The Hazards of Examining Just the Average

Authored by:

Paul F. Doruska (CF #1213)
College of Natural Resources
University of Wisconsin-Stevens Point

Upon Request of the
Society of American Foresters Forest Science and Technology Board

---

Consider a scenario where a landowner owns four tracts of forested land or a scenario where an agency operates four visitor centers. The landowner will probably be interested in knowing the average of a given attribute (say trees per acre) in each tract or an agency employee will probably be interested in knowing the average number of visitors per month to each of the centers. Some form of random sampling would then typically be completed in order to estimate the average trees per acre in each tract or the average number of visitors per month to each center.

The following table (Table 1) might be created to convey some of the results from the sampling endeavors. Based on the results of the sampling, the average number of trees per acre from each tract or the average number of visitors per month to the centers was identical (this is an extreme example, but points of discussion are often easier to explain with extreme examples in my opinion…)

Table 1. Average trees per acre in each of the four tracts or average number of visitors per month to each of the four visitor centers

Attribute	Tract A	Tract B	Tract C	Tract D
Average trees per acre	500	500	500	500
	Center A	Center B	Center C	Center D
Average visitors per month	500	500	500	500

With this table, then, the landowner or the employee has the information they wanted to know, namely the average trees per acre in each tract or the average number of visitors per month to each center. *But* *do they really have the information they wanted?* The averages are clearly presented; however, how one views an average is probably more important than the numeric values of the averages themselves.

With the information presented in Table 1, the landowner or the employee will probably want to conclude that the stands are similar with respect to trees per acre or the centers are similar with respect to visitors per month – after all, the averages are all the same within each scenario. This might then lead one to loosely conclude the tracts might be managed, or the centers staffed, in the same manner (please bear with me here, I am acutely aware that one would not jump to directly to these conclusions based on a myriad of other considerations – I am just using the example conclusions to underscore the purpose of this article.) However, it takes more than just the averages to draw conclusions using averages. Read the previous sentence again and remember it, we will come back to it again!)

The best description of the point being made in the previous paragraph is the following paraphrase of a quote I will attribute to Richard Oderwald (Virginia Tech) - I am almost positive he is who I heard it from, so he will get the credit until I am corrected...

> "If all you know is the average from a random sample, then here is what you really know: there is a 50% chance the correct (or true) average is larger than your number, a 50% chance the correct (or true) average is less than your number, leaving a 0% chance your number is correct."

Now there is a reassuring statement if there ever was one, right? One goes through a sampling endeavor to estimate an attribute's average *knowing* there is a 0% chance the answer will be correct. Now go back and reread that statement you were asked to remember. Based on the above paraphrase, one hopefully sees why one needs more than just the averages.

The additional information needed is a measure of the variation present in the dataset. Only when an average is viewed *in combination with* a measure of variation present in one's data does the average become more interpretable (or said another way: more useful.) Let us go back to our sampling scenarios of trees per acre or visitors per month and examine some underlying data (Table 2) that can lead to the results seen in Table 1. We will keep things fairly simple here (again, the point being made is more important than the actual data being used) using a sample size of five plots within each stand or five months for each visitor center.

Table 2. Randomly sampled data from each tract (trees per acre) or each visitor center (visitors per month).

Observation No.	Tract A or Center A	Tract B or Center B	Tract C or Center C	Tract D or Center D
1	200	450	350	0
2	400	450	400	0
3	600	500	400	0
4	600	500	650	0
5	700	600	700	2500

As shown in Table 1, the average of each column of numbers is exactly the same: 500; however, it should now be evident that these are very different tracts with respect to trees per acre or centers with respect to visitors per month. That conclusion might be completely missed though, if one focuses on solely on the average. To draw conclusions of how adequately an average describes a dataset, one must examine the variation present in the dataset in addition to the average. Without going into the all the details about sampling (again, keep the main point of this article in mind here), it becomes evident that the calculated average of 500 better "describes" some tracts or centers compared to others – an examination of the underlying data was required to make that statement. *It takes more than just the averages to draw conclusions using averages*.

So what else besides the average might one want to examine? Three measures of variation to be addressed herein instantly come to mind: standard deviation, standard error of the mean[1], and coefficient of variation. Another measure, confidence intervals (the combination of sampling uncertainty with some measures described herein), also come to mind, but I will save that for another installment in this series.

The first of the measures of variation to briefly discuss is standard deviation. The formula for it will not be shown here as one rarely, if ever, will actually use the formula. Functions built into spreadsheet software or many calculators will perform the calculation for you – how to interpret a standard deviation is far more important to this discussion.

