

What's the Point?

I've noticed a curious phenomenon. Students will complain that statistics is confusing and irrelevant. Then the same students will leave the classroom and happily talk over lunch about batting averages (during the summer) or the windchill factor (during the winter) or grade point averages (always). They will recognize that the National Football League's "passer rating"—a statistic that condenses a quarterback's performance into a single number—is a somewhat flawed and arbitrary measure of a quarterback's game day performance. The same data (completion rate, average yards per pass attempt, percentage of touchdown passes per pass attempt, and interception rate) could be combined in a different way, such as giving greater or lesser weight to any of those inputs, to generate a different but equally credible measure of performance. Yet anyone who has watched football recognizes that it's handy to have a single number that can be used to encapsulate a quarterback's performance.

Is the quarterback rating perfect? No. Statistics rarely offers a single "right" way of doing anything. Does it provide meaningful information in an easily accessible way? Absolutely. It's a nice tool for making a quick comparison between the performances of two quarterbacks on a given day. I am a Chicago Bears fan. During the 2011 playoffs, the Bears played the Packers; the Packers won. There are a lot of ways I could describe that game, including pages and pages of analysis and raw data. But here is a more succinct analysis. Chicago Bears quarterback Jay Cutler had a passer rating of 31.8. In contrast, Green Bay quarterback Aaron Rodgers had a passer rating of 55.4. Similarly, we can compare Jay Cutler's performance to that in a game earlier in the season against Green Bay, when he had a passer rating of 85.6. That tells you a lot of what you need to know in order to understand why the Bears beat the Packers earlier in the season but lost to them in the playoffs.

That is a very helpful synopsis of what happened on the field. Does it simplify things? Yes, that is both the strength and the weakness of any descriptive statistic. One number tells you that Jay Cutler was outgunned by Aaron Rodgers

in the Bears' playoff loss. On the other hand, that number won't tell you whether a quarterback had a bad break, such as throwing a perfect pass that was bobbled by the receiver and then intercepted, or whether he "stepped up" on certain key plays (since every completion is weighted the same, whether it is a crucial third down or a meaningless play at the end of the game), or whether the defense was terrible. And so on.

The curious thing is that the same people who are perfectly comfortable discussing statistics in the context of sports or the weather or grades will seize up with anxiety when a researcher starts to explain something like the Gini index, which is a standard tool in economics for measuring income inequality. I'll explain what the Gini index is in a moment, but for now *the most important thing to recognize is that the Gini index is just like the passer rating*. It's a handy tool for collapsing complex information into a single number. As such, it has the strengths of most descriptive statistics, namely that it provides an easy way to compare the income distribution in two countries, or in a single country at different points in time.

The Gini index measures how evenly wealth (or income) is shared within a country on a scale from zero to one. The statistic can be calculated for wealth or for annual income, and it can be calculated at the individual level or at the household level. (All of these statistics will be highly correlated but not identical.) The Gini index, like the passer rating, has no intrinsic meaning; it's a tool for comparison. A country in which every household had identical wealth would have a Gini index of zero. By contrast, a country in which a single household held the country's entire wealth would have a Gini index of one. As you can probably surmise, the closer a country is to one, the more unequal its distribution of wealth. The United States has a Gini index of .45, according to the Central Intelligence Agency (a great collector of statistics, by the way).¹ So what?

Once that number is put into context, it can tell us a lot. For example, Sweden has a Gini index of .23. Canada's is .32. China's is .42. Brazil's is .54. South Africa's is .65.* As we look across those numbers, we get a sense of where the United States falls relative to the rest of the world when it comes to income inequality. We can also compare different points in time. The Gini index for the United States was .41 in 1997 and grew to .45 over the next decade. (The most recent CIA data are for 2007.) This tells us in an objective way that while the United States grew richer over that period of time, the distribution of wealth grew more unequal. Again, we can compare the changes in the Gini index across

countries over roughly the same time period. Inequality in Canada was basically unchanged over the same stretch. Sweden has had significant economic growth over the past two decades, but the Gini index in Sweden actually fell from .25 in 1992 to .23 in 2005, meaning that Sweden grew richer *and* more equal over that period.

