Coping With The Tyranny of Statistics

By Jeff Siewert

Having spent 40+ years as an ammunition engineer, it's sometimes entertaining to listen to shooters grouse about how a particular load shot really good (e.g. small) groups, but on a subsequent occasion, shot bad (e.g. large) groups. You've heard it: "I put 4 shots into a really tiny group and the fifth was a flier." Then there's the "it's a half-MOA gun, when I do MY part." Or, there's other entertaining complaints about the impact point "wandering around" from group to group. All of these "conditions" are largely due to what I call the "Tyranny of Statistics".

Anytime bullets are fired down range, both the average impact point relative to the aimpoint and the group size are subject to the vagaries of statistical processes. As we fire more shots, we effectively continue to "roll the dice" with the processes controlling the bullet, the propellant, the barrel, the interaction among them, and external factors such as wind, all of which influence the flight path of the bullet. Are there things the shooter/reloader can do to reduce the variability? Absolutely. Can the variability be made zero? Absolutely, not. **Figure 1** shows the measured dispersion variation as a percent of the average dispersion (group size) as a function of the number of shots fired, along with the expected minimum of the dispersion variation, as a percent of the true dispersion (Assumed to be = 1.00 in this case).

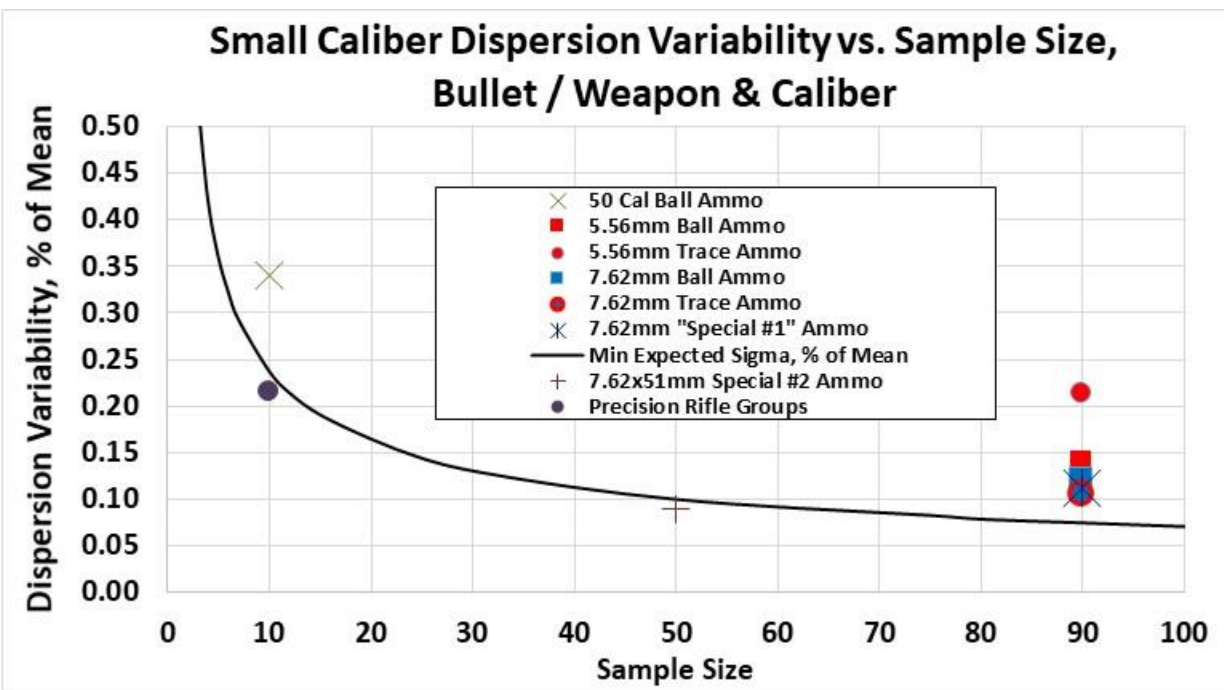

Figure 1: Small Caliber Dispersion Variation vs. Sample Size

It should be pointed out that the data presented is the average over many lots of ammunition. Where the actual dispersion variability is very close to the black line representing the minimum expected dispersion variability, this is evidence that the manufacturing process and ammunition testing process is "under control". Your testing might be under the black line in **Figure *1*** for a number of tests, but rest assured, it won't stay there forever.

As seen in **Figure 1**, the smallest you could expect to make the dispersion variation is shown by the black line, and the best any ammunition actually performs, in terms of variation, is quite near the line. On the right-hand side of this figure at n=90, the variation in dispersion for ammunition produced in very large volumes is shown. This is military ammunition, and it is tested in new, mid-life, and end of life barrels, so between that and the little tweaks made to tooling along the way, the actual measured dispersion variation is somewhat larger than the expected minimum for 90 shots fired.

Figure 2 shows the expected variation in the mean point of impact (MPI) as a function of the number of shots fired.

