1 Introduction to statistical literacy

Lynette Hoelter

Let's "set the stage" for thinking about data encountered in everyday media. With just a few key statistical concepts on your tool belt, your mindset can shift from simply accepting numbers as fact to questioning the data. For anyone who is afraid that this chapter will be math heavy – be reassured, it is not!

Quantitative literacy, *quantitative reasoning*, *statistical literacy*, and *numeracy* have become buzzwords in educational circles from K-12 through undergraduate training. Used interchangeably here, they all represent a core group of skills necessary to fully participate in today's information-rich society. Some have suggested nuances leading to unique meanings of each term, but whatever the buzzword, the goal is to boost high school students' abilities and comfort level with quantitative information.

Defining statistical literacy

For the purpose of this chapter, statistical literacy refers to a mindset or set of skills that are used in everyday life. That is, while some concepts might stem from the fields of math and science, the use of those ideas is equally, if not more, important. Statistical literacy, then, is the learning and using of quantitative skills *within a particular context*. Quantitative reasoning skills include:

- **the ability to read and interpret a chart or graph;**
- **calculating percentages;**
- **working within a scientific model;**

- **evaluating the data on which arguments are based and using data in making one's own decisions and arguments;**
- **knowing what kinds of data might be useful in answering particular questions.**

"Now more than ever, students need the intellectual power to recognize societal problems; ask good questions and develop robust investigations into them; consider possible solutions and consequences; separate evidence-based claims from parochial opinions; and communicate and act upon what they learn" (National Center for the Social Studies 2013, 6). The overall goal is to understand and critically evaluate the numbers encountered as part of everyday life.

The relationship between statistics and statistical literacy

There are two primary ways instructors teach statistics: with a focus on formulas and arithmetic or with a focus on key concepts and their applications. A quick scan of popular high school and college-level statistics textbooks shows a focus on formulas and arithmetic to be most common. Students often get overwhelmed by the *math* and focus only on formulas and calculations. Unfortunately, when they leave the course, they may not remember much about the analyses they did. It is difficult for instructors and students to overcome this mindset – even when students know exams are open book/open note and that their interpretations count more than their calculations, the tendency is still to grab on to the formulas because they are, in a sense, more *concrete*. Logic and application are not always straightforward. There are right and wrong answers in the number-crunching, formulas, and arithmetic, and sometimes instructors and students are most secure in that arena.

I would argue, however, that deep learning and statistical literacy can take place when students are introduced to and asked to make decisions about statistical tests and interpret the results beyond merely repeating the numbers back. It is exactly these kinds of skills that can be incorporated across the curriculum as well. That said, fundamental concepts from statistics are important in asking and answering the questions posed in quantitative reasoning. Such topics include more theoretical ideas, and the next section defines and provides examples of each. The numbers and examples are completely made up, unless otherwise noted.

Statistics review

A refresher about key topics is always helpful — whether it's been awhile since you took statistics, you have nightmares about the class, or you never took a formal statistics course. The organization of the *review* here is based on the order in which topics are typically covered in texts.

Variables

Sometimes the thing that is least on our minds when reading data and numbers is exactly what was measured, how, and by whom. In scientific language, the *what* is thought of as a *variable*. Asking how many students in a class are boys and how many are girls involves the variable *gender*. A variable is a trait that can vary from person to person (or across time or context) and about which you have or need information. If the classroom were made up entirely of boys, gender would not be a variable because it would not differentiate the students from each other.

How is this applied to statistical literacy?

Variables are important for two reasons.

- » **The media often present results as cause and effect.** Thinking about exactly what variables were included (and what might have been left out) leads to questioning this assumption of causality.
- » **Thinking more broadly about what was measured, how, and by whom brings forth questions** about quality of the data and potential biases depending on the source of the data or the report.

For example, a high school that says that 90% of its graduating seniors are going on to college seems impressive, until one digs further and finds that only 40% of the seniors are graduating or that *going on to college* in this case means that the student has reported an *intention* to attend college at some point in the future.

The way something is measured also affects the way that information can be used and presented later. Asking people about the highest degree they've earned, for instance, does not allow a researcher to later examine the impact on earnings of having some college but not finishing a degree, because the people who didn't go to college and the people who did but didn't earn a degree will be considered as having exactly the same level of education. A giant first step toward quantitative literacy is this understanding that the numbers we see represent variables, the data for which were measured in a specific way, by a specific person, for a specific reason.

