose of undercutting the demand for policy innovation.

## THE STRENGTHS AND WEAKNESSES OF THEORY, DATA, AND ANALYSIS ●

The good news is that good data, theory, and analysis have helped raise important problems on the public agenda and provided a road map for their solution. The bad news is that they don't always do this.

Evidence for the power of ideas can be found among the works of political scientists and sociologists of knowledge who have sought to understand the conditions under which ideas and sound reasoning (including statistical ones) make a positive difference in public and social policy. John Kingdon (2003), for example, has studied how issues rise on or fall from the public agenda at the national level. In his widely read book *Agendas, Alternatives, and Public Policies*, Kingdon persuasively demonstrates that sound analysis and good theories have made a substantial impact on bringing to national attention and providing solutions to a range of issues, from the deregulation of the airline industry to health care reform. That such analysis is more often at play in the primeval soup of policy alternatives than at the stage of setting the national agenda itself does not belie its importance.

Unfortunately, there's a flip side to this story. Consider, for example, the implementation of the Head Start program. Sheldon H. White and Deborah A. Phillips (2001) tell an interesting story about the role played by developmental psychologists and the research they brought to bear on this program's initial design. It's a story that illustrates the differences in cultures and operating procedures between the experts and practitioners in the evolution of Head Start and the limits of available data and analysis in answering the tactical questions that arise in the design and implementation of programs in specific local settings. As Jule M. Sugarman, Associate Director of Head Start in its early years, recollected,

> The Planning Committee had only established the policy. It was up to the administrators to set rules and guidelines. . . . [We] found that "experts" were not very deep in their knowledge. No one could tell us, based on real evidence, what the proper child-staff ratios or length of program should be. Despite their lack of depth, experts were vigorously committed to their point of view and often rejected other views in irrational and unproductive ways. Many of the decisions eventually had to be made by administrators because the professionals could not reach decisions among themselves. (Zigler & Valentine, 1979, pp. 118–119, as quoted in Featherman & Vinovskis, 2001, p. 89)

Tensions and differences in goals, as well as knowledge (or the lack thereof), also highlighted the relationships between the top-down, expert-guided central planners and the anti-elitist, local community action implementers. The objectives of developmental psychologists and pediatricians who saw child development as the program's purpose conflicted sharply with those who viewed Head Start in light of empowerment and mobilization of children's parents in poor communities.

Interestingly, White and Phillips (2001) alert us to another potential problem here as well. Developmental psychology at the time of the experiments that supported the creation of Head Start was heavily dominated by cognitive psychology and its focus on cognitive development in children (in contrast, e.g., to social and emotional development). On the other hand, preschool directors did not share this emphasis on cognitive development and, abetted by Head Start administrators who sought flexible and local adaptations, were cognizant of different local conditions among heterogeneous populations.

This is a point to which we will refer later, but one worth drawing your attention to now. That is to say, certain research designs—especially, sample surveys of individuals—have an often unrecognized tendency to focus explanations on the characteristics of individuals in contrast to the structure of circumstances and opportunities that broader social forces make available to some groups of individuals. This distinction—often referred to as the difference between agency and structure—runs deeply throughout the history of the social sciences and policy research. It will not be resolved in this book. I do, however, want to alert you to the fact that different research designs can unwittingly cause us to fall into one of these two camps, just as the fashion of a discipline at any point in time can lead us to explore some questions rather than others.

We can also approach the good news/bad news story about the influence of good theory, data, and analysis on the policy process as policy analysts by reframing this observation as a question. To wit, under what conditions is our analysis likely to make a positive difference?

## ● DESIGNING USEFUL RESEARCH

Judith Gueron, as head of the Manpower Demonstration Research Corporation, is one of the many analysts/administrators who have sought to draw lessons for those who wish to use sound research and analysis—in her case, social experiments—to inform public policy. Many of the lessons she has drawn from several decades of research on employment and training

programs appear obvious, although they are often breached in practice (e.g., diagnose the problem correctly, devise a reasonable treatment). But her success is probably predicated on

- designing a real-world test,
- addressing questions that the public cares about,
- contextualizing the results in the face of what is known about effective and ineffective programs,
- actively disseminating the results without overstating the case, and
- soliciting key partners throughout the process.