[1] Note that the words "mean" and "average" are essentially interchangeable, However, I am choosing to exclusively use the word average (and not mean) when referring to the sample data. Since the term "standard error of the mean" is a specific statistical term, I will not replace the word "mean" in "standard error of the mean" with the word "average".

Simply stated, a standard deviation is a numeric measure, in the same units as the average, of how observation values vary referenced to (or relative to) the average. Two important concepts are included in the previous sentence: (a.) observation values and (b.) referenced to the average. To better understand these concepts, we will use the visitor per month data from before (the same concepts apply to the trees per acre data, but at this point it will be easier to examine one example and not two) and examine the actual standard deviations calculated from each sample.

Table 3. Standard deviation of visitors per month for each of the four random samples

	Center A	Center B	Center C	Center D
Standard deviation of visitors per month	200	61	162	1,118

One thing to notice in this example is that the datasets that have more variation have larger standard deviations. This is a true statement *if and only if* the units of measure between the standard deviations being compared are identical. I cannot stress this fact enough! Since the units of measure on the four standard deviations being compared are visitors per month, respectively, they are indeed identical, allowing for the direct numerical comparison to be made. At this point then, we recognize that the data from Center D are far more variable than the data from Center B by examining the standard deviation (as opposed to having to scan the actual data) , but how does the standard deviation reference observation values to the average to give the user a sense for how variable the data are?

To answer that question, I do need to introduce just a bit of statistical fact (this will not be too painful, I promise) – in fact, I suspect you have heard the following phrases before.

If data are normally distributed, then

68% of all observations are within 1 standard deviation (in either direction) of the average,
95% of all observations are within 2 standard deviations (in either direction) of the average, and
99% of all observations are within 3 standard deviations (in either direction) of the average.

Allow me to apply that phrasing to our numbers from Center C.

68% of all observations from Center C (recall we presently have just a sample size of 5 observations) are expected to be in the range 500 $\pm$ 162,

95% of all observations from Center C are expected to be in the range 500 $\pm$ 324 (note that 324 is 2 times 162), and

99% of all observations from Center C are expected in the range 500 $\pm$ 486 (486 is 3 times 162).

Look back at the Center C data (Table 2.) Are not the above statements reasonable given what we know about Center C from the sample data?

By referencing the value of standard deviation from the sample to the value of the average from the sample, we have loosely placed “reasonableness bounds” on a given observation’s value – after all about 70% of the observations are expected to be within 1 standard deviation of the average. We can then loosely make the statement that if we took another observation from Center C, we reasonably know the range of values in between which it would probably fall.

So how does knowing the standard deviation help us make more informed statements about averages or compare averages? If the standard deviation is small relative to the average (i.e. Center B), one knows that the data do not vary much about the value of the average, and if the sample size was increased, the average probably would not change very much. Contrast that with the standard deviation from Center D. That standard deviation is large relative to the mean, so as a result, one does not have a good feel of what another observation’s value might be from Center D, and thus if the sample size was increased there is a pretty good chance the mean might move a great deal.

Let us know revisit the visitor center averages in Table 1, but now add a measure of variation, namely standard deviation, to the mix.

Table 4. Average and standard deviation of visitors per month for each of the four random samples

	Center A	Center B	Center C	Center D
Average visitors per month	500	500	500	500
Standard deviation of visitors per month	200	61	162	1,118

Merely by taking a look at the standard deviation in addition to the average, we know considerably more about the averages – namely how tightly packed the data may or may not be about that average. Even though the averages are the same from all four centers (remember that is where started many paragraphs ago), and one might think they are similar as a result, by looking at the standard deviation (a measure of variation) we now know the Centers are fairly different with respect to visitors per month despite the fact the averages are identical. *It takes more than just the averages to draw conclusions using averages.*

Standard deviation is one of three measures of variation that come to mind (there are more by the way) as mentioned earlier. If you understand how one can loosely interpret the standard deviation to provide more information about an average, is there really a need to consider others? I think so. Let me briefly explain.

Recall the standard deviation puts “reasonableness bounds” on an additional observation’s value and we discussed how one can use the standard deviation to judge to what extent an average might move (i.e. a little or a lot) if the sample size was increased. Why not then place “reasonableness bounds” on the average as opposed to an observation’s value. That in essence is a loose definition[2] of the standard error of the mean – “reasonableness bounds for an average”.