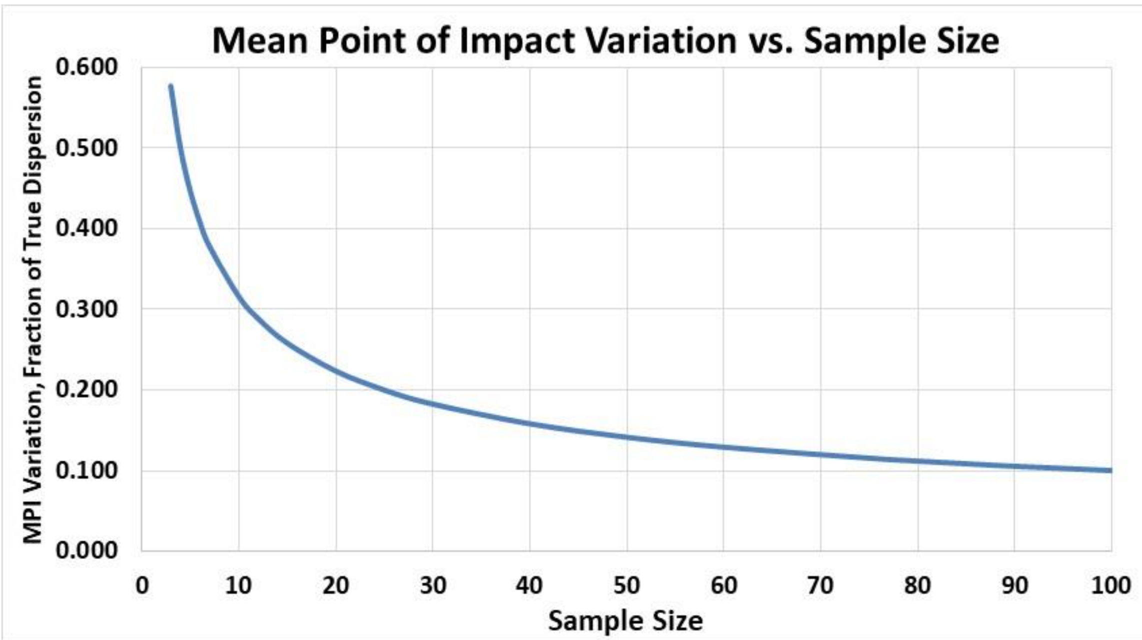

Figure 2: Mean Point of Impact Variation vs. Sample Size

What's seen in **Figure 2** shows the **short range** variation in average impact point of a group of bullets fired at the target as a fraction of the true dispersion. This is the amount of "jumping around" the average group impact point relative to the true average impact point, considered long term dispersion performance. If 5 shots are fired, and the dispersion (NOT extreme spread) of the ammunition being fired is 0.5 inch at 100 yards, the expected impact point variability group-to-group is 0.5 x 0.3 = 0.15 inches.

To effectively deal with the tyranny of statistics, we have to briefly open that can of worms. I'm not a statistician, or a mathematician, so I'm not going to list the equations that generate the probability density functions, you can research that yourself if you require more in-depth information. The Student's "T" distribution is a generalized form of the Gaussian distribution;

the Gaussian distribution represents the expected performance **the whole population** of specified items or characteristics like pressure, velocity, action time, etc. Once the number of samples gets to be above about 30, the difference between the Student's "T" distribution and the Gaussian Distribution differs by less than 0.1 percent. **Figure 3** compares the Student's T distribution to the Gaussian distribution; it can clearly be seen the Student's "T" distribution has a lower peak and heavier "tails" than the Gaussian distribution. If the user fires less than 30 shots in a group, the Student's "T" distribution is the appropriate choice for assessing the performance of the group. The shape of the Student's "T" distribution depends on the "Degrees of Freedom" (DoF) of the "system" being evaluated.

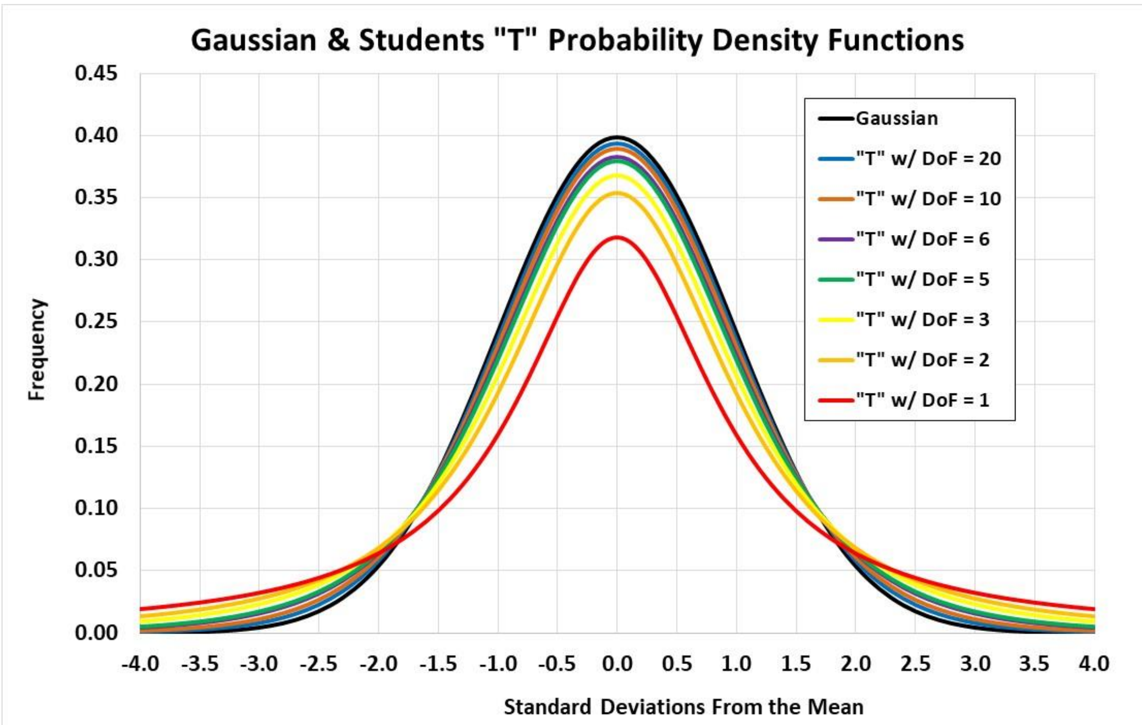

Figure 3: The Gaussian and Student's "T" Distributions

Dispersion Testing

Some shooters like to use the "extreme spread" metric for evaluating group size, others use different metrics to assess the performance of a new load. Taking the standard deviation of the impact points in the horizontal and vertical planes relative to the aim point provides a particularly useful metric for shooters as compared to extreme spread. The reason for this is quite simple: there is more data being acquired from the target. Instead of measuring and using **ONLY** the extreme spread of the most distantly separated impact points (one piece of data, with no weighting for the number of shots fired in the group), measuring the impact coordinates in the horizontal and vertical planes will provide you with 3 useful metrics. These are:

- Average impact point of the group with respect to the aim point, in Horizontal & Vertical axes
- the standard deviation of impact points in the horizontal plane
- the standard deviation of impact points in the vertical plane.