Percentages, rates, and percent change

Without context, numbers can seem astonishingly large or shockingly small. Standardizing numbers using percentages or rates provides context around the numbers. A percentage is calculated by dividing some part by the overall whole — taking the number of correct answers on an exam and dividing by the total possible points — and then multiplying by 100 to turn it into a percentage that ranges from 0–100. For example, a student earning 40 points out of a possible 50 points on a science quiz looks like this:

> 40/50 = .8 and .8 x 100 = 80%

Percentages provide a way to see how groups of very different sizes compare on a characteristic. Let's say School A has 1,010 females taking AP Calculus and School B has 76. This should not necessarily lead to the assumption that School A is better at getting females into advanced math classes. Suppose that those schools had 2,400 and 130 students in AP Calculus, respectively. Using percentages, we would see that School A's AP Calculus curriculum is 42% female while School B's is 58% female.

> **School A**
>
> 1010/2400 = 42% of students taking AP Calculus are female.
>
> **School B**
>
> 76/130 = 58% of students taking AP Calculus are female.

Taking it one step further: if math ability was present and encouraged in male and female students equally, we would expect each program to be close to 50% female. This example demonstrates that School A may not be supporting females in math as well as it could be, and School B is supporting females at least as much as males. The original 1,010 and 76 females did not provide any of the *story* that is possible by standardizing the numbers by converting them into percentages.

Similarly, rates are the number of occurrences of something divided by the number of *possibilities* for the phenomenon to occur, typically multiplied by 1,000 (or 100,000 for large populations). Violent crime rates reported for cities, then, are the number of crimes in a given time period divided by the number of people in the city at that same time, multiplied by 1,000. Like percentages, rates allow for the comparison of the chances of something happening basedon the size of the location in question.

> Let's say we know for a year:
>
> City A population = 5,000
>
> City A violent crimes = 37
>
> 37/5,000 = .0074 x 1,000 = 7.4
>
> This calculation tells us City A has a violent crime rate of 7.4 per 1,000 residents.

Related to percentages are *percentiles* and *percent change*. Percentiles are often used for things like children's heights and weights or standardized test scores. The number reported as the percentile is the percent of individuals above which that child's height, weight, or score falls. Someone scoring in the 90th percentile on the SAT would have a score above 90% of the people who took the test at that time. Percent change is a bit trickier because the denominator changes with each calculation. For example, a sale item marked *70% off* is cheaper than the same item marked *50% + an additional 20% off* because in the first case, the 70% is taken off the total purchase price but in the second case, the 50% is taken off first and then, using that amount as the new denominator, the 20% is taken off.

> $100 x 70% = $70 off, then $100-$70 = $30
>
> $100 x 50% = $50 off, then $50 x 20% = $10, then $50-$10 = $40
>
> An item that began as $100 in this example would be on sale for $30 and $40, respectively.

Likewise, large denominators mean that a small percent change can still produce a large number of people affected, whereas a large percent change to a small number will remain a small number.

How is this applied to statistical literacy?

Presenting numbers without context is a common strategy for making impressions, especially when defining issues as problems. Two thousand thefts sound like a lot, and a city might use that to push for more patrol officers on the streets. If the thefts were known to have occurred in a city of 450,000, though, people might be less panicked as they realize that just over four people in every 1,000 were victims of theft (especially if they can think of four people who leave their doors unlocked!).

The same "big deal" can be made by presenting a percentage without a sense of the base. Parma, Missouri, made the headlines when the election of its first black mayor caused 80% of the police force to resign. While the headlines were technically correct (in fact, 83.3% had resigned), this percentage was based on five resignations out of the total six officers. The facts are the same, but the emotional reaction to the resignation of 80% of the officers is different than knowing that five officers resigned.

Increases in percent change of contracting a particular disease often make headlines, especially when the percent change is large. It's important to remember that if the starting number is small, even a large jump should not be enough to send everyone to his or her physician for the latest test for that disease. It might be a 100% change, but that could mean going from one person to two affected out of thousands. On the flip side, a 1% change in the prison population could mean significant overcrowding, especially if the population has grown at the same time, because the base is large. As students become more statistically literate, they'll begin to think about the context in which numbers are presented (especially when there is no context at all), whether the right number was used as the denominator, and how large or small that denominator might be to begin.

