CHAPTER 1

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.

---

* 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.

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

---

* 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.

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.)*

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

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*.

CHAPTER 2

Descriptive Statistics

Who was the best baseball player of all time?

Let us ponder for a moment two seemingly unrelated questions: (1) What is happening to the economic health of America's middle class? and (2) Who was the greatest baseball player of all time?

The first question is profoundly important. It tends to be at the core of presidential campaigns and other social movements. The middle class is the heart of America, so the economic well-being of that group is a crucial indicator of the nation's overall economic health. The second question is trivial (in the literal sense of the word), but baseball enthusiasts can argue about it endlessly. What the two questions have in common is that they can be used to illustrate the strengths and limitations of descriptive statistics, which are the numbers and calculations we use to summarize raw data.

If I want to demonstrate that Derek Jeter is a great baseball player, I can sit you down and describe every at bat in every Major League game that he's played. That would be raw data, and it would take a while to digest, given that Jeter has played seventeen seasons with the New York Yankees and taken 9,868 at bats.

Or I can just tell you that at the end of the 2011 season Derek Jeter had a career batting average of .313. That is a descriptive statistic, or a "summary statistic."

The batting average is a gross simplification of Jeter's seventeen seasons. It is easy to understand, elegant in its simplicity—and limited in what it can tell us. Baseball experts have a bevy of descriptive statistics that they consider to be more valuable than the batting average. I called Steve Moyer, president of Baseball Info Solutions (a firm that provides a lot of the raw data for the *Moneyball* types), to ask him, (1) What are the most important statistics for evaluating baseball talent? and (2) Who was the greatest player of all time? I'll share his answer once we have more context.

Meanwhile, let's return to the less trivial subject, the economic health of the middle class. Ideally we would like to find the economic equivalent of a batting average, or something even better. We would like a simple but accurate measure of how the economic well-being of the typical American worker has been changing in recent years. Are the people we define as middle class getting richer, poorer, or just running in place? A reasonable answer—though by no means the "right" answer—would be to calculate the change in per capita income in the United States over the course of a generation, which is roughly thirty years. Per capita income is a simple average: total income divided by the size of the population. By that measure, average income in the United States climbed from $7,787 in 1980 to $26,487 in 2010 (the latest year for which the government has data).[1] Voilà! Congratulations to us.

There is just one problem. My quick calculation is technically correct and yet totally wrong in terms of the question I set out to answer. To begin with, the figures above are not adjusted for inflation. (A per capita income of $7,787 in 1980 is equal to about $19,600 when converted to 2010 dollars.) That's a relatively quick fix. The bigger problem is that the average income in America is not equal to the income of the average American. Let's unpack that clever little phrase.

Per capita income merely takes all of the income earned in the country and divides by the number of people, which tells us absolutely nothing about who is earning how much of that income—in 1980 or in 2010. As the Occupy Wall Street folks would point out, explosive growth in the incomes of the top 1 percent can raise per capita income significantly without putting any more money in the pockets of the other 99 percent.

In other words, average income can go up without helping the average American.

As with the baseball statistic query, I have sought outside expertise on how we ought to measure the health of the American middle class. I asked two prominent labor economists, including President Obama's top economic adviser, what descriptive statistics they would use to assess the economic well-being of a typical American. Yes, you will get that answer, too, once we've taken a quick tour of descriptive statistics to give it more meaning.

From baseball to income, the most basic task when working with data is to summarize a great deal of information. There are some 330 million residents in the United States. A spreadsheet with the name and income history of every American would contain all the information we could ever want about the economic health of the country—yet it would also be so unwieldy as to tell us nothing at all. The irony is that more data can often present less clarity. So we simplify. We perform calculations that reduce a complex array of data into a handful of numbers that describe those data, just as we might encapsulate a complex, multifaceted Olympic gymnastics performance with one number: 9.8.

The good news is that these descriptive statistics give us a manageable and meaningful summary of the underlying phenomenon. That's what this chapter is about. The bad news is that any simplification invites abuse. Descriptive statistics can be like online dating profiles: technically accurate and yet pretty darn misleading.

Suppose you are at work, idly surfing the Web when you stumble across a riveting day-by-day account of Kim Kardashian's failed seventy-two-day marriage to professional basketball player Kris Humphries. You have finished reading about day seven of the marriage when your boss shows up with two enormous files of data. One file has warranty claim information for each of the 57,334 laser printers that your firm sold last year. (For each printer sold, the file documents the number of quality problems that were reported during the warranty period.) The other file has the same information for each of the 994,773 laser printers that your chief competitor sold during the same stretch. Your boss wants to know how your firm's printers compare in terms of quality with the competition.