So, instead of 1 data point with "Extreme Spread", if we extract the bulleted data just above, there are 3 bits of useful data available when extracting the standard deviation of impacts in the horizontal and vertical planes. But that's not the only benefit to taking this data.

Let's dig into why using the average of the horizontal and vertical standard deviations is a good idea, compared to "extreme spread". When it comes to assessing bullet impacts on paper at relatively short ranges (<200 yards/meters), we're at the mercy of a wide variety of statistical problems that can really hamper our ability to interpret the results correctly. One of the largest problems encountered is what to do about a target with a group that is "skewed" (like a "flier"); that is the vertical axis is much larger or smaller than the horizontal axis? Figure 4 shows a couple of different analytical methods for dealing with a skewed target. In this plot, CEP stands for: Circular Error Probable. This is the radius of a circle containing 50% of the impacts and is generally considered to be the most reliable and objective method to measure short range group sizes. On the horizontal axis is the ratio of minimum sigma (standard deviation) to maximum sigma; on the left is a very skewed target, on the right a circular one. The vertical axis shows the ratio of CEP to maximum sigma (standard deviation). In the plot, several methods for computing dispersion are shown, but the line that is closest to "True CEP" for very skewed targets is an average of the horizontal and vertical standard deviations. For this reason, averaging the horizontal and vertical standard deviations is the most accurate, least variable method of calculating dispersion.

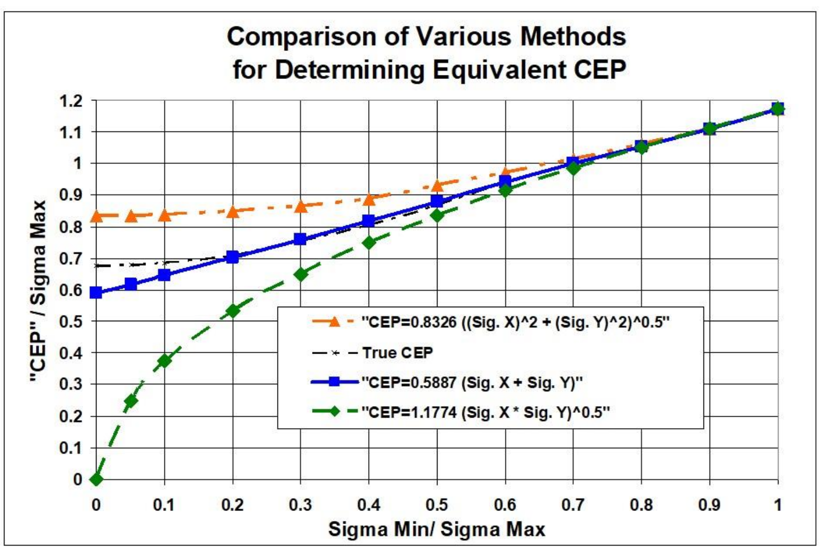

Figure 4: CEP vs. Axis Ratio Skewness

Calculating CEP is more complicated than taking the standard deviation of the impact points, and the downside to using only CEP is that it has the potential to miss calculations that would otherwise clearly point to problems with the gun/fixture. If your groups are consistently larger in the horizontal or vertical axis than the other one, the problem is with the test fixture, as will be discussed below.

Ammunition should shoot circular groups if groups are evaluated at short range. Stated another way, the horizontal and vertical standard deviations should be equal to one another if the sample size is large enough and we're evaluating at a range where winds, muzzle velocity variability and drag variability make only minor contributions to the targeting picture. So, if you're taking horizontal and vertical impact point data short range, and one axis consistently shows up larger or smaller than the other, it's a pretty sure bet that there's something going on with the "system". It could be a loose screw in the stock or the scope mount, scope parallax or differences in the way the shooter positions his or her eye behind the scope shot-to-shot, something about the rest on which the gun is situated, or something else **NOT** related to the ammunition.

Figure 5 shows the ratio of extreme spread to the true dispersion (as evaluated by standard deviation), as well as the dispersion variation (how much the group size is expected to change group-to-group) as a function of the number of shots in the group.

