Where Do Manufacturing Specifications Come From?

Donald J. Wheeler

Evidently Steven Ouellette did not like my June column. The adjectives he used were "complicated," "unhelpful," "backward," "confusing," "unnecessary," "crazy," and "disastrous." Yet, before he published his column he had in his possession the full mathematical explanation for the results I presented in that column. Without going into all of the calculus, this column will outline the justification for manufacturing specifications and explain their use.

For the record, my column in June had nothing to say about the important questions of process performance and measurement system acceptability, yet these are the only two questions addressed in Ouellette's reply. So in order to be clear on this point: The only way to avoid shipping some nonconforming product is to avoid making nonconforming product in the first place. To do this you will have to have a capable process and then you will need to operate that process predictably and on target. In my books I call this operating in the Ideal State. Moreover, in order to track process changes in a timely manner you will need a measurement system that is at least a Third Class Monitor. I will say more on this topic later.

For those who are not operating in the Ideal State there is still inspection, imperfect though it may be. This is where guardbanding is sometimes used. Historically, the various guardbanding schemes have been overly conservative, resulting in unnecessary costs for the supplier. Therefore, back in 1984, I used the appropriate probability models to determine how to create appropriate guardbands. The following is an outline of that argument and a summary of those results.

THE PROBABILITY OF CONFORMING PRODUCT

Let us begin by letting Y denote the value of an item in the product stream. When we measure this item we will get some observed value. Denote this observed value by X. The problem of measurement error is that X is seldom the same as Y. Since we have two variables here, we need to use a bivariate probability model. Moreover, since the normal distribution is the classic distribution for measurement error, we shall use a bivariate normal model. Thus, we can place the product values Y along the vertical axis, and our observed values X along the horizontal axis, and our bivariate model creates the ellipse in Figure 1. The better the measurement system, the thinner the ellipse and the stronger the correlation between the product values and the observations.

Now consider the experiment of measuring the same item repeatedly. The item being measured will have a value of Y = y. This value of Y will define a horizontal slice through the ellipse. The width of that slice will define a range of measurement values that will occur in conjunction with Y = y. The distribution shown on the horizontal axis and labeled $f(x|y)$ defines the conditional probability model for the measurements X given that Y = y. This conditional distribution of X given Y has a mean of:

$$MEAN(x|y) = y$$

thus the measurements will cluster around the product value. This distribution also has a standard deviation of:

$$SD(x \mid y) \;=\; \sigma_e \;=\; \textit{standard deviation for measurement error}$$

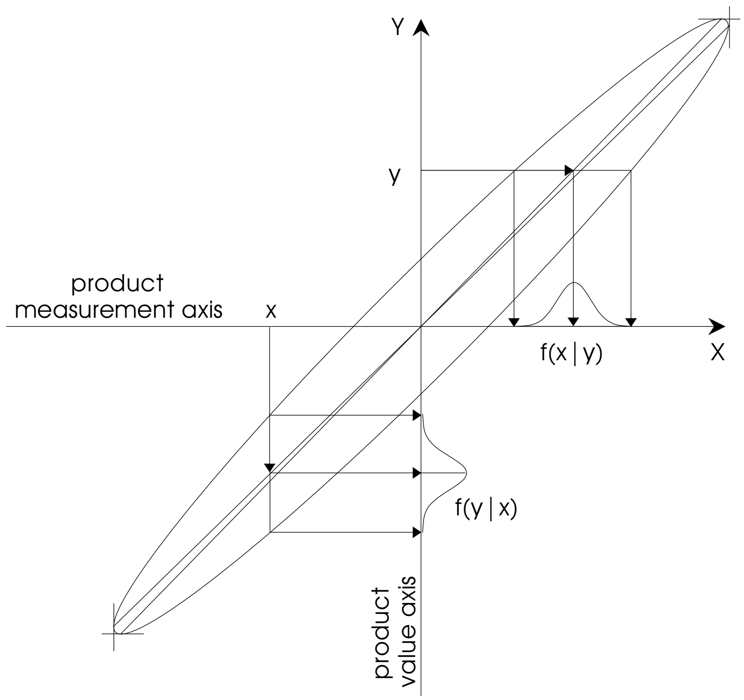

Figure 1: The Bivariate Distribution of Product Values Y and Observed Values X

The fact that the standard deviation of the conditional distribution of X given Y is the standard deviation of measurement error is the reason that all measurement system studies are built upon repeated measurements of a collection of product samples. However, the distribution of X given Y will not help in answering the question of whether of not an item is conforming.

