

The Parts Per Million Problem

More on the difference between a number and a computation

Donald J. Wheeler

Parts per million (ppm) is part of the language of six-sigma. It pervades the sales pitch and is used in all sorts of computations as a measure of quality. Yet what are the rules of arithmetic and statistics that govern the computation and usage of parts per million? To discover the answers read on.

REPORTING QUALITY LEVELS

The emphasis upon parts per million has generated some confusion in practice. I remember one factory visit where they told me they were “doing six-sigma.” As we went around the plant each department had a large sign hanging from the rafters that showed the department output during the past month along with the parts per million nonconforming level achieved. I did not have to actually ask how the six-sigma program was working—the values spoke for themselves. Every parts-per-million value in the plant was greater than 30,000! Say a department produced 7555 parts and had 368 nonconforming. While they reported this as 48,709 parts per million nonconforming, at the end of the month it was still 4.9 percent bad product! When your nonconforming product is at the parts per hundred level, nothing is gained by adding four extra digits and reporting parts per million nonconforming.

COMMUNICATING NUMERICAL QUANTITIES

We are used to parts per hundred and even parts per thousand. These numbers require no mental arithmetic to make sense. We can easily and naturally handle three digit numbers without pause. Whenever our measurement increment is small enough that a one-unit change (from n to $n+1$) is in the neighborhood of one part per hundred or one part per thousand of the total value, we will generally be comfortable that we know the essence the measurement. We report our weights in pounds which for adults results in numbers in the hundreds. Yet for infants we record weights in pounds and ounces. Since there are 112 ounces in seven pounds, this once again gives us knowledge to the nearest part per hundred level. Once we know a quantity to within one part per hundred or one part per thousand, using smaller measurement increments simply complicates the job of interpreting the measurement. So if you have a defect level of a few dozen parts per million, then ppm values are appropriate. But as soon as you get beyond 1000 ppm it is better to shift back to parts per thousand than to continue to use ppm increments.

Another place we see the importance of using two or three digit numbers to communicate is in financial reports where numbers are rounded and expressed as multiples of thousands or millions of dollars rather than being reported to the nearest dollar. We are comfortable with

numbers between 1 and 1000, and keeping the values within this range makes them easier to understand, easier to communicate, and easier to work with. Using values with five or six digits just increases the glaze factor for your audience.

SOMETHING FORGOTTEN

Sometime around the Fourth Grade you were taught the rule for the decimal expression of a fraction. There is a limit on how many digits are appropriate. This limit depends upon the denominator of the fraction. To determine the correct number of decimals to use for a fraction, you increase the denominator to the next power of ten, and use the number of zeroes in that power of ten as the number of decimals. (Except for denominators that are less than 10 where we use two decimal places as the default.) For the fraction used earlier of 368 out of 7555, the appropriate decimal expression for this fraction would be 0.0487, rather than the 0.04870946 found on your calculator. Since the denominator of 7555 is less than 10,000, four decimal places will be appropriate when converting 368/7555 to a decimal, resulting in 0.0487. This is saying that 368 out of 7555 is equivalent to 487 out of 10,000.

This rule helps you to avoid interpreting a number as being more precise than it actually is. (Note that 369/7555 is 0.0488, so a change of 1 in the numerator results in a change of about 1 part in 10,000. Using four decimal places avoids overstating the precision in this value.) While calculators and computers may well use the nine digit expression 0.04870946 to minimize round-off error in a string of computations, it is the four digit decimal expression you should use in communicating and interpreting the ratio of 368 to 7555. The nine digit 0.04870946 value is a *computation*. The 0.0487 value is a *number* that you and your audience can use, work with, and comprehend.

ESTIMATING THE FRACTION NONCONFORMING

When I began teaching statistics 45 years ago there were two approaches to the estimation of the fraction nonconforming: you could count the number nonconforming and divide by the number examined, or you could use the estimated mean and standard deviation for the process to obtain z-scores for the specification limits and use the standard normal distribution. Today it might seem like the number of options has increased, but all the new versions are variations on the second approach above. Before we discuss these modern options, we need to first look at the old manual approaches simply because they will illuminate some issues that get obscured by the software.

To have an example let us use the data collected by Adolphe Quetelet (in the mid Nineteenth Century) on the chest sizes for 5738 Scottish soldiers. These data are shown and listed in Figure 1 along with their histogram. This histogram is just about as good a bell-shaped curve as you could hope to find (given that the measurement increment is about one-half the size of the standard deviation).