Is the Gini index the perfect measure of inequality? Absolutely not—just as the passer rating is not a perfect measure of quarterback performance. But it certainly gives us some valuable information on a socially significant phenomenon in a convenient format.

We have also slowly backed our way into answering the question posed in the chapter title: What is the point? The point is that statistics helps us process data, which is really just a fancy name for information. Sometimes the data are trivial in the grand scheme of things, as with sports statistics. Sometimes they offer insight into the nature of human existence, as with the Gini index.

But, as any good infomercial would point out, *That's not all!* Hal Varian, chief economist at Google, told the *New York Times* that being a statistician will be “the sexy job” over the next decade.² I'll be the first to concede that economists sometimes have a warped definition of “sexy.” Still, consider the following disparate questions:

How can we catch schools that are cheating on their standardized tests?

How does Netflix know what kind of movies you like?

How can we figure out what substances or behaviors cause cancer, given that we cannot conduct cancer-causing experiments on humans?

Does praying for surgical patients improve their outcomes?

Is there really an economic benefit to getting a degree from a highly selective college or university?

What is causing the rising incidence of autism?

Statistics can help answer these questions (or, we hope, can soon). The world is producing more and more data, ever faster and faster. Yet, as the *New York Times* has noted, “Data is merely the raw material of knowledge.”^{3*} Statistics is the most powerful tool we have for using information to some meaningful end, whether that is identifying underrated baseball players or paying teachers more fairly. Here is a quick tour of how statistics can bring meaning to raw data.

Description and Comparison

A bowling score is a descriptive statistic. So is a batting average. Most American sports fans over the age of five are already conversant in the field of descriptive statistics. We use numbers, in sports and everywhere else in life, to summarize information. How good a baseball player was Mickey Mantle? He was a career .298 hitter. To a baseball fan, that is a meaningful statement, which is remarkable when you think about it, because it encapsulates an eighteen-season career.⁴ (There is, I suppose, something mildly depressing about having one's lifework collapsed into a single number.) Of course, baseball fans have also come to recognize that descriptive statistics other than batting average may better encapsulate a player's value on the field.

We evaluate the academic performance of high school and college students by means of a grade point average, or GPA. A letter grade is assigned a point value; typically an A is worth 4 points, a B is worth 3, a C is worth 2, and so on. By graduation, when high school students are applying to college and college students are looking for jobs, the grade point average is a handy tool for assessing their academic potential. Someone who has a 3.7 GPA is clearly a stronger student than someone at the same school with a 2.5 GPA. That makes it a nice descriptive statistic. It's easy to calculate, it's easy to understand, and it's easy to compare across students.

But it's not perfect. The GPA does not reflect the difficulty of the courses that different students may have taken. How can we compare a student with a 3.4 GPA in classes that appear to be relatively nonchallenging and a student with a 2.9 GPA who has taken calculus, physics, and other tough subjects? I went to a high school that attempted to solve this problem by giving extra weight to difficult classes, so that an A in an "honors" class was worth five points instead of the usual four. This caused its own problems. My mother was quick to recognize the distortion caused by this GPA "fix." For a student taking a lot of honors classes (me), any A in a nonhonors course, such as gym or health education, would actually pull my GPA down, even though it is impossible to do better than an A in those classes. As a result, my parents forbade me to take driver's education in high school, lest even a perfect performance diminish my chances of getting into a competitive college and going on to write popular books. Instead, they paid to send me to a private driving school, at nights over the summer.

Was that insane? Yes. But one theme of this book will be that an overreliance on any descriptive statistic can lead to misleading conclusions, or cause undesirable behavior. My original draft of that sentence used the phrase

“oversimplified descriptive statistic,” but I struck the word “oversimplified” because it’s redundant. Descriptive statistics exist to simplify, which always implies some loss of nuance or detail. Anyone working with numbers needs to recognize as much.