If the standard error of the mean is large relative to the average, the average would likely move a fair amount if the sample size was expanded. If the standard error of the mean is small relative to the average, the average would probably stay relatively constant if the sample size was increased. This is the exact same interpretation as the standard deviation, which makes perfect sense because the standard error of the mean is calculated by divided the standard deviation by the square root of the sample size. We could explore standard error of the mean further, but for the purposes of this article, it is not necessary to do so. The only tidbit to add is that standard error of the mean, just like the standard deviation, is unit dependent – comparing them to each other (with the larger standard error of the mean indicating more variable data) is possible *if and only if* the units of measure are the same. Let us now add standard error of the mean to our summary table of visitor center data (Table 5).

Table 5. Average, standard deviation, and standard error of the mean of visitors per month for each of the four random samples

	Center A	Center B	Center C	Center D
Average visitors per month	500	500	500	500
Standard deviation of visitors per month	200	61	162	1,118
Standard error of the mean of visitors per month	89[3]	27	72	500

[2] See also Oderwald, R.G. 1994. Getting more from your cruises. FORS Institute, Clemson, SC. 96p. (note FORS Institute is now defunct)
[3] Note that 200 divided by the square root of 5 equals 89 when rounded to an integer

Note that standard error of the mean conveys the same information as the standard deviation regarding the variability of the data about the average, it just does it in a slightly different manner. So just one of those measures, and not both, would suffice to provide more information about the average.

The final measure of variation to briefly mention is called the coefficient of variation. What distinguishes the coefficient of variation from the previous two measures of variation? Whereas the previous two measures were unit dependent, the coefficient of variation is not, and it is most often expressed as a percentage. Simply stated, the coefficient of variation is found by dividing the standard deviation by the average, then multiplying by 100%. The larger the coefficient of variation, the more variable the data about an average, and the more likely an average would move a fair amount if the sample size was increased. This time, though, there is no *if and only if* clause about requiring the units of measure to be the same when comparing variation across datasets – no such statement is necessary as the coefficient of variation is unitless. Let us add the coefficient of variation to our table of information for the visitor center.

Table 6. Average, standard deviation, standard error of the mean, and coefficient of variation of visitors per month for each of the four random samples

	Center A	Center B	Center C	Center D
Average visitors per month	500	500	500	500
Standard deviation of visitors per month	200	61	162	1,118
Standard error of the mean of visitors per month	89	27	72	500
Coefficient of variation of Visitors per month	40%[4]	12%	32%	224%

The benefit of using the coefficient of variation is actually lost in Table 6, after all, it provides the same information about the variability of the data relative to the average as either the standard deviation or

[4] Note that 200 divided by 500 then times 100 equals 40%

the standard error of the mean in this example. The benefits of the coefficient of variation become apparent when one is trying to compare the variability in data that are in *different* units.

To highlight this point, let us reexamine just the data from Center A. At present, the original data (Table 2) are displayed in visitors per month Allow me to loosely assign 4 weeks to a month for example purposes. Could not these data also be presented in units of visitors per two weeks after dividing the previous data by two? If we did this, our units would be different, but the level of variability about the average should be the same because the data are essentially the same, just expressed in different units or on different scales. (Think expressing volume from a timber inventory in MBF per acre versus BF/acre - we are using the same principle here). Let us know look at the measures of variation calculated from these two versions of the same data.

Table 7. Average, standard deviation, standard error of the mean, and coefficient of variation using visitors per month and visitors per 2 weeks data for Center A.

	Center A Visitors per month	Center A Visitors per 2 weeks ((/2 month)
Data	200	100
	400	200
	600	300
	600	300
	700	350
Average	500	250
Standard deviation	200	100
Standard error of the mean	89	45
Coefficient of variation	40%	40%

Note one might want to conclude that the visitors per month data are more variable about their average than the visitors per two weeks (or 1/2 month) are about their average because both the standard deviation and standard error of the mean are numerically larger for the former. You now know though that this comparison cannot even be made as the data are in different units! Once we remove the units from the picture, via the coefficient of variation, we correctly realize that the data are equally variable about their averages, as the coefficients of variation are equal – this has to be the case as these are the exact same data just expressed on different scales (i.e. in different units)

So where are we? Recall that we started by considering the averages of four random samples, from tracts or visitor centers, with those four averages being numerically equal. As a result, one might incorrectly conclude the four tracts or four visitor centers were similar. We then carefully examined the statement: *it takes more than just the averages to draw conclusions using averages*. Hopefully one now realizes that the combination of an average with a measure of variation is far more informative than using just the average alone. If one understands why seemingly similar means can actually result from very different data, than one might also be able to reason how seemingly different means might result from otherwise similar data. Read the previous statement again – it is an important one; we are essentially reversing the logic we have built to this point.

