
How Measurement Error Affects the Four Ways We Use Data

Donald J. Wheeler

Measurement error is generally considered to be a bad thing, and yet there is very little written about how measurement error affects the way we use our measurements. This column will consider these effects for four different uses of data. But first we need to describe how to characterize measurement error in both an absolute and a relative manner.

THE PROBABLE ERROR

If we repeatedly measured the same thing thousands of times using the same measurement system, and if our measurements had very small increments, we would end up with a histogram of measurements like Figure 1. Given such a collection of measurements, the intuitive estimate for the value of the thing being measured is the average of all of the measurements. But what does the dispersion of the histogram in Figure 1 represent?

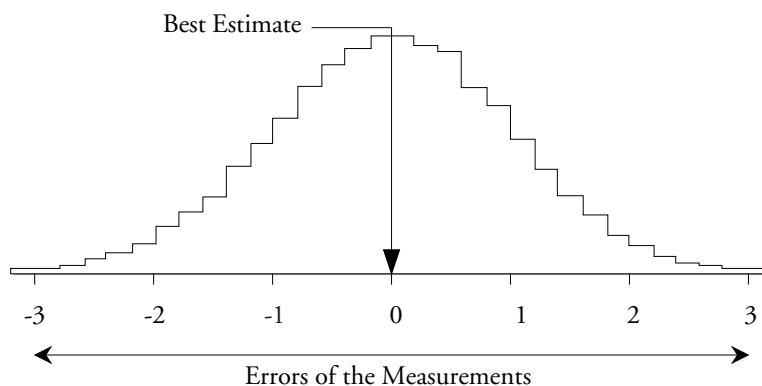


Figure 1: Repeated Measurements of the Same Thing

The variation in Figure 1 is due to measurement error. Without measurement error, all of the values would be equal to the average value. Thus, we can characterize measurement error by describing the variation in Figure 1. The National Institute of Standards and Technology recommends reporting the standard deviation statistic for the histogram. While this statistic will provide an efficient summary of the variation in Figure 1, it has the disadvantage of being impossible to explain in any intuitive manner. (For example, while the average is the balance point for the histogram, the standard deviation statistic can only be described as the square root of the rotational inertia for the histogram! So unless you like people looking at you like you are crazy, you will need to find another way to characterize the dispersion.)

Another way of characterizing the dispersion of a histogram is provided by the Probable Error. The Probable Error dates back to the 1820s, and it can be defined as a function of the standard deviation statistic:

$$\text{Probable Error} = 0.675 \text{ Standard Deviation Statistic}$$

The advantage of using the Probable Error is that it can be explained quite simply. The Probable Error effectively defines the median amount by which a measurement will err. Half the time a measurement will deviate from the best estimate by less than one Probable Error. Half the time a measurement will deviate from the best estimate by more than one Probable Error.

To illustrate this point we need to go back over 200 years. Toward the end of the Eighteenth Century Laplace twice tried to develop an intuitive model for measurement error. Since both of his models had problems, he abandoned this effort in 1777. In 1805, Gauss simply assumed that the normal distribution would provide a model for measurement error, but he offered no explanation for his choice. Then in May of 1810 Laplace published what came to be known as the Central Limit Theorem. In a footnote added as the paper went to publication, Laplace observed that his theorem justified Gauss's assumption and also solved his earlier quest for an intuitive model for measurement error. Since that time, the traditional model for measurement error has been the normal distribution.

A characteristic of the normal distribution is that the middle 50 percent of the area under the curve will be defined by the interval shown in Figure 2:

[Mean - 0.6745 Standard Deviation] to [Mean + 0.6745 Standard Deviation]