Inference

How many homeless people live on the streets of Chicago? How often do married people have sex? These may seem like wildly different kinds of questions; in fact, they both can be answered (not perfectly) by the use of basic statistical tools. One key function of statistics is to use the data we have to make informed conjectures about larger questions for which we do not have full information. In short, we can use data from the “known world” to make informed inferences about the “unknown world.”

Let’s begin with the homeless question. It is expensive and logistically difficult to count the homeless population in a large metropolitan area. Yet it is important to have a numerical estimate of this population for purposes of providing social services, earning eligibility for state and federal revenues, and gaining congressional representation. One important statistical practice is sampling, which is the process of gathering data for a small area, say, a handful of census tracts, and then using those data to make an informed judgment, or inference, about the homeless population for the city as a whole. Sampling requires far less resources than trying to count an entire population; done properly, it can be every bit as accurate.

A political poll is one form of sampling. A research organization will attempt to contact a sample of households that are broadly representative of the larger population and ask them their views about a particular issue or candidate. This is obviously much cheaper and faster than trying to contact every household in an entire state or country. The polling and research firm Gallup reckons that a methodologically sound poll of 1,000 households will produce roughly the same results as a poll that attempted to contact every household in America.

That’s how we figured out how often Americans are having sex, with whom, and what kind. In the mid-1990s, the National Opinion Research Center at the University of Chicago carried out a remarkably ambitious study of American sexual behavior. The results were based on detailed surveys conducted in person with a large, representative sample of American adults. If you read on, Chapter

10 will tell you what they learned. *How many other statistics books can promise you that?*

Assessing Risk and Other Probability-Related Events

Casinos make money in the long run—always. That does not mean that they are making money at any given moment. When the bells and whistles go off, some high roller has just won thousands of dollars. The whole gambling industry is built on games of chance, meaning that the outcome of any particular roll of the dice or turn of the card is uncertain. At the same time, the underlying probabilities for the relevant events—drawing 21 at blackjack or spinning red in roulette—are known. When the underlying probabilities favor the casinos (as they always do), we can be increasingly certain that the “house” is going to come out ahead as the number of bets wagered gets larger and larger, even as those bells and whistles keep going off.

This turns out to be a powerful phenomenon in areas of life far beyond casinos. Many businesses must assess the risks associated with assorted adverse outcomes. They cannot make those risks go away entirely, just as a casino cannot guarantee that you won’t win every hand of blackjack that you play. However, any business facing uncertainty can manage these risks by engineering processes so that the probability of an adverse outcome, anything from an environmental catastrophe to a defective product, becomes acceptably low. Wall Street firms will often evaluate the risks posed to their portfolios under different scenarios, with each of those scenarios weighted based on its probability. The financial crisis of 2008 was precipitated in part by a series of market events that had been deemed extremely unlikely, as if every player in a casino drew blackjack all night. I will argue later in the book that these Wall Street models were flawed and that the data they used to assess the underlying risks were too limited, but the point here is that any model to deal with risk must have probability as its foundation.

When individuals and firms cannot make unacceptable risks go away, they seek protection in other ways. The entire insurance industry is built upon charging customers to protect them against some adverse outcome, such as a car crash or a house fire. The insurance industry does not make money by eliminating these events; cars crash and houses burn every day. Sometimes cars even crash into houses, causing them to burn. Instead, the insurance industry

makes money by charging premiums that are more than sufficient to pay for the expected payouts from car crashes and house fires. (The insurance company may also try to lower its expected payouts by encouraging safe driving, fences around swimming pools, installation of smoke detectors in every bedroom, and so on.)

