

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.

We ought to begin with the crucial distinction between “precision” and “accuracy.” These words are not interchangeable. Precision reflects the exactitude with which we can express something. In a description of the length of your commute, “41.6 miles” is more precise than “about 40 miles,” which is more precise than “a long f—ing way.” If you ask me how far it is to the nearest gas station, and I tell you that it's 1.265 miles to the east, that's a precise answer. Here is the problem: That answer may be entirely inaccurate if the gas station happens to be in the other direction. On the other hand, if I tell you, “Drive ten minutes or so until you see a hot dog stand. The gas station will be a couple hundred yards after that on the right. If you pass the Hooters, you've gone too far,” my answer is less precise than “1.265 miles to the east” but significantly better because I am sending you in the direction of the gas station. Accuracy is a measure of whether a figure is broadly consistent with the truth—hence the danger of confusing precision with accuracy. If an answer is accurate, then more precision is usually better. But no amount of precision can make up for inaccuracy.

In fact, precision can mask inaccuracy by giving us a false sense of certainty, either inadvertently or quite deliberately. Joseph McCarthy, the Red-baiting senator from Wisconsin, reached the apogee of his reckless charges in 1950 when he alleged not only that the U.S. State Department was infiltrated with communists, but that he had a list of their names. During a speech in Wheeling, West Virginia, McCarthy waved in the air a piece of paper and declared, “I have here in my hand a list of 205—a list of names that were made known to the Secretary of State as being members of the Communist Party and who nevertheless are still working and shaping policy in the State Department.”¹ It turns out that the paper had no names on it at all, but the specificity of the charge gave it credibility, despite the fact that it was a bald-faced lie.

I learned the important distinction between precision and accuracy in a less malicious context. For Christmas one year my wife bought me a golf range finder to calculate distances on the course from my golf ball to the hole. The device works with some kind of laser; I stand next to my ball in the fairway (or rough) and point the range finder at the flag on the green, at which point the device calculates the exact distance that I'm supposed to hit the ball. This is an improvement upon the standard yardage markers, which give distances only to the center of the green (and are therefore accurate but less precise). With my Christmas-gift range finder I was able to know that I was 147.2 yards from the hole. I expected the precision of this nifty technology to improve my golf game. Instead, it got appreciably worse.

There were two problems. First, I used the stupid device for three months before I realized that it was set to meters rather than to yards; every seemingly precise calculation (147.2) was wrong. Second, I would sometimes inadvertently aim the laser beam at the trees behind the green, rather than at the flag marking the hole, so that my “perfect” shot would go exactly the distance it was supposed to go—right over the green into the forest. The lesson for me, which applies to all statistical analysis, is that even the most precise measurements or calculations should be checked against common sense.

To take an example with more serious implications, many of the Wall Street risk management models prior to the 2008 financial crisis were quite precise. The concept of “value at risk” allowed firms to quantify with precision the amount of the firm's capital that could be lost under different scenarios. The problem was that the supersophisticated models were the equivalent of setting my range finder to meters rather than to yards. The math was complex and arcane. The answers it produced were reassuringly precise. But the assumptions about what might happen to global markets that were embedded in the models were just plain wrong, making the conclusions wholly inaccurate in ways that destabilized not only Wall Street but the entire global economy.

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?

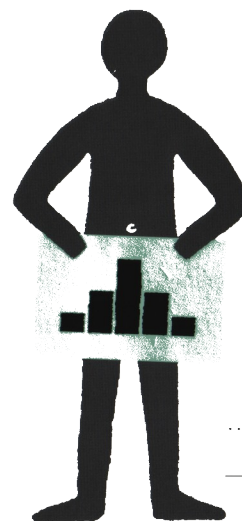
No. Smokers and nonsmokers are likely to be different in ways other than their smoking behavior. For example, smokers may be more likely to have other habits, such as drinking heavily or eating badly, that cause adverse health outcomes. If the smokers are particularly unhealthy at the twentieth reunion, 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.

NEW YORK TIMES BESTSELLER

naked statistics

STRIPPING THE DREAD FROM THE DATA



"Brilliant, funny
... the best math teacher
you never had."
—*San Francisco Chronicle*

charles wheelan

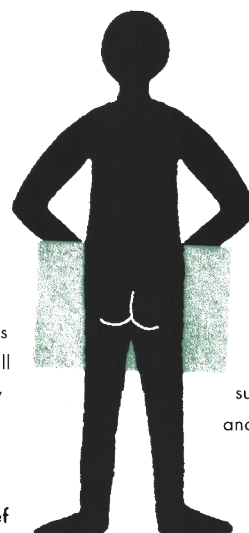
AUTHOR OF *NAKED ECONOMICS*

**"Sparkling
and intensely
readable."**

—*New York Times*

"Almost anyone
interested in sports,
politics, business, and
the myriad of other
areas in which statistics
rule the roost today will
benefit from this highly
readable, on target,
and important book."

—**Frank Newport,**
Gallup editor in chief



"[Wheelan] makes
statistics interesting and
fun. His book strips
the subject of its
complexity to
expose the sexy
stuff underneath."
—*The Economist*

"Charles Wheelan
strips away the
superfluous outergarments
and exposes the underlying
beauty of the subject."
—**Hal Varian, chief
economist at Google**

Charles Wheelan is the author of the internationally best-selling *Naked Economics* and a former correspondent for *The Economist*. He teaches public policy and economics at the Nelson A. Rockefeller Center at Dartmouth College and lives in Hanover, New Hampshire, with his family.