Fortunately the computer you've been using to read about the Kardashian marriage has a basics statistics package, but where do you begin? Your instincts are probably correct: The first descriptive task is often to find some measure of the "middle" of a set of data, or what statisticians might describe as its "central tendency." What is the typical quality experience for your printers compared with those of the competition? The most basic measure of the "middle" of a distribution is the mean, or average. In this case, we want to know the average number of quality problems per printer sold for your firm and for your competitor. You would simply tally the total number of quality problems reported for all printers during the warranty period and then divide by the total number of printers sold. (Remember, the same printer can have multiple problems while under warranty.) You would do that for each firm, creating an important descriptive statistic: the average number of quality problems per printer sold.

Suppose it turns out that your competitor's printers have an average of 2.8 quality-related problems per printer during the warranty period compared with your firm's average of 9.1 reported defects. That was easy. You've just taken information on a million printers sold by two different companies and distilled it to the essence of the problem: your printers break a lot. Clearly it's time to send a short e-mail to your boss quantifying this quality gap and then get back to day eight of Kim Kardashian's marriage.

Or maybe not. I was deliberately vague earlier when I referred to the "middle" of a distribution. The mean, or average, turns out to have some problems in that regard, namely, that it is prone to distortion by "outliers," which are observations that lie farther from the center. To get your mind around this concept, imagine that ten guys are sitting on bar stools in a middle-class drinking establishment in Seattle; each of these guys earns $35,000 a year, which makes the mean annual income for the group $35,000. Bill Gates walks into the bar with a talking parrot perched on his shoulder. (The parrot has nothing to do with the example, but it kind of spices things up.) Let's assume for the sake of the example that Bill Gates has an annual income of $1 billion. When Bill sits down on the eleventh bar stool, the mean annual income for the bar patrons rises to about $91 million. Obviously none of the original ten drinkers is any richer (though it might be reasonable to expect Bill Gates to buy a round or two). If I were to describe the patrons of this bar as having an average annual income of $91 million, the statement would be both statistically correct and grossly misleading. This isn't a bar where multimillionaires hang out; it's a bar where a bunch of guys with relatively low incomes happen to be sitting next to Bill Gates and his talking parrot. The sensitivity of the mean to outliers is why we should not gauge the economic health of the American middle class by looking at per capita income. Because there has been explosive growth in incomes at the top end of the distribution—CEOs, hedge fund managers, and athletes like Derek Jeter—the average income in the United States could be heavily skewed by the megarich, making it look a lot like the bar stools with Bill Gates at the end.

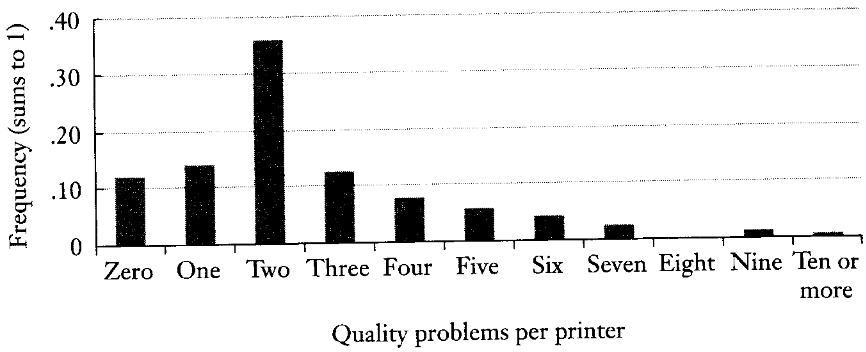
For this reason, we have another statistic that also signals the "middle" of a distribution, albeit differently: the median. The median is the point that divides a distribution in half, meaning that half of the observations lie above the median and half lie below. (If there is an even number of observations, the median is the midpoint between the two middle observations.) If we return to the bar stool example, the median annual income for the ten guys originally sitting in the bar is $35,000. When Bill Gates walks in with his parrot and perches on a stool, the median annual income for the eleven of them is still $35,000. If you literally envision lining up the bar patrons on stools in ascending order of their incomes, the income of the guy sitting on the sixth stool represents the median income for the group. If Warren Buffett comes in and sits down on the twelfth stool next to Bill Gates, the median still does not change.*

---

* With twelve bar patrons, the median would be the midpoint between the income of the guy on the sixth stool and the income of the guy on the seventh stool. Since they both make $35,000, the median is $35,000. If one made $35,000 and the other made $36,000, the median for the whole group would be $35,500.

For distributions without serious outliers, the median and the mean will be similar. I've included a hypothetical summary of the quality data for the competitor's printers. In particular, I've laid out the data in what is known as a frequency distribution. The number of quality problems per printer is arrayed along the bottom; the height of each bar represents the percentages of printers sold with that number of quality problems. For example, 36 percent of the competitor's printers had two quality defects during the warranty period. Because the distribution includes all possible quality outcomes, including zero defects, the proportions must sum to 1 (or 100 percent).

