Analytic Study? Or Mere Enumeration?

How does understanding the difference influence your thinking?

Bio: Terry received his master's in education in 2003, and is a Certified Quality Engineer and Six Sigma Greenbelt. More recently he completed a degree in Game Art & Design which he uses every day in teaching digital arts in Boulder, Colorado. Terry is an experienced teacher and parent. For more than twenty years he has brought his passion for continuous improvement into his professional and personal lives. As a Certified Quality Engineer he has experience improving quality in both industrial and software development environments as well as in the educational venue. He has studied improvement theory and practice for nearly 30 years, and was first exposed to Parenting with Love & Logic in 1996. In a twist of fate, he graduated high school with Charlie Fay in 1983. Terry brings a coherent mix of change theory, classroom skills, and relationship development to his classroom, and inevitably to the lives of those around him.

Abstract

In 1975 Dr. W. Edwards Deming authored an article in the American Statistician concerning the difference between enumerative study and analytic study. His intention was to provide clarity of function to practitioners of statistics, and especially to provide a motive force to include the principle in collegiate studies of statistics. He felt statistics texts in particular were woefully misleading in their proposed use of statistics in establishing causality using merely enumerative study. Deming's principle can be applied far more widely in human reasoning as an identifier of misattribution of cause. It assists us in avoiding the use of 'counting' as 'causality', and provides guidance for moving forward with analysis in order to form personal beliefs.

By Terry Rosen

I've been studying quality improvement for decades as a personal passion. As a teacher I often struggled with applications of process improvement in the classroom. For my masters in 2003 I wrote a thesis concerning application of process improvement methods in technology classrooms. Various roadblocks obviously exist in attempting to implement any bottom-up improvement efforts, especially in educational settings. As part of my personal effort I've read widely, and especially followed ideas and principles from Dr. W. Edwards Deming.

Dr. Deming did not write a lot about education specifically, but what he did write was very direct. In both Out of the Crisis and The New Economics he wrote about the fundamental crisis caused by misunderstanding the difference between analysis and enumeration. He reflected on it very critically, and included, in both his books, a reference to his own article, On Probability as a Basis for Action, (Deming, 1975).

I finally got tired of never having read this article, which was destined to be far over my mathematical head, and acquired the magazine itself from eBay in 2008. It was a difficult read, and it was not immediately clear to me what was being said. But in fairness to myself, it did not go completely over my head. I set it aside, and read it again the next day. And the next. Each day brought a bit more clarity.

To my benefit I had also begun studying human reasoning a la Aristotle, Plato, DesCartes, and later Black Swan (Taleb), who led to Kahmeman, (Thinking Fast and Slow). And this is the state I was in when I read Deming's article. Two weeks later it began to truly sink in.

So, what is enumeration? Enumeration is essentially counting something. What is the average lifespan for males in the United States? Or females? What were the company sales figures for 2013? Which counties in the United States have the highest (or lowest) per capita rate of kidney cancer? (Kahneman, 2011). First something has to be counted.

Sometimes the count is all we need. Occasionally we make decisions that require nothing else than the count. How many students are going on this field trip? Twenty six? Then we need twenty-six plus two boxed lunches, (two chaperones).

But sometimes we need to know more. Sometimes we need to know what 'caused' a number. For that we need analysis.

In Deming's article he references sending food to cities during WWII. For instance, how many people are there in Pittsburgh? Well, how exact of a number do we need? That may guide our sampling system for counting some of the people in order to guess at the total with some confidence. Or, if we have infinite resources, perhaps we could count literally every person in Pittsburgh at a given moment. But NO MATTER HOW ACCURATELY WE COUNT, we cannot say ANYTHING AT ALL about why those people are in Pittsburgh. (Deming, 1975.) This is his point. It's lost on the masses, almost entirely. We CANNOT merely count something, and use the count to establish ANYTHING causal.

