Statistics for Rifle Shooters

As a relative new comer to rifle shooting, there are many things which I have had to confront and master. Equipment selection, developing a position, managing natural point of aim, trigger control, breath control, wind reading and load development are only a few. On all occasions, I have always been able to find other very proficient shooters willing to share their wealth of knowledge and technical mastery of these basic skills. With this help comes another aspect the new shooter has to learn to negotiate while collecting all of this information: how to rectify the inevitable conflicting information provided by the experts -- and even the high masters. Having come from a background of engineering, I have been formally educated in the ways of physics, mathematics and even statistics. It is this last bit of education that forces me to write this paper. Having listened very carefully to the experts and the high masters, I am convinced that crimes against science are being perpetrated on a regular basis with a device known as the chronograph. As with all good engineers, I will attempt to first break you down into utter confusion by presenting a bunch of math without numbers as my wife likes to call it. From this, we will slowly uncover the mysteries and hidden pitfalls associated with using the statistical functions built into all chronographs and finally we will end with a practical guide to making useful, scientifically valid conclusions concerning load development with a chronograph.

Before we get to the math, let's first get you involved in the rifle, the load and the chronograph used. One thing a newcomer quickly learns is that it is important to have toys. I prepared 50 rounds at the same loading session using sorted .308 Lapua brass with primer pockets uniformed, trimmed to length and flash holes deburred; 46.9 grains of Varget in a 29" Obermeyer barrel on a trued Quadlock action; 155 SMK loaded to OAL of 2.775" to provide 0.025" jump in WTC95 chamber done by Al Warner; CCI-BR2 primers and my new CED Millennium chronograph. I fired the shots from prone at 200yds as part of my pre-season practice. The individual shots and their respective velocities in the order fired are plotted in Figure 1. The chronograph gave the following results for the 50 shot string:

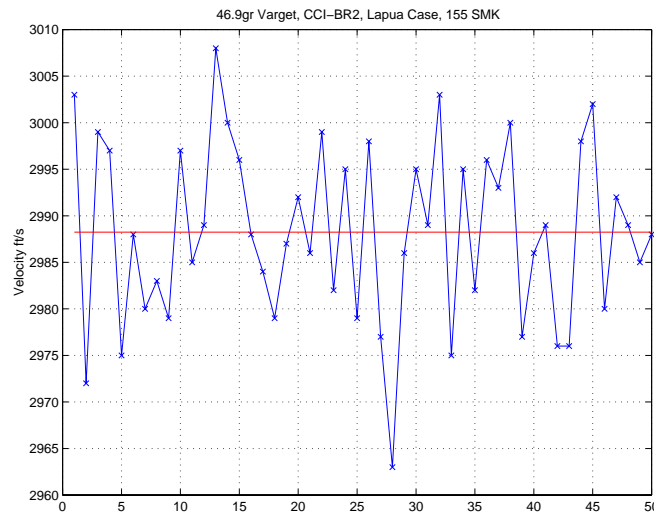MV=2988 ft/sec, SD = 9.6 ft/sec, ES = 46 ft/sec.

Figure 1: Velocity Data Gathered for 50 Shot String in Order Fired

So what exactly does the mean velocity, standard deviation and extreme spread given by the chronograph mean in terms of the load quality. The mean velocity is an average of all the shots. To calculate the mean, add all the shot velocities together and divide by the number of shots. A statistician would write

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

That looks intimidating – but what does it mean? It means add all the shot velocities together and divide by the number of shots. The mean indicates the typical velocity of the load and would be used to confirm the new load will have similar elevation and drift as the old one or used with a ballistics program to establish elevation zeros or confirm the bullet would remain supersonic at the target. The standard deviation is a measure of the average difference between each shot and the mean velocity. The numbers 4 and 6 have a mean of 5 as do the numbers 0 and 10, but their standard deviations are different. A statistician would write

