Cruise Percent Error

www.arkansastimber.info
Paul Doruska, Ph.D, RF, CF - UAM School of Forest Resources
Teddy Reynolds, BSF, RF, SR - Reynolds Forestry Consulting – RFC Inc.

Cruise or appraisal information catching a landowner's initial attention are:

1) Total Volumes or Tons.
2) Total Value.
3) Percent Cruise (example: sampled area of 10% using 1/10th acre plots with 1 plot/acre).

As a landowner, the above information, no matter how reputable the contracted cruiser, is only as valuable as one shoe (as pointed out in the preceding article). The other shoe is "percent error" (which is different from percent cruise mentioned above).

A professional cruise should always have the following information provided, and never accept or pay for less:

1) Total Tons, Volume and Tree Count by dbh and height.
2) Percent Error (preferably based on a 90% probability).

This article will illustratively walk you through the origin of percent error providing understanding into its practical use. The following cruise statistics will be chronologically discussed and related for easy understanding:

1) **Tons per Plot.**
2) **Average and Total Tons.**
3) **Standard Deviation.**
4) **Coefficient of Variation.**
5) **Confidence Interval (probability).**
6) **Error.**
7) **Percent Error.**

1) Tons per Plot:

Consider the following four stands, each of which are 30 acres in size and were inventoried using five plots. Typically, more than five plots are needed, but for the sake of illustration, five will be used. The timber weight (tons) per acre was estimated from each plot in each stand, and these numbers appear below.

Plot #	Stand A	Stand B	Stand C	Stand D
1	80	90	70	100
2	40	50	60	0
3	60	60	60	0
4	80	30	50	100
5	40	70	60	100

The first plot in Stand A was measured and from it we estimated a weight per acre equal to 80-tons per acre. The other four plots in Stand A, estimated at 40-tons per acre, 60-tons per acre, 80-tons per acre, and 40-tons per acre, respectively.

Now, whether estimating the average timber weight per acre in each stand or the total timber weight in each stand, we know that we first estimate the average weight per acre and then if we want to estimate total weight in the stand, we simply multiply the average weight per acre times the number of acres in the stand. Additionally, the correct acreage needs to be determined before being multiplied or extrapolated by the average weight or volume per acre (correct acres are the stand acres minus excluded acres in roads, power lines, well sites, pasture, buffer strips, etc.). The most accurate way to calculate the correct acreage is through digital orthographic quadrangle aerial photographs in ArcView software or by on ground GPS (global positioning system).

2) Average and Total Tons:

The same inventory numbers appear in the next table, just sorted from smallest to largest within each stand to simplify viewing. Beneath each column is the average of the numbers in that column – that's exactly how the average is estimated in an inventory – it's the average of the numbers from each plot. Beneath the respective averages appear the estimates of the stand total (the average weight per acre times the number of acres in the stand, 30 in each case for this example).

Plot #	Stand A	Stand B	Stand C	Stand D	
1	40	30	50	0	
2	40	50	60	0	
3	60	60	60	100	
4	80	70	60	100	
5	80	90	70	100	
Average	60	60	60	60	Tons per Acre
Total	1,800	1,800	1,800	1,800	Total Tons

In each case, the estimate of the average weight per acre equals 60 tons per acre and the estimate of the total tons in the stand is 1,800 tons. These numbers were setup to be equal to prove an invaluable forthcoming point.

Let's say you owned all five stands, had these inventories contracted, and were told only these results – the estimate of the average timber weight per acre was 60 tons per acre and since each stand was 30 acres in size, the estimate of tons in each stand is 1,800 tons. The result table would look like this:

	Stand A	Stand B	Stand C	Stand D	
Average	60	60	60	60	Tons per Acre
Total	1,800	1,800	1,800	1,800	Total Tons

If this is all you saw or all the information you were provided from the forest inventories, you would probably think all four stands were pretty similar, that is, all four stands looked about the same.

The enlightening fact is the above four stands are actually very different from each other.

For example, Stand C is the most uniform of the four, with the tons per acre estimated from each plot ranging from 50 to 70 tons per acre. If you were to have another plot measured in Stand C, what would be a good guess for what that plot's estimate of tons per acre would be? Given what I know about these five plots, I'd say it's most likely going to be between 50 and 70 tons per acre.

Let's take a closer look at Stand A. If you were to have another plot installed and measured in Stand A, what would you guess its estimate of tons per acre to be? I'd say somewhere between 40 and 80 tons per acre given what I already know about the stand. But considering the range of numbers, I am not as confident in my guess for Stand A as I was for Stand C.

