25

The *USNWR* rankings use sixteen indicators to score and rank America's colleges, universities, and professional schools. In 2010, for example, the ranking of national universities and liberal arts colleges used "student selectivity" as 15 percent of the index; student selectivity is in turn calculated on the basis of a school's acceptance rate, the proportion of the entering students who were in the top 10 percent of their high school class, and the average SAT and ACT scores of entering students. The benefit of the *USNWR* rankings is that they provide lots of information about thousands of schools in a simple and accessible way. Even the critics concede that much of the information collected on America's colleges and universities is valuable. Prospective students should know an institution's graduation rate and the average class size.

Of course, providing meaningful information is an enterprise entirely different from that of collapsing all of that information into a single ranking that purports to be authoritative. To critics, the rankings are sloppily constructed, misleading, and detrimental to the long-term interests of students. "One concern is simply about its being a list that claims to rank institutions in numerical order, which is a level of precision that those data just don't support," says Michael McPherson, the former president of Macalester College in Minnesota.[10] Why should alumni giving count for 5 percent of a school's score? And if it's important, why does it not count for ten percent?

According to *U.S. News & World Report*, "Each indicator is assigned a weight (expressed as a percentage) based on our judgments about which measures of quality matter most."[11] Judgment is one thing; arbitrariness is another. The most heavily weighted variable in the ranking of national universities and colleges is "academic reputation." This reputation is determined on the basis of a "peer assessment survey" filled out by administrators at other colleges and universities and from a survey of high school guidance counselors. In his general critique of rankings, Malcolm Gladwell offers a scathing (though humorous) indictment of the peer assessment methodology. He cites a questionnaire sent out by a former chief justice of the Michigan Supreme Court to roughly one hundred lawyers asking them to rank ten law schools in order of quality. Penn State's was one of the law schools on the list; the lawyers ranked it near the middle. *At the time, Penn State did not have a law school.*[12]

For all the data collected by *USNWR*, it's not obvious that the rankings measure what prospective students ought to care about: How much learning is going on at any given institution? Football fans may quibble about the composition of the passer index, but no one can deny that its component parts—completions, yardage, touchdowns, and interceptions—are an important part of a quarterback's overall performance. That is not necessarily the case with the *USNWR* criteria, most of which focus on inputs (e.g., what kind of students are admitted, how much faculty are paid, the percentage of faculty who are full-time) rather than educational outputs. Two notable exceptions are the freshman retention rate and the graduation rate, but even those indicators do not measure learning. As Michael McPherson points out, "We don't really learn anything from U.S. News about whether the education they got during those four years actually improved their talents or enriched their knowledge."

All of this would still be a harmless exercise, but for the fact that it appears to encourage behavior that is not necessarily good for students or higher education. For example, one statistic used to calculate the rankings is financial resources per student; the problem is that there is no corresponding measure of how well that money is being spent. An institution that spends less money to better effect (and therefore can charge lower tuition) is punished in the ranking process. Colleges and universities also have an incentive to encourage large numbers of students to apply, including those with no realistic hope of getting in, because it makes the school appear more selective. This is a waste of resources for the schools soliciting bogus applications and for students who end up applying with no meaningful chance of being accepted.

26

I began writing this chapter many days ago. Since then, I've had a chance to watch the documentary film *Bhutto*. Wow! This is a remarkable film about a remarkable family. The original footage, stretching all the way from the partition of India and Pakistan in 1947 to the assassination of Benazir Bhutto in 2007, is extraordinary. Bhutto's voice is woven effectively throughout the film in the form of speeches and interviews. Anyway, I gave the film five stars, which is pretty much what Netflix predicted.

At the most basic level, Netflix is exploiting the concept of correlation. First, I rate a set of films. Netflix compares my ratings with those of other customers to identify those whose ratings are highly correlated with mine. Those customers tend to like the films that I like. Once that is established, Netflix can recommend films that like-minded customers have rated highly but that I have not yet seen.

That's the "big picture." The actual methodology is much more complex. In fact, Netflix launched a contest in 2006 in which members of the public were invited to design a mechanism that improved on existing Netflix recommendations by at least 10 percent (meaning that the system was 10 percent more accurate in predicting how a customer would rate a film after seeing it). The winner would get $1,000,000.

Every individual or team that registered for the contest was given "training data" consisting of more than 100 million ratings of 18,000 films by 480,000 Netflix customers. A separate set of 2.8 million ratings was "withheld," meaning that Netflix knew how the customers rated these films but the contest participants did not. The competitors were judged on how well their algorithms predicted the actual customer reviews for these withheld films. Over three years, thousands of teams from over 180 countries submitted proposals. There were two requirements for entry. First, the winner had to license the algorithm to Netflix. And second, the winner had to "describe to the world how you did it and why it works."[3]

In 2009 Netflix announced a winner: a seven-person team made up of statisticians and computer scientists from the United States, Austria, Canada, and Israel. Alas, I cannot describe the winning system, even in an appendix. The paper explaining the system is ninety-two pages long.* I'm impressed by the quality of the Netflix recommendations. Still, the system is just a super fancy variation on what people have been doing since the dawn of film: find someone with similar tastes and ask for a recommendation. You tend to like what I like, and to dislike what I dislike, so what did you think of the new George Clooney film?

That is the essence of correlation.