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{N - 1}}$$

The extreme spread is just the difference between the highest and lowest shot velocities. Again our statistician would write

$$R = \max(x) - \min(x) .$$

The standard deviation and extreme spread are measures of the shot to shot uniformity of the loads and typically are used to estimate the potential accuracy of long range loads. The equations look very intimidating, but fortunately the chronograph already has these in its memory and we don't need to know how to calculate the numbers. We should, however, understand roughly what they mean and how to properly use the results provided by the chronograph. A statistician generally likes to look at data plotted in what is known as a histogram. A histogram plots the velocity range and the number of times a shot was measured at that velocity. A histogram of the 50 shots is shown in Figure 2. To illustrate, of the 50 shots fired, there where 7 between 2987 and 2990 ft/sec.   The mean attempts to estimate what velocity is most likely to occur or the peak of the histogram. The standard deviation and extreme spread attempt to measure how wide the histogram is. Figure 2 is a very common profile for histograms that statisticians refer to as "normal". Note that most shots were near the middle of the range and the farther away from the middle or mean velocity, the fewer shots occurred. Generally, for a normal distribution, the width or extreme spread of the histogram is equal to about six times the standard deviation.
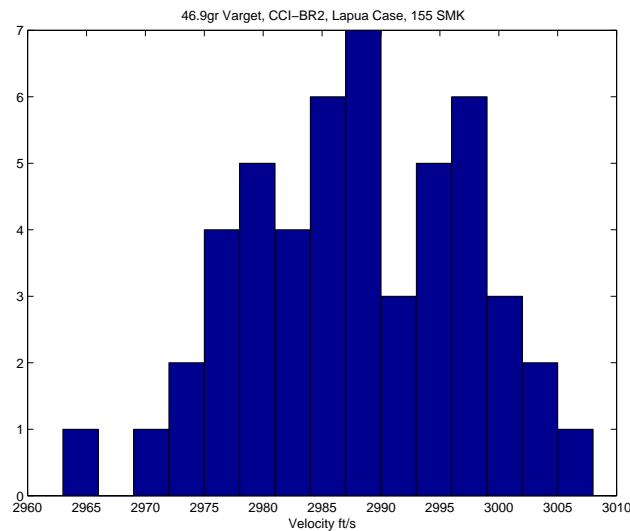

Figure 2: Histogram of 50 Shot String Velocities

Great. The complex equations are now through and if I am a good engineer you are now sitting there wondering if there is a point to all this and when am I going to get to it. One way of finding the best load for my rifle is to load 50 rounds of every case, primer, powder weight, seating depth, bullet combination I can think of. Fire them all, carefully collect the statistics from the chronograph and figure out which has the best numbers. Right – and with the remaining 4 or 5 ounces of powder from the 2 eight pound jugs and the 75 or so shots I have left in the new barrel before it gives out, I will go shoot a 600yd match before I start all over again. Obviously, we need a more practical way of estimating the potential quality of loads with much smaller numbers of tested rounds.

We will investigate scientifically sound methods of using 5, 10 or 20 shot strings to characterize a load combination. To do this, artificial strings will be chosen from the 50 shots already fired. For example, forty five 5 shot strings can be produced as follows:

String 1: Shots # 1,2,3,4,5
String 2: Shots # 2,3,4,5,6
…
String 45: Shots # 46,47,48,49,50.

Note that these were actual sequences fired and if only 5 shots from the 50 were fired, it could have easily been any one of these. Similarly forty 10 shot strings and thirty 20 shot strings could be produced.

Examining these hypothetical strings from the real measurements will allow investigation of the various methods of characterizing the performance of the 50 rounds. Figure 3 shows the statistics for the possible strings. The top plot shows the mean velocity for the strings. For each string, the corresponding 5(red), 10(green) and 20(blue) shot mean velocity can be looked. For example the fifth 5 shot string had a mean velocity of 2984 ft/sec and the tenth 10 shot string had a mean velocity of 2992 ft/sec. Similarly the standard deviations and extreme spreads for the strings can be read from the second and third plots in Figure 3.
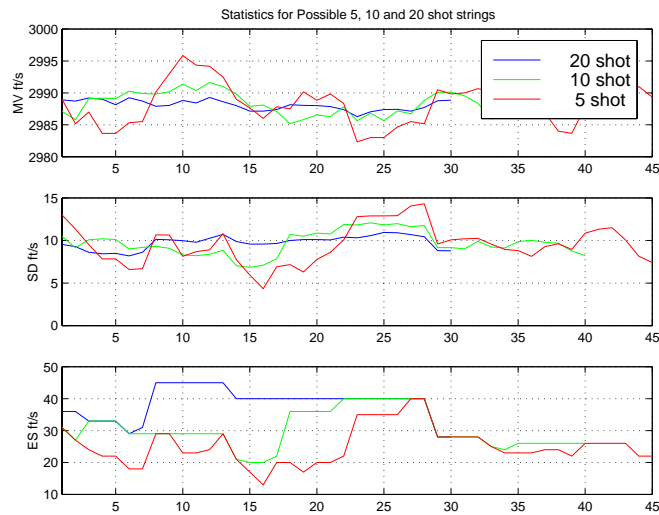