When you are standing at the end of the production line holding an item that you have just measured, the question of interest is, "Given this observed value X, is it likely that the product value Y is within the specifications?" To answer this question we begin with a single observed value X = x. This value for X creates a vertical slice through the ellipse and defines a range of product values Y that could have given rise to the observed value X = x. The conditional distribution of Y given that X = x is labeled as $f(y \mid x)$ and shown on the vertical axis in Figure 1. This conditional distribution defines the probability model for this range of values for Y. This distribution is a normal distribution with mean of:

$$MEAN(y \mid x) \quad = \quad \rho\, x \;+\; (\,1-\rho\,)\; MEAN(X)$$

and a standard deviation of:

$$SD(y \mid x) \quad = \quad \sqrt{\rho}\;\sigma_e$$

where ρ denotes the intraclass correlation coefficient. (This intraclass correlation coefficient is the square of the correlation between X and Y, and may be interpreted as the correlation between two measurements of the same thing.) The mean of this conditional distribution immediately establishes the intraclass correlation as the metric for use in evaluating the acceptability of a measurement system simply because it defines how the mean value for Y becomes less and less dependent upon the value for X as the intraclass correlation drops.

The conditional distribution of Y given X is the distribution we have to use in answering the question: "Is this item likely to be conforming?" Specifically, the probability that the measured item will be conforming may be found by integrating the conditional distribution of Y given X, with respect to Y, between the lower watershed specification limit and the upper watershed specification limit.

WHY USE WATERSHED SPECIFICATIONS?

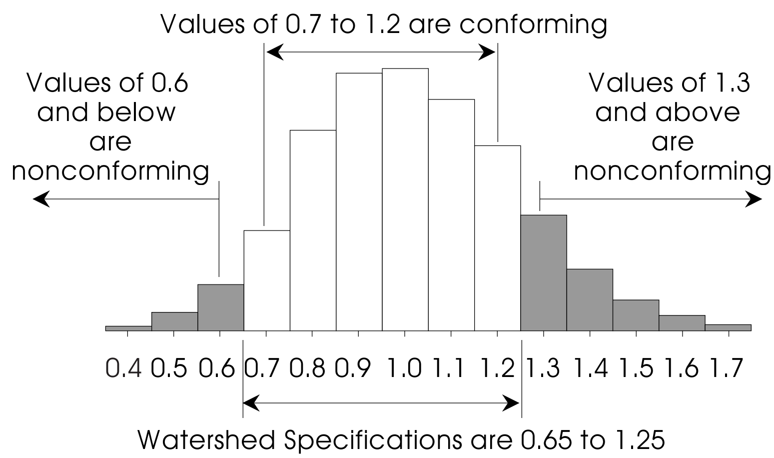

Figure 2 Watershed Specifications

The integral above is going to treat the Y axis as a continuum. In practice, our X values are discrete, with each value rounded off to a specific measurement increment. We have to make an adjustment for this discrepancy between our discrete measurements and the underlying continuum from which they came. According to general practice, specifications are stated in terms of A to B where both A and B are acceptable values. Say A = 0.7 and B = 1.2 and our measurements are recorded to the nearest 0.1 unit. Under these conditions the first nonconforming values would be 0.6 and 1.3. Thus, our watershed specification values are 0.65 to 1.25. This is the portion of the continuum that corresponds to the acceptable values of 0.7 to 1.2, as shown in Figure 2.

THE RESULTS

However we define our manufacturing specifications it should be clear that the most extreme values within those specifications are the ones that are most likely to represent an item that might be nonconforming. Therefore we evaluate the different options that follow by looking at the most extreme values for X that fall within the manufacturing specifications. For these values we evaluate the probability of conforming product using the integral defined above. Of course, these probabilities will vary depending upon the process capabilities. (If that was the point of Ouellette's article, then he was correct in stating that I did not cover this aspect of the problem in the earlier article. I did that for the sake of simplicity.)

If we use the stated specifications as our manufacturing specifications, and if we get an observed value that is either the minimum acceptable value, or the maximum acceptable value, and we calculate the probability of conforming product, we will get the curve shown in Figure 3 as a function of the process capability.

