More Capability Confusion

A little algebra, done correctly, can take you places that do not exist

Donald J. Wheeler

Here we take a serious look at some nonsensical ideas about capability ratios. Following a quick review of predictability and capability, and a brief discussion of the traditional ways of characterizing capability and performance, we will consider the shortcomings of four bits of capability confusion found in some of today's software.

The predictable operation of a process is an achievement rather than being an intrinsic property of that process. This means that the notion of a process mean, a process standard deviation, or a process capability is not well-defined until that process is operated predictably. Given data from a predictable process we can use statistics for location and dispersion, along with specification values, to estimate the process mean, the process standard deviation, and the process capability in the manner outlined in Figure 1.
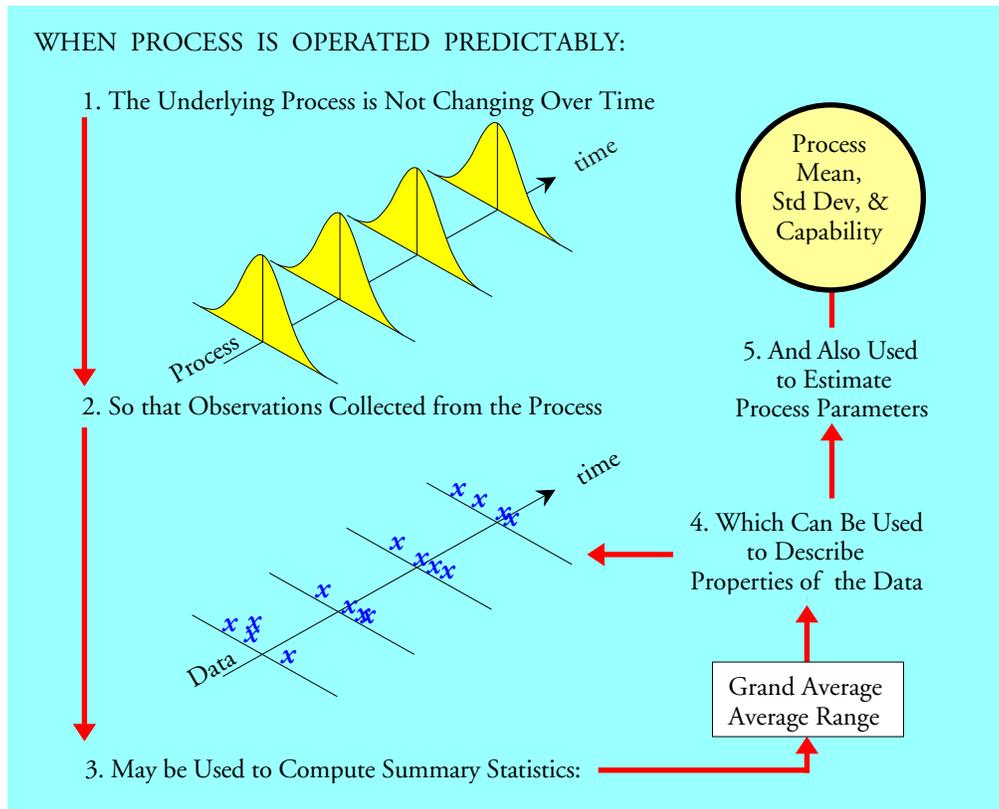