To illustrate the different approaches to estimating the fraction of nonconforming product let us consider what would happen if the Scottish army only stocked shirts with chest sizes from 35 to 45 inches. What percentage of their soldiers would have uniform shirts that did not fit properly?

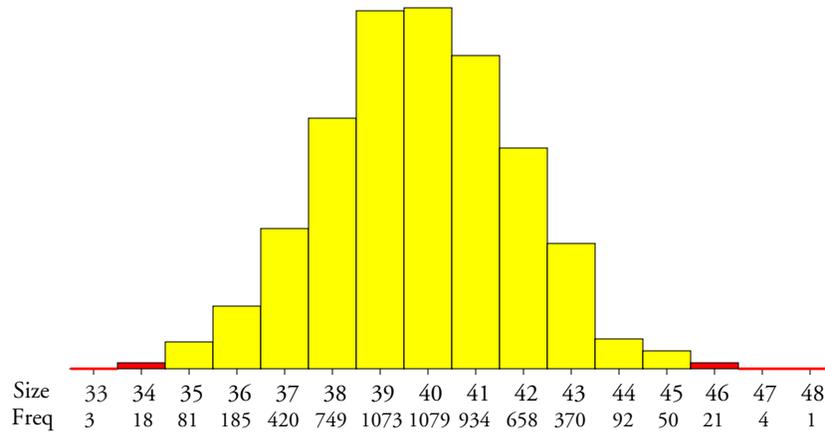


Figure 1: Chest Sizes for 5738 Scottish Soldiers

Binomial Point Estimate Approach: From the data in Figure 1 we see that 21 soldiers would not have shirt sizes small enough, and 26 soldiers would not have shirt sizes large enough. This results in the ratio of $47/5738 = 0.0082$, or about 8 per thousand. The 95% Agresti-Coull interval estimate for this proportion would be 0.0061 to 0.0110, or about 6 to 11 per thousand. Notice that this approach does not impose any assumed distribution upon the data. It simply lets the data speak for themselves. (And by using the Agresti-Coull interval estimate we can work with counts all the way down to zero, unlike the more common Wald interval estimate.)

Normal Probability Model Approach: Using the normal probability model we start by using the average and standard deviation statistics to estimate the mean and standard deviation parameters. These data have an average of 39.83 inches and a standard deviation statistic of 2.05 inches. Using the continuity correction factor our cut-offs are 34.5 inches and 45.5 inches, which result in z-scores of -2.60 and 2.77 . Thus, from tables of the standard normal distribution we would estimate that $0.0047 + 0.0028 = 0.0075$ or about 7.5 soldiers per thousand would not have a shirt that fit. So we get essentially the same answer with either approach.

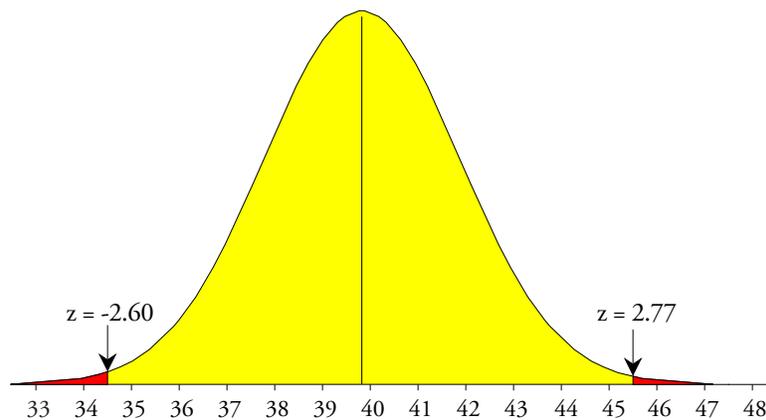


Figure 2: Normal Model for Chest Sizes for 5738 Scottish Soldiers

The binomial point estimate approach always yields an *unbiased* estimate of the fraction

nonconforming. The simple use of the normal distribution illustrated above, or the similar use of other probability models, will always yield *biased* estimates of the fraction nonconforming. (As shown in the literature over 50 years ago, with normally distributed data, the *unbiased* point estimate of the fraction nonconforming will require the use of a symmetrical beta distribution. For citations regarding this see my paper "The Variance of an Estimator in Variables Sampling," *Technometrics*, v.12, 1970, pp. 751-755.)