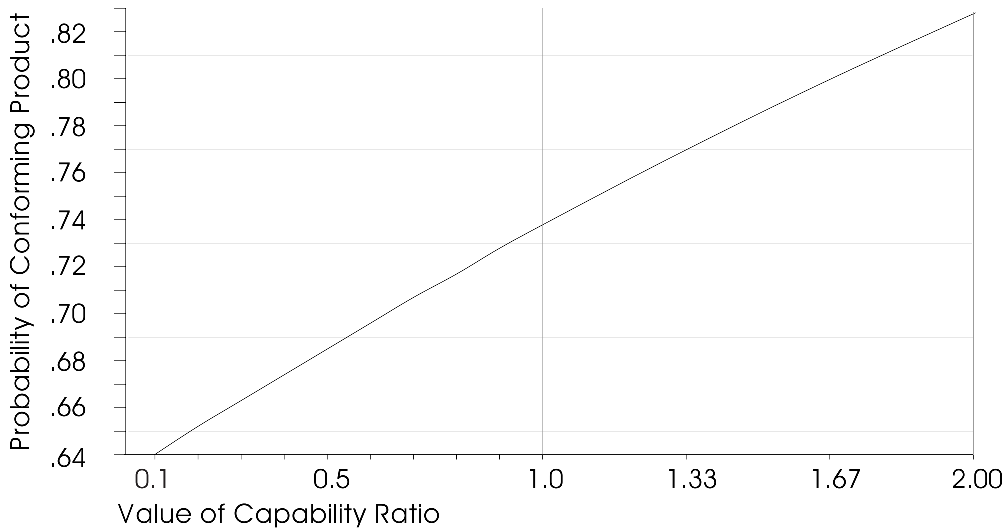

Figure 3: The Minimum Probabilities of Conforming Product Using the Stated Specifications

Thus, without regard for the capability, when you ship using the stated specifications, you can be sure that there is at least a 64% chance that the shipped material will be conforming. With a Capability of 1.0 this minimum probability goes up to at least 74%. With a Capability of 2.0 this minimum probability goes up to at least 83%. If your customer is happy with these numbers, then guardbanding is not for you. Figure 3 is the basis for saying that the watershed specifications define 64% manufacturing specifications.

GUARDBANDING

For those who are not willing to live with the risks of Figure 3, there is always the option of tightening the specifications by some amount. Most schemes for doing this do not take advantage of the mathematics above, and as a result they end up tightening the specifications too much. In terms of what increment to use in defining different options, I used the Probable Error because it is a function of the standard deviation of measurement error, and it also defines the amount of round-off that is appropriate for the measurements.

$$\textit{Probable Error} = 0.675 \; \sigma_e$$

First I considered what would happen if the watershed specifications were tightened by one Probable Error. Looking at the most extreme observed values within these tightened limits and calculating the probability of conforming product we get the curve in Figure 4.

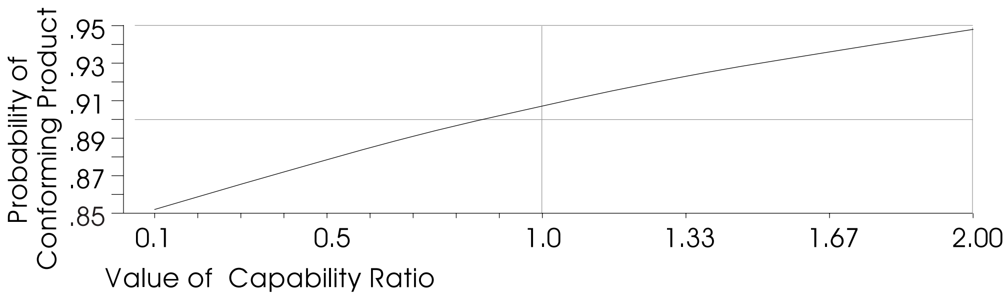

Figure 4: The Minimum Probabilities of Conforming Product when Both Specifications are Tightened by One Probable Error

When the watershed specifications are tightened by one Probable Error on each end you will have at least an 85% chance of conforming product. With a capability of 1.0 this will go up to at least 91%. With a capability of 2.0 this will go up to at least 95%. Thus, when your manufacturing specifications are the watershed specifications tightened by one Probable Error you will have at least an 85% chance of conforming product.

When the watershed specifications are tightened by two Probable Errors on each end you will have at least a 96% chance of conforming product. With a capability of 1.0 this will go up to at least 97.8%. With a capability of 2.0 this will go up to at least 99%. This curve is the bottom of the three curves in Figure 5. Thus, when your manufacturing specifications are the watershed specifications tightened by two Probable Errors you will have at least a 96% chance of conforming product.