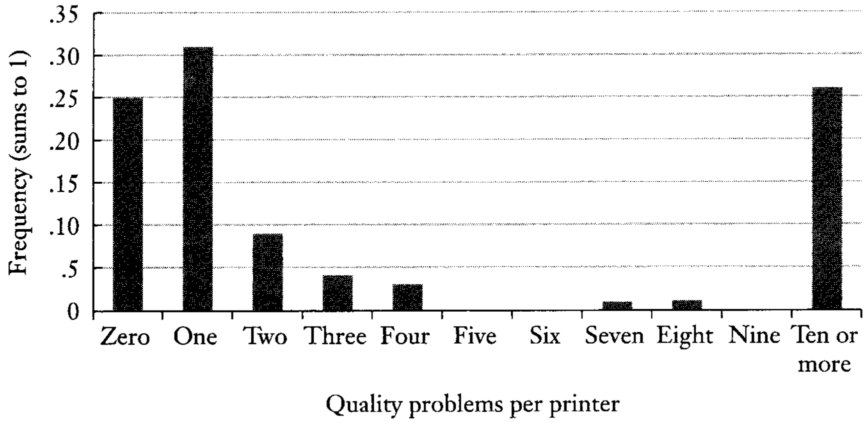
Frequency Distribution of Quality Complaints for Competitor's Printers



Because the distribution is nearly symmetrical, the mean and median are relatively close to one another. The distribution is slightly skewed to the right by the small number of printers with many reported quality defects. These outliers move the mean slightly rightward but have no impact on the median. Suppose that just before you dash off the quality report to your boss you decide to calculate the *median* number of quality problems for your firm's printers and the competition's. With a few keystrokes, you get the result. The median number of quality complaints for the competitor's printers is 2; the median number of quality complaints for your company's printers is 1.

Huh? Your firm's median number of quality complaints per printer is actually *lower* than your competitor's. Because the Kardashian marriage is getting monotonous, and because you are intrigued by this finding, you print a frequency distribution for your own quality problems.

Frequency Distribution of Quality Complaints at Your Company



What becomes clear is that your firm does not have a uniform quality problem; you have a "lemon" problem; a small number of printers have a huge number of quality complaints. These outliers inflate the mean but not the median. More important from a production standpoint, you do not need to retool the whole manufacturing process; you need only figure out where the egregiously low-quality printers are coming from and fix that.*

Neither the median nor the mean is hard to calculate; the key is determining which measure of the "middle" is more accurate in a particular situation (a phenomenon that is easily exploited). Meanwhile, the median has some useful relatives. As we've already discussed, the median divides a distribution in half. The distribution can be further divided

* Manufacturing update: It turns out that nearly all of the defective printers were being manufactured at a plant in Kentucky where workers had stripped parts off the assembly line in order to build a bourbon distillery. Both the perpetually drunk employees and the random missing pieces on the assembly line appear to have compromised the quality of the printers being produced there.

into quarters, or quartiles. The first quartile consists of the bottom 25 percent of the observations; the second quartile consists of the next 25 percent of the observations; and so on. Or the distribution can be divided into deciles, each with 10 percent of the observations. (If your income is in the top decile of the American income distribution, you would be earning more than 90 percent of your fellow workers.) We can go even further and divide the distribution into hundredths, or percentiles. Each percentile represents 1 percent of the distribution, so that the 1st percentile represents the bottom 1 percent of the distribution and the 99th percentile represents the top 1 percent of the distribution.

The benefit of these kinds of descriptive statistics is that they describe where a particular observation lies compared with everyone else. If I tell you that your child scored in the 3rd percentile on a reading comprehension test, you should know immediately that the family should be logging more time at the library. You don't need to know anything about the test itself, or the number of questions that your child got correct. The percentile score provides a ranking of your child's score relative to that of all the other test takers. If the test was easy, then most test takers will have a high number of answers correct, but your child will have fewer correct than most of the others. If the test was extremely difficult, then all the test takers will have a low number of correct answers, but your child's score will be lower still.

Here is a good point to introduce some useful terminology. An "absolute" score, number, or figure has some intrinsic meaning. If I shoot 83 for eighteen holes of golf, that is an absolute figure. I may do that on a day that is 58 degrees, which is also an absolute figure. Absolute figures can usually be interpreted without any context or additional information. When I tell you that I shot 83, you don't need to know what other golfers shot that day in order to evaluate my performance. (The exception might be if the conditions are particularly awful, or if the course is especially difficult or easy.) If I place ninth in the golf tournament, that is a relative statistic. A "relative" value or figure has meaning only in comparison to something else, or in some broader context, such as compared with the eight golfers who shot better than I did. Most standardized tests produce results that have meaning only as a relative statistic. If I tell you that a third grader in an Illinois elementary school scored 43 out of 60 on the mathematics portion of the Illinois State Achievement Test, that absolute score doesn't have much meaning. But when I convert it to a percentile—meaning that I put that raw score into a distribution with the math scores for all other Illinois third graders—then it acquires a great deal of meaning. If 43 correct answers falls into the 83rd percentile, then this student is doing better than most of his peers statewide. If he's in the 8th percentile, then he's really struggling. In this case, the percentile (the relative score) is more meaningful than the number of correct answers (the absolute score).