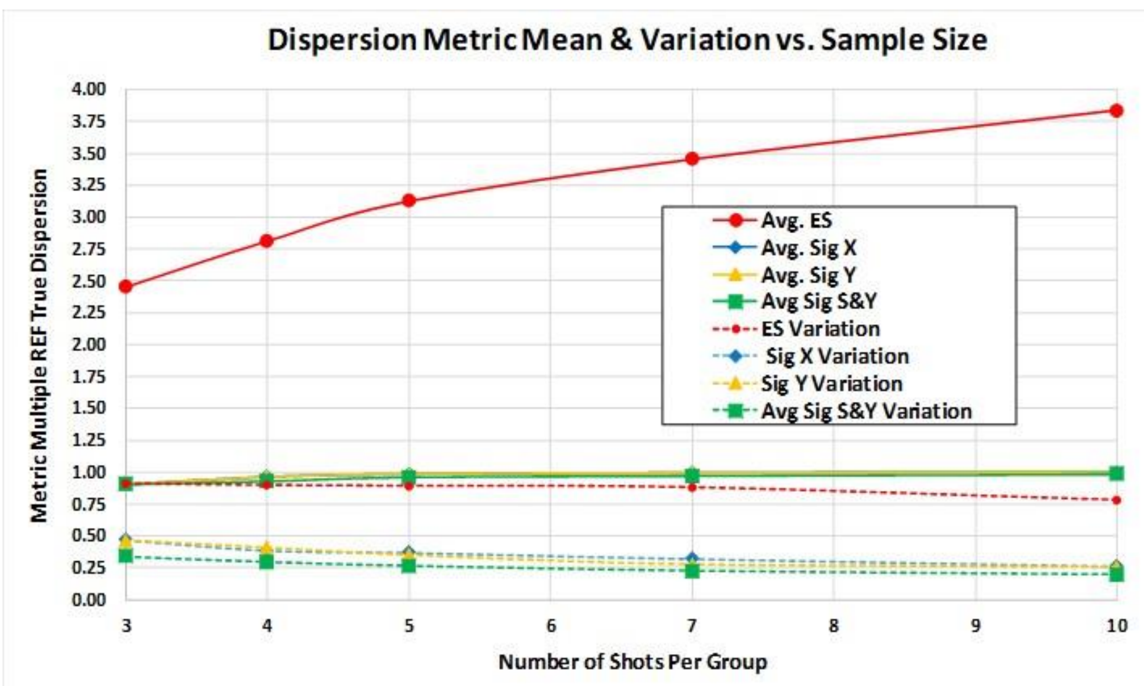

Figure 5: Comparison of Extreme Spread & Standard Deviation vs. # of Shots

In Figure 5, it can be seen that the “Extreme Spread” dispersion metric does nothing but grow in relation to the true dispersion when the number of shots is increased. This is decidedly NOT what we want, we want to have MORE confidence in our dispersion measurements when we fire more shots, not less. Further, the variation in averaging the horizontal and vertical standard deviation grows smaller with increased number of shots.

In a dispersion assessment conducted by some friends a few years ago, the purpose was to assess the effect of overall loaded cartridge length on group size (e.g. dispersion, not “accuracy”). For this test, the center of the projectile impact locations on the paper targets were measured with a drafting scale ruler relative to the aiming point. Once the impact points were measured, each impact coordinate was entered into an Excel spreadsheet for subsequent analysis. The data entry for a typical 10 shot group is shown below in Figure 6.

		X	Y
	3.24"	0.094	-0.656
		0.688	-1.000
Mean X:	-1.547	-2.688	-1.906
Mean Y:	-1.109	-3.000	-2.125
Sigma in mils	0.785	-1.469	-1.000
Sigma X:	1.558	-3.813	-3.844
Sigma Y:	1.268	-0.313	-0.438
Sigma Avg.	1.413	-1.313	-0.656
		-3.219	-0.313
		-0.438	0.844

Figure 6: Impact Point Data Entry in Spreadsheet

Figure 7 shows the equation contained in Cell V18 to compute the mean (average) for the impact points in the X coordinate. This equation is for the Cell range W16 to W25 (Column W, Rows 16 thru 25).

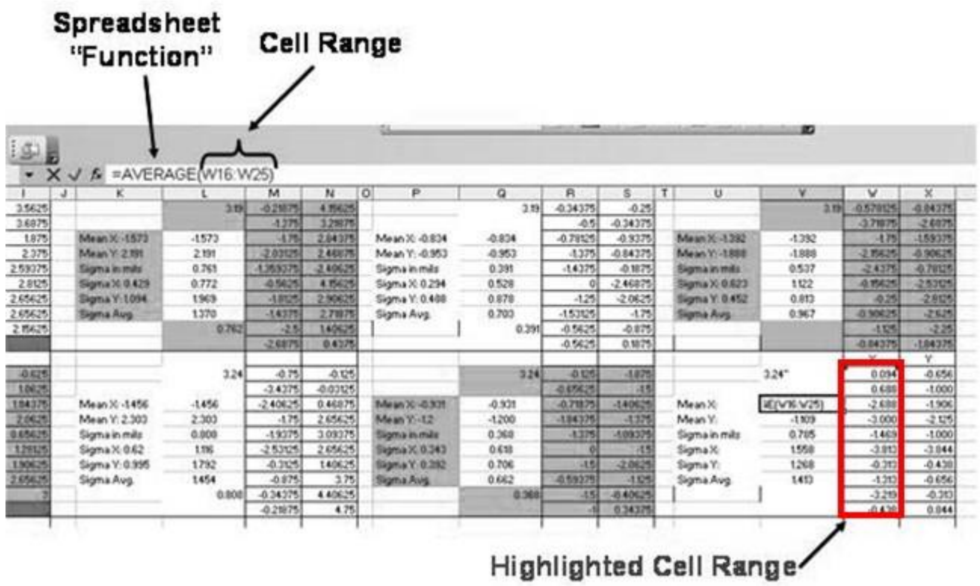

Figure 7: Formula for Computing Average (Mean) Impact Points

The equation for computing the standard deviation (sigma) for the impact points in the X coordinate for the cell range W16 to W25 is shown in Figure 8.