Now, let's consider Stand D. If you were to have a sixth plot measured in this stand, do you care to guess what its estimate of tons per acre would be? I sure would not want to guess given what I've seen in the previous plots.

Which gets me back to the invaluable point, presently all you know from the forest inventory is the average volume or weight per acre (or total in the stand), and consequently you don't know enough to judge the results of the forest inventory. If you owned these four stands and all you knew were the average tons per acre, you would be inclined to think all four stands were similar. Now, let's back up one step and assume you owned just one stand, and were told the average tons per acre in your stand was 60 tons per acre and the estimate of the total in the stand was 1,800 tons. Given just this information, you don't know if your stand is more like Stand C or more like Stand D. Don't you think you should and deserve to know more? That is your first invaluable point.

3) **Standard Deviation:**

The reality about any estimate of an average volume or weight per acre obtained from a forest inventory that uses plots is that there is a 50% chance the true, but unknown, average is larger than your estimate; and there is a 50% chance the true, but unknown, average is less than your estimate. That is a sobering thought. If all you know is the estimate of the average, that does not do you much service. We need to know how good the estimate of the average is.

An estimate of how our plots vary about the average needs to be obtained. If the plots don't vary much above and below the average as in Stand C, you should feel pretty good about your average. If the plots vary a great deal above and below the average as in Stand D, you should be worried about the estimate of the average. Look back at the previous example and see if you agree with that logic.

There is a statistical attribute called standard deviation that represents plot-to-plot variation about an average. Simply stated, the larger a standard deviation calculated from a number of plots, compared to the average calculated from those plots, the more worried about that average you should be. Let's go back to our four stands below and have a look. Beneath the average is included the standard deviation calculated from the plots in each stand. We will stop discussing stand totals at this point because they are completely dependent on the averages. Clearly, if the estimate of the average is good, the estimate of the total in the stand will be good (and vice versa).

	Stand A	Stand B	Stand C	Stand D	
	40	30	50	0	
	40	50	60	0	
	60	60	60	100	
	80	70	60	100	
	80	90	70	100	
Average	60	60	60	60	Tons per Acre
Standard Deviation	20	22	7	55	Tons per Acre

First, notice the units on the standard deviation will always be the same as the average (tons per acre in this case). That's good because it's easy to relate one to the other since they are in the same units of measure. Remember I asked you to guess what the estimate of tons per acre would be from a theoretical sixth plot put in each stand? Well, that's really what standard deviation tells us without having to look at the raw data. Subtract the standard deviation from the estimate of the average, and that's the lower limit (bound) on what another plot would reasonably be expected. Add the standard deviation to the estimate of the average and that's the upper limit on what another plot would reasonably be expected.

Once we look at the standard deviation (a measure of variation) in combination with our average, we know how good our estimate of the average is. That is why I said if all you look at in a forest inventory is the average, you do not know enough. If the standard deviation is small relative to the average (Stand C), we should be happy with our estimate of the average. If the standard deviation is large relative to our average (Stand D), our estimate of the average is not worth the paper it's written on. Always look at the average and a measure of variation to judge the applicability of your estimate of the average.

4) Coefficient of Variation (CV):

There is another measure of variation that one can look at that combines the average and standard deviation. The reason you might want to look at this new one is the standard deviation is somewhat dependent on the units of measure used. A funny thing happens with regards to standard deviation if the numbers from the plots are "large"; the standard deviation tends to also be a "large number". If the numbers from the plots tend to be "small", the standard deviation from those plots tends to be a "small number". Let's look at an example using Doyle Log Rule Board Feet per acre (a measure of volume, abbreviated BF). Look at the following estimates of board feet per acre resulting from eight plots measured as part of a forest inventory (these plots are unrelated to the previous on-going plot illustration).

Plot #1	**5,000**	**BF/acre**
Plot #2	**8,000**	**BF /acre**
Plot #3	**3,500**	**BF /acre**
Plot #4	**6,000**	**BF /acre**
Plot #5	**9,500**	**BF /acre**
Plot #6	**4,000**	**BF /acre**
Plot #7	**6,000**	**BF /acre**
Plot #8	**7,000**	**BF /acre**

The standard deviation calculated from these eight plots equals about 2,000 board feet per acre. That's a large number, but does that imply our estimate of the average is wild?

Since board feet per acre numbers tend to be large, they are often reduced in terms of thousand board feet per acre (abbreviated MBF/acre) which equals board feet per acre divided by 1000. For example, 10,000 BF/acre equals 10 MBF/acre. Let's rewrite our list of eight plots in terms of MBF/acre instead of BF/acre and recalculate the standard deviation.