Another statistic that can help us describe what might otherwise be a jumble of numbers is the standard deviation, which is a measure of how dispersed the data are from their mean. In other words, how spread out are the observations? Suppose I collected data on the weights of 250 people on an airplane headed for Boston, and I also collected the weights of a sample of 250 qualifiers for the Boston Marathon. Now assume that the mean weight for both groups is roughly the same, say 155 pounds. Anyone who has been squeezed into a row on a crowded flight, fighting for the armrest, knows that many people on a typical commercial flight weigh more than 155 pounds. But you may recall from those same unpleasant, overcrowded flights that there were lots of crying babies and poorly behaved children, all of whom have enormous lung capacity but not much mass. When it comes to calculating the average weight on the flight, the heft of the 320-pound football players on either side of your middle seat is likely offset by the tiny screaming infant across the row and the six-year-old kicking the back of your seat from the row behind.

On the basis of the descriptive tools introduced so far, the weights of the airline passengers and the marathoners are nearly identical. *But they're not.* Yes, the weights of the two groups have roughly the same "middle," but the airline passengers have far more dispersion around that midpoint, meaning that their weights are spread farther from the midpoint. My eight-year-old son might point out that the marathon runners look like they all weigh the same amount, while the airline passengers have some tiny people and some bizarrely large people. The weights of the airline passengers are "more spread out," which is an important attribute when it comes to describing the weights of these two groups. The standard deviation is the descriptive statistic that allows us to assign a single number to this dispersion around the mean. The formulas for calculating the standard deviation and the variance (another common measure of dispersion from which the standard deviation is derived) are included in an appendix at the end of the chapter. For now, let's think about why the measuring of dispersion matters.

Suppose you walk into the doctor's office. You've been feeling fatigued ever since your promotion to head of North American printer quality. Your doctor draws blood, and a few days later her assistant leaves a message on your answering machine to inform you that your HCb2 count (a fictitious blood chemical) is 134. You rush to the Internet and discover that the mean HCb2 count for a person your age is 122 (and the median is about the same). Holy crap! If you're like me, you would finally draft a will. You'd write tearful letters to your parents, spouse, children, and close friends. You might take up skydiving or try to write a novel very fast. You would send your boss a hastily composed e-mail comparing him to a certain part of the human anatomy—IN ALL CAPS.

None of these things may be necessary (and the e-mail to your boss could turn out very badly). When you call the doctor's office back to arrange for your hospice care, the physician's assistant informs you that your count is within the normal range. But how could that be? "My count is 12 points higher than average!" you yell repeatedly into the receiver.

"The standard deviation for the HCb2 count is 18," the technician informs you curtly.

What the heck does that mean?

There is natural variation in the HCb2 count, as there is with most biological phenomena (e.g., height). While the mean count for the fake chemical might be 122, plenty of healthy people have counts that are higher or lower. The danger arises only when the HCb2 count gets excessively high or low. So how do we figure out what "excessively" means in this context? As we've already noted, the standard deviation is a measure of dispersion, meaning that it reflects how tightly the observations cluster around the mean. For many typical distributions of data, a high proportion of the observations lie within one standard deviation of the mean (meaning that they are in the range from one standard deviation below the mean to one standard deviation above the mean). To illustrate with a simple example, the mean height for American adult men is 5 feet 10 inches. The standard deviation is roughly 3 inches. A high proportion of adult men are between 5 feet 7 inches and 6 feet 1 inch.

Or, to put it slightly differently, any man in this height range would not be considered abnormally short or tall. Which brings us back to your troubling HCb2 results. Yes, your count is 12 above the mean, but that's less than one standard deviation, which is the blood chemical equivalent of being about 6 feet tall—not particularly unusual. Of course, far fewer observations lie two standard deviations from the mean, and fewer still lie three or four standard deviations away. (In the case of height, an American man who is three standard deviations above average in height would be 6 feet 7 inches or taller.)

