Is "Group Size" the Best Measure of Accuracy?

by John E. Leslie III

PREMISE

Despite the recent increase in articles applying mathematics and statistics to the everyday problems of handloaders and target shooters, few have attempted to answer the really big question: "Which statistics should be used to measure ammunition/firearm accuracy?"

In his article, Craig[1] pointed out the problems associated with using three or five shot groups to try to determine the best load for a particular firearm. He used a computer simulation to demonstrate that the laws of probability can cause a randomly chosen shot group fired with less consistent ammunition to be smaller than another random shot group fired with ammunition which was, in fact, more consistent.

This apparition is much more likely to occur in shot groups containing fewer shots than in shot groups containing a large number of shots. As more shots are fired, the laws of probability catch up with the looser grouping ammunition and show it up for what it really is.

I have been doing research to try to identify the best statistical measure of shot group dispersion or "tightness." Thus, I decided to recreate Mr. Craig's computer simulation and include the statistics which I have been researching.

THE SIMULATION

The first computer simulation examined the success rate of various statistics at identifying the shot group fired by the tighter grouping load out of four possible choices. Each ammunition "load" was 20% less consistent than the previous load. To get an accurate representation of the statistics' success rate, 65,000 sets of shot groups were created for each number of shots.

In addition to the commonly used "group size" measure (what statisticians would call extreme spread), I tested four other statistics: the figure of merit, the diagonal, the mean radius, and the radial standard deviation[2]. A graph showing these statistics' success rates for correctly determining the tightest grouping load is included as Figure 1. These percentages are not absolute numbers, as we will see later in this article. Greater or lesser differences between the ammunition loads will change the success rates.

EXTREME SPREAD

Extreme spread is the most widely used measure of shot group dispersion. It is defined as the maximum distance between the center of any two shots within the group.

There are, however, several problems with extreme spread -- most

---

[1] Peter Craig, "Accuracy Testing," *Precision Shooting* 39, no. 4 (August 1993): p58-62.

[2] Source of formulas: Frank E. Grubbs, Ph.D., *Statistical Measures of Accuracy for Riflemen and Missile Engineers*, 3d printing, (Havre De Grace, MD: By the author, 4109 Webster Road, 1991), 15-26 passim.
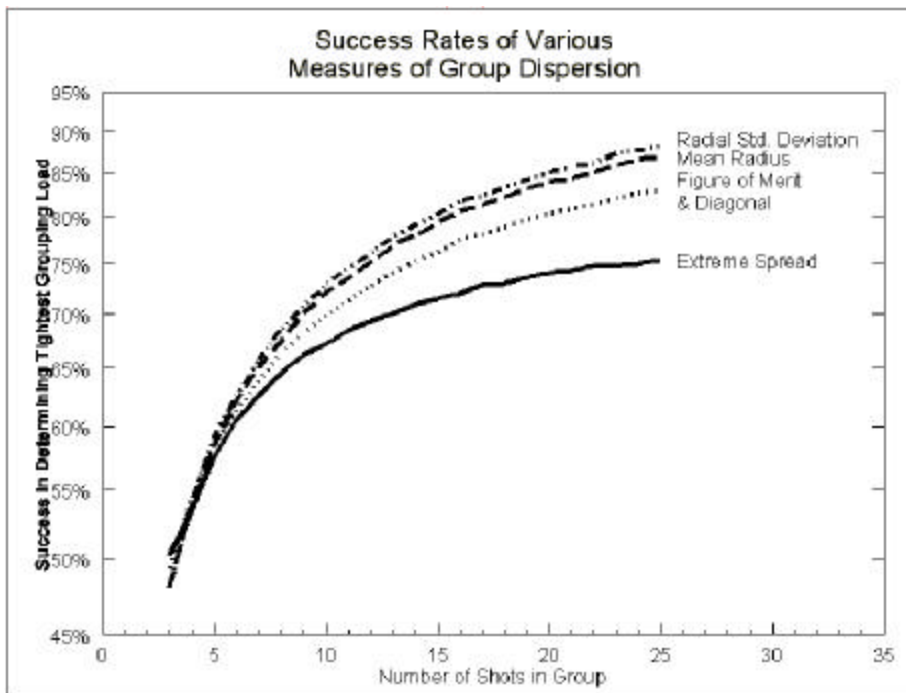
notably, the measure's domination by the group's outliers. By definition, outliers are shots which have a low probability of occurrence; otherwise they would not stand out so. Since extreme spread measures the distance between extreme shots, it really measures the spread between the least likely to be repeated shots in the group. Also, by only using data from two shots within the group, it ignores the data represented by the other, more likely to be repeated, shots.