But we do anyway. From Thinking Fast and Slow we discover the narrative fallacy. It's the inescapable urge to hear a fact and then guess something plausible about why that fact is true. "The United States counties with the highest per capita incidence of kidney cancer have these qualities: rural, sparsely populated, mostly republican." (Kahneman, 2011). Why is this? Why do you think it's true?

When I first read Kahneman's book I fell for every example of every cognitive bias. I likely fell for most of them the second time through as well. I won't ruin the story, it's worth the read, (perhaps the most important book I'll ever read). But for our purposes, the point is clear. Counting kidney cancer by the county CANNOT tell us anything causal about the cancer. To identify cause we MUST do an analysis.

Dr. Deming demonstrates one way to analyze processes that are prone to exhibiting causal elements: Shewhart's control chart. (Deming, 1994) Used well, a control chart can reveal many facets of abnormal variation. But at its heart it is still 'counting'. How does Deming use the chart to go beyond mere enumeration?

Deming guides us to use the chart to measure variation. Then he mentors us in finding ways to reduce the variation, (system variation, or common cause), until most or all of the variation is within +/- 3 standard deviations. These are PRECONDITIONS to analysis. To reiterate, no analysis is possible of a process that is not in control, (within +/- 3 standard deviations).

Once our process is stable, we can change a variable. With the variable changed, we can observe its effect on our chart, (at least twenty time series measurements, preferably 30). At which time we can assess if there was a change, and whether we benefited from the change or not. More importantly, we can say that the change 'caused' the shift on the chart.

Analysis (for cause) requires a stable system, a time series measure, a change to a variable, and another time series measure.

Not insignificantly, the data in the chart must have been either an entire population, or a random sampling of enough items from which to make a conclusion of confidence, (which applies only to the population from which the samples were randomly chosen).

Schools do not have data like this. Nor does human experimentation often lend itself to this level of study. Yet this question comes up every day in class, in various ways.

If Johnny gets a 65% on his quiz, what 'caused' this score? What percent of the content did Johnny know? Why is attendance an average of 2% higher for girls than for boys? (Or is it the other way around?) Do we need to provide free breakfast in our cafeteria every day? If so, why? What time should school start each day in order to promote student success?

These help illustrate the complexity of using numbers to find causes, but do not display how wide the problem of 'assumption' is, nor the extent to which it enters into decision making for adults of all kinds. Nor does it clearly illustrate the nature of the problem at hand.

It is clear from personal experience and 1,000s of animal studies, that animals of all kinds can co-relate events. The most behavioristic studies show that bees can be trained to roll a ball to a spot in order to get sugar. This is done by repeated co-relation of behavior to receipt of sugar. And most people are aware of the caution of using correlation to 'imply' causation. Some even go as far as to refute an argument merely because correlation is being used in a claim of causation, (for reasons we are reviewing here, this is not a bad policy in general). Yet, generally, almost universally, items with a causal nature ALSO co-relate in some way. The problem is not that correlation does not EQUAL causation, but that correlation is not SUFFICIENT to conclude causation.

Indeed, study of Deming's article led me to the following principle:

No causal relationship can be perceived unless first there was perceived a 'non-causal' relationship.

In the experimental description above, researchers will conclude causation by relating the change in the data to the experimental change in the variable, one kind of controlled experiment.

Examples abound everywhere of instances where correlation was badly assumed to be causal. My favorite, described above, involves kidney cancer. Recall "The United States counties with the HIGHEST per capita incidence of kidney cancer have these qualities: rural, sparsely populated, mostly republican", (Kahneman, 2011). Now, mysteriously, we must also learn that "The United States counties with the LOWEST per capita incidence of kidney cancer have these qualities: rural, sparsely populated, mostly republican." How can the lowest, AND highest per capita instance be in counties with the same descriptors? The way the study was performed, regression to the mean caused these results. Sparse population resulted in some counties (at random) having higher per capita values. In other words, those counties appeared 'at random', an effect of randomness was the cause. Randomness, CAUSED, that data distribution. (Taleb, 2005.)

