

## So What is the Fraction Nonconforming?

How to know what the data are really telling you

Donald J. Wheeler

One of the most common questions about any production process is “What is the fraction nonconforming?” Many different approaches have been used to answer this question. This paper will compare the two most widely used approaches and define the essential uncertainty inherent for all of these approaches.

In order to make the following discussion concrete we will need an example. Here we shall use a collection of 100 observations obtained from a predictable process. These values are the lengths of pieces of wire with a spade connector on each end. These wires were used to connect the horn button on a steering wheel assembly. For purposes of our discussion let us assume the upper specification for this length is 113 mm.

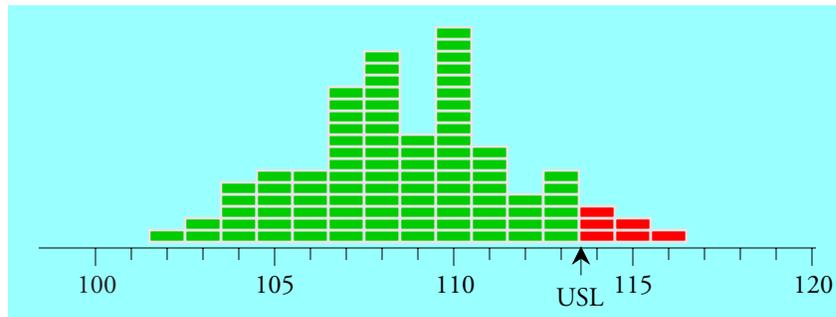


Figure 1: Histogram of 100 Wire Lengths

### THE EMPIRICAL APPROACH

The oldest and simplest approach to estimating the fraction nonconforming is arguably the best approach. We simply divide the number of items that are outside the specifications,  $Y$ , by the total number of items inspected,  $n$ . The name commonly used for this ratio is the binomial point estimate:

$$p = \text{Binomial Point Estimate} = \frac{\text{No. Nonconforming}}{\text{Total No. Observed}} = \frac{Y}{n}$$

This ratio provides an unbiased estimate of the process fraction nonconforming. For the data of Figure 1, six of the observed values exceed 113, so  $Y = 6$  while  $n = 100$ , and our binomial point estimate for the fraction nonconforming is  $p = 0.06$  or 6%.

(Note that calling the ratio above the binomial point estimate is simply a label. It identifies the formula. If the two counts above satisfy certain conditions then this ratio would provide a point estimate of a binomial parameter. Calling the ratio the binomial point estimate does not imply any assumption about a probability model for either the counts or the measurements upon which they are based.)

If the data come from 100% inspection, then there is no uncertainty in the descriptive ratio above. The 6% is the fraction rejected at inspection, and the only uncertainty is the uncertainty of an item being misclassified.

However, if we are using the data of Figure 1 to *represent* product not measured, or to *predict* what might be made in the future, then we will need to be concerned with the uncertainty involved in the extrapolation from the product measured to the product not measured. Here, because the production process was being operated predictably, this extrapolation makes sense.

When data are used for representation or prediction we will need to use an interval estimate in addition to the point estimate for the fraction nonconforming. The interval estimate will define the range of values for the *process* fraction nonconforming that are *consistent* with the observed point estimate.

Most textbooks will give a formula for an approximate 95% interval estimate that is centered on the binomial point estimate  $p$ .

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

This formula is commonly referred to as the Wald interval estimate (even though it was first published by Laplace in 1812). While this simple approximation is satisfactory when the proportions are in the middle of the range between 0.00 and 1.00, it does not work well for proportions that are close to either 0.00 or 1.00. Since the fraction nonconforming will hopefully be near 0.00, we will need to use the more robust Agresti-Coull interval estimate.

The 95% Agresti-Coull interval estimate uses a formula similar to the Wald formula above, but it uses the Wilson point estimate in that formula. For a 95% interval estimate the Wilson point estimate is approximated by adding two successes and two failures:

$$\tilde{p} = \frac{Y + 2}{n + 4}$$

With this adjustment we obtain a 95% interval estimate that works all the way down to  $Y = 0$ . In our example, using  $Y = 6$  and  $n = 100$ , the Wilson point estimate is 0.0769, and the 95% Agresti-Coull interval estimate for the process fraction nonconforming is:

$$\tilde{p} \pm 1.96 \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.0769 \pm 0.0512 = 0.026 \text{ to } 0.128$$

So, the data in Figure 1 give us a binomial point estimate of 6%, and a 95% Agresti-Coull interval estimate for the process fraction nonconforming of 2.6% to 12.8%. While 6% nonconforming is our best point estimate, the uncertainty of the extrapolation from our observed values to the underlying process means that our observed value of 6% nonconforming is consistent with a process that is producing anywhere from 2.6% to 12.8% nonconforming.