Figure 3: Statistics for Possible 5, 10 and 20 Shot Strings

Recall that the statistics for the entire 50 shot string given by the chronograph were:

MV=2988 ft/sec, SD = 9.6 ft/sec, ES = 46 ft/sec.

These represent the very best characterization of the velocity uniformity of the 50 rounds. Any analysis based upon smaller 5, 10, or 20 shot samples must be judged based upon its ability to reproduce these

numbers. The first and most important fact presented in this paper is that random 5, 10 or 20 shot sequences all result in different statistics than the ones calculated for all 50 shots. In statistics, we refer to these smaller sequences as samples of the 50 shot population and the statistics generated by the samples are referred to as estimates of the population statistics. Thus, if during the winter off season I load 1000 rounds for competition, I may go to the range in the spring to estimate the expected performance of the total 1000 round population by randomly selecting 5, 10 or 20 rounds to test. I am not interested in the statistics of this sample shot sequence so much as we are interested in estimating the performance of the entire 1000 rounds. As can be seen from the data presented in Figure 3, it is very unlikely that the statistics from this sample will be exactly that of the 1000. To be clear, if you fire 5 shots from the 1000 you made over the winter over your chronograph and it tells you the SD = 4, then the standard deviation calculated for those 5 shots is exactly 4. But the standard deviation of all 1000 rounds is unlikely to be 4. The number presented on the chronograph is just an estimate of the total performance and the chronograph will not tell you how good a "guess" it is. So how do we know what is going on?

The mathematics of statistics was designed to answer these very questions. Statistics strives to prevent random chance or "luck" from influencing conclusions made using samples. Statisticians might ask questions like "based upon a sample, how good is the likely performance of the population?" or "what is the likelihood that the population represented by one sample is better or worse than the population represented by another sample?" Similarly, competitive rifle shooters often ask questions like "does my new lot of Varget yield the same velocity as my last lot?", "will Winchester or CCI BR primers provide the most uniform velocities?", or of more critical interest, "does Ray always win because he has something better than I do?"

To begin the investigation, let's take a closer look at the statistics for the hypothetical strings in Figure 3. For the 5 shot strings the lowest mean velocity was 2982 ft/sec and the highest was 2996 ft/sec. Thus, no matter which 5 rounds we chose, we would have been no more than 8 ft/sec away from the true answer we were seeking. Notice that the 10 shot strings had a minimum of 2886 and a maximum of 2992. Similarly, the 20 shot strings had a minimum of 2987 and a maximum of 2989. Notice a trend? The larger the sample, the closer we get to the true answer. Statisticians use what are known as confidence intervals to estimate how close our sample statistics are likely to vary from the true population statistics. They would want to make statements like, "I am 95% confidant that the true mean of the population is between 2980 and 3000 ft/sec." So how do we estimate the confidence intervals for the mean velocity? A useful rule of thumb for rifle shooters would be 95% confidence intervals on the mean velocity are given as:

$$5 \text{ shots: } MV_{sample} - \frac{5}{4} \ SD_{sample} < MV_{population} < MV_{sample} + \frac{5}{4} SD_{sample}$$

$$10 \text{ shots: } MV_{sample} - \frac{7}{8} SD_{sample} < MV_{population} < MV_{sample} + \frac{7}{8} SD_{sample}$$

$$20 \text{ shots: } MV_{sample} - \frac{5}{8} SD_{sample} < MV_{population} < MV_{sample} + \frac{5}{8} SD_{sample}$$