Some distributions are more dispersed than others. Hence, the standard deviation of the weights of the 250 airline passengers will be higher than the standard deviation of the weights of the 250 marathon runners. A frequency distribution with the weights of the airline passengers would literally be fatter (more spread out) than a frequency distribution of the weights of the marathon runners. Once we know the mean and standard deviation for any collection of data, we have some serious intellectual traction. For example, suppose I tell you that the mean score on the SAT math test is 500 with a standard deviation of 100. As with height, the bulk of students taking the test will be within one standard deviation of the mean, or between 400 and 600. How many students do you think score 720 or higher? Probably not very many, since that is more than two standard deviations above the mean.

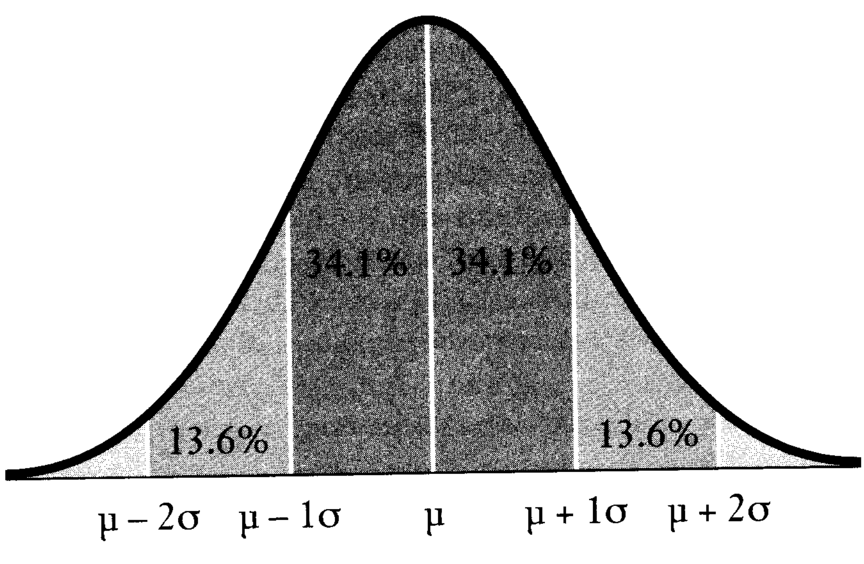
In fact, we can do even better than "not very many." This is a good time to introduce one of the most important, helpful, and common distributions in statistics: the normal distribution. Data that are distributed normally are symmetrical around their mean in a bell shape that will look familiar to you.

The normal distribution describes many common phenomena.

Imagine a frequency distribution describing popcorn popping on a stove top. Some kernels start to pop early, maybe one or two pops per second; after ten or fifteen seconds, the kernels are exploding frenetically. Then gradually the number of kernels popping per second fades away at roughly the same rate at which the popping began. The heights of American men are distributed more or less normally, meaning that they are roughly symmetrical around the mean of 5 feet 10 inches. Each SAT test is specifically designed to produce a normal distribution of scores with mean 500 and standard deviation of 100. According to the *Wall Street Journal*, Americans even tend to park in a normal distribution at shopping malls; most cars park directly opposite the mall entrance—the "peak" of the normal curve—with "tails" of cars going off to the right and left of the entrance.

The beauty of the normal distribution—its Michael Jordan power, finesse, and elegance—comes from the fact that we know by definition exactly what proportion of the observations in a normal distribution lie within one standard deviation of the mean (68.2 percent), within two standard deviations of the mean (95.4 percent), within three standard deviations (99.7 percent), and so on. This may sound like trivia. In fact, it is the foundation on which much of statistics is built. We will come back to this point in much great depth later in the book.

The Normal Distribution



The mean is the middle line which is often represented by the Greek letter μ. The standard deviation is often represented by the Greek letter σ. Each band represents one standard deviation.

Descriptive statistics are often used to compare two figures or quantities. I'm one inch taller than my brother; today's temperature is nine degrees above the historical average for this date; and so on. Those comparisons make sense because most of us recognize the scale of the units involved. One inch does not amount to much when it comes to a person's height, so you can infer that my brother and I are roughly the same height. Conversely, nine degrees is a significant temperature deviation in just about any climate at any time of year, so nine degrees above average makes for a day that is much hotter than usual. But suppose that I told you that Granola Cereal A contains 31 milligrams more sodium than Granola Cereal B. Unless you know an awful lot about sodium (and the serving sizes for granola cereal), that statement is not going to be particularly informative. Or what if I told you that my cousin Al earned $53,000 less this year than last year? Should we be worried about Al? Or is he a hedge fund manager for whom $53,000 is a rounding error in his annual compensation?

In both the sodium and the income examples, we're missing context. The easiest way to give meaning to these relative comparisons is by using percentages. It *would* mean something if I told you that Granola Bar A has 50 percent more sodium than Granola Bar B, or that Uncle Al's income fell 47 percent last year. Measuring change as a percentage gives us some sense of scale.