The concept mentioned at the end of the previous paragraph often arises in scientific literature and/or reports and/or presentations when the averages from various treatments in a particular experiment are being compared (left unaddressed herein are the concepts of sampling uncertainty and confidence intervals, which also play a role in the comparisons being described). I bet there are times when you have examine a summary table or chart that looks like the following (note no specific units are provided – it does not matter for the point being made),

Table 8. Comparison of averages between treatments in experiment XYZ: Averages followed by the same letter were not significantly different.

Treatment	Average
One	14.6 a
Two	10.5 a
Three	3.2 b

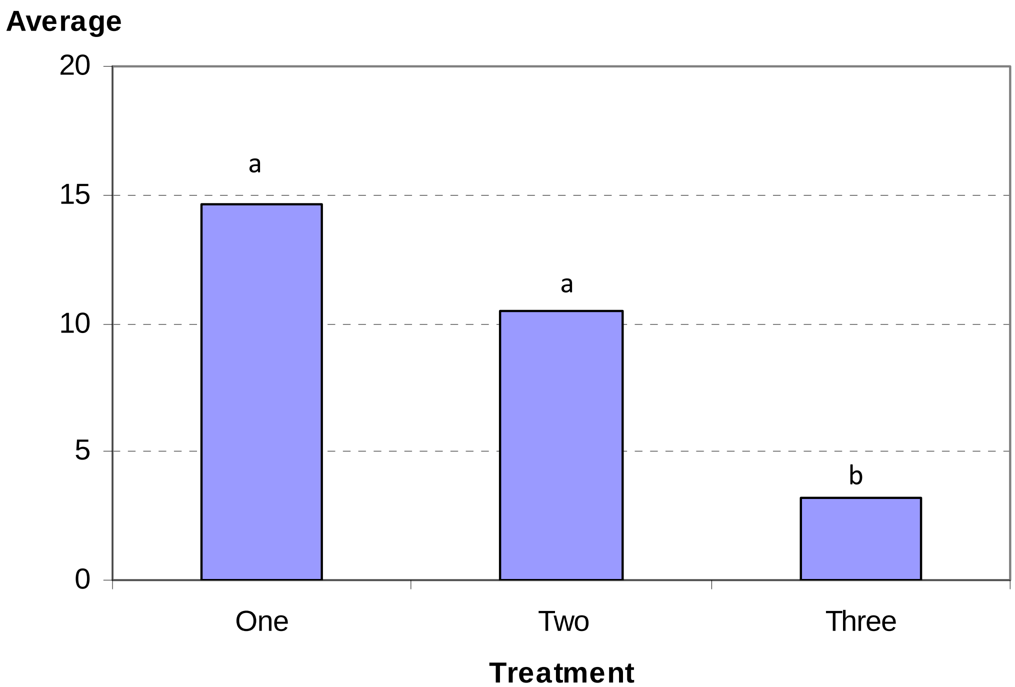

Figure 1. Comparison of averages between treatments in experiment XYZ: averages denoted by the same letter were not significantly different.

In both cases, the conclusions are as follows. Treatment three led to a significantly lower average than either treatment one or two, Also, there is no significant different in the averages of treatments one and two. How can the averages of 14.6 and 10.5 not be different from one another? The answer is rather simple: the level of variation present in the data (though not presented directly in the table or the figure) indicate one or both averages might move a great deal if the sample size was increased (do note sampling uncertainty and probabilities of errors in hypothesis testing both unaddressed in this article, also play roles here). *It takes more than just the averages to draw conclusions using averages!*

I hope you found this article both interesting and useful. Pease keep in mind there is a large amount of material I left out, but in my opinion, what I have left out is not needed to gain a basic understanding of the statement: *it takes more than just the averages to draw conclusions using averages* and the realization that averages and measures of variation should always be examined in unison.

I have avoided using any symbols in this article in an effort to improve understanding. The reader should be aware that symbols do exist for many of the terms used herein. For example, the coefficient of variation is often abbreviated CV. Note that Greek letters are used to represent the true but unknown population values, whereas English letters are used to represent the sampling derived estimates of those true but unknown values. The following table summarizes some of the most commonly used symbols.

Table 9. Common symbols for terms used herein.

	True but unknown population value	Sampling derived estimate of the true but unknown population value
Average	μ (the Greek letter mu; pronounced "mew")	either $\overline{x}$ or $\overline{y}$ (read as " x bar" or "y bar")
Standard deviation	σ (the Greek letter sigma)	s
Standard error of the mean	$\sigma_{\overline{x}}$ or $\sigma_{\overline{y}}$	$s_{\overline{x}}$ *or SE*

Author's note: the next installment in this series is expected to be address sampling uncertainty and confidence intervals. Each was mentioned but left unaddressed in this article.