So, if I measure a five shot sequence over the chronograph and it reports a MV = 2988 with a SD=10, then we estimate that the true mean of the population is between MV_{min} = 2988 – 1.25*10 = 2976 and MV_{max}=2988+1.25*10= 3000. If we get the same statistics from a 20 shot string, we would estimate the true mean velocity was between 2982 and 2994. Thus more shots yields better estimates. So how good is this theory in practice? Figure 4 shows the mean velocities produced for all the hypothetical samples from the 50 shot population with their respective 95% confidence intervals. Note that the different samples roam around within the interval but do not exceed it and that the confidence intervals gradually get smaller going from 5 to 10 to 20 shot samples.

So what have we learned?
- The chronograph gives us precise statistics for the samples we provide.
- These statistics are very likely different from the statistics we are interested in.
- By using a simple calculator, we can bound the true answer we are interested in.
- The larger the samples we use, the better our estimates or the tighter the bounds become.
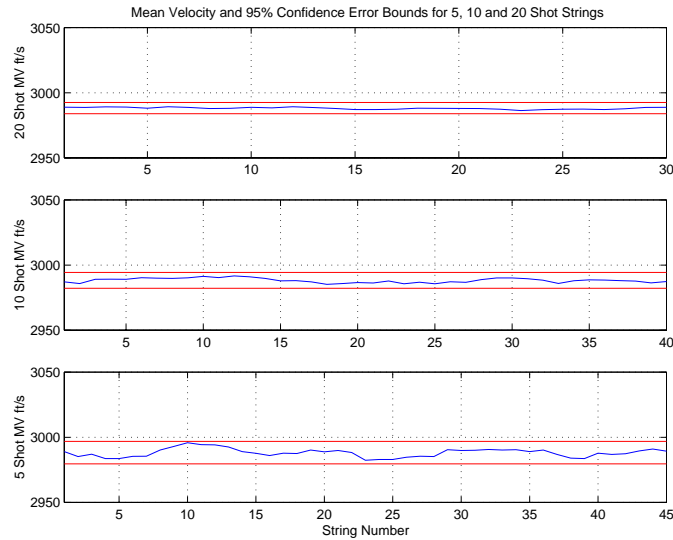
Figure 4: Mean Muzzle Velocities and 95% Confidence Intervals

In fact, if you are asking basic questions like "will I be supersonic at 1000yds" or "is the new lot going the same velocity as the old lot", 5 or 10 shots may very well give you an estimate which is sufficiently close to the true answer that you can proceed with confidence. 20 shots looks like overkill for estimating mean velocity. Great you say – that is about what I normally do with the chronograph.

Alas, very few of us are content with estimating only the mean velocity of the population. What we want to know is how much variation can we expect around the mean velocity. This is only natural -- less velocity variation leads to tighter elevation groups resulting in fewer dropped points. Besides, the chronograph provides us this information for free when we measure velocity by calculating extreme spread and standard deviation. Correct – but just as with mean velocity, the chronograph gives us the precise statistics for the sample but does not tell us how to use it to estimate the properties of the population which is what we really care about.

Let's consider the hypothetical strings from Figure 3 again. For five shot groups, the lowest standard deviation was 4 ft/sec for String # 16. Now that's a good load! However, String #28 yielded a SD of almost 15 ft/sec. That load needs some work. But wait – they are the same load -- all loaded at the same time with precisely the same components. How can this be? Bottom line is String #16 was lucky and String #28 was not. Lets look at what happened with the 10 shot strings. String #15 had a SD of 7 ft/s and String#25 had a SD of 12 ft/sec. That sounds better. What about the 20 shot strings? String #6 had a SD of 8 ft/s and String#25 had a SD of 11 ft/sec. There is that trend again. More data means a more accurate estimation of the population statistics. Statisticians can estimate confidence intervals for the standard deviation as well. A useful rule for rifle shooters would be to have 90% confidence:

5 shots: $\quad \frac{1}{2} SD_{sample} < SD_{population} < 2 SD_{sample}$

10 shots: $\quad \frac{2}{3} SD_{sample} < SD_{population} < \frac{3}{2} SD_{sample}$

20 shots: $\quad \frac{3}{4} SD_{sample} < SD_{population} < \frac{4}{3} SD_{sample}$.