You probably learned how to calculate percentages in fourth grade and will be tempted to skip the next few paragraphs. Fair enough. But first do one simple exercise for me. Assume that a department store is selling a dress for $100. The assistant manager marks down all merchandise by 25 percent. But then that assistant manager is fired for hanging out in a bar with Bill Gates,* and the new assistant manager raises all prices by

* Remarkably, this person was one of the ten people with annual incomes of $35,000 who were sitting on bar stools when Bill Gates walked in with his parrot. Go figure!

25 percent. What is the final price of the dress? If you said (or thought) $100, then you had better not skip any paragraphs.

The final price of the dress is actually $93.75. This is not merely a fun parlor trick that will win you applause and adulation at cocktail parties. Percentages are useful—but also potentially confusing or even deceptive. The formula for calculating a percentage difference (or change) is the following: (new figure – original figure)/original figure. The numerator (the part on the top of the fraction) gives us the size of the change in absolute terms; the denominator (the bottom of the fraction) is what puts this change in context by comparing it with our starting point. At first, this seems straightforward, as when the assistant store manager cuts the price of the $100 dress by 25 percent. Twenty-five percent of the original $100 price is $25; that's the discount, which takes the price down to $75. You can plug the numbers into the formula above and do some simple manipulation to get to the same place: ($100 – $75)/$100 = .25, or 25 percent.

The dress is selling for $75 when the new assistant manager demands that the price be raised 25 percent. That's where many of the people reading this paragraph probably made a mistake. The 25 percent markup is calculated as a percentage of the dress's new reduced price, which is $75. The increase will be .25($75), or $18.75, which is how the final price ends up at $93.75 (and not $100). The point is that a percentage change always gives the value of some figure *relative to something else*. Therefore, we had better understand what that something else is.

I once invested some money in a company that my college roommate started. Since it was a private venture, there were no requirements as to what information had to be provided to shareholders. A number of years went by without any information on the fate of my investment; my former roommate was fairly tight-lipped on the subject. Finally, I received a letter in the mail informing me that the firm's profits were 46 percent higher than the year before. There was no information on the size of those profits in absolute terms, meaning that I still had absolutely no idea how my investment was performing. Suppose that last year the firm earned 27 cents—essentially nothing. This year the firm earned 39 cents—also essentially nothing. Yet the company's profits grew from 27 cents to 39 cents, which is technically a 46 percent increase. Obviously the shareholder letter would have been more of a downer if it pointed out that the firm's cumulative profits over two years were less than the cost of a cup of Starbucks coffee.

To be fair to my roommate, he eventually sold the company for hundreds of millions of dollars, earning me a 100 percent return on my investment. (Since you have no idea how much I invested, you also have no idea how much money I made—which reinforces my point here very nicely!)

Let me make one additional distinction. Percentage change must not be confused with a change in percentage points. Rates are often expressed in percentages. The sales tax rate in Illinois is 6.75 percent. I pay my agent 15 percent of my book royalties. These rates are levied against some quantity, such as income in the case of the income tax rate. Obviously the rates can go up or down; less intuitively, the *changes* in the rates can be described in vastly dissimilar ways. The best example of this was a recent change in the Illinois personal income tax, which was raised from 3 percent to 5 percent. There are two ways to express this tax change, both of which are technically accurate. The Democrats, who engineered this tax increase, pointed out (correctly) that the state income tax *rate* was increased by *2 percentage points* (from 3 percent to 5 percent). The Republicans pointed out (also correctly) that the state income tax had been raised by *67 percent*. [This is a handy test of the formula from a few paragraphs back: $(5 - 3)/3 = 2/3$, which rounds up to 67 percent.]

The Democrats focused on the absolute change in the tax rate; Republicans focused on the percentage change in the tax burden. As noted, both descriptions are technically correct, though I would argue that the Republican description more accurately conveys the impact of the tax change, since what I'm going to have to pay to the government—the amount that I care about, as opposed to the way it is calculated—really has gone up by 67 percent.

Many phenomena defy perfect description with a single statistic. Suppose quarterback Aaron Rodgers throws for 365 yards but no touchdowns.

Meanwhile, Peyton Manning throws for a meager 127 yards but three touchdowns. Manning generated more points, but presumably Rodgers set up touchdowns by marching his team down the field and keeping the other team's offense off the field. Who played better? In Chapter 1, I discussed the NFL passer rating, which is the league's reasonable attempt to deal with this statistical challenge. The passer rating is an example of an index, which is a descriptive statistic made up of other descriptive statistics. Once these different measures of performance are consolidated into a single number, that statistic can be used to make comparisons, such as ranking quarterbacks on a particular day, or even over a whole career. If baseball had a similar index, then the question of the best player ever would be solved. Or would it?