If we changed the upper specification for our example to be 114 mm, then  $Y$  would be 3, the binomial point estimate would be 3%, and the 95% Agresti-Coull interval estimate would be  $0.0481 \pm 0.0411$ . Thus, an observed value of 3% nonconforming would be consistent with a process fraction nonconforming between 0.7% and 8.9%.

Now consider what would happen if the upper specification was 116 mm. Here  $Y$  would be

0, the binomial point estimate would be 0%, and yet the 95% Agresti-Coull interval estimate would be  $0.0192 \pm 0.0264$ . Thus, our observed value of 0.0% nonconforming would be consistent with a process fraction nonconforming between 0.0% and 4.6%.

Since  $Y$  can not get any smaller than zero, this last interval estimate reflects the limitations of the inferences that can be drawn from 100 observed values. Processes producing less than 4.6% nonconforming can, and will, produce some 100 piece samples that have zero nonconforming items!

Thus, the Agresti-Coull interval estimate provides us with a way to characterize the process fraction nonconforming based on the observed data. It defines the uncertainty that is inherent in any use of the data to estimate the process fraction nonconforming.

#### THE PROBABILITY MODEL APPROACH

Sometimes, rather than using the data to directly estimate the fraction nonconforming, a probability model is fitted to the histogram and used to compute the tail areas beyond the specification limits. While the data are used in fitting the probability model to the histogram, the estimate of the fraction nonconforming will be obtained from the fitted model rather than directly from the data. As an example of this approach we will fit a normal distribution to the Wire Length Data.

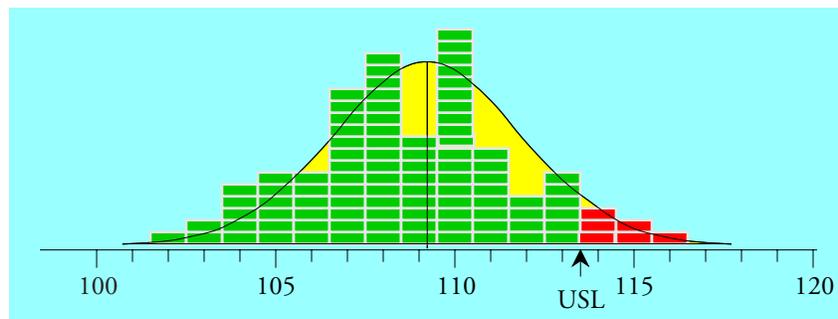


Figure 2: A Normal Distribution Fitted to the Wire Length Data

The Wire Length Data have an average of 109.19 mm, a standard deviation statistic of 2.82 mm, and the process behavior chart shows no evidence of unpredictable operation while these values were obtained. A normal probability model having a mean of 109.19 and a standard deviation parameter of 2.82 is shown superimposed on the Wire Length Data in Figure 2.

As before, we assume that the upper specification limit is 113 mm. Since the measurements were made to the nearest whole millimeter, this upper spec becomes 113.5 mm in the continuum used by the model. When we standardize 113.5 mm we obtain a  $z$ -score of 1.53, and from our standard normal table we find that this corresponds to an upper tail area of 0.0630. Thus, using a normal probability model we obtain a point estimate of the process fraction nonconforming of 6.3%, which is essentially the same as the binomial point estimate found earlier.

So, just how much uncertainty is attached to this estimate? Judging from the few cases where the interval estimate formulas are known for model-based point estimates, we can say that if the probability model is appropriate, then this estimate is likely to have a slightly smaller interval estimate than the empirical approach. However, if the probability model is not appropriate, then

this estimate can have substantially more uncertainty than the empirical estimate. Which brings us to the first problem with using a probability model: Any choice of a probability model will, in the end, turn out to be an unverifiable assumption. It amounts to nothing more than an assertion made by the investigator.