Probability can even be used to catch cheats in some situations. The firm Caveon Test Security specializes in what it describes as “data forensics” to find patterns that suggest cheating.⁵ For example, the company (which was founded by a former test developer for the SAT) will flag exams at a school or test site on which the number of identical *wrong answers* is highly unlikely, usually a pattern that would happen by chance less than one time in a million. The mathematical logic stems from the fact that we cannot learn much when a large group of students all answer a question correctly. That’s what they are supposed to do; they could be cheating, or they could be smart. But when those same test takers get an answer wrong, they should not all consistently have *the same wrong answer*. If they do, it suggests that they are copying from one another (or sharing answers via text). The company also looks for exams in which a test taker does significantly better on hard questions than on easy questions (suggesting that he or she had answers in advance) and for exams on which the number of “wrong to right” erasures is significantly higher than the number of “right to wrong” erasures (suggesting that a teacher or administrator changed the answer sheets after the test).

Of course, you can see the limitations of using probability. A large group of test takers might have the same wrong answers by coincidence; in fact, the more schools we evaluate, the more likely it is that we will observe such patterns just as a matter of chance. A statistical anomaly does not prove wrongdoing. Delma Kinney, a fifty-year-old Atlanta man, won \$1 million in an instant lottery game in 2008 and then another \$1 million in an instant game in 2011.⁶ The probability of that happening to the same person is somewhere in the range of 1 in 25 trillion. We cannot arrest Mr. Kinney for fraud on the basis of that calculation alone (though we might inquire whether he has any relatives who work for the state lottery). Probability is one weapon in an arsenal that requires good judgment.

Identifying Important Relationships (Statistical Detective Work)

Does smoking cigarettes cause cancer? We have an answer for that question—but the process of answering it was not nearly as straightforward as one might think. The scientific method dictates that if we are testing a scientific hypothesis, we should conduct a controlled experiment in which the variable of interest (e.g., smoking) is the only thing that differs between the experimental group and the control group. If we observe a marked difference in some outcome between the two groups (e.g., lung cancer), we can safely infer that the variable of interest is what caused that outcome. We cannot do that kind of experiment on humans. If our working hypothesis is that smoking causes cancer, it would be unethical to assign recent college graduates to two groups, smokers and nonsmokers, and then see who has cancer at the twentieth reunion. (We can conduct controlled experiments on humans when our hypothesis is that a new drug or treatment may improve their health; we cannot knowingly expose human subjects when we expect an adverse outcome.)*

Now, you might point out that we do not need to conduct an ethically dubious experiment to observe the effects of smoking. Couldn't we just skip the whole fancy methodology and compare cancer rates at the twentieth reunion between those who have smoked since graduation and those who have not?

No. Smokers and nonsmokers are likely to be different in ways other than their smoking behavior. For example, smokers may be more likely to have other habits, such as drinking heavily or eating badly, that cause adverse health outcomes. If the smokers are particularly unhealthy at the twentieth reunion, we would not know whether to attribute this outcome to smoking or to other unhealthy things that many smokers happen to do. We would also have a serious problem with the data on which we are basing our analysis. Smokers who have become seriously ill with cancer are less likely to attend the twentieth reunion. (The dead smokers definitely won't show up.) As a result, any analysis of the health of the attendees at the twentieth reunion (related to smoking or anything else) will be seriously flawed by the fact that the healthiest members of the class are the most likely to show up. The further the class gets from graduation, say, a fortieth or a fiftieth reunion, the more serious this bias will be.

We cannot treat humans like laboratory rats. As a result, statistics is a lot like good detective work. The data yield clues and patterns that can ultimately lead to meaningful conclusions. You have probably watched one of those impressive police procedural shows like *CSI: New York* in which very attractive detectives and forensic experts pore over minute clues—DNA from a cigarette butt, teeth marks on an apple, a single fiber from a car floor mat—and then use the

evidence to catch a violent criminal. The appeal of the show is that these experts do not have the conventional evidence used to find the bad guy, such as an eyewitness or a surveillance videotape. So they turn to scientific inference instead. Statistics does basically the same thing. The data present unorganized clues—the crime scene. Statistical analysis is the detective work that crafts the raw data into some meaningful conclusion.