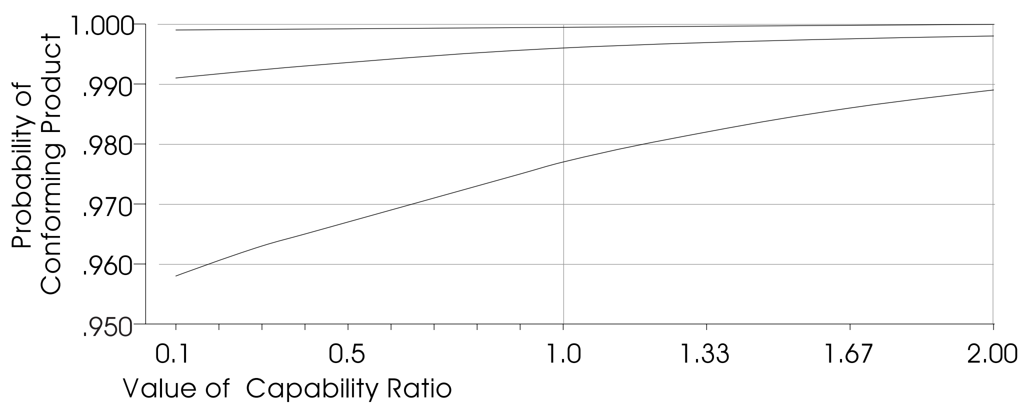

Figure 5: The Minimum Probabilities of Conforming Product when Both Specifications are Tightened by 2, 3, and 4 Probable Errors

When the watershed specifications are tightened by three Probable Errors on each end you will have at least a 99% chance of conforming product. With a capability of 1.0 this will go up to at least 99.6%. With a capability of 2.0 this will go up to at least 99.8%. This curve is the middle of the three curves in Figure 5. Thus, when your manufacturing specifications are the watershed specifications tightened by three Probable Errors you will have at least a 99% chance of conforming product. Notice that three Probable errors will be approximately $2\sigma_e$ rather than the more common, but incorrect, 99% guardband value of $3\sigma_e$.

When the watershed specifications are tightened by four Probable Errors on each end you will have at least a 99.9% chance of conforming product regardless of your capability. This curve is the top curve in Figure 5.

All of these adjustments are much smaller than the traditional values commonly used in guardbanding, which saves the supplier money while providing the protection needed.

THE INTRACLASS CORRELATION

As part of operating in the Ideal State as a way to guarantee 100% conforming product you will need to have a measurement system that can give a timely warning of any process excursion. It turns out that the value of the intraclass correlation defines the relationships and therefore provides the appropriate metric for judging the acceptability of a measurement system for a given application.

The intraclass correlation ρ defines that proportion of the variation in the measurements that can be attributed to the variation in the product stream. The complement of the intraclass correlation ($1 - \rho$) defines that amount of variation in the measurements that is attributable to the measurement system. The intraclass correlation statistic is commonly computed according to:

$$r = 1 - \frac{\textit{estimated variance of measurement error}}{\textit{estimated variance of product measurements}}$$

The estimated variance of measurement error would be the square of our estimate of σ_e from some measurement error study. The estimated variance of product measurements should be obtained from some within-subgroup measure of dispersion using measurements drawn from the product stream. (Global measures of variation should be avoided here.)

An explanation of what the intraclass correlation is and does is given in my book *EMP III: Evaluating the Measurement Process and Using Imperfect Data.* The following is a synopsis of the results established there, although the expression of some of these results has been updated here.

Any signal of a change in the production process will be attenuated by measurement error. This attenuation is characterized by:

$$\textit{Attenuation of Signals from Production Process} = 1 - \sqrt{\rho}$$

The limits on a process behavior chart will be inflated by measurement error. This inflation can be characterized by:

$$\textit{Inflation of Process Behavior Chart Limits} = \frac{1}{\sqrt{\rho}} - 1$$

Of course, between the signal attenuation and the inflation of the limits, measurement error will affect the ability of a process behavior chart to detect process changes in a timely manner. The traditional way of characterizing the sensitivity to a signal is to consider the Average Run Length = the number of subgroups between the point when a signal occurs and the point when it is detected. Here we look at a process shift equal to three Sigma(Y) and consider using the four detection rules of the Western Electric Zone Tests. Table 1 gives the Average Run Lengths (ARL) as a function of the Intraclass Correlation for different combinations of detection rules. These ARL curves are shown in Figure 6.