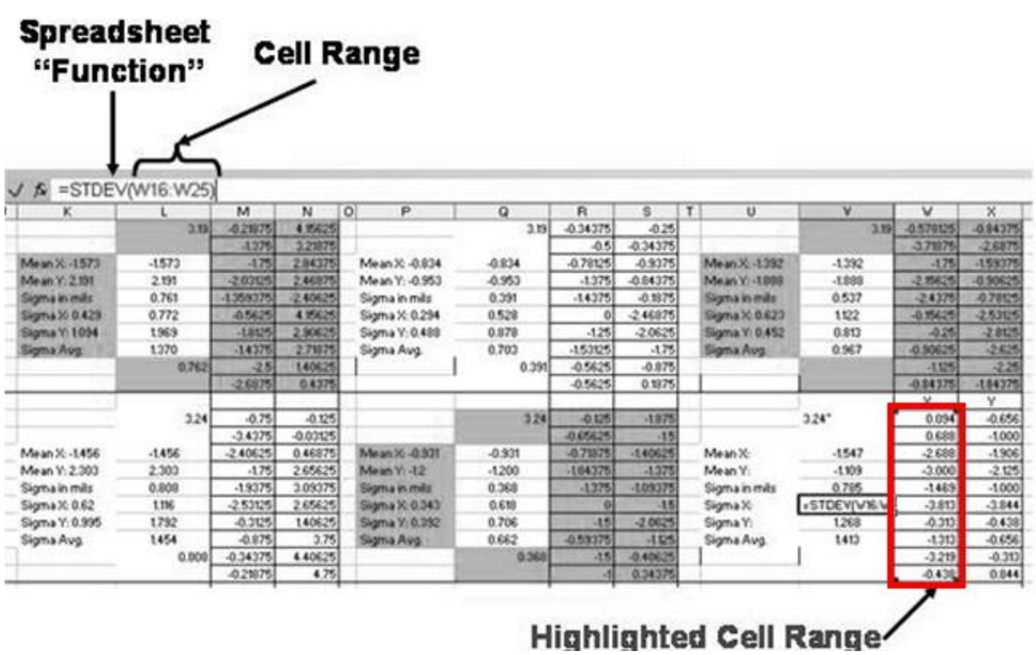

Figure 8: Formula for Computing Standard Deviation (Sigma) of Impact Points

The average for X & Y coordinates as well as the standard deviations for both coordinates were computed for each combination of bullet type/weight and loaded cartridge length. The results were then collected to generate plots of dispersion vs. bullet type/weight and loaded length, as well as average impact points as a function of the cartridge loaded length and bullet weight.

To estimate the dispersion of each load, the standard deviation in horizontal and vertical projectile impact locations were averaged. This approach was taken because as shown above in **Figure *4***, averaging is more representative of the true ammunition dispersion as the ratio of minimum standard deviation to maximum standard deviation (whether horizontal is greater than vertical or vice versa). As dispersion patterns become heavily "skewed" (horizontal pattern much larger or smaller than the vertical), alternate methods for computing CEP (circular error probable) lines on the left hand side of Figure 4 increasingly diverge from the blue "average" line. Thus, we chose to average the dispersion in the horizontal and vertical axis to prevent a "flier" from adversely skewing the data.

The dispersion of the tested cartridges as a function of projectile type and cartridge loaded length are shown in **Figure 9**.

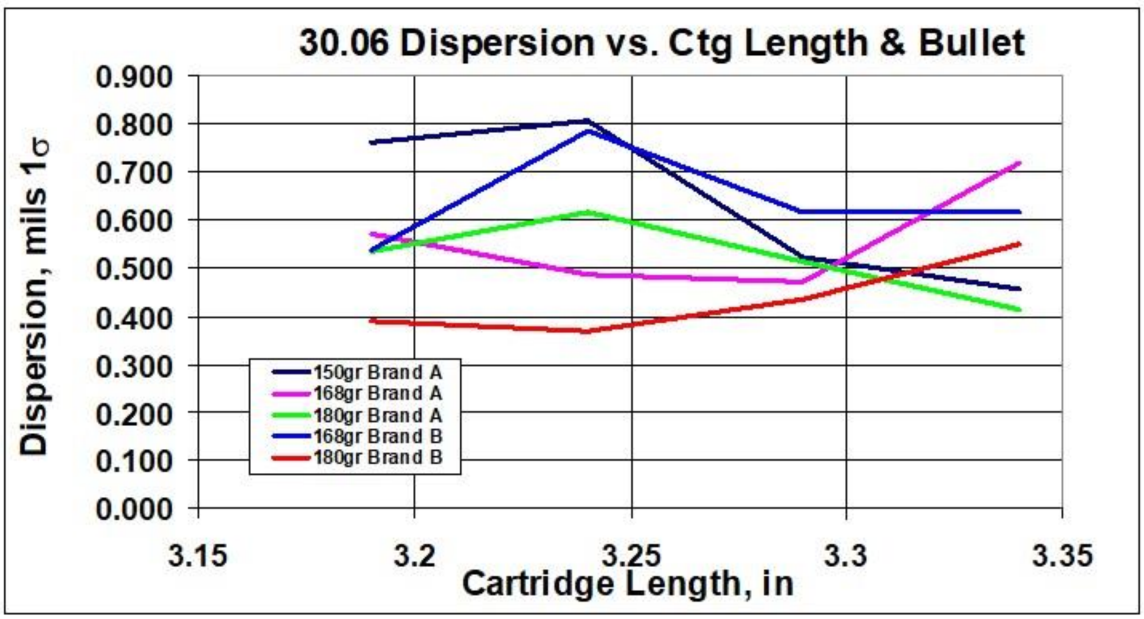

Figure 9: Dispersion vs. Loaded Cartridge Length and Bullet Brand & Type

For this plot, the raw projectile impact point data at the target as measured in inches was divided by the distance to the target from the gun muzzle as measured in inches to express the group size in terms of angle. This result was divided by 1000 to convert the result to an angle measurement commonly used in military application called "gunners mils", or mils for short. This conversion allows for easy extrapolation of results to any reasonable range (up to about 300 yards).

An inspection of the impact points for this test leads to additional, interesting observations. The impact points are shown in **Figure 10**.