Often a co-relation of variables may indicate a place to look for a causal connection. And a valid study may reveal significant relationships between two or more variables. But mere counting (enumeration) is not sufficient to show causality.

So, too, it is insufficient to count two or more things, to establish causality. We cannot give students a battery of 12 different tests and then say we've performed an analysis that indicates causality. We cannot simply test 100x as many students and say 'now we have enough data to indicate causality. Even if we test every single student in the country, we cannot say anything at all about what caused those scores. So it is that comparing students in two classes from different schools to each other makes no sense whatsoever. Students in science, from two different schools, represent two non-random samples, (or two distinct populations).

Whenever we examine data this way we can say with virtual certainty that the numbers we find will be different, but we will have nothing to say about WHY they are different. This would be true even in two factories owned by the same company using the same brand of machines and same raw materials on the same date. The two sets of data, no matter what differences they were measuring, would not reveal the cause(s) of the differences. Yet all over the country there are school grading services rating schools based on test scores and other enumerated values, presented to suggest that schools with higher scores are 'better' schools, attempting to attract more students to those schools. This is similar to the idea that students with higher IQ will make better students because they are 'smarter' students. Even here we have the difficulty of using an enumeration to make a causal judgment, (actually the issue with IQ is far more dramatic than I present here).

Again, from Kahneman, we become aware of the use of the law of small numbers, (Kahneman, 2011). This ‘law’ is a fallacy that says when we study a small sample from a population that the numbers will be equivalent to the numbers we would get if we studied the whole population. By using statistics we can use our desired certainty level to define how many samples we need to select ‘at random’ from the population in order to perform our study in a valid way. So, if I inspect a box with 50 items in it, I need to check 13 of them, at random, to be fairly sure there are no defects. If I select less than 13, or, if I select them non-randomly (for instance picking the top 13 items), then I’ve destroyed the validity of the study, and can then make no assumptions about the quality as a whole. Or, worse, if I select just one, and then assume the rest are good, I’ve ruined any claim to the quality of the population in the box. I call this enumeration with a sample of one. Worse yet, we might assume that since the last 12 boxes were good, we’ll pass this box without looking in it at all, an enumeration with a sample of zero. (Or, we could even use some sort of irrational bias, conscious or unconscious, to skip a box. I call this enumeration with a sample of -1.)

In Black Swan, Taleb discusses the near ludicrous nature of stock market reports. On one day in December, 2003, soon after the capture of Saddam Hussein, the U.S. Treasuries rose at 13:01 with the Bloomberg headline: “U.S. TREASURIES RISE; HUSSEIN CAPTURE MAY NOT CURB TERRORISM”. Thirty minutes later U.S. treasuries fell, prompting a new Bloomberg headline: “U.S. TREASURIES FALL; HUSSEIN CAPTURE BOOSTS ALLURE OF RISKY ASSETS”, (Taleb, 2007). The same cause, exactly the same, caused the market to go up, and thirty minutes later, to come down. This was a mere correlation with no real basis in fact, a classic post hoc fallacy. (Treasury values fluctuate during the day as a normal occurrence.)

This may sound silly, yet we make judgments this way every day, and hear others doing it everywhere we go. And why shouldn’t we? In general we do no harm making this type of error, and only in special cases should we be truly concerned. But realize this, we perceive a level of certainty that is equal to having performed a scientific study ourselves. And in a field of experts, paid well to make decisions for us, we hope these experts know how to assess their data. Yet I’ve only ever met one person that had even heard of this difference between enumeration and analysis of data, and they claimed it applied in very few situations. (From personal conversation with a professional quality consultant, c. 2005.)

