

7 Habits of Highly Numerate People

BY DOUG BERDIE

We are inundated daily with assertions supposedly supported by solid data. This torrent will only increase, given the political climate and upcoming off-year election. We must be intelligent consumers of these claims if we are to avoid being misled or duped. Based on 40-plus years of sifting through all sorts of data, I've identified seven common fallacies that pop up frequently in news reports, advocacy pitches and other summaries. Critically examining data-based claims in light of these potential pitfalls can help you draw more valid conclusions — or at least fewer invalid ones.

1. Don't project group data to individual people.

It may be true that "the percentage of Trump voters who favor a border wall is larger than the percentage of non-Trump voters who do." But this tells us nothing about any specific Trump voter's views on such a wall. To assume that someone you know who voted for President Donald Trump favors a border wall would be naive.

Common sense should be sufficient to dismiss spurious generalizations from groups to individuals. The 2017 World Happiness Report, for example, lists Norway, Denmark and Iceland as the three countries where residents are the happiest, and it lists the U.S. as 14th. Yet few of us would conclude that "all" residents in those Nordic countries are happier than "any" U.S. resident.

Nonetheless, some communities these days object to having mosques built in their neighborhoods because of anxieties about terrorism. Such fears are based on making broad generalizations apply to individuals and are nothing more than crass stereotyping that cannot be justified.

2. Don't blindly accept "averages."

The two most commonly reported types of "averages" are:

- The "mean" (the arithmetic average, where the total quantity of a thing observed is divided by the number of individual observations).
- The "median" (the midpoint, at which half of the observations are larger and the other half smaller).

Mean and median values often differ dramatically, and each is more useful in specific situations. The mean is most useful when the values of individual observations do not vary a great deal. When they do vary sharply, the median is more useful.

For example, the 2014 median household income in the U.S. was \$53,657; the mean income was \$72,641.

The mean was so much higher because of the influence of the "mega-wealthy" — outliers such as billionaires whose incomes pulled up the calculated mean. That distorting effect of extreme riches is why the median is used in reputable discussions of income.

But again, no average can tell you much about an individual within a group.

For example, a group of five annual incomes of \$30,000, \$40,000, \$50,000, \$70,000 and \$200,000 would yield a "median" income of \$50,000. That median value only matches the income of one of the five families, and does not even come close to reflecting the highest of the five. Yet the "mean" income — \$78,000 — would be even less representative.

So examine averages carefully — and, please, don't project those values to individual people.

3. Watch out for the “disparity fallacy.”

We hear frequently of disparities between and among groups. These claims are made by countries in regard to other countries and by various activist groups in reference to subparts of the population. But when disparities exist, it is not always the case that they represent a problem.

What I call the “disparity fallacy” follows when the following blanket conclusion is drawn: “There’s a disparity, which means there is a serious problem, which means something must be done to eliminate the disparity.”

Some claims of disparity aren’t even intelligible. Others may indicate a real underlying problem. Some are invoked merely to further political causes. And some seem of little interest to anyone.

An example of a disparity claim that’s not even intelligible occurred when Trump, lamenting his view of inequities in how member nations fund NATO, claimed that the U.S. should contribute only “its proportion” to funding the organization.

The question is, “proportion of what?” Did he mean our contribution should be proportionate to our GDP, to our median household income, to our balance-of-trade ratio with other NATO members, or some other benchmark?

Underlying the confusion in this example, and many others where so-called disparities are criticized, is what I call “that dang denominator.” In statistics, one can prove almost anything by shuffling around different denominators — measuring things “as a proportion of” strategically chosen benchmarks — until the desired result is found.

Trump didn’t even tell us what denominator he was referencing. Therefore, his assertion that an unjust disparity exists is basically meaningless.

Studies, advocate communications and news reports often use “proportion of residents” in a geographic area as the benchmark denominator when examining disparities. A Feb. 10 Star Tribune article (“Police stops scrutinized in St. Paul”) indicated that 33 percent of St. Paul police traffic stops involve black drivers, whereas blacks only constitute 14 percent of St. Paul’s driving population.

Yet, in a Feb. 15 Star Tribune article about a different study and a different police department (“Police study finds no racial bias”), University of Minnesota Prof. Michelle Phelps rightly pointed out a major difficulty with using this denominator: People are not always residents of the town in which they are stopped by police.

Researchers have been testing alternatives to using residential census counts as the benchmark to determine the extent to which racial profiling may exist in traffic stops. One alternative is to prorate stops with police presence. Obviously, a higher “police-to-resident” presence in a given neighborhood (likely because of high rates of service calls there) would result in more stops or arrests because a larger proportion of infractions would be seen.

Other researchers are examining ways to assess possible police bias by examining traffic camera data in situations where the officer would not have been able to discern a person’s race before initiating a traffic stop.

The disparities between black and nonblack drivers’ stop and arrest rates do seem problematic enough to warrant further investigation. But the easy reliance on resident population as the base for calculations oversimplifies and obscures the truth.

An example where the disparity fallacy clearly seems at work appeared in a 2016 Star Tribune article (“The color of music”) that cited the proportion of black Americans in orchestras being smaller than their proportion in the population. The remainder of the article dealt with how to resolve this assumed “problem.” But those

data do not necessarily indicate any problem at all. At most they raise a question: “Hmmm, I wonder why that difference exists? Let’s take a deeper look. There might be a problem here.”