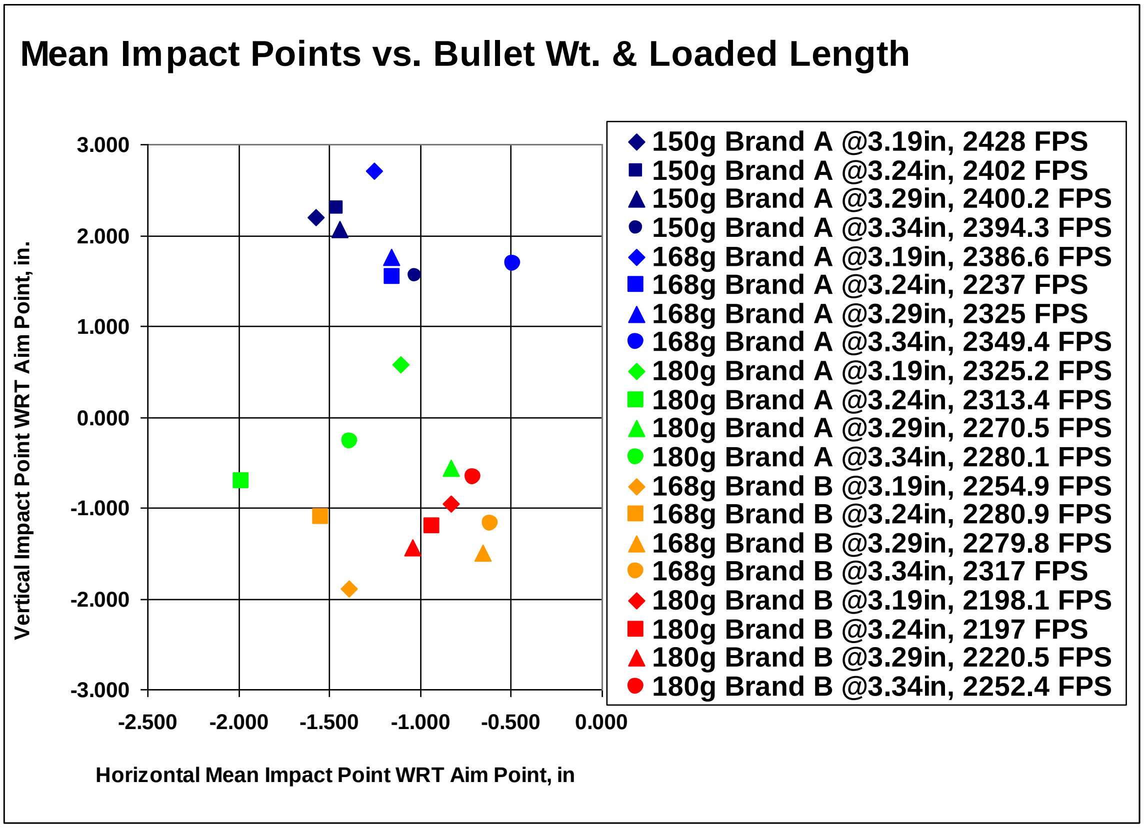

Figure 10: Impact Point for Dispersion Testing by Bullet Type & COAL

There are several interesting observations to be made in **Figure 10**.

The first is, for Brand A 150g and 168grain bullets, the 3.34" loaded length, the longest loaded length evaluated, the average impact points are distinctly different than for the same bullets loaded at shorter cartridge lengths.

The second interesting observation is that for the Brand B 180g bullets, the impact point doesn't shift much as a function of loaded length. This means that selection of this bullet will require the hand loader to do less "fiddling" to develop an accurate load (small dispersion) in this rifle, and that variations in the projectile seating depth won't have a huge effect on the "zero" setting of the scope.

Velocity Testing

The “Satterlee” approach to load development relies heavily on the assumption that when no change in muzzle velocity is seen with increasing charge weigh for single shots, that is indicative of muzzle velocity at which barrel movement is minimized. Further, the “Satterlee” test relies not only on “perfect” correlation between pressure and muzzle velocity, it also relies on reducing the shot-to-shot engraving pressure variation going to zero while the test is being conducted. This is a physical impossibility, because there are shot-to-shot variations in bullet engraving force prior to achieving peak pressure. If the shooter is firing only 1 shot per charge weight (e.g. for a “Satterlee” test), and measuring only muzzle velocity, the shooter is by default, selecting the 1 DoF distribution “T” distribution from the **Figure 3**. This means there is a large amount of uncertainty in the measured result. **Figure 11** shows a simulation of a Satterlee test with 4 “velocity nodes”.