After Chapter 11, you will appreciate the television show I hope to pitch: *CSI: Regression Analysis*, which would be only a small departure from those other action-packed police procedurals. Regression analysis is the tool that enables researchers to isolate a relationship between two variables, such as smoking and cancer, while holding constant (or “controlling for”) the effects of other important variables, such as diet, exercise, weight, and so on. When you read in the newspaper that eating a bran muffin every day will reduce your chances of getting colon cancer, you need not fear that some unfortunate group of human experimental subjects has been force-fed bran muffins in the basement of a federal laboratory somewhere while the control group in the next building gets bacon and eggs. Instead, researchers will gather detailed information on thousands of people, including how frequently they eat bran muffins, and then use regression analysis to do two crucial things: (1) quantify the association observed between eating bran muffins and contracting colon cancer (e.g., a hypothetical finding that people who eat bran muffins have a 9 percent lower incidence of colon cancer, controlling for other factors that may affect the incidence of the disease); and (2) quantify the likelihood that the association between bran muffins and a lower rate of colon cancer observed in this study is merely a coincidence—a quirk in the data for this sample of people—rather than a meaningful insight about the relationship between diet and health.

Of course, *CSI: Regression Analysis* will star actors and actresses who are much better looking than the academics who typically pore over such data. These hotties (all of whom would have PhDs, despite being only twenty-three years old) would study large data sets and use the latest statistical tools to answer important social questions: What are the most effective tools for fighting violent crime? What individuals are most likely to become terrorists? Later in the book we will discuss the concept of a “statistically significant” finding, which means that the analysis has uncovered an association between two variables that is not likely to be the product of chance alone. For academic researchers, this kind of statistical finding is the “smoking gun.” On *CSI: Regression Analysis*, I envision a researcher working late at night in the computer lab because of her daytime

commitment as a member of the U.S. Olympic beach volleyball team. When she gets the printout from her statistical analysis, she sees exactly what she has been looking for: a large and statistically significant relationship in her data set between some variable that she had hypothesized might be important and the onset of autism. She must share this breakthrough immediately!

The researcher takes the printout and runs down the hall, slowed somewhat by the fact that she is wearing high heels and a relatively small, tight black skirt. She finds her male partner, who is inexplicably fit and tan for a guy who works fourteen hours a day in a basement computer lab, and shows him the results. He runs his fingers through his neatly trimmed goatee, grabs his Glock 9-mm pistol from the desk drawer, and slides it into the shoulder holster beneath his \$5,000 Hugo Boss suit (also inexplicable given his starting academic salary of \$38,000 a year). Together the regression analysis experts walk briskly to see their boss, a grizzled veteran who has overcome failed relationships and a drinking problem .

..

Okay, you don't have to buy into the television drama to appreciate the importance of this kind of statistical research. Just about every social challenge that we care about has been informed by the systematic analysis of large data sets. (In many cases, gathering the relevant data, which is expensive and time-consuming, plays a crucial role in this process as will be explained in Chapter 7.) I may have embellished my characters in *CSI: Regression Analysis* but not the kind of significant questions they could examine. There is an academic literature on terrorists and suicide bombers—a subject that would be difficult to study by means of human subjects (or lab rats for that matter). One such book, *What Makes a Terrorist*, was written by one of my graduate school statistics professors. The book draws its conclusions from data gathered on terrorist attacks around the world. A sample finding: Terrorists are not desperately poor, or poorly educated. The author, Princeton economist Alan Krueger, concludes, “Terrorists tend to be drawn from well-educated, middle-class or high-income families.”⁷

Why? Well, that exposes one of the limitations of regression analysis. We can isolate a strong association between two variables by using statistical analysis, but we cannot necessarily explain why that relationship exists, and in some cases, we cannot know for certain that the relationship is causal, meaning that a change in one variable is really causing a change in the other. In the case of terrorism, Professor Krueger hypothesizes that since terrorists are motivated by political goals, those who are most educated and affluent have the strongest

incentive to change society. These individuals may also be particularly rankled by suppression of freedom, another factor associated with terrorism. In Krueger's study, countries with high levels of political repression have more terrorist activity (holding other factors constant).