Average/central tendency

Most people know that when they hear about the *average* or *central tendencies* of something, they should think about who is represented by that statement. If Starbucks talks about the amount of coffee an average person drinks per week, it is useful to know whether they are reporting the average for their customers or for a representative sample of adults in the U.S., as those two numbers could be quite different.

There are actually several types of *averages* in statistics and using one over the other can affect interpretation of results.

Mode – Sometimes it is only possible to know what trait or number is the most common among a group of people or items – like knowing that July and August are the most popular birthday months in the U.S. or that blue is most commonly reported as a favorite color. The mode, or most commonly occurring value, is the only meaningful average that can be used for variables that are measured by putting people into categories (like race, gender, or religion).

Median – The median, another type of average, is the value that splits a distribution of people or things exactly in half. The median should be used for variables that result in ordered responses, such as one's highest degree earned or questions answered on a scale from strongly agree to strongly disagree or by categories (e.g., age cohorts).

Mean – Lastly, the mean represents the literal arithmetic average of something, like when someone asks everyone in a room for his or her age, adds those numbers up, and divides by the number of people in the room. The value of each individual's score is used when calculating the mean, therefore one or two extremely high or low scores can shift the resulting value.

While these seem straightforward, an agenda can be supported by choosing to present results using one average rather than another.

How is this applied to statistical literacy?

Being statistically literate involves asking which average is being presented and whether using another might present a more accurate picture. Schools present their *average* scores on achievement tests. Those schools where one or two students have a bad day and score very low are going to want to report the median; schools where one or two students score very high will likely want to report the mean to take advantage of those high scores. One should think about whether conclusions would be different depending which measure of central tendency was used.

Best practice is for both the mean and the median to be reported. Even better is when the average is accompanied by information about how *spread out* the scores in the distribution are (e.g., standard deviation). Suppose students are given an exam scored 0–100, and both the mean and the median come out to be 75. Knowing whether the scores clustered around 75 or were spread across the 100-point continuum provides more information about student learning than knowing the mean or median alone.

Sampling

Sampling is confusing for a lot of people. Part of this confusion is because sampling is largely a theoretical concept. Part comes from the way terms are used in referring to sample designs seems contradictory to the use of those same terms in everyday language. It may be easy to convince people that results based on just a few people might not be reliable. What about a sample of 100 people? Is that enough? Sometimes the harder battle is convincing them that a sample of 100 can actually be large enough to draw conclusions — but it is harder to get them to think about the fact that the size is not the only characteristic to consider when looking at a sample. A sample's design, rather than its size, is the key characteristic in thinking about whether results can be meaningful for a larger group of people. That is, for a sample to represent some larger group, the individuals in a sample must be chosen at random.

These designs, called probability designs, range from something that could be as straightforward as putting everyone's name in a hat and selecting a certain number (a simple random sample) to complex designs that involve multiple layers of selection, some of which include an element of chance and others that do not. The chance of selection is key. In reliable data reporting, the chance of selection, or probability design, is transparent and explained clearly to the reader.

If one wants to be able to say something about high school students in the U.S., one cannot simply take a random sample of the students in one's school because that school is not likely representative of all high schools in the country. This is a very different definition of *random* than one students bring from everyday vocabulary. To some, *random* could have a negative connotation and implies no structure about who is in or out of the sample. In contrast, in sample selection, random or probability designs require the structure of having a list of possible sample units and using some rather systematic method for selecting individuals (or households or forest plots) from that list. Standing in the hallway and on the campus quad and choosing people *at random* does not qualify as a probability sample because there is no way to know the parameters of the population of interest, and selection is not usually as *random* as we think – one might be more likely to talk to those who look at the one collecting the data or the data collector might be more comfortable talking to women than to men. This type of sample is called a *convenience sample* and is one of the many non-probability designs. A *non-probability sample* is when the data collector uses his or her own judgment about who to include in the sample and who not to include. The fact that results of non-probability samples cannot be generalized does not mean that studies based on those designs should not be done. Non-probability sample designs are the most effective way to study harder-to-find populations – imagine trying to study homelessness using a national probability sample – or topics that require understanding a particular situation or phenomenon in more depth than would be possible from a survey.