Figure 1: Estimating Parameters for a Predictable Process

Whenever a process is operated unpredictably we have to think about the process mean, the process standard deviation, and the process capability as properties that are changing over time. However, the statistics we compute are simply data plus arithmetic. So while we can always compute our statistics, the arithmetic does not make the underlying notions real. Our statistics can only provide estimates of process parameters when the process is operated predictably. When our process is operated unpredictably the process parameters are changing and are therefore divorced from the statistics as shown in Figure 2.
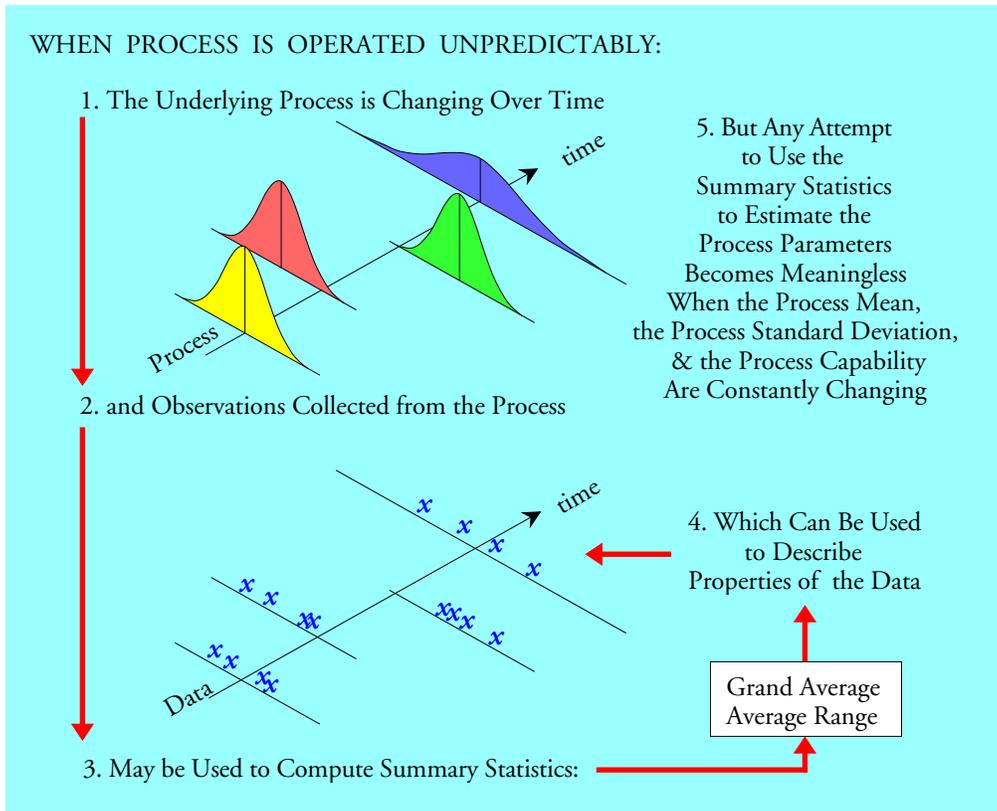
WHEN PROCESS IS OPERATED UNPREDICTABLY:

1. The Underlying Process is Changing Over Time

time

5. But Any Attempt to Use the Summary Statistics to Estimate the Process Parameters Becomes Meaningless When the Process Mean, the Process Standard Deviation, & the Process Capability Are Constantly Changing

Process

2. and Observations Collected from the Process

time

4. Which Can Be Used to Describe Properties of the Data

Data

Grand Average
Average Range

3. May be Used to Compute Summary Statistics:

Figure 2: Parameters Are Not Well-Defined for Unpredictable Processes

So on the one hand we have process parameters that are only well-defined when a process is operated predictably, and on the other we have statistics which we can always compute. Clear thinking therefore requires that we make a distinction between statistics and parameters. In the words of Walter Shewhart, "…measurements of phenomena in both social and natural science for the most part obey neither deterministic nor statistical laws until assignable causes of variability have been found and removed." While we can always compute our statistics, we do not always have parameters for those statistics to estimate.

CAPABILITY AND PERFORMANCE INDEXES

The four standard indexes in common use today are the following statistics:

$$\text{Capability Ratio} = C_p = \frac{USL - LSL}{6\,Sigma(X)} = \frac{\text{space available}}{\text{space required}}$$

$$\text{Performance Ratio} = P_p = \frac{USL - LSL}{6\,s} = \frac{\text{space available}}{\text{space used in past}}$$

$$\text{Centered Capability Ratio} = C_{pk} = \frac{2\,DNS}{6\,Sigma(X)} = \frac{\text{effective space available}}{\text{space required}}$$

$$\text{Centered Performance Ratio} = P_{pk} = \frac{2\,DNS}{6\,s} = \frac{\text{effective space available}}{\text{space used in past}}$$

The quantities in these formulas are defined as follows. The difference between the specification limits, $USL - LSL$, is the specified tolerance. As the voice of the customer it defines the *total space available* for the process.

The distance to the nearer specification, *DNS*, is the distance from the average to the nearer specification limit. Operating with an average that is closer to one specification than the other effectively narrows the space available to the process. It is like having a process that is centered within limits that have a specified tolerance = $2\,DNS$. Thus, the numerator of both of the "centered" indexes characterizes the *effective space available* due to the fact that the process is not centered within the actual specification limits.

Sigma(X) denotes any one of several within-subgroup measures of dispersion. One such measure would be the average of the subgroup ranges divided by the appropriate bias correction factor. The quantity denoted by $6\,Sigma(X)$ represents the voice of the process and is the *generic space required by a process* when that process is operated up to its full potential.