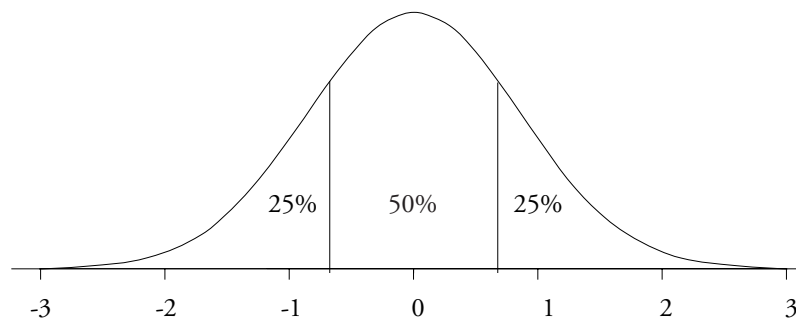


Figure 2: The Middle 50% of the Normal Distribution

So what does all of this mean in practice? Given that we have measured an item, what can we say about that observed value? Using the generic model for measurement error shown in Figure 3, our observed value will fall within the middle zone of Figure 2 half the time, and it will fall outside the middle zone half the time. When it falls within the middle zone, the error of the measurement will be less than one Probable Error. When it falls outside the middle zone, the error of the measurement will be greater than one Probable Error. While we will never know which is the case, we can say that the median amount by which we have erred is defined by one Probable Error.

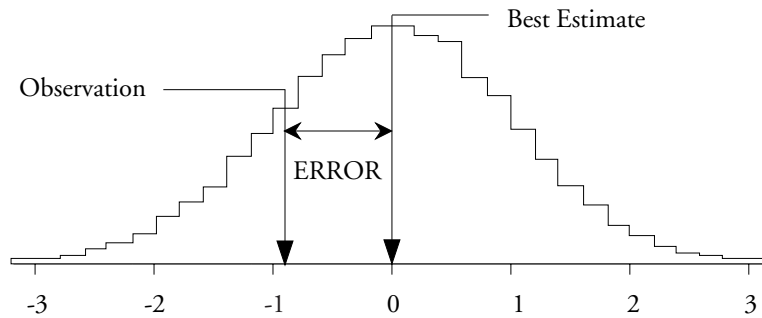


Figure 3: The Error of a Single Measurement

Thus, when we use the Probable Error to describe the uncertainty of a measurement, we are providing our listeners with a description of the median error of a single measurement. Since people intuitively understand averages and medians much better than they understand any measure of dispersion, this approach has the advantage of actually communicating with your listeners. (Moreover, since the Probable Error can be expressed in terms of the standard deviation statistic, it is fully equivalent to the recommended approach.)

While the Probable Error provides a way to characterize the uncertainty in a measurement in an absolute sense, we will also need a way to characterize the relative utility of a measurement.

THE INTRACLASS CORRELATION

My December column described the traditional measure of relative utility for a measurement, the Intraclass Correlation Coefficient (ICC). This ratio compares the variance of the stream of product values with the variance of the stream of product measurements:

$$ICC = \frac{\text{Variance of Product Values}}{\text{Variance of Product Values} + \text{Variance of Measurement Errors}}$$

This ratio will describe that proportion of the variation in the product measurements that is directly attributable to the variation in the product stream. As measurement error gets larger relative to the product variation this ratio will decrease. Thus, this ratio will characterize the relative contributions of product variation and measurement error to the overall uncertainty in a measurement.

In answering the question of how measurement error affects the ways in which we use measurements we shall make use of both the Probable Error and the Intraclass Correlation Coefficient. In doing this it is helpful to consider four different ways that we use measurements.

DESCRIPTION

One use of a measurement is to describe the measured item. The need to describe might be motivated by idle curiosity, by a need to be informed, or by a desire to have the data for future use; but whatever the motivation, this use of a measurement is to answer the question of how many or how much. When a value is used to describe the measured item it is important to understand the limitations of the value used—the uncertainty attached to the value itself. The source of this uncertainty will primarily be the uncertainty in the measurement process.

Let us assume that there is a jig to hold the part shown in Figure 4, gauges to measure the dimensions L and D , and a procedure for loading the part into the jig and reading the gauges. Without getting lost in the particulars for Figure 4, this measurement process is like most in that it has both a measuring device and a procedure for using it.