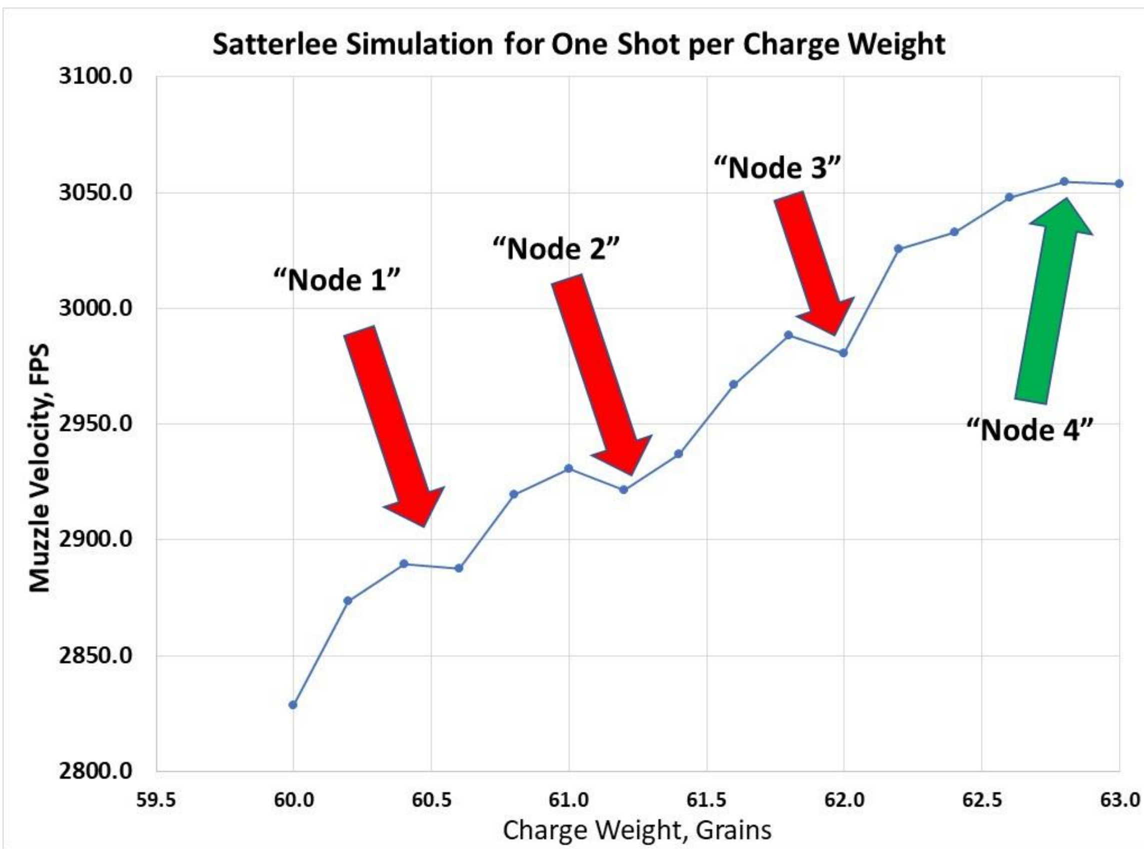

Figure 11: Simulated Satterlee Velocity Test With 0.2grain Charge Increments

In the best-case scenario, with the user shooting one shot per charge weight, there is an approximately 32% chance of the velocity hitting the actual “mean” for that charge weight. It also means, for 68% of the tests, the measured muzzle velocity will **be something other than** the true average velocity for that charge weight! If a big pile of these less-than-representative single shot muzzle velocity test shots are strung together while incrementing the charge weight, eventually a “velocity node” will appear. The location of the “velocity node” is caused by completely random shot-to-shot changes in engraving pressure, which are totally out of the control of the reloader/shooter. As a result, this “velocity node” is completely meaningless when it comes to selecting a charge weight that result in minimum sensitivity to changes in charge weight because it’s the result of purely random shot-to-shot muzzle velocity variations for a single load test series, the results of which would likely **not** be repeated if the test were to be replicated multiple times.

A change in engraving force prior to the shot attaining peak pressure affects the peak chamber pressure developed, and hence, muzzle velocity. **Figure 12** shows the measured engraving force variation as a fraction of the measured average peak engraving force for the bullets listed. The engraving force testing was done on an “Instron” load testing machine at a local university; a sample of 15 bullets of each type were pressed through barrel sections, at 1.0 inch per minute, and the engraving force vs. travel for each bullet was measured.

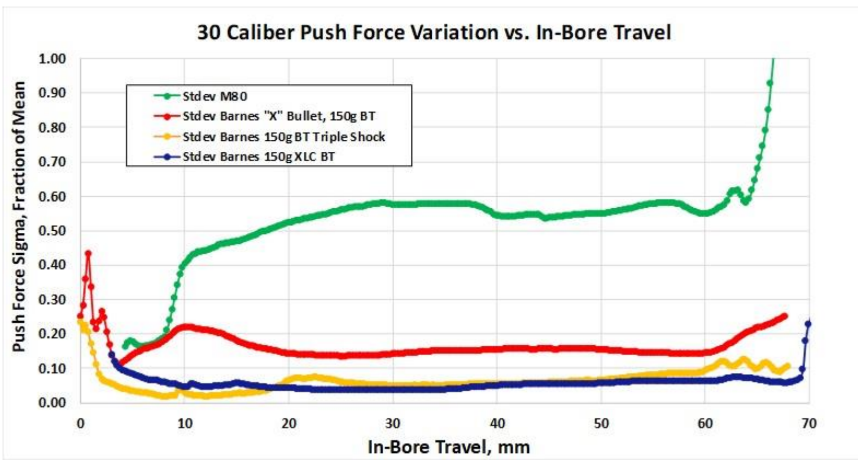

Figure 12: Engraving Force Variation vs. Travel For Several Bullets

While the firing event is considerably different than a quasi-static push test in a laboratory, the data gives some idea of the shot-to shot variation that might be expected as a normal course of events when firing.

Figure 13 shows the peak pressure versus muzzle velocity results for a 50-shot lot acceptance test (LAT) of a 220g lead core bullet fired in a 300 Win Mag cartridge **with a single charge weight**; the measured pressure-velocity data points are the blue diamond’s running upward moving from left to right. The correlation between pressure and velocity for this cartridge is particularly good compared to some other (notably straight-walled) cartridges, so this correlation might be as good as we might expect to achieve for “factory” ammunition. I’ve superimposed the Gaussian distribution (in black) and the Student’s T distribution for 1 DoF (in red) on the pressure-velocity behavior with the peak at the “average” of the data, and the distributions are “stretched” to accurately represent the muzzle velocity standard deviation (AKA “sigma”) of the LAT data. It’s clear there are a fair number of shots clustered near the observed mean, but there are also shots clustered below and above the mean which the shooter might happen to “hit” in a Satterlee single shot, ramped charge weight test. It should be mentioned that the correlation between pressure and muzzle velocity is very likely to not be as good as shown in **Figure 13** with the case less than about 95% full. This is another factor weighing heavily against the use of a single shot to determine "average" muzzle velocity for a given charge weight. Any one of the blue diamonds could represent your shot taken at any charge weight during the charge ladder test.