The global standard deviation statistic, *s*, is the descriptive statistic introduced in every introductory statistics class. Since it is computed using all of the data, it effectively treats the data as one homogeneous group of values. This descriptive statistic is useful for summarizing the past, but if the process is not being operated predictably the changes in the process will tend to inflate this global measure of dispersion. Thus, this measure of dispersion simply describes the past without respect to whether the process has been operated up to its full potential or not. Thus the denominators of the two performance ratios define the *space used by the process in the past*.

If we think of the underlying capability of the process as a parameter that we are trying to estimate, these four statistics will be four estimates of the same underlying parameter *only when that process is operated predictably and is centered within the specifications*.

As a process is operated unpredictably the two performance indexes will become smaller than the two capability indexes. As a process is operated off-center the two "centered" indexes will become smaller than the other two indexes. So, these four statistics may be estimates of one parameter, four estimates of two different quantities, or four estimates of four different quantities. While the formulas do not change, the interpretation of what each index represents will depend upon the behavior of the underlying process.

All of the information regarding the relationship between the process and the specifications

that can be incorporated into numerical summaries is contained in these four values. Additional numerical summaries are unnecessary. To learn more about these four indexes see my columns "The Gaps between Performance and Potential" **QDD**, Sept. 1, 2010, and "The Problem of Long-Term Capability" **QDD,** July 8, 2013.

The capability and performance ratios make sense because they are ratios of quantities that are logically comparable. They compare the voice of the customer (the space available and the effective space available) with the voice of the process (the process potential and the process performance). These four ratios are also index numbers—values greater than 1.00 are favorable and values less than 1.00 are unfavorable. As a result, no external reference value is needed to understand these capability and performance indexes.

CAPABILITY CONFUSION PART ONE

As soon as people start computing index numbers someone is going to come along with an alternative computation. Back in the 1980s, when capability ratios were new, Richard Lyday kept a file of these alternative formulas. He eventually had over 80 different formulas for capability ratios. Whenever one of these formulas would correctly describe some aspect of the relation between the process and the specifications, it would usually turn out to be equivalent to of one of the four traditional indexes listed above. However, some of these alternate formulas turned out to be fallacious, and these erroneous formulas are the source of much of the capability confusion. We shall look at four of these in what follows.

Two confusing capability ratios are the so called Taguchi capability ratios (even though Dr. Taguchi had nothing to do with the creation of these ratios):

$$\text{Taguchi Capability Ratio} = C_{pm} = \frac{USL - LSL}{6\sqrt{MSD(\tau)}} = \frac{\textit{space available}}{6 \ (\textit{radius of gyration about target})}$$

$$\text{Centered Taguchi Capability Ratio} = C_{pmk} = \frac{2\,DNS}{6\sqrt{MSD(\tau)}} = \frac{\textit{effective space available}}{6 \ (\textit{radius of gyration about target})}$$

where $MSD(\tau)$ is the "mean squared deviation about the target value τ." It is the sum of the within- subgroup variance plus the square of the bias, and it is commonly computed as follows:

$$MSD(\tau) = \left[\, (\textit{Sigma(X)})^2 \ + \ (\bar{\bar{X}} - \tau)^2 \,\right]$$

The square root of the $MSD(\tau)$ is known as the "radius of gyration about the target." While the radius of gyration about the target does characterize a property of the process, there is absolutely no rationale for multiplying this quantity by the number six. (While there is a reason to multiply the standard deviation by six, the denominator of the ratios above does not represent any physical property or characteristic of the process.) Thus, the first problem here is that while each component can be described and explained individually, the ratios above defy rational interpretation.

The rationale commonly given for using the Taguchi capability ratios is that they do a better job of characterizing the degree to which the process is on-target (than do the traditional capability ratios). So how well do they do this?

With the traditional capability ratios it is the size of C_{pk} relative to C_p that describes how close you have come to operating at the mid-point of the specifications. Examination of the traditional formulas will show that this ratio is simply:

$$\frac{C_{pk}}{C_p} = \frac{2\,DNS}{USL - LSL} = \frac{\text{effective space available}}{\text{space available}}$$

With the Taguchi capability ratios it is also the size of C_{pmk} relative to C_{pm} that describes how close you have come to operating at the mid-point of the specifications. Examination of the formulas will show that this ratio is simply:

$$\frac{C_{pmk}}{C_{pm}} = \frac{2\,DNS}{USL - LSL} = \frac{\text{effective space available}}{\text{space available}}$$

So, if we are looking for a "better way" to characterize the degree to which the process is operating on target, why use a more convoluted pair of ratios to end up making exactly the same comparison?

A third problem with the Taguchi capability ratios is that they are not index numbers. When the traditional capability ratios, C_p and C_{pk}, exceed 1.00 you are getting into the zone where you can expect 100% conforming product. When the traditional performance ratios, P_p and P_{pk}, exceed 1.00, you are likely to have produced 100% conforming product in the past. And when these traditional indexes are less than 1.00 the potential for nonconforming product exists.