In other words, the effectiveness of your analysis is as much a political process as it is a statistical one.

Peter Szanton (2001) addresses the question of how to increase the value of analysis through a different approach. He asks why policy prescriptions—based on appropriate designs and statistics—are so often ignored by public officials. More specifically, Szanton asks why the advice given to local public officials—whether from social scientists based in universities, think tanks, or consulting firms—was so often useless and unused in helping resolve the urban problem in the United States in the 1960s and 1970s. There's plenty of blame to go around in his account. But unlike most research that focuses on the supply side of advice, Szanton argues persuasively that the lion's share of the problem lies in city governments themselves. In short, they lack the incentives and capacities to accept and act on sound advice.

Although Szanton (2001) is careful to circumscribe the lessons he draws from cities, I believe that its lessons extend well beyond that arena, as the parallel between this work and Judith Gueron's attests. Szanton helps us better understand the conditions under which policy research is most (and least) likely to achieve the aim of "informing" the design and implementation of "better" public policies and programs.

Szanton (2001) draws several conclusions based on a variety of case studies in which scholarly advice was ignored (and, occasionally, followed). Ineffective advice was more likely to occur when

- structures and relationships between analysts and policymaker were formal (e.g., lodged in urban "centers" and "institutes") rather than personal or informal,
- goals were national and ambitious rather than specifically local, and
- efforts were funded directly and exclusively by third parties (e.g., foundations) rather than by local decision makers.

Efforts to impose a technical solution to a political problem were invariably doomed and short-lived.

Conversely, successful collaborations between a policy analyst and local government tended to be based on less visible, less formal, lower-level ad hoc relations. Under these conditions, advice was more likely to be acted on and more likely to endure. Success was also more likely to result when the advisor was nonthreatening, persistent, flexible, committed, and willing and able to take the blame for any shortcoming in results and, conversely, redirect the light of success on public officials. Success was also furthered by selecting problems that were amenable to quick and effective solutions, for which a demonstration or experiment could be conducted. They were problems that city officials or bureaucrats thought important.

These insights move Szanton (2001) to describe the following difficulties that the prospective consumers of advice face:

> Innovation in city agencies must negotiate an obstacle course of civil service regulations, line-item budgets, collective bargaining requirements, community sensitivities, an attentive press, and the charges of a political opposition. A local government agency, in short, is deeply embedded in a local social setting and tightly constrained by it. (p. 113)

Szanton (2001) concludes with nuggets of advice for the three sets of institutions in this drama. There are far too many to summarize here, but a selected few are listed:

- Advisors and foundations should seek to augment the capacity of local governments to accept and act on advice as much as provide advice (p. 125). Plenty of good ideas for change and improvement exist; it's the lack of political, managerial, or fiscal capacity that's the problem.

- Avoid the search for universal truths and generalities. Although this is helpful for building theory (and in securing tenure as a university professor), a community's needs are likely to be unique. They will want advisors to address their particular needs. When funders and advisors insist on searching for more universal truths, clients become disinterested, uncooperative, and resistant—characteristics that do not create a fertile bed for constructive change.

- Findings, advice, and recommendations will tend to favor the interests of some person or group at the expense of another and are, therefore, inherently political. Being political, those who give it should brace themselves for attacks against their results, motives, costs, and methods. If you can't take the heat, stay out of the kitchen!

- Advice, if calling for substantial change, will always require time to bring about that change.
- For consumers of advice, don't ask for it if you don't want it!

## THE GRAMMAR OF STATISTICS ●

Pick up any textbook in statistics, and you will soon realize why statistics seem so foreign. It is. The language for many is not only new but mysterious. This is so for several reasons. The language of statistics is

- paradoxically precise yet probabilistic;
- slightly askew from everyday usage and downright misleading in some instances;
- replete with instances in which the same word takes on substantially different meanings, even in a statistical context; and
- replete with double negatives (e.g., rejecting the null hypothesis is one of my favorites).