Plot #1	**5.0**	**MBF/acre**
Plot #2	**8.0**	**MBF /acre**
Plot #3	**3.5**	**MBF /acre**
Plot #4	**6.0**	**MBF /acre**
Plot #5	**9.5**	**MBF /acre**
Plot #6	**4.0**	**MBF /acre**
Plot #7	**6.0**	**MBF /acre**
Plot #8	**7.0**	**MBF /acre**

Using the MBF/acre estimate from our 8 plots, we get a standard deviation of about 2 MBF/acre. Two sure sounds like a better standard deviation than 2,000. Well, we can't really say that because our units of measure are different. This is what I was talking about when I said, "large numbers" from plots tend to produce "large" standard deviations and "small numbers" from plots tend to result in "small" standard deviations. In this case, the amount of variation represented by each of the two standard deviations is identical – after all they came from the same eight plots – just the units of measure on the volumes were different! Standard deviations must always be examined relative to the averages calculated from the same numbers.

In the case of the BF/acre numbers above, the average found from these eight plots is 6,125 BF/acre. For the MBF/acre numbers above, the estimate of the average is 6.125 MBF/acre. To be correct, we need to look at the standard deviation of 2,000 BF/acre in conjunction with the average of 6,125 MBF/acre, and we need to look at the standard deviation of 2 MBF/acre in conjunction with the average of 6.125 MBF/acre.

And that's where this next important measure of variation comes into play. It's called the Coefficient of Variation (CV), and it is found by dividing the standard deviation from a series of plots by the average calculated from those plots, and then multiplying the result by 100%. It takes the units of measure out of the picture! It is a relative measure of variation expressed as a percent. Simply stated, the larger the CV, the poorer and less reliable your estimate of the average and the more worried you should be.

Following are the BF/acre and MBF/acre examples side by side.

	5,000 BF/acre	**5.0 MBF/acre**
	8,000 BF /acre	**8.0 MBF/acre**
	3,500 BF /acre	**3.5 MBF/acre**
	6,000 BF /acre	**6.0 MBF/acre**
	9,500 BF /acre	**9.5 MBF/acre**
	4,000 BF /acre	**4.0 MBF/acre**
	6,000 BF /acre	**6.0 MBF/acre**
	7,000 BF /acre	**7.0 MBF/acre**
Average	**6,125 BF/acre**	**6.125 MBF/acre**
Standard Deviation	**2,000 BF/acre**	**2.0 MBF/acre**
CV	**33%**	**33%**

Even though the averages and the standard deviations are different, the CV's are the same. That's because CV takes the units of measure out of the picture and relates the standard deviation directly to the average. Since the volumes are the same in both scenarios (they are just expressed in different units) the CV's are the same.

Let's go back to our original four stands, and calculate the CV for each inventory.

Stand A	**Stand B**	**Stand C**	**Stand D**
40	**30**	**50**	**0**
40	**50**	**60**	**0**
60	**60**	**60**	**100**
80	**70**	**60**	**100**
80	**90**	**70**	**100**

Average	**60**	**60**	**60**	**60**	**Tons per Acre**
Standard Deviation	**20**	**22**	**7**	**55**	**Tons per Acre**
CV	**33%**	**37%**	**12%**	**92%**	

Do you see how CV combines the standard deviation and average together to instantaneously tell us if we have a good estimate of the average? If the CV is "small" we have a good estimate of the average. If our CV is "large", then we have a poor estimate of the average.

If a CV equals 100%, that means the variation present from plot to plot in the stand is as large as the average. That is a lot of variation. The average from that inventory is basically useless. If the CV equals 50%, that tells us the variation from plot to plot in that stand is about half as large as the average. A CV of 50% is surely better than one equal to 100%, but even with a CV of 50% there is still quite a bit of variation present and we should not feel good about our average. Since forest stands will always have variation in them, CV's from 15%-30% are about as low as they can reasonably be. So if you have a CV in that ballpark, you should feel pretty good about that estimate of the average

Clearly, we want the CV calculated in any inventory to be "small" before we feel good about the average calculated from that inventory. Again, that invaluable point, we need more than just the estimate of the average to get the full picture of any stand.

5) Confidence Interval and Probability:

There is one more aspect of a forest inventory that we must account for when looking at inventory results – and that aspect is uncertainty.

Most forest inventories using plots follow a grid system within the stand. That is, a series of lines along which plots are measured are theoretically placed in the stand, with each line being a certain distance apart from the other. Plots are installed at points along these lines with the plots being the same distance apart.