Given the declining participation in surveys and other research, scientists are debating generalizability of some sampling designs. Market research, for example, is often comfortable with a large sample no matter how the individuals in the sample were (or were not) selected. Some polling firms are using large opt-in Internet panels from which they then draw random samples and generalize results. The important thing to think about is whether the people who were included in a given sample might have characteristics that make them different from those who were not included in important ways.

Random-digit dialing is a probability sample. The results from this kind of sample could be generalized, but what if the dialing hits only landline numbers? The resulting sample will be very different than if cell phones were included. In the former case, the generalizability could only extend to people with landlines, not to the general population.

How is this applied to statistical literacy?

The numbers presented in the media and other outlets should be approached with *who* questions:

» **who was surveyed;**
» **who participated in the survey;**
» **who might have been systematically excluded; and**
» **to whom, if anyone, can the results be generalized?**

A red flag should go up if no information about the sample is provided. Similarly, seeing that the sample included a large number of people is a good first step toward credibility, but information about how the sample was selected will allow for decisions about whether the reported findings might be true for anyone beyond those who were included. The *who* is also important when the target population of a study is different than the one to whom the results are being applied. Questioning every fifth person in line at a randomly selected Starbucks at a given time is not likely to yield answers about coffee preferences and consumption that are representative of all adults in that city, even though a probability design (systematic random sampling) was employed.

Margin of error/confidence

The use of a sample of people, households, or Starbucks locations to estimate the value of a characteristic for the population of those items encompasses a bit of uncertainty. Statisticians will often report their results followed by a *margin of error* or *confidence interval* because there is a chance that any one sample chosen might not accurately represent the population. This creates an interval, or range, where the researcher can be more certain that the true value falls rather than using the exact value found in the sample. Wider intervals mean more certainty that the true value is within the range. Smaller intervals have higher chances of missing the true value. The size of the interval required to reach a given level of confidence is also related to the size of the original sample. A narrower interval can be used to achieve the same level of confidence with a larger sample than would be necessary if the sample were smaller. Often the margin of error is presented as something like "54% ±2%" and it can be interpreted to mean that the estimate of the population value falls between 52% (the reported number minus two percent) and 56% (the reported number plus two percent).

How is this applied to statistical literacy?

Providing a margin of error is one way researchers and journalists can help data consumers know how much confidence to put in the numbers reported. A large margin of error means that the numbers themselves should be taken with a grain of salt, whereas smaller margins of error engender more trust. In an effort to present *news*, media outlets have a tendency to present the figures and maybe mention the margin of error but not consider whether the resulting intervals overlap. Political polling is one place in which this happens — it might be reported that support for Clinton is 52% and for Trump is 48%, ±3%, so Clinton is ahead in the polls. However, once the margin of error is taken into consideration, the candidates essentially report they are tied because the ranges overlap (49%-55% for Clinton, and 45%-51% for Trump). We now know that had voters taken the margin of error more seriously, they might have seen earlier that either candidate could have won.

Correlation

A *correlation* describes the connection between two variables. It represents the strength and direction of the relationship between them. Correlations can be either positive or negative. A *positive correlation* means that higher values on the first variable are related to higher values on the second. A positive correlation also means that lower values on the first are related to lower values on the second. For example, height and shoe size are positively correlated – in general, taller people have larger shoe sizes than shorter people. A *negative correlation* means that as one variable increases, the other variable decreases. For example, the relationship between education and racial prejudice might be a negative correlation. People who have more education tend to score lower on scales of prejudice than those with lower levels of education.

There are two very important things to keep in mind when considering correlation.

- **A correlation or association between two variables does not mean they are causally related.** Growing taller does not cause your shoe size to be bigger. It is likely genetics, nutrition, and other factors that are causing growth in both height and size of feet. An association between two things is necessary to say that one causes the other, but it is not sufficient. Causality requires proper time ordering (does the first variable actually occur in time before the second), a logical reason for why one should cause the other, and checking that no third factor is causing the relationship between the first two, like genetics in the example of height and shoe size.
- **A correlation only picks up a linear association between two variables.** For example, there is a relationship between the length of time that a person is married and marital satisfaction, but it is not a linear relationship. Marital satisfaction is highest at the beginning of a marriage and later in the marriage, with a drop in between, so that the pattern is U-shaped.

How is this applied to statistical literacy?