All these characteristics, of which there will be ample examples throughout, get in the way of understanding and communicating statistics, but they help make a decent wage for the statisticians who invent the jargon, use it, and criticize others' misuse of it. This book tries to ease your pain but won't eliminate it entirely. You will have to learn the language in order to understand what others are saying or writing and in order to translate that language for your colleagues who haven't read this book and completed the exercises that accompany it.

Statisticians appear to be a rather negative group on the whole. In addition to a preoccupation with rejecting the null hypothesis, they're also fond of focusing a great deal of attention on errors. Indeed, they've developed quite an assortment of them. They begin with "errors of observation" (see Groves et al., 2004). These are defined as differences between your conceptual or theoretical constructs (what you want to measure) and what you actually measure, say, in respondents' answers to your questions. You were asking about "profits." The respondent thought you were asking about "prophets"!

There are also "errors of nonobservation" (Groves et al., 2004, p. 60). How can I err in something I don't observe, you may ask? Such errors may be "sampling errors," that is, the differences between a statistic calculated from a sample of the population and its "true" value in the population as a whole (this true value is referred to as a population **parameter**,* in

*Words in bold font can be found in the glossary to this book.

contrast to a "statistic" or "estimate," which is based on a small set of observations taken from a population). These errors may also arise from coverage errors, where some members of the population of interest are excluded or underrepresented in the study, say, because they're less likely to be home during the hours in which interviews take place or less likely to have access to a phone in a telephone survey. And errors may arise because those people who did respond are unlike those who did not—the topic of **nonresponse bias**, to which we will return later. There are plenty of errors to go around.

All this attention on errors leads many readers to despair. You need not. If the errors are not systematic, they may behave much like "white noise." They'll surely reduce the size of some statistics that seek to explain the variation in the concept you would like to understand better. But a little noise never hurt anyone. It's the missed messages that are disguised by the noise that should concern us. We will return to this issue under the rubric of concepts such as **residuals**. Yes, it is a different language.

Errors of measurement are commonplace. Consider the task of measuring someone's height. You may not have precise tools for this task, or your angle for viewing your colleague's height may have led you to misread the ruler that you tape to the wall. You may measure her height with or without shoes, with or without the end of a level to rest on the top of your friend's head, and so on. But, on the whole, your measures aren't likely to be terribly wrong. For most purposes, they're probably quite good enough. It's the potential for systematic errors in measurement that might cause problems.

Can there be such errors in measuring someone's height? Yes. It turns out that we're all a little taller in the morning after a good night's sleep in a horizontal position than we are in the evening after having stood on our feet or sat on our butts for an entire day. Why? Our vertebrae are cushioned by intervertebral disks made up of fibers and a gel the consistency of Jello. These disks compress somewhat during the day and expand while we're asleep. Surely "errors" in measuring height caused by these small daily changes won't be too bad either, will they? Again, probably not. But it's something we should always think about.

The statisticians' preoccupation with errors can be a downer. Get over it. If you're led into such a funk by this text or others, remember the following three points:

1. Don't throw the baby out with the bathwater. Use common sense, even in the face of fancy statistics. A little error, especially in comparison with the magnitude of an actual effect and the uses to which you'll put the measure, may be quite alright.

2. Don't make the opposite mistake, however, of thinking that errors aren't all that bad. They can be, although not necessarily so.

3. Statistics may shroud themselves in a fog of apparent precision with, say, ***p* values** of .0134. This is an illusion. Decimal points do not confer precision. Don't be taken in by them or purposefully try to mislead others by overusing them yourself.

## BE CAREFUL WHAT YOU WISH FOR: THE POTENTIALLY PERVERSE EFFECTS OF NUMBERS ●

As you well know, the need for quantifiable results can produce perverted outcomes. Many accuse the No Child Left Behind (NCLB) law of creating too great an emphasis on standardized tests, which lead teachers to "teach to the test." Some worry that teachers are repeatedly drilling students on standardized tests on subjects for which they and their principals are being held accountable (e.g., math and reading to the neglect of subjects such as history and science and activities such as art and physical exercise). Of course, teaching to the test need not be bad if the tests are well designed and push students to learn what is known to be useful and important.