Let's say you and I each independently conducted two forest inventories in a single stand using the exact same number of plots. Let's further assume we both measure all trees correctly. The only difference is that I put my plots 20 feet west of your plots. After finishing the fieldwork, we both return to the office and work up the results. Will we get the same estimates of the average weight per acre for this stand? We would like to think so because we put the same number of plots in the same stand. However, the reality is that our estimates of the average weight per acre will be slightly different from each other. This is the direct result of each of us measuring a slightly different series of plots (remember, my plots were all 20 feet west of your plots). Does that mean my estimate is more correct than yours, or that your estimate is more correct than mine? No, so what's going on here? The answer is uncertainty.

One important end result of any forest inventory is something called a confidence interval. A confidence interval puts an upper and lower bound on our estimate of the average, based on the average we obtained, a measure of variation about the average (the standard deviation), the number of plots we actually measured, and a factor for uncertainty. The confidence interval, then, combines the three important statistical factors involved with a forest inventory: the average, a measure of variation, and uncertainty. If the upper and lower bounds are close to the average, we should feel good about the estimate. If the upper and lower bounds are far away from the average, we should not feel good about the average.

Confidence intervals have a percentage attached to them called "probability". For example, we might calculate a 95% probability for our confidence interval, or a 90% probability for our confidence interval, or an 80% probability for our confidence interval, and so on. The probability percentage chosen for a confidence interval basically answers the question "how sure are you?" For example, let's say the results of a forest inventory tells us we have an average of 35 cords per acre, a standard deviation of 17 cords per acre, and we have an 80% probability for our confidence interval estimate of the average cords per acre; and it equals [28 , 42] cords per acre. Note that confidence intervals are written [lower bound, upper bound] followed by the units. We interpret this interval as follows: we are 80% sure the true, but unknown, average cords per acre in that stand is somewhere between 28 and 42 cords per acre. It could be 28, it could be 42, and it could be 30. All we are saying is that we are 80% sure it's at least 28 and no more than 42 cords per acre. Now instead of only an average and a measure of variation, we also have established a percentile range that we can be certain contains the true, but unknown average.

Here is another illustration for confidence intervals and probability: You could perform 100 inventories in a stand and for 80 out of those 100 inventories the estimate of the average cords per acre will be between 28 and 42 cords per acre. The average will always be the center of the confidence interval, so that's how we can judge whether our bounds (limits) are "close" to the average. As a result, the average minus the lower bound will always equal the upper bound minus the average (we'll call that quantity "error" as you will see in a moment). For the example, the average is 35 cords per acre, and the lower and upper bounds of the confidence interval are both 7 cords per acre away from the average (that's pretty good, as 28 and 42 are reasonably "close" to 35). The upper and lower limits of the confidence intervals put reasonable bounds on the estimate on the average itself.

6) Error:

Let's take a quick peek at how confidence intervals are calculated and put all the pieces together. A confidence interval from a forest inventory, involving plots, is calculated as follows:

Average $\pm$ Error

The estimate of the average centers the interval on a number line, the error component sets the bounds (subtract to get the lower, add to get the upper) of the interval. You can think about the confidence interval as putting your arms around the estimate of the average. For an estimate of an average to be good, the bounds (your arms) should put a pretty tight squeeze, if you will, on the average. For the [28, 42] cords per acre example our average is 35 and our error is 7 so we have 35 $\pm$ 7 for average $\pm$ error.

The real "guts" to a confidence interval is that error component. The average is literally just along for the ride. So let's take a closer look at the error component.

$$\text{Error} = (\text{something called a T value}) \times \frac{\text{standard deviation}}{\text{square root of the number of plots}}$$

The quantity called the "T value" is how uncertain enters the picture (how sure you want to be or a measure of probability). The more sure you want to be, the larger the T value will be. The surer you want to be, everything else equal, the wider the interval must be. Think about it, if you want to be 95% sure (capture the average 95 out of 100 times if you performed 100 inventories), the interval should be wider than if you only wanted to be 80% sure (capture the average only 80 of 100 times).

Here is a table of easy to use T values based on how you sure want to be, or said another way, how confident you want to be. Note, these are not exact values, for that you'll need a complete table found in any statistical textbook, but for our uses here, these numbers will work just fine.

How Sure Are You (Probability)	T value to Use
99%	2.6
95%	2.0
90%	1.6
80%	1.3
70%	1.0
60%	0.8

The more sure you want to be, then, the larger the T value and the larger the error component.