Figure 1. Success Rates of Various Measures of Group Dispersion

While extreme spread outperformed all of the other measures for the three shot groups, it was the worst statistic of all for groups of four or more shots.

FIGURE OF MERIT

The figure of merit (FOM) is the average of the maximum horizontal group spread and the maximum vertical group spread.

This measure uses data from at least two shots but more likely three or four shots. Since it uses more data points (shots), the effect of an outlier gets diluted: it now has a 25% influence rather than 50% as with extreme spread. In the simulation, the FOM proved superior to extreme spread in correctly choosing the tighter grouping load for groups of four or more shots. I believe this is due to the use of twice as many data points.

DIAGONAL

The diagonal statistic uses inputs similar to the FOM. It is calculated by taking the square root of the sum of the maximum horizontal spread squared and the maximum vertical spread squared. The success ratios for the diagonal were almost identical to those of the FOM; in fact, these two measures are shown on the same line on the graph in Figure 1. I believe these results reflect the similarity of their inputs. All of the advantages mentioned above for the FOM also apply to the diagonal.

MEAN RADIUS

The mean radius, as the name implies, is simply the average distance of all of the shots of the group from the group center.

This measure uses data from every shot, not just two or four shots. Here once again, additional information helped improve the accuracy of the statistic: the mean radius was a more reliable predictor of the smallest load than either the extreme spread or FOM/diagonal statistics.

RADIAL STANDARD DEVIATION

The radial standard deviation (RSD) is similar to the standard deviations we are all familiar with except that it is two-dimensional. It is calculated by taking the square root of the sum of the horizontal variance and the vertical variance.

Like the mean radius, this statistic uses all of the available data points from the shot group. The RSD was the most accurate measure I examined for determining the tightest grouping load.
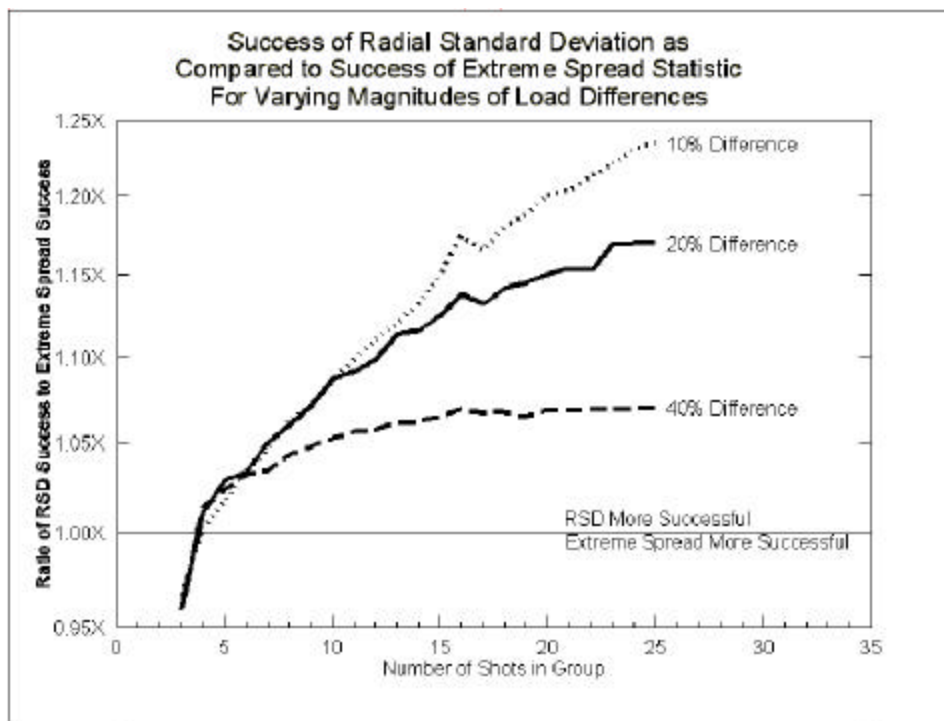
DIFFERENT SIZED LOADS