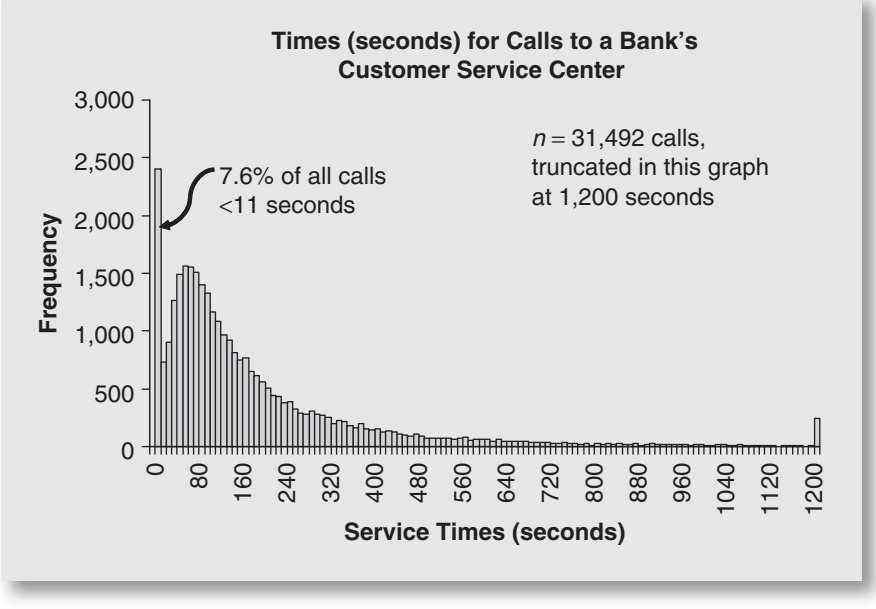
Because the measurement of proficiency in NCLB is left to each state to determine, you can find some states dumbing down their tests to ensure that high percentages of their students achieve proficiency. The current winner of the award for statistical shame is Mississippi, which has designed a fourth-grade test of reading, for example, in which about 90% of their children score "proficient or better." This score ties Mississippi for the best score in the country with Nebraska. When Mississippi fourth graders, however, are tested by a national standardized reading test in the National Assessment of Educational Progress (NAEP), only about 18% of these children achieve reading proficiency or better. Mississippi's kids drop from 1st to 50th in rank among the states, falling 71 percentage points lower on the NAEP than on their home-grown tests (Wallis & Steptoe, 2007, p. 39).

Equally pernicious, tests and statistics can lead teachers and school administrators to cheat, as we will see later in our examination of test score data from school districts in Texas. And there is ample evidence that the high-stakes standardized tests have led some principals, whose jobs depend on showing high proficiency levels and/or improving scores, to hold back (i.e., "retain") children in, say, the 9th grade, knowing that the students face an important test in 10th grade. These students, stigmatized as "dropbacks," often drop out of school (McNeil, Coppola, Radigan, & Heilig, 2008).

The perverse incentives that numbers can create are not, of course, limited to public schools in the United States. A study of the length of telephone calls to a customer service center at a small bank demonstrates the unintended consequences of numeric goals and numeric measurement systems (Shen, 2003). Figure 1.1 (a histogram generated in Excel) shows a quite striking fact: A large number (and percentage) of calls to the service center were terminated in 10 seconds or fewer. You wouldn't think a service employee could satisfactorily respond to a customer in less than 10 seconds. Indeed, the most frequent length of response was 2 seconds! What's happening here?

It turns out that the bank had a policy when these data were collected to penalize service representatives if the average length of their calls was "too long." How would you make sure you didn't violate company policy if you were a customer service rep? Hang up quickly on a few callers, of course. "Oops, sorry about that. I hit the wrong key. May I help you?" After these data were shown to the bank, the bank changed its policy.

**Figure 1.1** The Perverse Consequences of Numeric Performance Guidelines




SOURCE: Based on Shen (2003), as referenced in Moore and McCabe (2006, pp. 10–11).

The good news lesson of this example is that good data can chase away bad data (and associated behaviors). (Not always, of course. We're still waiting on Mississippi.)