Being aware that correlation does not necessarily mean causation is the important piece here. There is a great website (www.tylervirgen.com/spurious-correlations) that demonstrates strong correlations between variables for which there is no reasonable causal relationship, such as a .993 correlation between the divorce rate in Maine and the per capita consumption of margarine for any given year. These correlations, while entertaining, are the result of a computer sifting through massive amounts of data to find things that have similar patterns. Hopefully one would know to question a report about divorce rates and margarine consumption and not infer any kind of causal relationship, but sometimes it is not that clear. A report on the *New York Times* blog, *The Upshot,* suggested that "heavier babies do better in school," based on a "study of children in Florida [that] found that those who were heavier at birth scored higher on math and reading tests in the third to eighth grades" (Leonhardt and Cox, 2014). The time ordering is there – birth weight surely comes before third and eighth grade test scores, and the two variables (weight and test scores) are associated. The authors even note that education, race, and age of the mother were taken into account (*controlled*) and the relationship held. Before mothers of small babies seek out extra tutoring for their children, it pays to consider other factors that might be causing the relationship between weight and scores. For example, better nutrition and general health likely account for both larger babies and better cognitive functioning as measured by test scores. Understanding what a correlation means and what it does not is critical in using quantitative reasoning with such stories.

Significance

Statistical significance refers to the ability to say that a reported result would only happen x% of the time by chance. Typically these percentages are set at .1%, 1%, or 5% for them to be compelling. This precise meaning allows researchers to report findings as statistically significant when they reach these thresholds. Keep in mind that the identification of something as *significant* in statistical terms doesn't necessarily mean that there is a momentous finding. Calculating statistical significance is based in part on sample size. When there is a large enough sample, it is easy to produce findings that are statistically significant even when a relationship is weak or change is small. If a large sample of students who took a standardized test showed that there is what a statistician would call a statistically significant difference between female students who achieved an 88% score and male students who achieved an 87%, one might ask oneself whether that 1% difference really mattered. The key is to think about whether that statistical significance translates into real significance or importance.

How is this applied to statistical literacy?

Journalists are always hunting for the next big story, and significant research results can grab readers' attention, but the difference between statistical and substantive or *real* significance might be lost in the process. Quantitative reasoning involves finding out what the actual difference or relationship is and considering whether the size of that effect is important in the big picture. Reported differences in mean levels of marital satisfaction for men and women might seem like a big deal, and one probably could easily think of anecdotal evidence to support a finding that says men have significantly higher levels of marital satisfaction than do women. But what if that *difference* was that men scored 90.2 and women scored 88.4 on a 100-point index of marital satisfaction? Does that less-than-two-point difference mean men and women really experience marriage differently? Wouldn't a more accurate story be one that reports both men and women are really pretty satisfied with their marriages? Questions such as these about the implications of statistically significant findings are another sign of quantitative reasoning.

Raising statistical literacy in the classroom

The mindset of statistical literacy is developed through practice. Exposing students to news articles and guiding them through questions about how the variables were measured, who was in the sample, and whether the results are worthy of such attention is one way to delve deeper. This can be used to focus students on the task at hand or to convey the content of the day. There are a number of blogs and news sources that provide quantitative information so it is relatively easy to find something related to most topics. Getting students to think quantitatively can be done without the math behind the statistics. Perhaps a science course is looking at the effects of climate change on policy (or vice versa) and having students write a paper that includes statistics about the average temperature each year or the amount of carbon monoxide emissions eliminated by a given policy offers the opportunity to remind students about credibility of sources and types of averages.

Students in many high school courses are asked to convey information as an infographic rather than writing a more traditional research paper. Each of these assignments allows students to actively engage with research and quantitative information, offering the scaffolding needed for quantitative literacy to become habit. This chapter provides a foundation for understanding fundamental concepts from statistics so that we become more statistically literate and critical evaluators of numbers, results, and claims we encounter.

Resources

Arum, Richard and Josipa Roksa. 2010. *Academically Adrift: Limited Learning on College Campuses.* Chicago: University of Chicago Press.

Leonhardt, David and Amanda Cox. 2014. "Heavier Babies Do Better in School." *New York Times: The Upshot.* http://www.nytimes.com/2014/10/12/upshot/heavier-babies-do-better-in-school.html?_r=0

National Council for the Social Studies (NCSS). 2013. The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for Enhancing the Rigor of K-12 Civics, Economics, Geography, and History. Silver Spring, MD: National Council for the Social Studies.