The advantage of any index is that it consolidates lots of complex information into a single number. We can then rank things that otherwise defy simple comparison—anything from quarterbacks to colleges to beauty pageant contestants. In the Miss America pageant, the overall winner is a combination of five separate competitions: personal interview, swimsuit, evening wear, talent, and onstage question. (Miss Congeniality is voted on separately by the participants themselves.)

Alas, the disadvantage of any index is that it consolidates lots of complex information into a single number. There are countless ways to do that; each has the potential to produce a different outcome. Malcolm Gladwell makes this point brilliantly in a *New Yorker* piece critiquing our compelling need to rank things.[2] (He comes down particularly hard on the college rankings.) Gladwell offers the example of *Car and Driver*'s ranking of three sports cars: the Porsche Cayman, the Chevrolet Corvette, and the Lotus Evora. Using a formula that includes twenty-one different variables, *Car and Driver* ranked the Porsche number one. But Gladwell points out that "exterior styling" counts for only 4 percent of the total score in the *Car and Driver* formula, which seems ridiculously low for a sports car. If styling is given more weight in the overall ranking (25 percent), then the Lotus comes out on top.

But wait. Gladwell also points out that the sticker price of the car gets relatively little weight in the *Car and Driver* formula. If value is weighted more heavily (so that the ranking is based equally on price, exterior styling, and vehicle characteristics), the Chevy Corvette is ranked number one.

Any index is highly sensitive to the descriptive statistics that are cobbled together to build it, and to the weight given to each of those components. As a result, indices range from useful but imperfect tools to complete charades. An example of the former is the United Nations Human Development Index, or HDI. The HDI was created as a measure of economic well-being that is broader than income alone. The HDI uses income as one of its components but also includes measures of life expectancy and educational attainment. The United States ranks eleventh in the world in terms of per capita economic output (behind several oil-rich nations like Qatar, Brunei, and Kuwait) but fourth in the world in human development.[3] It's true that the HDI rankings would change slightly if the component parts of the index were reconfigured, but no reasonable change is going to make Zimbabwe zoom up the rankings past Norway. The HDI provides a handy and reasonably accurate snapshot of living standards around the globe.

Descriptive statistics give us insight into phenomena that we care about. In that spirit, we can return to the questions posed at the beginning of the chapter. Who is the best baseball player of all time? More important for the purposes of this chapter, what descriptive statistics would be most helpful in answering that question? According to Steve Moyer, president of Baseball Info Solutions, the three most valuable statistics (other than age) for evaluating any player who is not a pitcher would be the following:

1. On-base percentage (OBP), sometimes called the on-base average (OBA): Measures the proportion of the time that a player reaches base successfully, including walks (which are not counted in the batting average).
2. Slugging percentage (SLG): Measures power hitting by calculating the total bases reached per at bat. A single counts as 1,

a double is 2, a triple is 3, and a home run is 4. Thus, a batter who hit a single and a triple in five at bats would have a slugging percentage of (1 + 3)/5, or .800.

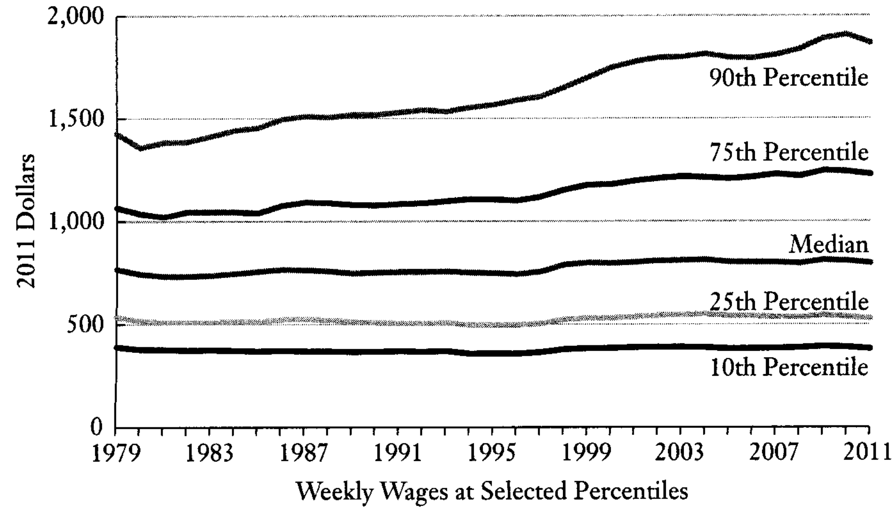
3. At bats (AB): Puts the above in context. Any mope can have impressive statistics for a game or two. A superstar compiles impressive "numbers" over thousands of plate appearances.