Table 1: Average Run Lengths for a Three Sigma Process Shift

Intraclass Correlation	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05
Rule 1 ARL	2.0	2.2	2.6	3.3	4.2	5.5	7.6	11.3	20.5	50	101
Rules 1, 2, 3, & 4 ARL	1.7	1.8	2.0	2.2	2.4	2.7	3.2	4.1	5.8	9.5	16.2

These characterizations of how measurement error will affect the ability of a process behavior chart to detect process changes allow us to fully characterize the relative utility of a measurement system for a given application. In doing this we end up with four classes of measurement systems.

When the intraclass correlation is between 1.00 and 0.80 you will have a First Class Monitor. Here any signals from the production process will be attenuated by less than 10%. Using Detection Rule One, the Average Run Length for detecting a three-sigma shift will be less than 2.6 subgroups (compared to 2.0 subgroups for a perfect measurement system).

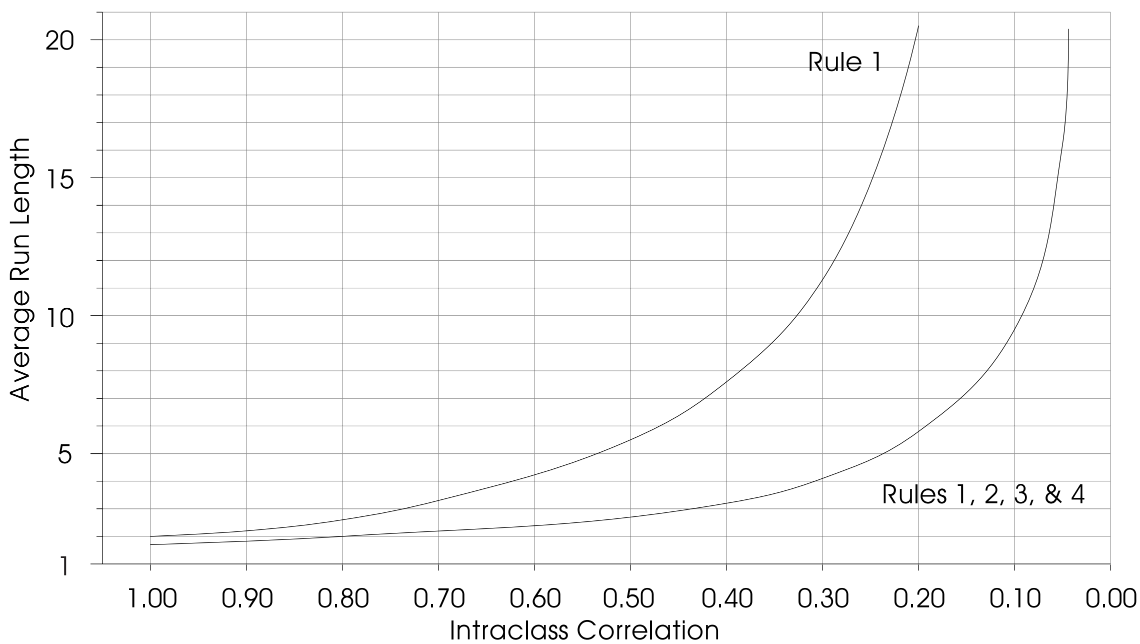

Figure 6: Average Run Lengths for a Three Sigma Process Shift

When the intraclass correlation is between 0.80 and 0.50 you will have a Second Class Monitor. Here any signals from the production process will be attenuated by 10% to 30%. Using Detection Rule One, the Average Run Length for detecting a three-sigma shift will be less than 5.5 subgroups. Using Detection Rules One, Two, Three and Four the Average Run Length for detecting a three-sigma shift will be less than 2.7 subgroups.

When the intraclass correlation is between 0.50 and 0.20 you will have a Third Class Monitor. Here any signals from the production process will be attenuated by 30% to 55%. Using Detection Rules One, Two, Three and Four the Average Run Length for detecting a three-sigma shift will be less than 5.7 subgroups.

When the intraclass correlation is less than 0.20 you will have a Fourth Class Monitor. Here the measurement system is on the ropes and should only be used in desperation. Signals from the production process are attenuated by more than 55%, and the ability to detect process signals rapidly vanishes as measurement error completely dominates the observations.

OUELLETTE'S EXAMPLES

Steven Ouellette used the measurement system described in my June column as the basis for three examples. That measurement system recorded viscosities to the nearest 10 centistokes. The Probable Error for a single reading was found to be 37 cs, and the standard deviation for measurement error was 54.4 cs. In these examples Ouellette assumed that single determinations would be used to characterize each batch of product.