Model-based estimates tend to show slightly less uncertainty than the binomial point estimate approach. Our 95% interval estimate for the mean value is 39.77 to 39.88. Our 95% interval estimate for the standard deviation parameter is 2.02 to 2.08. As a result, the upper cut-off has a z-score somewhere between 2.70 to 2.84, while the lower cut-off has a z-score somewhere between -2.53 to -2.66. Using these values, our estimate on the low-end tail area under the curve varies from 0.0057 to 0.0039, while our estimates of the high-end tail area under the curve varies from 0.0035 to 0.0023. Thus, based on the normal probability model, our estimate of the number of soldiers who will not have a shirt that fits will vary from 0.0062 to 0.0092, or 6 to 9 per thousand. This tighter interval estimate is a consequence of using the assumed probability model to limit the possibilities. Of course, this tighter result is only as good as the assumed normal model. A different model might result in different estimates.

In the past we used the normal distribution simply because it is the distribution of maximum entropy. The middle 90 percent of a normal distribution is spread out to the maximum extent possible, making the outer ten percent as far, or further, away from the mean than the outer ten percent of *any* other probability model. So, when we use a normal model we end up with reasonable, worst-case values for the fraction nonconforming.

Today this model-based approach has been "updated" by having the computer fit any one of several different probability models to the data prior to computing an estimate for the fraction nonconforming. Nevertheless, the overall approach remains the same. In this case the only reasonable model for Figure 1 is a normal distribution, so we will not attempt to use any other model. Whenever the fraction nonconforming is in the parts per hundred or parts per thousand range, the use of an appropriate probability model to estimate the fraction nonconforming is a reasonable approach. We will generally get estimates that are approximately right. Of course, when operating at this level of nonconformity, the estimates based on probability models will turn out to be essentially the same as those from the binomial point estimate approach, but then some people do prefer complexity to clarity. For more on this see my column from June 2011, "Estimating the Fraction Nonconforming."

PARTS PER MILLION

If the Scottish army stocked shirts in sizes ranging from 33 to 48 inches, what fraction of the soldiers would not have a shirt that fits?

The binomial point estimate would be zero to four decimal places, 0.0000. (See if you can figure out why.) The corresponding 95% Agresti-Coull interval estimate would range from 0.0000 to 0.0008. What does this interval mean in practice? It would mean that less than one soldier in a thousand would be without a shirt that fit. This computation is based on the data, it is consistent with the data, it makes no assumptions regarding the data, and it is quick and easy to compute from the original data.

The normal distribution approach would result in the z-scores of -3.58 and 4.23 . These result in a point estimate of 186 ppm. Using the interval estimates for location and dispersion as before, we estimate that this fraction might range from 137 parts per million to 254 parts per million. However, these exceedingly precise values are all built on the assumption that the infinite tails of the normal probability model correctly describe the data *out in the region where there are no data*. So is this assumption likely to be true?

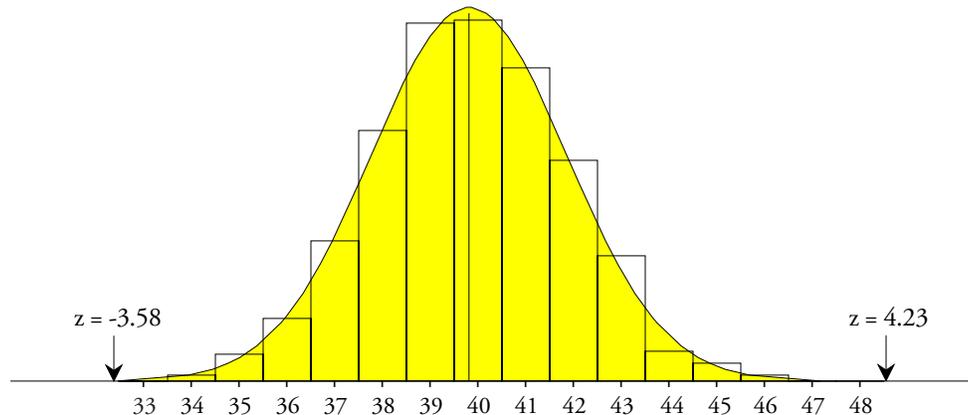


Figure 3: Normal Model for Chest Sizes for 5738 Scottish Soldiers

The first problem with these parts per million computations is that the simple, old-fashioned, chi-square lack-of-fit test for the normal distribution in Figure 3 has a lack-of-fit statistic of 47.44. The reference chi-square distribution with 11 degrees of freedom has a 99th percentile of 24.73. Thus, at the 1% level, there is a detectable lack of fit between our data and the normal probability model *in the region where there are data*.