I recently read a book whose author claimed that humans had evolved to discern causal connections. This prompted further examination of my own view that correlation works as causation, most of the time. Indeed, my own feeling is that the mistake of treating correlation as causation generally causes people to be ‘too careful’ rather than ‘too risky’. Kahneman’s work provides support for this showing that fear of loss outweighs desire for gain by a factor of roughly 2x, (Kahneman, 2011). And being too careful rarely gets you killed. This could lead one to believe that ‘correlation as causation’ is an evolutionary trait. Perhaps. But causality? I don’t think there’s any reason at all to believe that discerning causality (as distinct from mere correlation), occurred as an evolutionary adaptation. Indeed, discerning actual causation is so difficult that most people do not fully grasp the known complexities of it. Actuaries seem to be immune to many of the cognitive biases of most everyone else. As they relate to probability and causality I mean, (Kahneman, 2011), but actuaries are made, not born. And actuarial knowledge is born of statistics.

Statistical naiveté leads to the claim that ‘people can lie about anything with statistics’. For this to be true the audience must be partially ignorant of statistical reasoning. I would thus phrase it this way:

> Statistically naïve people can fool statistically naïve people (including themselves), with statistics.
>
> Statistically savvy people can fool statistically naïve people with statistics.
>
> Statistically savvy people cannot fool statistically savvy people with statistics.

Sadly, the exact same kind of statement can be made of unconscious biases. If you don’t know about them, you can’t protect yourself from them. This can be seen everywhere in sales and marketing. Even if you DO know about them, you may still be vulnerable because the nature of unconscious bias is that it is ‘unconscious’.

Put more simply, basic reasoning and logic reveal the prevalent use of logical fallacies in everyday discourse. Some are statistically related, or bias related, and others are just not widely known. Yet logical fallacies have been well understood for over 2000 years. It seems to me that all these pitfalls of reasoning should be regularly included in the education of children. Of course, that would mean teachers, (and parents), would have to know about them. And they do not.

Although many people are aware of fallacies, most people I ask cannot name two. No one I’ve met had already heard of Kahneman or his book about unconscious biases, yet we know that some of these principles are used in sales, (priming is a wonderful example). Only one person I’ve met has ever claimed to know about enumeration vs. analysis. And this person was a professional quality consultant, (implying an understanding of statistical theory and process improvement in particular.) And, he claimed it was an interesting idea but with very few actual uses, and then proceeded to present a lecture on unconscious bias of Kahneman’s studies, without having heard of Kahneman’s book. [Yes, truly bizarre.]

Critical thinking, with reason, logic, and statistical reasoning, are important in evaluating a rationale in decision making. Indeed, fallacies often relate directly to the same concept rendered as a mathematical, philosophical or statistical principle. Quality thinking can be seen as two things, good ideas (rationale), and bad ideas (including fallacies). Reducing bad thinking will improve the effect of your rationale, but avoiding mistakes does not equate to a good idea.

As for this idea, enumeration vs. analysis, does it have any value? Does learning it affect our thinking? It affected mine, but I am an enumerative sample of one. And I was willing to learn it, and dive deep to understand it, and apply it to my teaching as I know Dr. Deming did. Will it affect your thinking?

I hope so. And either way, I'd love to hear from you about its effect, for good or bad. One kind of analytic experiment.

References

W. Edwards Deming (1975) On Probability as a Basis for Action, The American Statistician, 29:4, 146-152, DOI: 10.1080/00031305.1975.10477402

W. Edwards Deming, 2000. "Out of the Crisis," MIT Press Books, The MIT Press, edition 1, volume 1, number 0262541157, January.

W. Edwards Deming, 2000. "New Economics for Industry, Government, Education". MIT Press, 1994.

Kahneman, D. (2011). Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

Taleb, N. N. (2007). The Black Swan: The Impact of the Highly Improbable. New York: Random House.

Taleb, N. N. (2005). Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets. New York: Random House.

This article can be found on LinkedIn @ https://www.linkedin.com/pulse/analytic-study-mere-enumeration-terry-rosen/?trackingId=Kt%2F4tmf1%2B9uNId7P9dCeng%3D%3D