This discussion leads me back to the question posed by the chapter title: What is the point? The point is not to do math, or to dazzle friends and colleagues with advanced statistical techniques. The point is to learn things that inform our lives.

Lies, Damned Lies, and Statistics

Even in the best of circumstances, statistical analysis rarely unveils “the truth.” We are usually building a circumstantial case based on imperfect data. As a result, there are numerous reasons that intellectually honest individuals may disagree about statistical results or their implications. At the most basic level, we may disagree on the question that is being answered. Sports enthusiasts will be arguing for all eternity over “the best baseball player ever” because there is no objective definition of “best.” Fancy descriptive statistics can inform this question, but they will never answer it definitively. As the next chapter will point out, more socially significant questions fall prey to the same basic challenge. What is happening to the economic health of the American middle class? That answer depends on how one defines both “middle class” and “economic health.”

There are limits on the data we can gather and the kinds of experiments we can perform. Alan Krueger's study of terrorists did not follow thousands of youth over multiple decades to observe which of them evolved into terrorists. It's just not possible. Nor can we create two identical nations—except that one is highly repressive and the other is not—and then compare the number of suicide bombers that emerge in each. Even when we can conduct large, controlled experiments on human beings, they are neither easy nor cheap. Researchers did a large-scale study on whether or not prayer reduces postsurgical complications, which was one of the questions raised earlier in this chapter. *That study cost \$2.4 million.* (For the results, you'll have to wait until Chapter 13.)

Secretary of Defense Donald Rumsfeld famously said, “You go to war with the army you have—not the army you might want or wish to have at a later time.” Whatever you may think of Rumsfeld (and the Iraq war that he was explaining), that aphorism applies to research, too. We conduct statistical

analysis using the best data and methodologies and resources available. The approach is not like addition or long division, in which the correct technique yields the “right” answer and a computer is always more precise and less fallible than a human. Statistical analysis is more like good detective work (hence the commercial potential of *CSI: Regression Analysis*). Smart and honest people will often disagree about what the data are trying to tell us.

But who says that everyone using statistics is smart or honest? As mentioned, this book began as an homage to *How to Lie with Statistics*, which was first published in 1954 and has sold over a million copies. The reality is that you *can* lie with statistics. Or you can make inadvertent errors. In either case, the mathematical precision attached to statistical analysis can dress up some serious nonsense. This book will walk through many of the most common statistical errors and misrepresentations (so that you can recognize them, not put them to use).

So, to return to the title chapter, what is the point of learning statistics?

To summarize huge quantities of data.

To make better decisions.

To answer important social questions.

To recognize patterns that can refine how we do everything from selling diapers to catching criminals.

To catch cheaters and prosecute criminals.

To evaluate the effectiveness of policies, programs, drugs, medical procedures, and other innovations.

And to spot the scoundrels who use these very same powerful tools for nefarious ends.

If you can do all of that while looking great in a Hugo Boss suit or a short black skirt, then you might also be the next star of *CSI: Regression Analysis*.

* The Gini index is sometimes multiplied by 100 to make it a whole number. In that case, the United States would have a Gini Index of 45.

* The word “data” has historically been considered plural (e.g., “The data are very encouraging.”) The singular is “datum,” which would refer to a single data point, such as one person’s response to a single question on a poll. Using the word “data” as a plural noun is a quick way to signal to anyone who does serious research that you are conversant with statistics. That said, many authorities on grammar and many publications, such as the *New York Times*, now accept that “data” can be singular or plural, as the passage that I’ve quoted from the *Times* demonstrates.

* This is a gross simplification of the fascinating and complex field of medical ethics.