The amount of plot to plot variation present enters the picture through standard deviation. The only other component is the square root of the number of plots used. Hopefully you can see the impact of variation on the confidence interval. The more variable the plots, the larger the standard deviation will be. Since the standard deviation is in the numerator of the fraction, the larger the standard deviation, the larger the error component will be and the farther away from the average the bounds will be. Recall that if our bounds are far away from our average, we should not put much faith in the average.

Now let us look at confidence intervals for our four stand examples.

	Stand A	Stand B	Stand C	Stand D	
Average	60	60	60	60	Tons per Acre
Standard Deviation	20	22	7	55	Tons per Acre
90% Interval	[46, 74]	[44, 76]	[55, 65]	[21, 99]	Tons per Acre
Error	$\pm$ 14	$\pm$ 16	$\pm$ 5	$\pm$ 39	

80% Interval	**[48, 72]**	**[47, 73]**	**[56, 64]**	**[28, 92]**	**Tons per Acre**
Error	**± 12**	**± 13**	**± 4**	**± 32**	
70% Interval	**[51, 69]**	**[50, 70]**	**[57, 63]**	**[35, 85]**	**Tons per Acre**
Error	**± 9**	**± 10**	**± 3**	**± 25**	

Move up or down any stand and you'll see the impact of changing "how sure you want to be". Move between stands for any given interval to see the impact of changing the level of variation.

For our estimate of the average from our inventory to be good, and one we could literally take to the bank, we want a "small" error at an acceptable level of confidence (degree of "how sure you want to be"). We've been saying that our average from Stand D was weak and the confidence interval again confirms this. We've said all along our average for Stand C was one we should feel good about and the confidence interval again confirms this. If all we looked at were the averages, though, we would have missed this boat completely, as the average for Stand C and for Stand D are both equal to 60 tons per acre.

7) Percent Error:

One final thought. Many folks want to discuss the error associated with a forest inventory. Many people assume the error associated with a forest inventory to be equal to what is called the "percent cruise" of the inventory. The percent cruise for an inventory is the total acreage actually measured divided by the total acres in the stand and then multiplied by 100%. So for example, if the acreage of a stand equals 200 acres and the total area measured across all plots equals 20 acres (200 1/10 acre plots or 1 plot per acre), then the percent cruise for this inventory is (20/200) times 100% equals a 10% cruise. Unfortunately, many people mistakenly assume that the percent cruise is also the percent error for the inventory. Sorry, but this is just not the case.

The good news is you already know what the error associated with any inventory is: it's the plus or minus component of the confidence interval (average ± error). The correct percent error for an inventory then, is the following:

Percent Error = (Error/Average) x 100%

Let's look at the "percent error" associated with 90% confidence intervals for our four stands:

	Stand A	**Stand B**	**Stand C**	**Stand D**	
Average Tons	**60**	**60**	**60**	**60**	**tons per acre**
Total Tons	**1,800**	**1,800**	**1,800**	**1,800**	**tons in the stand**
Standard Deviation	**20**	**22**	**7**	**55**	**tons per acre**
CV	**33%**	**37%**	**12%**	**92%**	
90% Interval	**[46, 74]**	**[44, 76]**	**[55, 65]**	**[21, 99]**	**tons per acre**
Error	**± 14**	**± 16**	**± 5**	**± 39**	
Percent Error	**23.3%**	**26.7%**	**8.3%**	**65.0%**	

The larger the percent error, the poorer estimate of the average. The smaller the percent error, the better estimate of the average. Our estimate of the average tons per acre for Stand C still looks great, whereas the average tons per acre calculated for Stand D proves to be useless. Would you have known this if all you had to examine was the average or total for tons or volumes? Hopefully, now you know the rest of the story.

We have come a long way, and the moral to the story is to look at more than just the average and total when considering the results of a forest inventory based on plots. To get a complete picture of the average volume or weight per acre in your stand, it's important to consider the "percent error". Other optional statistics to view are the confidence interval and associated error, the coefficient of variation, and the relative magnitude of the standard deviation compared to the average.

When purchasing a cruise, pre-negotiate it contingent on the following:

Intended Use	Percent Error	Probability
Management Planning < 100 acres	10% - 20%	80%
Timber Sale < 100 acres	5%	90%
Timber Sale > 100 acres	5%	95%

If a 5% cruise error for a timber sale cannot be achieved based on a 90% - 95% probability due to a high coefficient of variation, then a 100% tree-count (marking) should be performed instead of a cruise (100% tree-counts for timber sales are always statistically more accurate than a cruise).

Inventories should be planned and conducted to obtain a useful interval as in Stand C and not an inadequate interval as in Stand D; and you will only know that, if you rank the cruise results with the invaluable “percent error”. Now you have a pair.

.