The Vagaries of Averages

The following is a fictional, but not unlikely conversation between the president of a distributorship and the sales manager.

"How did the customer survey come out?"

"Not bad. Overall, we averaged about a 4.7."

"What does that mean?"

"It's somewhere between neutral and satisfied. Better than between neutral and dissatisfied."

"I guess. What are we going to do next?"

"We'll do the survey again next year."

Unfortunately this is the sort of action we take after we accumulate a lot of data; we accumulate a lot more data, but we don't take any concrete steps to use it to define problems and improve our business. Often it's because the answers we get don't point to the actions we need to take.

There are a number of reasons for that. We ask the wrong questions. We ask the wrong people. We tabulate and report the data in the wrong way. In this article we're going to look at a specific case of this last reason, the use of averages where averages really don't apply. Where they are, at best, unclear and, at worst, misleading.

It's easy to understand why we're so fond of averages. We use them in financial reports, employee surveys, customer surveys and almost anything else that has to do with numbers. They turn a mass of numbers into a single, understandable figure. They're easy to do; just dump the data into Excel and use the AVERAGE() function. And they look so official with four or five digits to the right of the decimal.

But the fact that we like them and the fact that they're easy doesn't necessarily make them useful.

In some cases averages are perfectly appropriate. All through school you were concerned about averages; your progress in school and probably your privileges at home depended on them. An 85 average on a course meant that, assuming the tests were fair, you had retained and regurgitated 85 percent of the material. It didn't matter whether the 85 average was two tests with scores of 85 or two tests with a 100 and a 70. The average answer the question: what part of the material did the student learn? It's based on a logically incrementing scale (zero to 100). And it provides a basis for appropriate action. If your average is high enough, you move on to the next course. If not, you repeat the course.

We're so used to this, that we might assume it works everywhere. But whether it does is based on the three things mentioned in the paragraph above.

1. Is the scoring based on a continuous or logically incrementing scale?
2. Does it provide an answer to the proper question?
3. Does the answer indicate what type of action you should take?

Let's take a look at each of the three questions in the light of common business situations.

Is it based on a continuous or logically incrementing scale?

This is so obvious you might wonder why it's even being mentioned. There are some things that we always average (such as test grades), and some things that can't be averaged (e.g.: Which of the following fruits do you like best? Apples, Oranges, Kumquats, Grapes, or Bananas.) The best you can do is calculate the frequency of each answer. But some commonly used answer templates are not so obvious.

A common format for employee and customer surveys presents respondents with a statement and asks them to indicate their agreement on a scale ranging from "Disagree Strongly" to "Agree Strongly." Typically the scale consists of seven points, and the midpoint is "Neither Agree nor Disagree."

A logically incremented scale will always move from one extreme to the other in even intervals. The first three points on this scale (and the last three) do that. It's a predictable step from "Disagree Strongly" to "Disagree" to "Disagree Somewhat." But is it a predictable step from "Disagree Somewhat" to "Neither Agree nor Disagree?" At that point the scale goes from measuring a reaction to the statement to measuring no reaction. The answer may mean that the respondent doesn't have enough information for an opinion, or — more likely — just doesn't care.

A better solution here, if you insist on interpreting the data in averages, would be having a scale ranging from "Absolutely Untrue" to "Absolutely True." However, that might encounter problems in the next section.

Does it provide an answer to the proper question?

In a rational world all tasks requiring resources would be done with a purpose. In other words, if we going to spend time and money on something, we should get something out of it. With information gathering, that "something" should be a tool that we can use in measuring our current status and determining actions to improve it. That status may have to do with financials, customer or employee satisfaction, or internal processes.

The problem is that too often we do not accurately define the question we need to answer. For instance, what's informally known as "AR Days Out" and more formally known as "Average Collection Period" is always reported as an average.

The average collection period number answers the question, "What is our average collection period in days?" That's useful for month-to-month comparisons and for comparing your average to the industry average shown in your industry's PAR.

However, it doesn't provide a meaningful answer to the question: How effective are our credit procedures? Based on the average, you don't know whether the 54 days out means that all of your customers are over 30 days, or what combination of 30-day, 60-day, and 90-day (or, heaven forbid, 120-day) collections make up the 54-day average.

To really answer the question, "How effective our credit procedures?" you need to know the shape of the frequencies — how many accounts fall into each category.

In the customer/employee survey, a comparison of averages from one period to the next might give some indication of movement, it doesn't answer the more compelling question: How many of our customers (or employees) buy into this goal statement?

Similarly, average inventory turns (created using dollars) provides a reasonably useful snapshot of what's happening to your inventory in terms of dollars, but it doesn't answer any questions regarding specific parts of your inventory. And it doesn't provide a specific indication of the action you should take to improve your situation.

That's the third and possibly the most important question.

Does the answer indicate what type of action you should take?

Properly tabulated and presented the statistics should not only tell you what is happening, but give you an idea of what you should do to improve the situation. If the statistics are not actionable, they're not very valuable.

Going back to the examples used above, looking at the number of accounts in each buckets (30-day, 60-day, 90-day, and 120-day) might point you to possible changes in your credit policy or in your customer qualification.

In the customer survey, what we are usually trying to do is determine the level of satisfaction of our customer base. A number of studies have shown that a customer who is simply satisfied is much less likely to remain a customer than one who is extremely satisfied; so we are interested in the number (or percent-age) of respondents that rate us at the top end of the scale compared to the number or percentage that rate us in the middle or at the lower end. (The real difference between a customer being "Very Satisfied" and being "Extremely Satisfied" is debatable; in practice, we usually lump the two answers at either end of the scale together.)

In nearly fifty years of running numbers for businesses, I've encountered a great deal of skepticism regarding statistics (and have heard the "lies, damned lies, and statistics" line more times than I wanted). However, we come back to the fact that the numbers are simply a tool, and — as with any tool — it must be used properly.

This article has been aimed at just one of our common, but improper uses of numbers — the easy, very official looking, and often useless presentation of information using averages, or the even more pernicious "average of averages."

Instead of loading everything into the spreadsheet and having it calculate the average, take a few more minutes and examine the data and what you want from it by asking and answering these three questions:

Can the data, considering the answer form, really be averaged? Is it based on a continuous or logically incremented scale?

What is the real question, and what information do I have to have to get a real answer?

When I get the answer, will it point to the action I need to take to improve the situation?