27

Probability is the study of events and outcomes involving an element of uncertainty. Investing in the stock market involves uncertainty. So does flipping a coin, which may come up heads or tails. Flipping a coin four times in a row involves additional layers of uncertainty, because each of the four flips can result in a head or a tail. If you flip a coin four times in a row, I cannot know the outcome in advance with certainty (nor can you). Yet I *can* determine in advance that some outcomes (two heads, two tails) are more likely than others (four heads). As the folks at Schlitz reckoned, those kinds of probability-based insights can be extremely helpful. In fact, if you can understand why the probability of flipping four heads in a row with a fair coin is 1 in 16, you can (with some work) understand everything from how the insurance industry works to whether a pro football team should kick the extra point after a touchdown or go for a two-point conversion.

Let's start with the easy part: Many events have known probabilities. The probability of flipping heads with a fair coin is $\frac{1}{2}$. The probability of rolling a one with a single die is $\frac{1}{6}$. Other events have probabilities that can be inferred on the basis of past data. The probability of successfully kicking the extra point after touchdown in professional football is .94, meaning that kickers make, on average, 94 out of every 100 extra-point attempts. (Obviously this figure might vary slightly for different kickers, under different weather circumstances, and so on, but it's not going to change radically.) Simply having and appreciating this kind of information can often clarify decision making and render risks explicit. For example, the Australian Transport Safety Board published a report quantifying the fatality risks for different modes of transport. Despite widespread fear of flying, the risks associated with commercial air travel are tiny. Australia hasn't had a commercial air fatality since the 1960s, so the fatality rate per 100 million kilometers traveled is essentially zero. The rate for drivers is .5 fatalities per 100 million kilometers traveled. The really impressive number is for motorcycles—if you aspire to be an organ donor. The fatality rate is thirty-five times higher for motorcycles than for cars.[3]

In September of 2011, a 6.5-ton NASA satellite was plummeting to earth and was expected to break apart once it hit the earth's atmosphere. What were the chances of being struck by the debris? Should I have kept the kids home from school? The rocket scientists at NASA estimated that the probability of any individual person's being hit by a part of the falling satellite was 1 in 21 trillion. Yet the chances that anyone anywhere on earth might get hit were a more sobering 1 in 3,200.* In the end, the satellite did break apart on reentry, but scientists aren't entirely certain where all the pieces ended up.[4] No one reported being hurt. Probabilities do not tell us what will happen for sure; they tell us what is *likely to happen* and what is *less likely to happen*. Sensible people can make use of these kinds of numbers in business and life. For example, when you hear on the radio that a satellite is plummeting to earth, you should not race home on your motorcycle to warn the family.

28

Probability also enables us to calculate what might be the most useful tool in all of managerial decision making, particularly finance: expected value. The expected value takes basic probability one step further. The expected value or payoff from some event, say purchasing a lottery ticket, is the sum of all the different outcomes, each weighted by its probability and payoff. As usual, an example makes this clearer. Suppose you are invited to play a game in which you roll a single die. The payoff to this game is \$1 if you roll a 1; \$2 if you roll a 2; \$3 if you roll a 3; and so on. What is the expected value for a single roll of the die? Each possible outcome has a $\frac{1}{6}$ probability, so the expected value is:

$$\tfrac{1}{6}(\$1) + \tfrac{1}{6}(\$2) + \tfrac{1}{6}(\$3) + \tfrac{1}{6}(\$4) + \tfrac{1}{6}(\$5) + \tfrac{1}{6}(\$6) = \tfrac{21}{6}, \text{ or } \$3.50.$$

At first glance, the expected value of \$3.50 might appear to be a relatively useless figure. After all, you can't actually earn \$3.50 with a single roll of the die (since your payoff has to be a whole number). In fact, the expected value turns out to be extremely powerful because it can tell you whether a particular event is "fair," given its price and expected outcome. Suppose you have the chance to play the above game for \$3 a throw. Does it make sense to play? Yes, because the expected value of the outcome (\$3.50) is higher than the cost of playing (\$3.00). This does not guarantee that you will make money by playing once, but it does help clarify which risks are worth taking and which are not.

We can take this hypothetical example and apply it to professional football. As noted earlier, after a touchdown, teams have a choice between kicking an extra point and attempting a two-point conversion. The former involves kicking the ball through the goalposts from the three yard line; the latter involves running or passing it into the end zone from the three yard line, which is significantly more difficult. Teams can choose the easy option and get one point, or they can choose the harder option and get two points. What to do?

Statisticians may not play football or date cheerleaders, but they can provide statistical guidance for football coaches.[9] As pointed out earlier, the probability of making the kick after a touchdown is .94. This means that the expected value of a point-after attempt is also .94, since it equals the payoff (1 point) multiplied by the probability of success (.94). No team ever scores .94 points, but this figure is helpful in quantifying the value of attempting this option after a touchdown relative to the alternative, which is the two-point conversion. The expected value of "going for two" is much lower: .74. Yes, the payoff is higher (2 points), but the success rate is dramatically lower (.37). Obviously if there is one second left in the game and a team is behind by two points after scoring a touchdown, it has no choice but to go for a two-point conversion. But if a team's goal is to maximize points scored over time, then kicking the extra point is the strategy that will do that.

Excerpts from Wheelan, C. (2013). *Naked Statistics*. New York, NY. W. W. Norton and Company.