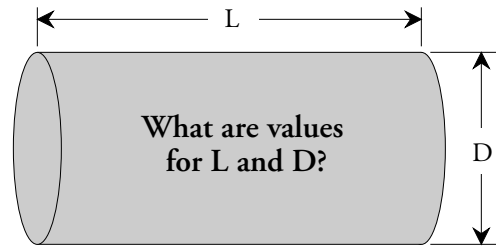


Figure 4: Using a Measurement for Description

So, when you have your measurement, how many digits do you record? Say the instrument readout for L is 2.003248 inches. Here the measurement increment is one millionth of an inch. But what if the Probable Error for these measurements of L is 0.001 inch? From above, the measurement will err by more than 0.001 inch at least half the time. This will make the last three digits in the readout for L complete noise. To record the value for L beyond three decimal places is to record noise. Thus, the Probable Error tells us when we are recording too many digits.

On the other hand, assume that the conventional wisdom is that the diameters D are only good to the nearest tenth of an inch. Our readout value for D might be set to only show the first decimal place, and our value for D might be recorded as 1.1 inches. But what if the Probable Error for measurements of D is also 0.001 inches? The measurements could be good to a thousandth, but they are recorded to a tenth! Rounding the measurements off to the nearest tenth of an inch will introduce round-off error and degrade the quality of the measurements. Thus, Probable Error also tells us when we need to add digits to our recorded values.

Therefore, the first thing we learn is that knowledge of the Probable Error will help you to record the proper number of digits. In general, you will want to have a measurement increment that is approximately the same size as the Probable Error. Useful guidelines are:

Your measurement increment should not be larger than 2 Probable Errors.

Your measurement increment should not be smaller than 0.2 Probable Error.

When your measurement increment falls outside the range defined above you will either be throwing away useful information in the round-off or recording noise. While the second of these two mistakes may be preferable to the first, both are misleading and inappropriate. Knowing the Probable Error allows you to detect and avoid both mistakes.

CHARACTERIZATION

A second way in which we use measurements is to characterize the item measured relative to some specification. Of course, any attempt to characterize an item will involve the description of that item, thus the comments regarding the number of digits to record given above will also apply here. In fact, knowing the appropriate number of digits to record can be essential in

knowing when specifications are unrealistically tight. Returning to the example above, specifying the dimension for L as plus or minus 0.0001 inches when the measurements have a probable error of 0.001 inch is a formula for trouble.

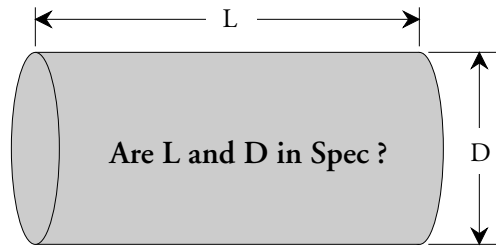


Figure 5: Using a Measurement for Characterization

When using a measurement for characterization there is the inevitable problem of misclassification. How can we be sure that the item is conforming when measurement says it is? This problem was addressed in my June and July columns of 2010. There we found that the Probable Error provides the essential piece of information for removing the uncertainty associated with this use of a measurement. To summarize those columns, a measurement cannot begin to be used to certify that an item is conforming until the specified tolerance is greater than 5 or 6 Probable Errors.

REPRESENTATION

Whenever we describe an item or characterize an item relative to specifications we are focused on the item measured. Since the Probable Error provides an absolute characterization of measurement error it is sufficient to answer the questions in these two uses. However, when we begin to use our measurements to describe or characterize the product *that was not measured* we will need more than a absolute characterization of measurement error.

Representation is the name I use to refer to the act of characterizing a batch, lot, or portion of production relative to specifications. This happens in all sorts of environments where a sample is taken and used to decide if the product produced in the same batch, lot, or time period is, or is not, conforming. In order to make this determination we will have to extrapolate from the product that was measured to the product that was not measured. And all such extrapolations will depend upon the assumption that the sample is representative of the whole.