Having established that the RSD was superior at selecting the best load in the above simulation, I wanted to determine the effect of varying magnitudes of differences among the loads. My first simulation used four loads that were progressively 20% larger than the previous load. I decided to run the simulation twice more - once using half of that difference between loads (the 10% difference loads) and once using twice the original difference between the loads (the 40% difference loads). Both simulations showed the same rank order between the statistics as my first simulation, but the amount of the improvement of the RSD (and the other statistics) over extreme spread varied greatly. My comparison of the RSD's accuracy relative to the extreme spread's accuracy is shown in Figure 2.

This study showed that these statistics' success rates were sensitive to the amount of variation between the loads. While the RSD was always more accurate than extreme spread, its advantage shrunk when faced with identifying the more obviously differing loads (40% differences). Conversely, the RSD was dramatically better than extreme spread at distinguishing among the more difficult to differentiate loads (10% differences). When the going got tough, the RSD clearly demonstrated its superiority.

DIFFERENT NUMBERS OF LOADS

A final dimension of the RSD versus extreme spread question that I examined was whether the statistics' ranking would be effected by distinguishing between two loads rather than the four loads used in the other simulations. The results of the two-load simulation were identical, in both rank order and magnitude, to the results of the four-load simulation.

Success of Radial Standard Deviation as Compared to Success of Extreme Spread Statistic For Varying Magnitudes of Load Differences

CONCLUSION

This exercise has proven that the extreme spread statistic, which we all put so much faith in, is only marginally adequate for the task. The radial standard deviation can distinguish between loads with fewer shots fired and a higher degree of confidence.

The consequences of this finding are important for all shooters, not just reloaders. Position shooters cannot only match their ammunition to their firearm more reliably using the RSD but they can also use this statistic to judge the effect of changes in their position construction. If the RSD of their groups declined significantly after the change, they would know that they should keep the modification. Shooters can also use this measure to evaluate alterations to their equipment: How much of an improvement did I get from fire lapping my barrel? Did my new stock really improve my accuracy? Is it worthwhile for me to separate my rimfire ammunition by rim thickness? Does it matter, from an accuracy point of view, how thoroughly I clean my firearm?

The major drawback to using the RSD is the hassle of calculating it. First you must determine the cartesian (x & y axes) coordinates of all the shots in the group. Then you must average all of the x values and all of the y values separately to find the coordinates of the group center. Next you would calculate the variances of the group in the x and y directions. Finally, you would add the two variances and find the square root of the total. After you have done this a few times, you realize why extreme spread is still so popular -- it is much easier to calculate!

Fortunately, the personal computer revolution comes to the rescue. There are several PC programs, including one named ScorStat that I wrote for IBM compatibles, which can help you do some or all of the necessary calculations. I would expect to see additional programs become available as this type of statistical analysis becomes more popular.

One interesting fact that I learned from my research was that the U.S. military has been using mean radius and radial standard deviation to measure shot group dispersion for a long time. The earliest reference to these statistics that I have found so far, was a World War I study that used mean radius to compare the relative accuracy of the M1903 and M1917 rifles to that of the Moisin-Nagant[3].

It seems to me that with all of the time, money and effort that shooters put into developing the proper loads, equipment, and positions, they should be using the most reliable and accurate statistics available to judge their results.

John E. Leslie III
j.leslieiii@sprynet.com

---

[3] Steven Trask, "Testing the Moisin-Nagant," *American Rifleman*, (September 1918) reprinted in *American Rifleman* 141, no. 9 (September 1993): p112.