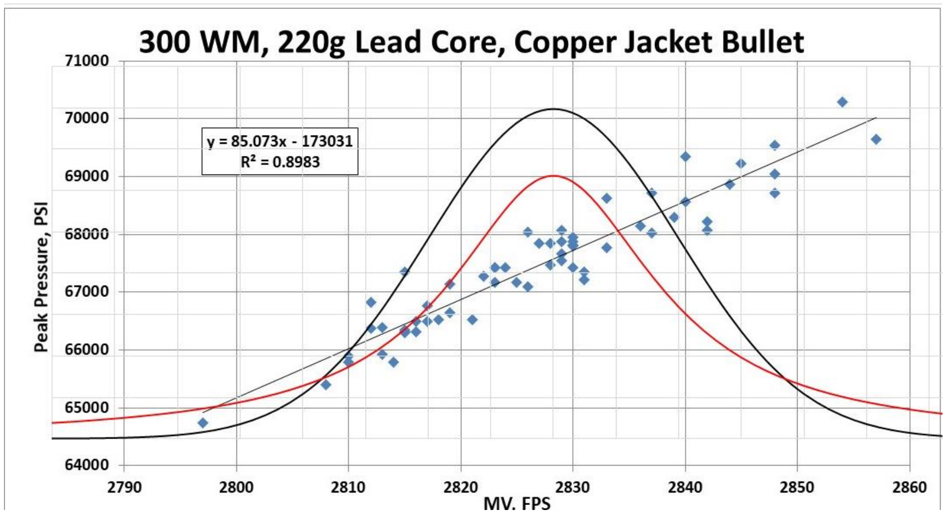

Figure 13: Measured Pressure-Velocity Data with Superimposed Gaussian & Students T Distributions

One of the key fit parameters in **Figure 13** is the R^2 number; here it's 0.8983. That means there is pretty good, but not perfect, correlation between peak pressure and muzzle velocity. Perfect correlation would have an R^2 of 1.00, and all the data points would fall on the line.

Figure 14 shows what might happen if you repeat the single shot, ramped charge weight test 100 times with 0.2 grain charge increments with a muzzle velocity standard deviation of 10 FPS at all charge weights. As previously mentioned, the shot-to-shot muzzle velocity variation arises from the differences in engraving pressure variation, along with things like shot start pressure, primer output, case volume variation, and factors **<u>other than</u>** charge weight. A "band" of possible muzzle velocity vs. charge weight solutions becomes evident with a large number of replications of single shot charge weight velocity assessments, and no one charge weight will result in any less sensitive velocity performance than any other. Yes, data is simulated using typical powder performance, but it only assumes a muzzle velocity scatter of 10 FPS for each charge increment.

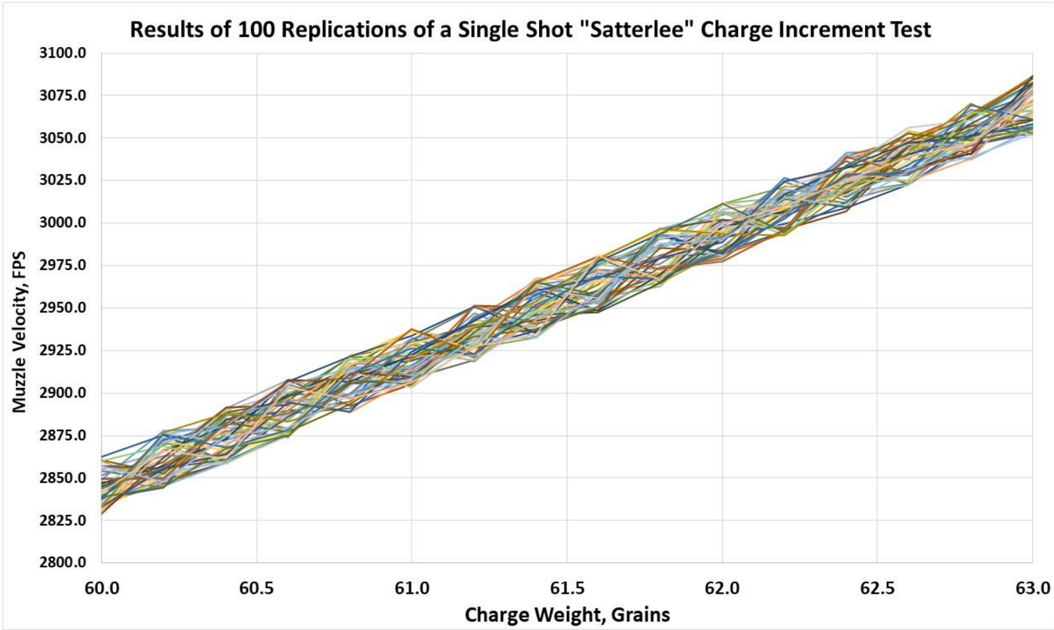

Figure 14: 100 Replications of Single Shot Satterlee Test with 0.2g Charge Increments

If you believe the incremental charge weight method using ONLY velocity with <u>no</u> targets can provide a "optimum" charge weight at which velocity sensitivity to changes in charge weight is minimized, you have answer the question: "why this test completely free from the normal shot-to-shot muzzle velocity variation from changes in engraving pressure?" for yourself when you do this test. Due to bullet engraving pressure variability, zero pressure/velocity variability is a physical impossibility. Thus, relying on some random "flat spot" in the charge weight vs. muzzle velocity data ONLY, without a target, isn't likely to lead you to select an optimum load for your gun, in my opinion.

As a result of the normal shot-to-shot engraving force variation, it's my contention that the muzzle velocity variability can't go to zero under any circumstances for a reasonable number of shots (n>5). So, the only thing the ramped charge weight test can actually do with any reliability is give you an approximation of the slope of velocity with charge weight for a particular cartridge, bullet, powder and primer combination. Further, It's also my opinion that you'd be much better served firing 3 groups of five shots each at the three different charge weights if you're looking to hit a particular muzzle velocity or barrel transit time, because by the time you've fired 5 shots, the calculated "average" muzzle velocity differs by no more than about 7.5 FPS from the true population average for that charge weight, on average, at a 90% confidence level.

So, as much as we'd like to have a convenient "shortcut" to load development, simply measuring muzzle velocity vs. charge weight alone, without shooting targets, isn't the way, <u>in my opinion</u>. Gathering velocity data <u>with</u> holes in the target is much more likely to lead you to a load that shoots small groups.