In Ouellette's first example he postulated specifications of 2500 ± 175 cs and a process with a capability of 1.00. In computing his watershed specifications Ouellette made two mistakes. First he used 0.1 times the probable error for his adjustment, rather than half of the measurement increment, and then he tightened the specifications rather than widening them!

In defining the capability to be 1.00 he defined the standard deviation for the product measurements to be 58.33 cs. This leads to an intraclass correlation of:

$$r = 1 - \frac{(54.4)^2}{(58.33)^2} = 0.13$$

This means that the measurement system in this example is a Fourth Class Monitor. Only 13% of the variation in the product measurements actually come from variation in the product stream. Here the Average Run Length for detecting a three-sigma process shift with Rule One is 42 subgroups. With all four rules it is still 8 subgroups.

So, this measurement system will not detect process changes in a timely manner, but can it be used to decide whether or not to ship product? The specifications are 2500 ± 175 = 2325 to 2675. Using these stated specifications will allow you to ship virtually all of your batches, but all that you can say for certain about the marginal batches is that they have at least a 64% chance of conforming. But doesn't Figure 3 show 74% for a capability of 1.0? Yes, it does. But with a Fourth Class Monitor you are not likely to know when your process changes, hence the minimum likelihood of 64% for the marginal batches.

However, guardbanding your specifications by two Probable Errors will result in manufacturing specs of 2394 to 2606. Here you will have to blend the marginal batches (3.5% from each end) but you can assure your customer that the shipped batches have at least a 96% chance of conforming to the stated specifications of 2500 ± 175.

Thus, depending upon what risks you and your customer are willing to take, this Fourth Class Monitor might still be useful in deciding what batches to ship. (Which I believe was Ouellette's point.) However, the fact that a Fourth Class Monitor cannot track process changes in a timely manner means that this process could go on walkabout and you would not know it for quite some time. Here the guardbanding protects you from the limitations of the weak measurement system.

In Ouellette's second example the measurement system is still a Fourth Class Monitor with an intraclass correlation of 0.13. However the specifications were changed to 2500 ± 350, which boosts the capability up to 2.0. Guardbanding the specs by four Probable Errors will give manufacturing specs of 2293 to 2707. Virtually all of the product will get shipped, and even if the process changes you can still assure your customer that the batches you ship have at least a 99.9% chance of conforming. Thus, guardbanding protects the supplier and the customer here in spite of the inability of the measurement system to track process changes.

In Ouellette's third example he postulated specifications of 2500 ± 88 cs and a process with a capability of 0.50. The measurement system is still a Fourth Class Monitor. Using the stated specifications you will have 64% manufacturing specs, and about 13.5% of the batches will be rejected and will have to be blended. About 20% of the stuff you end up shipping to your customer will have about one chance in three of being nonconforming. This is not a pretty picture, but at least we can quantify the risks inherent in using a weak measurement system with tight specs.

However, if we used the average of four determinations, rather than using a single determination, we could turn this Fourth Class Monitor into a high-end Second Class Monitor. Here the Probable Error would be:

$$\textit{Probable Error for Average of Four Readings} = \frac{PE}{\sqrt{n}} = \frac{36.7 \text{ cs}}{\sqrt{4}} = 18 \text{ cs}$$

and the intraclass correlation would be:

$$r = 1 - \frac{(27.2 \text{ cs})^2}{(58.33 \text{ cs})^2} = 0.78$$

Now the measurement system can track process changes and also help in improving the production process. Guardbanding the specs by one Probable Error would result in manufacturing specifications of 2425 to 2575. This would increase the number of blended batches from 13.5% to about 19.5%, but now you could assure your customers that the shipped batches would have at least an 88% chance of conforming.

All of which reminds me of one of my clients who told me that: "We never make any nonconforming product here."

"Oh, really?"

"Yeah, if a batch doesn't qualify, it will always qualify as 'Base Fluid'."

"So what does that do for you?"

"Well, at one point we had a two-year supply of Base Fluid on hand."

Guardbanding does not solve the problems of bad measurement processes, nor does it make the production process any better. It simply buys a piece of insurance at the point of shipping the product. It can be used with good measurement systems and poor measurement systems. It can be used with processes having small capabilities and also with those having large capabilities. It is a technique for quality assurance, rather than one for quality improvement. While it is always better to avoid burning the toast, once it is burned it is time to think about how to scrape it.