So what happens when the Taguchi capability ratios change from less than 1.00 to greater than 1.00? Do we move from a region of potential nonconforming product to the region where nonconforming product is unlikely? The answer to this question is no. The size of the Taguchi capability ratio says nothing about the potential for having nonconforming product, and the change from greater than 1.00 to less than 1.00 does not correspond to any interesting transition point for the process relative to the specifications.

Combining all of the above, we see that the Taguchi capability ratios *do not tell you anything useful or new about the capability of the process*. This means that they are not, in any logical sense, *capability ratios*. They are simply an example of statistical purgatory—computations that have no utility in practice (except to confuse the uninitiated).

CAPABILITY CONFUSION PART TWO

Another attempt to come up with a novel capability ratio involves the decomposition of the traditional capability ratio into a "process component" and an "error component." The idea here is to estimate the "true" process capability by isolating and removing that portion of the capability index that is due to measurement error. To do this we start with the idea that a product measurement, X, is the sum of a product value, Y, and a measurement error, E:

$$\text{Product Measurement} = \text{Product Value} + \text{Measurement Error}$$
$$\text{or}$$
$$X = Y + E$$

Since the product values are logically independent of measurement error, the rules of probability theory tell us that the variance for X, $V(X)$, is equal to the sum of the variance for the product stream, $V(Y)$, and the variance of the measurement system, $V(E)$.

$$V(X) \;=\; V(Y) \;+\; V(E)$$

We cannot observe *V(Y)* directly, so we have to work with estimates of *V(X)* and *V(E)*.   If we think of the "true process capability" that is to be estimated as:

$$\boldsymbol{CPY} \;=\; \frac{USL - LSL}{6\;SD(Y)} \;=\; \frac{USL - LSL}{6\sqrt{V(X) - V(E)}}$$

Then a biased estimate of this quantity could be obtained using:

$$Estimate\ of\ \boldsymbol{CPY} \;=\; C_{py} \;=\; \frac{USL - LSL}{6\sqrt{[Sigma(X)]^2 - [Sigma(E)]^2}}$$

where *Sigma(E)* is an estimate of the standard deviation of the measurement system. While more complex formulas for this quantity exist, the problem here is not in the computation, but rather in the idea of a "true process capability."

Whether we are talking about what the process might produce, or inspecting product relative to the specifications, we cannot do so without reference to a measurement system. Thus, the concept of a "true process capability" is an idea that lives right next door to "a line has zero width," and down the street from "the constant pi has an infinite number of decimals." On the mathematical plane all of these ideas are valid, but in this world these theoretical concepts can only be approximated—all lines have measureable width, we always compute things using only a finite number of digits for pi, and the process capability is always observed and estimated using an imperfect measurement system. Any discussion of the "true process capability" is the statistical equivalent of arguments about how many angels can dance on the head of a pin. Or, as Dr. Deming used to put it, "There is no true value of anything. Change the measurement method and you will change the value observed."

While we may sometimes be able to compute the statistic C_{py}, this value has no practical meaning. While it attempts to characterize the product stream for a predictable process relative to the specifications, we can never observe that product stream without using some less than perfect measurement system. And the traditional capability ratio *already* characterizes the *observed* product stream relative to the specifications.

Yet there is an even more serious problem of interpretation for this capability ratio. Since the next piece of confusion shares this problem, it will be explained below.

CAPABILITY CONFUSION PART THREE

Our third example of capability confusion involves the use of the language of capability to describe something completely different. Rather than focusing on the relationship between the voice of the process and the voice of the customer, this ratio harks back to the pre-1984 guidelines which sought to define a good measurement system by comparing it with the specified tolerance.

The statistic C_g uses an arbitrary formula:

$$C_g \;=\; \frac{Arbitrary\ Fraction\ of\ Space\ Available}{6\;Est.\ SD(E)} \;=\; \frac{K\ [USL - LSL]}{6\;Sigma(E)}$$

The denominator is six times the estimated standard deviation of the measurement system, while the numerator is some arbitrary proportion of the specified tolerance for the product measurements. This arbitrary proportion is a user selected value that supposedly represents the

maximum portion of the specifications to be "consumed" by measurement error.

Of course, the arbitrary nature of the numerator above prevents us from coming up with a standard way to interpret this ratio. Yet there is even a more fundamental problem here. Measurement error does not directly consume a portion of the specifications in the manner implied by the ratio above.