To illustrate, if I measure a five shot sequence over the chronograph and it reports a MV = 2988 with a SD=10, then we estimate that the true standard deviation of the population is between ½*10= 5 and 2*10=20. If we get the same statistics from a 20 shot string, we would estimate the true standard deviation was between 7.5 and 13.3. Lets be perfectly clear, if you take 5 shots of a load with Winchester primers and measure a SD=6 and then shoot 5 shots of the same load with CCI primers and get a SD=19, what can

you say about the difference between the load's potential accuracy? Nothing… That variation is within the normal changes expected of 5 shot samples from a population having SD=10 ft/sec and therefore the only scientifically sound conclusion is they are similar. In fact, there is better than an 80% chance that the populations represented by the two samples are similar in performance. Nonsense you say? Consider the 5 shot samples from our measurements. We know that the rounds were all identical to within my ability to reload and the entire 50 shot population had a SD=9.6 -- yet String #16 had a SD=4 and String #28 had a SD of nearly 15. In fact, if I were really unlucky and chose rounds #1,2,13,27 and 32 as a five shot string, I would have got a SD over 20. Rounds #12, 16, 31, 41 and 48 would yield a SD less than 1 ft/sec. Given that it would be possible to get either 1 or 20 from the actual rounds fired, the fact that theory predicts you are 90% likely to be greater than 5 or less than 20 is not at all out of line.

Figure 5 shows the SD estimates from the hypothetical strings along with the 90% confidence intervals. Generally the data behaves as the theory predicts. It is interesting to note that Sample#16 from the 5 shot samples is actually below the bound. 90% confidence means 1 in 10 chances may fail. String#16 was a very lucky string, but it was the only one of 45 possible strings outside the bounds which is much better than 90% success rate. Want better than 90% certainty? Either you have to make the confidence intervals wider or use more shots in the sample. There is no free lunch here – estimating standard deviations from samples is hard work.
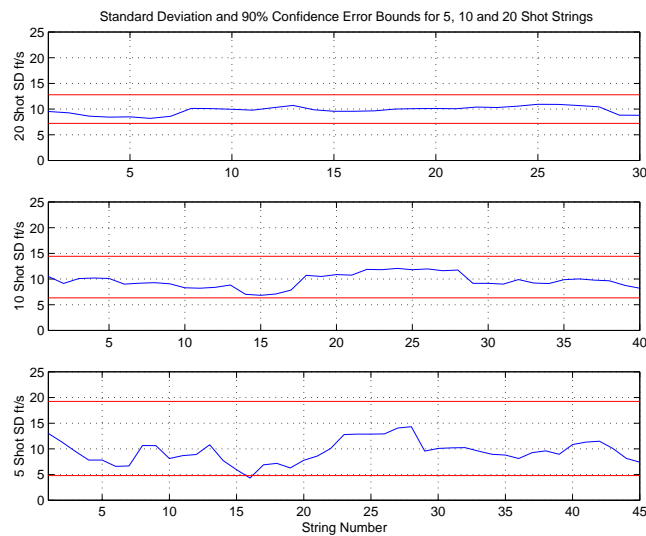

Figure 5: Standard Deviation and 90% Confidence Estimates for 5, 10 and 20 Shot Strings

So what have we learned?
- The bounds on the true population standard deviation can be estimated with a simple calculator.
- The larger the samples we use, the better our estimates or the tighter the bounds become.
- To be scientifically confident that a load has a better SD than another requires considerably more shots to be fired than are normally used.
- Most figures we read and hear about with respect to SD have very little meaning…