While lack-of-fit tests may sometimes allow us to rule out a probability model, no test will ever allow us to *validate* a particular probability model. Moreover, given a sufficient amount of data, you will *always* detect a lack of fit between your data and any probability model you may choose. This inability to validate a model is the reason that it is traditional to use the normal distribution when converting capabilities into fractions nonconforming. Since the normal distribution is a maximum entropy distribution, its use amounts to performing a generic, worst-case analysis. (It is important to note that this use of a normal distribution is a matter of convenience, arising out of a lack of information, and is not the same as an *a priori* requirement that the data "be normally distributed.")

To illustrate the generic nature of estimates based on the normal distribution we will use the Ball Joint Socket Thickness Data shown in Figure 3. There we have 96 values collected over the course of one week while the process was operated predictably. The average is 4.656 and the standard deviation statistic is 1.868.

The first probability model fitted to these data is a normal distribution having a mean of 4.656 and a standard deviation parameter of 1.868. There is a detectable lack of fit between the histogram and this normal distribution.

The second probability model fitted to these data is a gamma distribution with  $\alpha = 6.213$  and  $\beta = 0.749$ . This model has a mean of 4.654 and a standard deviation of 1.867. There is no detectable lack of fit between this gamma distribution and the histogram.

The third probability model fitted to these data is a Burr distribution with  $c = 1.55$  and  $k = 58.55$  that has been shifted to have a mean of 4.656 and stretched to have a standard deviation parameter of 1.868. There is no detectable lack of fit between this Burr distribution and the histogram.

So we have two models that "fit" these data and one model that "does not fit" these data.

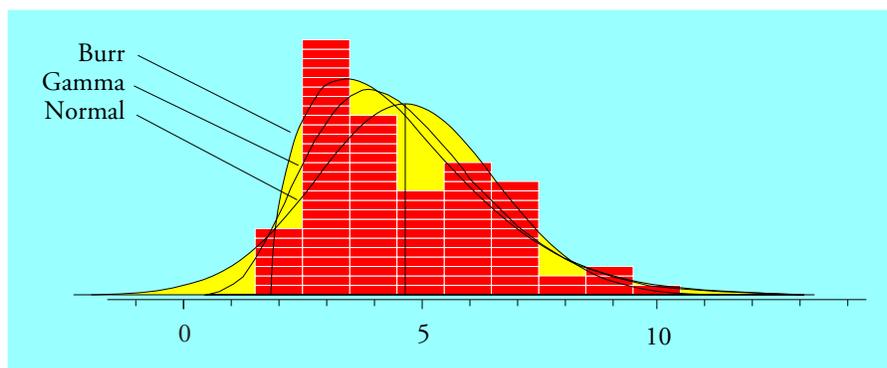


Figure 3: Two Probability Models Fitted to the Socket Thickness Data

Skewed histograms usually occur when the data pile up against a barrier or boundary condition. As a result we are commonly concerned with the areas in the elongated tail. So here we will consider the upper tail areas defined by the cutoff values of 5.5, 6.5, 7.5, etc. Figure 4

shows these upper tail areas computed four ways: (1) using the normal distribution, (2) using the fitted gamma distribution, (3) using the fitted Burr distribution, and (4) using the empirical binomial point estimate. In addition, we find that the 95% Agresti-Coull intervals bracket all four estimates for each cut-off value.

Cutoff Value	Y	Agresti Coull 95% Lower Bound	Normal Model Estimate	Gamma Model Estimate	Burr Model Estimate	Binomial Point Estimate	Agresti Coull 95% Upper Bound
5.5	32	.247	.320	.288	.276	.333	.433
6.5	18	.122	.153	.156	.151	.188	.278
7.5	6	.027	.057	.078	.078	.063	.133
8.5	4	.014	.016	.036	.037	.042	.107
9.5	1	0	.0036	.016	.017	.010	.063

Figure 4: Four Different Point Estimates of Upper Tail Areas for the Socket Thickness Data

In spite of the differences between the four estimates in each row, all four estimates fall within the 95% Agresti-Coull interval. This illustrates what will generally be the case: *The uncertainty inherent in the data will usually be greater than the differences between the various model-based estimates.*

This makes any discussion about which model-based estimate is best into an argument about noise. When we are estimating the fraction nonconforming, the uncertainty in our estimates will generally overwhelm the differences due to our choice of probability model. This uncertainty even covered the estimates when there was a detectable lack of fit for the normal distribution.

*The numbers you obtain from a probability model are never really as precise as they look.*

This is why the generic ballpark values obtained by using a normal distribution are generally sufficient. The ballpark is so large that the normal distribution will get you in the right neighborhood even when there is a detectable lack-of-fit.