## A PREVIEW OF TAKEAWAYS FROM THE BOOK ●

I will shortly argue that you should begin your research at the point you want to end. That is to say, what decisions do you (or someone else) need to make for which statistics will matter? So let us begin with a preview of where you can expect to land at the end of this flight. As a result of reading the chapters and completing the exercises, you will become painfully aware of the following:

- **Correlation** is not causation, although you can come close with a little help from friends such as temporal order, strong theory, consistent evidence from multiple studies, and the elimination of competing alternative explanations.
- Single statistical measures can be misleading. Base important decisions on evidence that has been replicated and findings from a variety of studies and methods that converge on the same conclusions.
- Sound analysis begins by displaying the data and getting a feel of its shape, character, and idiosyncrasies.
- Variability characterizes every human process and the data we use to describe those processes.
- Don't mistake chance variation for causation, patterns, or trends. Chance can offer a "fool's gold" of apparent regularity.
- Statistics may help reduce uncertainty but not eliminate it. Become comfortable with probabilities.
- A statistically significant difference does not necessarily mean a meaningful difference.
- Research (i.e., data collection, analysis, and presentation) of the kind described in this book should be focused on helping people decide among alternative courses of action. Check your idle curiosity at the door. It's too expensive for most governments, public agencies, and nonprofit organizations.
- Research design is as important, if not more so, than statistical tests. Sophisticated statistics cannot overcome poorly designed and executed data collection efforts.

Intermediate statistics such as regression require a fairly extensive knowledge of the assumptions on which they are based and the consequences of and remedies for their violation. Equally important, they require special effort on the part of the policy analyst to translate their results into "stories" that others—not trained in their use—can readily understand.

- A great deal of what we think we know actually rests on assumptions for which we have no solid research.

- Two data points do not make a trend.

- Effective presentations begin with your message and follow with selected evidence to support it.

- Statistics may reveal the truth, but simple, unexpected, concrete, credible, emotional stories are more likely to persuade others to act or think differently (Heath & Heath, 2007).

- Statistics are often best communicated after you transform them into scales or analogies around which your audiences can wrap their hearts and minds. People have a harder time imagining what $2.3 trillion is (the costs of health care in the United States in 2007) than $7,600 (the average costs of health care for every man, woman, and child).

- Every research design has its strengths and weaknesses. Knowing them will help you use these tools to advance an argument, make a better decision, disarm an opponent, or spot a statistical charlatan.

- There are appropriate statistical tools for nearly every task. Use a hammer to drive a nail, a screwdriver to turn a screw.

- It's not only okay but sometimes a good idea to transform data and then reanalyze them. Some transformations are indispensable—for example, a percent.

- Nearly every statistic rests on one or more assumptions when making a statement about a process taking place in a population from which you only have drawn a sample (i.e., statistical **inference**). A careful statistician will check to see if these assumptions are met and then modify the data or select a different tool in response to that detective work. An even better statistician will help ensure that the right data are collected in the first place and the results presented in ways that target audiences can understand.

- The field of statistics includes a lot of polysyllabic terms (e.g., **heteroscedasticity**) that may impress friends, family, and colleagues, but get in the way of making points that "stick."

## RHETORIC VERSUS PHILOSOPHY ●

As David Brooks wrote in an op-ed in the *New York Times* of March 2, 2006, philosophy is the search for truth, while rhetoric is the persuasive argumentation of a position. These two pursuits are often in conflict. It is the hope of this book that we can bring them into harmony.

Mark Twain attributes this quote to the British statesman Disraeli: "There are three kinds of lies: lies, damned lies, and statistics." An erudite American statistician, Fred Mosteller, quipped in response: "It's easy to lie with statistics, but it's easier to lie without them." Considerable research in cognitive psychology and decision theory has demonstrated repeatedly that our guts, hearts, and heads play tricks on us that good data and statistics can help protect us from. Our objective here is to learn how to make persuasive and principled statistical arguments and identify those that are not.

Note for those of you reading this text as part of an applied course on statistics: This book is part of a fully integrated instructional package that includes a student workbook (available at http://www.sagepub.com/pearsonsp/) in which you will find weekly exercises. The first one of these exercises is to be completed after reading the first three chapters of this book. This is pretty easy stuff so far. There's no need to stop here. Read on.