In Moyer's view (without hesitation, I might add), the best baseball player of all time was Babe Ruth because of his unique ability to hit and to pitch. Babe Ruth still holds the Major League career record for slugging percentage at .690.[4]

What about the economic health of the American middle class? Again, I deferred to the experts. I e-mailed Jeff Grogger (a colleague of mine at the University of Chicago) and Alan Krueger (the same Princeton economist who studied terrorists and is now serving as chair of President Obama's Council of Economic Advisers). Both gave variations on the same basic answer. To assess the economic health of America's "middle class," we should examine changes in the median wage (adjusted for inflation) over the last several decades. They also recommended examining changes to wages at the 25th and 75th percentiles (which can reasonably be interpreted as the upper and lower bounds for the middle class).

One more distinction is in order. When assessing economic health, we can examine income or wages. They are not the same thing. A wage is what we are paid for some fixed amount of labor, such as an hourly or weekly wage. Income is the sum of all payments from different sources. If workers take a second job or work more hours, their income can go up without a change in the wage. (For that matter, income can go up even if the wage is falling, provided a worker logs enough hours on the job.) However, if individuals have to work more in order to earn more, it's hard to evaluate the overall effect on their well-being. The wage is a less ambiguous measure of how Americans are being compensated for the work they do; the higher the wage, the more workers take home for every hour on the job.

Having said all that, here is a graph of American wages over the past three decades. I've also added the 90th percentile to illustrate changes in the wages for middle-class workers compared over this time frame to those workers at the top of the distribution.



Source: "Changes in the Distribution of Workers' Hourly Wages between 1979 and 2009," Congressional Budget Office, February 16, 2011. The data for the chart can be found at http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/120xx/doc12051/02-16-wagedispersion.pdf.

A variety of conclusions can be drawn from these data. They do not present a single "right" answer with regard to the economic fortunes of the middle class. They do tell us that the typical worker, an American worker earning the median wage, has been "running in place" for nearly thirty years. Workers at the 90th percentile have done much, much better. Descriptive statistics help to frame the issue. What we do about it, if anything, is an ideological and political question.

• • •

APPENDIX TO CHAPTER 2

Data for the printer defects graphics

	Zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten or more
Frequency of competitor's defects	12	14	36	13	8	6	5	3	0	2	1
	Zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten or more
Frequency of your defects	25	31	9	4	3	0	0	1	1	0	26

Formula for variance and standard deviation

Variance and standard deviation are the most common statistical mechanisms for measuring and describing the dispersion of a distribution. The variance, which is often represented by the symbol σ^2, is calculated by determining how far the observations within a distribution lie from the mean. However, the twist is that the difference between each observation and the mean is squared; the sum of those squared terms is then divided by the number of observations.

Specifically:

For any set of n observations $x_1, x_2, x_3 \ldots x_n$ with mean μ,

$$\text{Variance} = \sigma^2 = [(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \ldots (x_n - \mu)^2]/n$$

Because the difference between each term and the mean is squared, the formula for calculating variance puts particular weight on observations that lie far from the mean, or outliers, as the following table of student heights illustrates.

Group 1	Height (μ = 70 inches)	Distance from the mean = Absolute value of $(x_n - \mu)$*	$(x_n - \mu)^2$	Group 2	Height (μ = 70 inches)	Distance from the mean = Absolute value of $(x_n - \mu)$*	$(x_n - \mu)^2$
Nick	74	4	16	Sahar	65	5	25
Elana	66	4	16	Maggie	68	2	4
Dinah	68	2	4	Faisal	69	1	1
Rebecca	69	1	1	Ted	70	0	0
Ben	73	3	9	Jeff	71	1	1
Charu	70	0	0	Narciso	75	5	25
		Total = 14	Total = 46			Total = 14	Total = 56
			Variance = 46/6 = 7.7				Variance = 56/6 = 9.3
			Standard deviation = $\sqrt{7.7} = 2.8$				Standard deviation = $\sqrt{9.3} = 3$

* Absolute value is the distance between two figures, regardless of direction, so that it is always positive. In this case, it represents the number of inches between the height of the individual and the mean.

Both groups of students have a mean height of 70 inches. The heights of students in both groups also differ from the mean by the same number of total inches: 14. By that measure of dispersion, the two distributions are identical. However, the variance for Group 2 is higher because of the weight given in the variance formula to values that lie particularly far from the mean—Sahar and Narciso in this case.

Variance is rarely used as a descriptive statistic on its own. Instead, the variance is most useful as a step toward calculating the standard deviation of a distribution, which is a more intuitive tool as a descriptive statistic.

The standard deviation for a set of observations is the square root of the variance:

For any set of n observations $x_1, x_2, x_3 \ldots x_n$ with mean μ, standard deviation = σ = square root of this whole quantity =

$$\sqrt{[(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \ldots (x_n - \mu)^2]/n}$$