“But using a fitted probability model will let us compute tail areas for capability indexes greater than 1.00.”

Yes, it will, and that is the second problem with the probability model approach. No matter how many data you have, there will always be a discrepancy between the extreme tails of your probability model and the tails of your histogram. This happens simply because histograms always have finite tails.

#### HOW HISTOGRAMS GROW WITH $n$

Figure 5 shows the average number of standard deviations between the average value for a histogram and the most extreme value of that histogram. As a histogram grows to include more data the maximum and minimum values move away from the average value. Figure 5 shows the average size of the finite tails of histograms involving different amounts of data.

$n$ for Histogram	Average Distance	$n$ for Histogram	Average Distance	$n$ for Histogram	Average Distance
30	2.04	90	2.47	500	3.04
40	2.16	100	2.51	1000	3.24
50	2.25	150	2.65	2000	3.43
60	2.32	200	2.75	3000	3.53
70	2.38	300	2.88	4000	3.61
80	2.43	400	2.97	5000	3.65

Figure 5: Average Number of Standard Deviations Between the Average and the Most Extreme Value for Histograms of  $n$  Data

Once you get beyond 200 data, the tails of the histogram grow ever more slowly with the increasing number of data. While most of the values in Figure 5 have been known since 1925, this aspect of data analysis has seldom been taught to our students. Histograms with less than 1000 data will rarely have points more than 3.3 standard deviations away from the average. This means that the major discrepancies between a probability model and a histogram are going to occur in the region out beyond three standard deviations on either side of the mean.

These discrepancies undermine all attempts to use probability models to compute meaningful fractions nonconforming when the capability indexes get larger than 1.10. The values in Figure 5 show that when the capability indexes get larger than 1.10 you will commonly have no data points outside the specifications. As a result your count  $Y$  will generally be zero, the point binomial estimate will be zero, and the Agresti-Coull interval estimate will depend solely upon the number of data in the histogram.

However, when we use a probability model to estimate the fraction nonconforming for capability indexes larger than 1.00 we will have to compute *infinitesimal areas under the extreme tails of the assumed probability model*. Here the result will depend upon our assumption rather than depending upon the data.

To illustrate this dependence Figure 6 will extend Figure 4 to cutoff values beyond three sigma. Here the upper tail areas are given in parts per million.

Cutoff Value	Agresti Coull 95% Lower Bound	Normal Model Estimate	Gamma Model Estimate	Burr Model Estimate	Binomial Point Estimate	Agresti Coull 95% Upper Bound
10.5	0	588	6778	7400	0	47440
11.5	0	72	1083	3100	0	47440
12.5	0	7	414	1200	0	47440
13.5	0	0.453	154	478	0	47440
14.5	0	0.023	56	180	0	47440
15.5	0	0.001	20	24	0	47440

Figure 6: Additional Upper Tail Areas for the Socket Thickness Data in Parts Per Million

Both the gamma model and the Burr model showed no detectable lack of fit with these data. Yet the upper tail areas from these two "fitted" models differ by as much as a factor of three. So, which model is right?

Say the upper specification limit for the Socket Thickness Data is 13.5. Which estimate from

Figure 6 should you tell your boss?

1. One-half part per million nonconforming?
2. One hundred fifty-four parts per million nonconforming?
3. Four hundred seventy-eight parts per million nonconforming?
4. Or something less than 4.7% nonconforming?

Only the fourth answer is based on the data. The first three values are imaginary answers based on the infinitesimal areas in the extreme tails of assumed probability models.

When the uncertainty interval is zero to 47,000 parts per million, any model you pick, regardless of whether or not it "fits the data, will manage to deliver an estimate that falls within this interval.

So while our assumed probability models allow us to compute numbers out to parts per million and even parts per billion, *the data themselves will not support an estimate that is more precise than something less than five parts per hundred.* Think about this very carefully.

When you compute tail areas out beyond three sigma you are computing values that are entirely dependent upon the assumed probability model. These tail areas will have virtually no connection to the original data. Because of the inherent discrepancy between the tails of the histogram and the tails of the probability model, the conversion of capability indexes that are larger than 1.00 into fractions nonconforming will tell you more about the assumed model than it will tell you about either the data or the underlying process. This is why such conversions are complete nonsense.