In this case, the calculation of the disparity is based, again, on the total number of black Americans in the population. If one had, instead, used as the denominator “the total number of black Americans who have any interest in being in an orchestra and who have undertaken required training,” the results may have been very different.

Many other disparities seem to be largely ignored, perhaps because they do not support a particular social cause, such as when a national network announced that 70 percent of NFL players are black, four or five times the African-American percentage of the population. What about the low proportion of Asian-Americans playing professional basketball? I have heard no one lament such disparities.

Always obtain clarity as to what the denominator is when disparities are presented so you can be sure the comparison makes sense and warrants serious attention.

4. Don’t assume “statistically significant” means, necessarily, “practically significant.”

News reports and advocates revel in citing research that reports “statistically significant differences” among groups of people. This kind of significance (it basically means the difference is unlikely to be accidental) can be little more than an artifact of how large the compared samples are. With large-enough samples, one is almost guaranteed of finding “statistically significant” differences. Yet the size of the differences may be so small as to have no practical implication.

For example, researchers recently found, when comparing the IQs of hundreds of thousands of firstborn and later-born individuals, that the firstborns had higher average IQs — but only between 1 and 3 points higher, hardly of much practical interest.

When group differences of any kind are reported, the first question should always be: “How large is the difference between the groups?” And if it’s an irrelevant difference, who cares if it’s statistically significant?

A March 12 Star Tribune article reported that the odd white posts in the middle of the road at 2nd Street at 3rd Avenue reduced “average” traffic speed from 32 miles per hour to 30 mph (“Posts that slow you down are meant to”). Because the article did not say anything about the number of vehicles included in the pre-post samples, it is impossible to assess whether the speed reduction was statistically significant, although one could certainly wonder if the 2 mph drop had practical significance.

5. Beware of claims asserting extreme statistical precision.

Survey and poll reports, and other research, often tout a “plus-or-minus” number that, purportedly, shows “how good” the data are. This number, often referred to as the “margin of error,” shows the “sampling precision” of the data and depends on the number of people in the research sample, the number of people in the population from which the sample was drawn, the variability within the population of the characteristic being measured and the “confidence level” (usually, but not necessarily, 95 percent).

Although sampling precision is an arithmetic calculation, true precision would require that two underlying assumptions be met: (1) that a true random sample was obtained for the research, and (2) that 100 percent of that random sample actually participated in the survey. Because these assumptions are rarely met perfectly, survey practitioners often “weight” data in an attempt to make the sample more representative of the population of interest. However, given that most surveys are interested in opinions, weighting by demographic factors (gender, race, income, age, etc.) is far from foolproof, and it is likely that the margins of error presented are, at best, fair estimates.

6. Don't confuse "social science" with "hard science."

Much of the probability theory underlying reported data arose in the hard sciences — particularly the agricultural research of R.A. Fisher and other statistical pioneers. Those researchers could measure all the seeds that were planted and did not face the "nonresponse bias" that social researchers confront when people refuse to participate. What's more, the measurements in agriculture were physical, such as the height and weight of the crop, and the measurement techniques were consistent. Much medical and scientific research reported these days closely approximates these kinds of controls.

In social research, however, people are often asked questions, and how they interpret the question, their fears of social chastisement depending on how they answer and other psychosocial factors interfere with consistent measurements.

For example, questions using emotionally loaded words and phrases (e.g., "liberal," "tax increase," "gun control"), questions that only present one side of an issue (e.g., "Do you favor increasing the minimum wage to \$15 an hour?" as opposed to "Do you favor or oppose increasing the minimum wage to \$15 an hour?"), leading questions (e.g., "Do you favor more spending on schools to improve student learning?") and questions with other wording flaws will not elicit accurate information. Hence, blindly accepting claims arising from surveys and opinion polls is naive. One should always carefully review the actual questions asked to assess how responses may have been skewed.

7. Beware generalizations that data providers make.

Sound social and economic data are collected from random samples of a well-defined population and projected to that population. Misleading (or self-serving) data are collected from one population and projected to a different one.

If one wishes to project election results, one must find and survey a random sample of people who will vote. This is difficult, because some who voted previously will not vote in an upcoming election — and vice versa.

Techniques must be used to identify new voters and to vet existing voters by asking about the likelihood that the person will vote in the future election. And different types of voters prefer participating in surveys in varying ways. Older voters may be more likely to have landlines. Younger voters may prefer mobile phones or text surveys or other types of new technology. To capture a truly representative sample of likely voters, one must carefully sample proportionately from all these groups. Such sampling and data collection is expensive and requires great care.

As a consumer of survey data, you should ask how a sample was collected and look for signs that it is representative of the intended population. Reputable pollsters go to great lengths to identify representative samples of people and to convince them to participate. Pollsters who may wish to obtain data to support preconceived positions (e.g., political parties and other special-interest groups) often start with their own lists of people believed likely to support their views and contact only such people. Obviously, data obtained from these biased lists are not generalizable to the general public. So always look at how the sampling was done before placing faith in the results.

It's unfortunate that so many politicians, businesses, activists and others are more interested in finding data to support their positions than they are in finding solid data that would inform decision making effectively. Yet, that's how it is. It's incumbent on all of us to carefully examine what underlies data before accepting or touting it.

Doug R. Berdie, of Minneapolis, is a semiretired marketing executive and researcher.

Source: 7 HABITS OF HIGHLY NUMERATE PEOPLE - Star Tribune, 2018-04-22 Commentary.

Hosted with permission from Doug Tice, Commentary Editor, Minneapolis Star and Tribune.