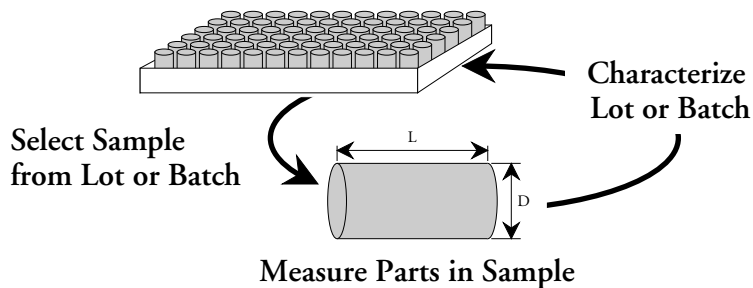


Figure 6: Using Measurements for Representation

The uncertainty of this extrapolation involves more than the uncertainty of measurement

error. A second component of this uncertainty is fact that the product measured will not be exactly the same as the product not measured. If we can account for this difference between the sample and the remainder of the batch, then we can make our extrapolation with some assurance of being right. However, if we cannot account for the difference between the sample and the remainder of the batch, our attempt at using the sample to represent the batch will be little more than wishful thinking. Thus, we come to a dilemma. We can use a sample to represent a batch only if we understand the differences between the sample and the batch. Failing this, our attempts at representation will fail. Since the question of when can you extrapolate will also arise with the next use of data, the answer to this dilemma will be discussed in the next section.

PREDICTION

A fourth way that we use measurements is to predict what a production process is likely to produce in the future. Here we are extrapolating from the product measured to the product not yet made. Such an extrapolation is called for every time a new process is “verified” or “certified.”

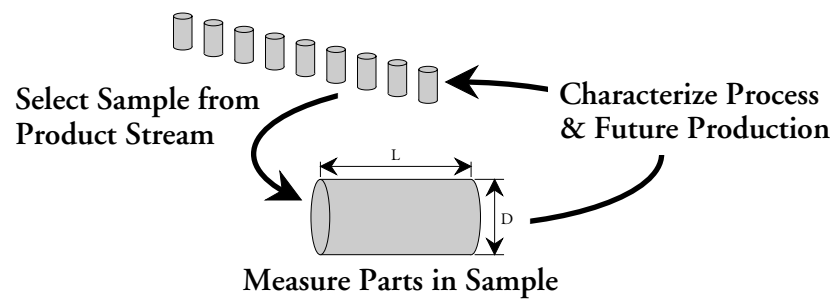


Figure 7: Using Measurements for Prediction

The uncertainty of this extrapolation involves more than the uncertainty of measurement error. AHere we have the uncertainty introduced by the fact that the product measured will be different from the product not measured, and in addition, we also have the uncertainty that comes from the fact that the product produced in the future might be different from the product already produced. If we can account for these additional sources of uncertainty, then we can make a reasonable prediction. If we cannot account for these additional sources of uncertainty, then any predictions we make will be little more than wishful thinking.

Thus, both the problem of representation and the problem of prediction depend upon the representativeness of the sample. How can we have any confidence that the variation in the sample is the same as the variation in the batch or in the production process over time? The answer is that we are going to have to have some way of evaluating the homogeneity and consistency of the production process. If we have a homogeneous product stream, then any sample drawn from that product stream will provide a reasonable basis for representation and prediction. However, if we do not have a homogeneous and consistent product stream, if the product stream is changing over time, then our sample is going to mislead us and our attempts at representation and prediction will fail.

Now you already know that the only way to determine if a product stream is homogeneous and consistent over time is to use a process behavior chart on that process. If the process has been operated predictably in the past, then it is reasonable to assume that it is likely to continue to be operated predictably in the foreseeable future. When this is the case the Natural Process

Limits will fully characterize the variation within the product stream, or within the batches, and this characterization will allow you to extrapolate from the measured product to the product not measured.

However, if the process has not been operated predictably in the past, then it is highly unlikely that it will be operated predictably in the future. Here you will know that the process is changing at sundry times and in unpredictable ways. These changes will completely undermine any attempt to extrapolate from the sample to the product stream or to the batches.

Thus, the only reliable basis for extrapolating from a sample to a batch or to the product stream is to have a production process that is being operated predictably. The predictable operation will guarantee the homogeneity that is required in order to assume that the sample is representative of the batch or product stream. And the only way to determine if a process is operated predictably is to have a process behavior chart for that process. As a result, the question of how measurement error can affect your ability to use your measurements for representation or prediction turns into a question about how measurement error affects your ability to use a process behavior chart to detect a process that is being operated unpredictably.