If we have a detectable lack of fit in the region where we have data, how can we justify an extrapolation to the region where we have no data?

The second problem with these parts per million values is that, with 5738 data, we simply do not know anything about parts per million events. With 5738 data the values above are only known to parts per ten-thousand. So, if we had no lack of fit problem, the values of 137 ppm to 254 ppm found using the normal distribution could only be interpreted as meaning one or two parts per ten-thousand, even though the computations contain more decimal places.

A GRADUATE LEVEL SECRET

No histogram has infinite tails.

Every data set has both a minimum and a maximum.

Therefore, when you fit a probability model to your data and that model has an infinite tail, there will always be some point where the histogram and the model part company. Even when the model passes every lack-of-fit test your software can throw at it, any model with an infinite tail will always cease to track the data at some point. For more on this topic see my column for June, 2013, "Why We Keep Having 100 Year Floods."

Because of this unavoidable discrepancy between models and data, whenever you compute *infinitesimal areas* under the *extreme tails* of an *assumed probability model* out in the *region where you*

have no data you will end up with values that have no contact with reality. When you attempt to use such infinitesimal values you are dealing with computations rather than numbers. The values you compute will be *artifacts* of the fitted model rather than being *characteristics* of the underlying process. While fiction can be entertaining, it should never be confused with reality.

When your data set contains dozens or hundreds of values, your knowledge is limited to parts per hundred or parts per thousand values. While the computations can have more digits, those extra digits have no meaning in practice, no matter what model was used to find them.

When your data set contains thousands of values, your knowledge is limited to parts per ten-thousand (four decimal places). Experience shows that when your data set gets this large you will begin to find, almost every time, a detectable lack of fit regardless of the probability model you choose. (Try fitting some other probability model to the data of Figure 1 and see what happens. A location and scale shifted chi-square distribution with 278 degrees of freedom has a detectable lack of fit even though it matches the first four moments quite nicely.) With enough data you will always find a detectable lack-of-fit.

This lack-of-fit phenomenon is due to the underlying mathematics. All probability models are *limiting* characteristics for an *infinite* sequence of *independent and identically distributed* random variables. In consequence, a probability model *cannot* be said to be a characteristic of any *finite portion* of that infinite sequence. In the real world, your *data* are never distributed according to any probability model.

A small data set might be considered to be a short finite sequence of values, and as such it could be part of many different infinite sequences having many different limiting probability models. This is why you seldom find a detectable lack-of-fit with small data sets. Larger data sets would comprise longer finite sequences, and these will fit in with fewer infinite sequences, and therefore will have fewer possible limiting distributions. This is the region where you will start to find a detectable lack of fit. Eventually, as we get more and more data in our sequence of observations, it will become less and less likely that our observations will all have come from one and the same unchanging system. As the assumption of *identically distributed* random variables breaks down we will find that *no single model* will ever fit our data. This is why, when your data set gets large enough, you will always find a detectable lack of fit.

So when can you actually use parts per million values? It is only when you have hundreds of thousands of values that you actually have knowledge at the parts per million level. For example, a student from a phone company observed that they kept track of the number of dropped calls each day. With a denominator in the millions of calls, they truly had parts per million *numbers*. (Notice also that they were using the unbiased binomial point estimate.)

COMPUTATIONS VERSUS NUMBERS

Using probability models for parts-per-hundred and parts-per-thousand computations is reasonable. This has been done for well over two hundred years. However, using a few thousand data to fit a probability model and then using that probability model to compute ultra-precise parts-per-million estimates of the fraction nonconforming is simply a triumph of computation over common sense. (Doing this while using hundreds of data is even more absurd.) Thus, our computers have allowed us to take a standard technique and apply it in an inappropriate way. While probability models can be tortured to produce parts-per-million

computations, believing that these computations are meaningful *numbers* is nothing less than an outright hallucination.

The truth is given by the binomial point estimate and the Agresti-Coull 95% interval estimate. At the parts-per-hundred and parts-per-thousand levels, estimates based on probability models will generally be in reasonable agreement the binomial point estimate.

When we compute model-based estimates at the parts-per-million level we will be extrapolating into the region where we have no data. Such computations will be artifacts of the probability model used rather than being a characteristic of the process producing the data.

So when does a *computation* become a *number*? Only when it makes sense in context. You learned the difference between computations and numbers that made sense in the Fourth Grade. Are you as smart today as you were then?