You say wait a minute, what about extreme spread? It must be the better indicator… If there is one thing you have learned from this article, what would it be? I hope it is that the more samples used in a calculation, the more accurate it becomes. Consider this: if I fire two shots and calculate ES, how many data points have I used in the calculation? Answer: 2. If I fire 50 shots and calculate ES, how many shots have I used in the calculation? Answer: 2 – the highest and the lowest. What can we infer about a statistician's opinion of ES as a measure of variability? Don't believe me? – look back at Figure 3. ES figures for 5 shots ranged from 12 to 40 ft/sec. For 10 shots, they ranged from 20 to 40 ft/sec. For 20 shots, they ranged from 30 to 46 ft/sec. They are slowly getting closer to the true answer, but there is no guarantee that they will converge to the true extremes of the population within any given sample. In fact, it is unlikely that our 50 shot sample would be indicative of the true extreme spread of the 1000 rounds I

made this winter. How do I know that? Recall that the true extreme spread of a population is about 6 times the standard deviation. My 50 shots showed a SD=9.6 which would predict an ES of 58 ft/sec. If I measured the whole 1000, I would probably cover that spread. I would also burn out my barrel , do a lot of reloading and never make it to a match. Extreme Spread does NOT converge as a predictable function of sample size and is not a reliable statistical indicator of load velocity variations. It is a measure of how lucky or unlucky you were when you chose your sample. PERIOD. Now, with that said, if you fire a string and the extreme spread is higher than 6 times the SD you are looking for – STOP. That load isn't going to get you there. ES can be used to reject a load, but it will not reliably help you choose the right one.

So what have we learned?
- ES is not a reliable statistical indicator
- The best indicator of velocity variations is the standard deviation.
- Most figures we read and hear about with respect to SD and ES have very little meaning…

OK. Now that I have shown two loads with a SD of 19 and 6 must be treated as equal, your favorite chronograph number, ES, is worthless and virtually everyone who has ever cited their magic chrono numbers was statistically misguided; what do you do? Throw the chronograph in the junk heap with the other toys that haven't panned out? Well… No. The chronograph is a useful tool if used appropriately. Just as there are basic elements to correctly aiming a rifle or operating a lathe, – there are basic techniques required to draw meaningful conclusions from experiments with a chronograph. These techniques come from statistics and require only a simple calculator to check.

So what have we learned and how do I pick a good load?

- The more data you include in a calculation, the more statistically accurate it will be. So with respect to component choices, start with the same components the winners are shooting. There is far more data in all the rounds they fire collectively with good results than you can possibly collect on your chronograph. If it works for them, it will work for you.
- Start about 10% below maximum charges and work your way up to the desired velocity in small powder increments using 5 shot groups checking for pressure signs as you go. The goal is to get up to the desired velocity and the theory has shown that reasonably accurate estimates can be obtained with these small sample sizes.
- Testing for velocity variations requires larger sample sizes – 5 shot samples will not yield reliable results. For my rifle, 20 shot samples showing a standard deviation between 7 and 13 would seem to be an achievable standard based upon the current analysis and would suggest that the true population standard deviation was near 10 ft/sec.
- New combinations should be tested with 20 shots and should be considered an improvement only if the sample SD is below 7. Even then it could be a lucky group and should not be considered a true improvement until firing a second 20 shot string with SD below 9 confirms the performance.
- Similarly, new combinations tested with 20 shot strings should not be considered inferior unless the sample SD is greater than 13. 20 shot strings resulting in a SD between 7 and 13 have better than 80% chance of equaling the performance of our 10 ft/sec population benchmark.

Following these basic guidelines, you can use your chronograph to develop long range loads similar in capability to those used by the nations finest shooters. I believe that you will find that by making statistically sound judgments, many loads produce statistically similar results and loads in general are not as finicky as current conventional wisdom would lead us to believe. Time with the chronograph is best spent confirming you do not have a bad load rather than spent searching for the magic combination. In reality, your shooting performance will probably improve since you can spend more time practicing with your set load and less time playing around with load combinations that have little or no statistical significance in terms of shooting performance. Finally, an additional caveat: the true measure of the accuracy potential of a load should be measured not on the chronograph but on the target paper at the intended full range. However, when evaluating group sizes, please remember to use sufficient samples to draw a statistically valid conclusion…

Possible Future Articles

So what are reliable figures for SD used by the top long range shooters? ENTER Ray's data here… Does Ray beat me because he has magic loads?

Do good chronograph statistics have anything to do with good group sizes? Can we correlate shot velocity variations, sample statistics and group sizes at range?

Can I believe the measurements from my chronograph?