Figure 8 provides the answer to this latter question. The derivation of Figure 8 was explained in my December column. Here we are interested in the consequences of this graph.

As we move from left to right along the horizontal axis the measurement error increases and the Intraclass Correlation Coefficient drops from 1.00 to 0.00. The vertical axis shows the probability of detecting a three standard error shift in location. The curves show how these probabilities change as the measurement error increases.

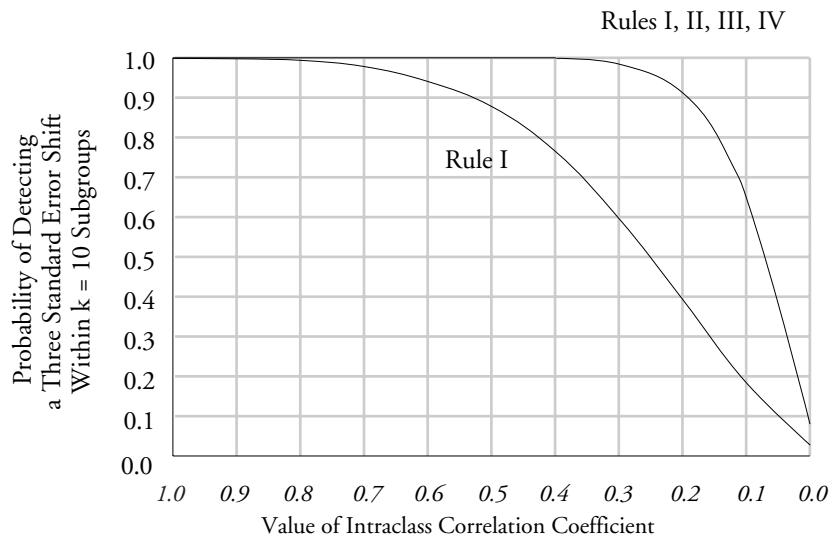


Figure 8: How Measurement Error Affects Our Ability to Detect a Three Sigma Shift

The first curve shows what happens when we interpret a point outside the three-sigma limits as a signal of a process change (detection rule one). Looking at this curve we see that measurement error will have to get large before there is any appreciable loss in our ability to detect process changes. For example, when the ICC value drops to 0.50 the measurement error will be the same size as the variation in the process stream, and yet we will still have an 88% chance of detecting our three standard error shift.

The second curve shows what happens when we use the popular set of detection rules

known as the Western Electric Zone Tests. With all four of these detection rules in play, we can maintain better than a 90% chance of detecting a three standard error shift all the way down to an ICC value of 0.20. At this point the variation due to measurement error will be *four times larger* than the variation in the product stream!

Thus, both theory and practice combine to tell us that even large amounts of measurement error will not have a large impact upon our ability to use measurements for representation and prediction.

SUMMARY

The Probable Error provides an absolute characterization of measurement error. It is easy to explain to those who need to use and understand it. It defines the proper number of digits to record for any measurement, and it provides a yardstick for creating manufacturing specifications when needed. When measurements are used to describe or characterize the items measured the Probable Error will answer all of the questions regarding the impact of measurement error.

When measurements are used to extrapolate from the sample to the batch or the product stream we will need to have some basis for justifying the extrapolation. The only justification that has ever been demonstrated to be effective in practice is to have a production process that has been operated predictably, and the only technique for determining when this happens is the process behavior chart. Since the representative and predictive uses of data involve more sources of uncertainty than just measurement error, we have to use the traditional measure of relative utility for a measurement, the Intraclass Correlation Coefficient, to characterize how measurement error affects these two uses of data. As shown, measurement error has to become very large before it can hinder either the representative or predictive uses of data.

So while measurement error may be undesirable, it is a fact of life. Moreover, it is a fact of life that we can live with. A proper understanding of Probable Error and the Intraclass Correlation Coefficient will allow us to quantify the uncertainty in our measurements, and to know when our ability to use the measurements has been compromised. Fortunately, we can use our imperfect data to make correct decisions even in the presence of large amounts of measurement error.