#### SUMMARY FOR MODEL-BASED ESTIMATES

Thus, there are two problems with using a probability model to estimate the process fraction nonconforming. The first is that any choice of a probability model is essentially arbitrary, and the second is that the use of a probability model encourages you to extrapolate beyond the tails of the histogram to compute imaginary quantities.

Given the uncertainty attached to any estimate of the process fraction nonconforming, the choice of a probability model will usually make no real difference as long as the capability indexes are less than 1.00. Here the use of a generic normal distribution will provide reasonable ballpark values, and there is little reason to use any other probability model.

However, when the specifications fall beyond the tails of the histogram, and especially when the capability indexes exceed 1.10, no probability model will provide credible estimates of the process fraction nonconforming. Computing an infinitesimal area under the extreme tails of an assumed probability model is an exercise that simply has no contact with reality.

#### SO WHAT DO WE REALLY KNOW?

What we know depends upon how many data we have and whether or not those values were collected while the process was operated predictably. Moreover, the only way to determine if the data were obtained while the process was operated predictably is by using a process behavior chart with rational sampling and rational subgrouping.

If the data show evidence that the process was changing while the data were collected, then the process fraction nonconforming may well have also changed, making any attempt at estimation moot. (In the absence of a reasonable degree of predictability, all estimation is futile.)

If the data show no evidence of unpredictable operation, then we may use the binomial point estimate to characterize the process fraction nonconforming. In addition we may also use the 95% Agresti-Coull interval estimate to characterize the uncertainty in our point estimate. This approach is quick, easy, robust, and assumption free.

When no observed values fall beyond a specification limit our count  $Y$  becomes zero and the binomial point estimate goes to zero. However, the 95% Agresti-Coull interval estimate will still provide an upper bound on the process fraction nonconforming. These upper bounds will depend solely upon the number of observations in the histogram,  $n$ . Selected values are shown in Figure 7.

$n$ for Histogram	95% Upper Bound	$n$ for Histogram	95% Upper Bound	$n$ for Histogram	95% Upper Bound
30	13.8%	90	5.0%	500	0.9%
40	10.7%	100	4.6%	1000	0.5%
50	8.7%	150	3.1%	2000	0.24%
60	7.4%	200	2.3%	3000	0.16%
70	6.4%	300	1.6%	4000	0.12%
80	5.6%	400	1.2%	5000	0.10%

Figure 7: The Uncertainty in Estimates of Process Fraction Nonconforming when  $Y = 0$

The upper bounds listed in Figure 7 define the essential uncertainty in ALL estimates of the fraction nonconforming that correspond to a count of  $Y = 0$  nonconforming. This means that when you use a probability model to compute a tail area that is beyond the maximum or the minimum of your histogram, then regardless of the size of your computed tail area, the process fraction nonconforming can be anything up to the upper bound listed in Figure 7. There we see that it takes over 1000 data to get beyond the parts per hundred level of uncertainty.

Say, for example, that you have a histogram of 70 data collected while the process was operated predictably and on-target with an estimated capability ratio of 1.33. Say that these data are suitably bell-shaped and that you use a normal distribution to estimate the process fraction nonconforming to be 64 parts per million. Figure 7 tells us that with only 70 data all you really know about the process fraction nonconforming is that it is probably less than 6.4% nonconforming. This is 1000 times greater than the computed value of 64 ppm! With this amount of uncertainty how dogmatic should you be in asserting that the process fraction nonconforming is 64 parts per million?

Until you have hundreds of thousands of data collected while the process is operated predictably you simply do not have any basis for claiming that you can estimate the fraction nonconforming to the parts per million level.

One day a client from a telephone company observed that they computed the number of dropped calls each day in parts per million. However, in this case they were using the empirical approach and the denominator was in the tens of millions. With this amount of data the uncertainty in the computed ratio was less than 0.5 ppm, and reporting this number to the parts per million level was appropriate.

But for the rest of you, those who have been using a few dozen data, or even a few hundred data, as the basis for computing parts per million nonconforming levels, I have to tell you that the

numbers you have been using are no more substantial than a mirage. The uncertainties in such numbers are hundreds or thousands of times larger than the numbers themselves.

The first lesson in statistics is that all statistics vary. Until you understand this variation you will not know the limitations of your computed values. Using parts per million numbers based on probability models fitted to histograms of a few hundred data without regard to their uncertainties is a violation of this first lesson of statistics. It is a sign of a lack of understanding regarding statistical computations. Hopefully, now you know how to estimate the fraction nonconforming and how to avoid some of the snake oil that is out there.