Both C_{py} and C_g make comparisons that are logically flawed. Even though the algebra may be done correctly, the comparisons made by these ratios do not make sense. The idea of an index number is that you are comparing like things in the ratio. Both of the ratios C_{py} and C_g make comparisons between things that are not comparable. Returning to the equation for variances given above:

$$V(X) \;=\; V(Y) \;+\; V(E)$$

This relationship is shown graphically on the left side of Figure 2. In consequence, we can only show the relationship between the standard deviation parameters using the sides of a right triangle. (To do otherwise would be equivalent to denying the Pythagorean theorem.)

Now it is always the product measurements, *X*, that get compared to specifications. Thus, we draw a line representing the specified tolerance below the hypotenuse of the right triangle because the hypotenuse represents *SD(X).*
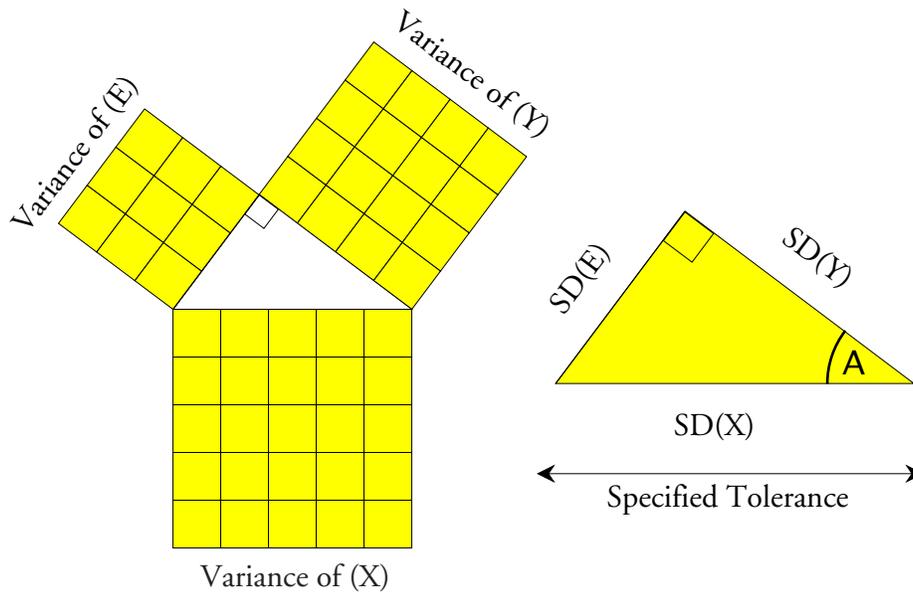

Figure 2: Relationships Between Variances and Standard Deviations

When we compute the traditional capability ratio, C_p we are comparing the specified tolerance with a multiple of an estimate of *SD(X).* The space available is defined by the specified tolerance, the space required is a multiple of *SD(X),* and the capability ratio compares these two spaces directly. This comparison is logical and reasonable. As the value of *SD(X)* changes the process capability will change, and the traditional ratio will describe this change.

However, both *CPY* and C_g seek to compare an estimate of one of the sides of the right

triangle with the specified tolerance. Specifically, The formula for *CPI* can be rewritten as:

$$C_{py} \;=\; \frac{USL - LSL}{6\ SD(Y)} \;=\; \frac{USL - LSL}{6\ SD(X)}\ \frac{SD(X)}{SD(Y)} \;=\; C_p\ \text{secant(A)}$$

Letting *K* denote the arbitrary fraction of the specified tolerance used, C_g can be written as:

$$C_g \;=\; \frac{K\ [USL - LSL]}{6\ SD(E)} \;=\; K\,\frac{USL - LSL}{6\ SD(X)}\ \frac{SD(X)}{SD(E)} \;=\; K\ C_p\ \text{cosecant(A)}$$

Now the problem with these two formulas is that capability ratios are supposed to be indexes. Indexes are commonly expressed and interpreted as percentages. While this interpretation holds true for the traditional capability index, C_p, it does not hold true for either C_{py} or C_g. When you multiply a percentage by a trigonometric function you no longer have a percentage. Interpreting trigonometric functions as percentages will get you hopelessly lost. Whenever you seek to interpret C_{py} and C_g as percentages you are essentially ignoring the consequences of the Pythagorean theorem. And regardless of whether or not you had trouble with trigonometry, the Pythagorean theorem is still true.

Just because two quantities have meaning in their own right does not mean that you can form a ratio of those two quantities and end up with a meaningful result. For each of the formulas given here we can describe in English what each numerator and denominator represents. However, that does not guarantee that the ratios themselves make sense.

And that is how a little algebra, done correctly, can take you to places that have no connection to the real world.