White Paper

Process Capability Metrics

Determining Confidence Intervals for the Z-score

SIX SIGMA QUALTEC

Process Capability Metrics

… and how to get them to start

A White paper by Six Sigma Qualtec

“We’re two Sigma.” “We’re three Sigma.” “We’re negative Sigma.”

Good or bad, I always see a look of satisfaction in the face of a new Black Belt the first time that they calculate the “Sigma level” for their process. As the namesake for the Six Sigma initiative, new practitioners are always drawn to the Sigma level metric when studying capability analysis for the first time. Prior to embarking on a capability study, students of Six Sigma should also be taught the importance of validating the integrity, relevance, accuracy and precision of the metrics and measurement system that they are using to draw conclusions and make recommendations. So it is only fitting that the validity of the Sigma-level metric be understood as well. To accomplish this we will determine the confidence interval for the Sigma level as an assessment of its precision. As we will see, a significant level of uncertainty can exist in your Sigma-level metric depending on how you carry out your capability analysis.

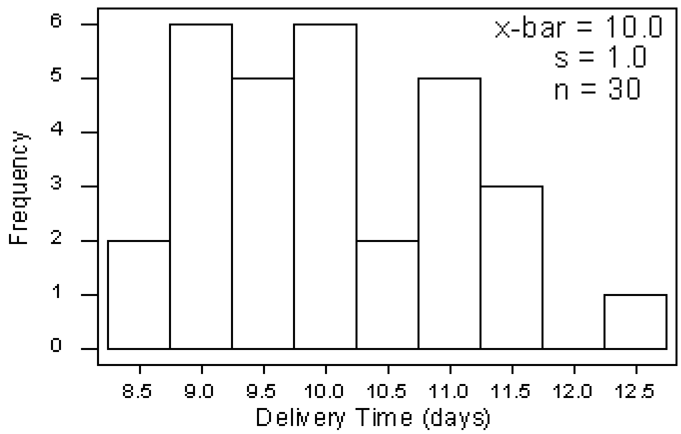
Students of statistics are all familiar with the concept of the confidence interval for the sample mean and standard deviation.[1] Similarly, the confidence interval concept can be applied to other statistics such as the Z-score. The Z-score, often referred to as the “Sigma Level,” is a measure of process or product capability commonly linked with Six Sigma process improvement projects.[2] Given the importance of this statistic, it is critical for Black Belts, process owners and management to understand the uncertainty associated with the determination of a Z-score.

The following example illustrates the derivation of a confidence interval for the Z-score. The derivation is based on the confidence intervals for the mean and standard deviation, which are the two parameters found in the Z-score formula. (See Equation 1.) The approach also makes use of Monte Carlo simulation, which will be described in more detail later. (The example involves continuous time data. A similar outcome results from the analysis of Z-scores derived from attribute data. However, that analysis is beyond the scope of this paper.)

Example

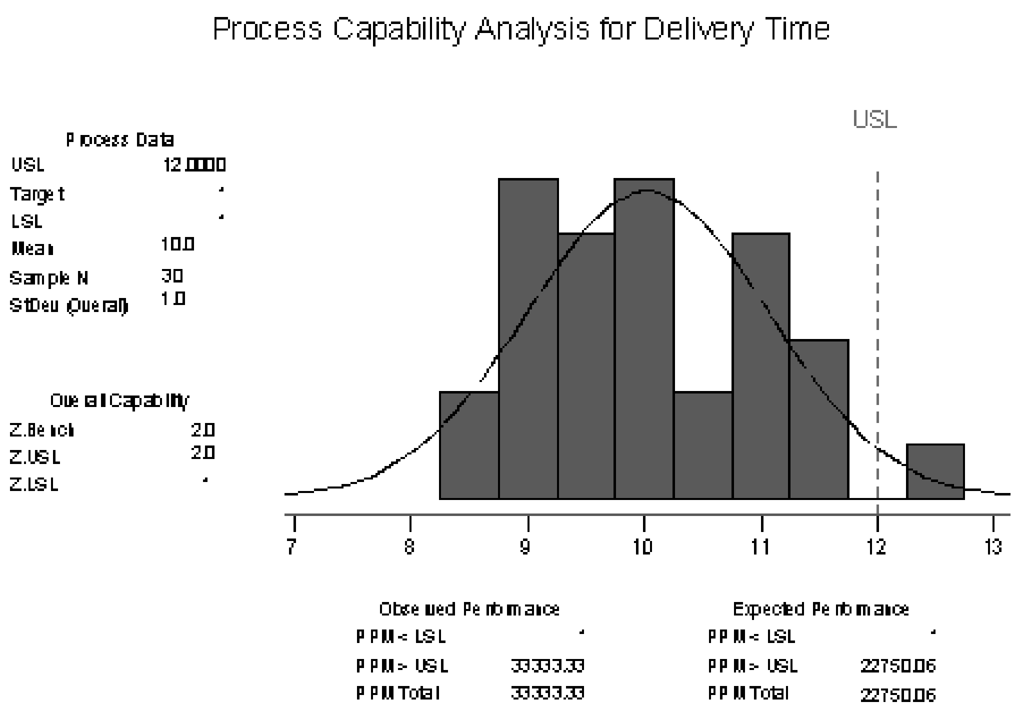
Management at ABC Corporation is interested in determining the capability of their order fulfillment process vs. customer requirements. The key customer metric is shipment time – the time that it takes an order to be received by a customer after the order has been placed. Shipment time is of particular interest because a Black Belt team completed a project in this area six months ago and claimed significant improvement (Prior to the project the Sigma level was 1.5). Consequently, an audit team completed a capability study to see if the claimed improvements have been sustained. ABC Corporation is a low volume producer and only ships five orders per month. Therefore, delivery times for thirty shipments are available since the Black Belt project was completed six months ago. In the interest of time a decision was made to base the audit on this sample of 30 shipments. Based on customer feedback, delivery times of less than 12 days are acceptable. Results are shown next page in Figure 1:

Figure 1
Delivery Time Capability Study Histogram



Based on this sample, average delivery time was 10.0 days with a standard deviation of 1.0 day. Using this sample and a customer stated upper specification limit (USL) of 12 days capability analysis was completed and is shown in Figure 2:

Figure 2
Process Capability Analysis for Delivery Time



Based on this analysis Zusl = 2.0 which corresponds to 22750.06 ppm. (For simplicity sake, let us ignore any discussion about the "1.5 Sigma shift" [2]. Also note that for this example unbiasing constants were not used for the estimate of the process standard deviation and a subgroup size of 1 was used.[3])

At this point most audit teams would stop here and report that the delivery process is "two Sigma" and the Black Belt team would take their reward or punishment depending on how this "number" was judged. Recalling that the Z-score before the Black Belt project was 1.5 it does appear that an improvement was made. But is a Z-score of 1.5 significantly different than 2.0, statistically speaking?

We must remember that the formula for Zusl (Equation 1) contains and s, which are only estimates of the population mean and standard deviation. We must think about and s relative to their confidence intervals, which are calculated using Equations 2 and 3.[1]

Equation 1

$$Z_{USL} = \frac{USL - \bar{x}}{s}$$

Equation 2

$$\bar{x} \pm t_{\nu,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Equation 3

$$s\sqrt{\frac{n-1}{\chi^2_{\nu,1-\frac{\alpha}{2}}}} \le \sigma \le s\sqrt{\frac{n-1}{\chi^2_{\nu,\frac{\alpha}{2}}}}$$

What if the true population mean is closer to the USL than ? The Z-score will be lower reflecting poorer capability than reported. On the other hand, if the population standard deviation is close to the sample's lower confidence bound, the Z-score will be higher than reported.

To understand just how much higher or lower the true process capability might be (i.e., how much uncertainty exists in the calculation of the Z-score) let us determine a confidence interval for the Z-score based on the confidence intervals for and s using a Monte Carlo simulation. The Monte Carlo approach is a numerical method that utilizes a sequence of random numbers to perform the simulation. [4] Random numbers are selected in a series of trials from probability distribution functions (PDFs) that are assumed to represent the underlying behavior of the system. For example, many manufacturing process variables can be described using a normal (or Gaussian) PDF.

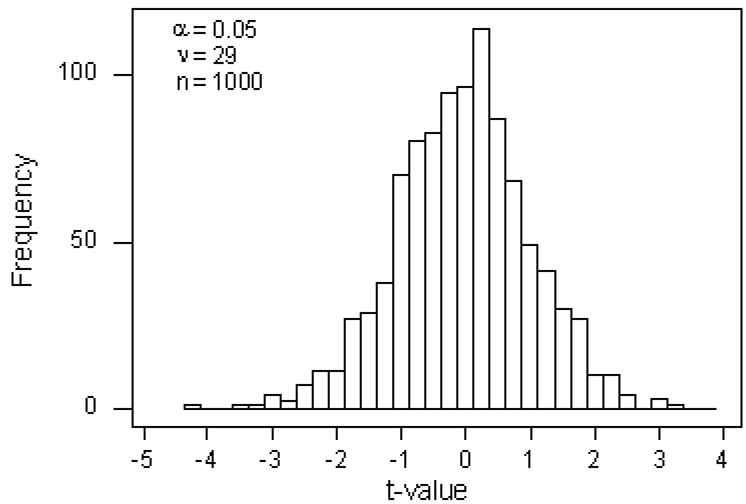
Returning to our example, we begin by redefining ZUSL by combining equations 1-3 to give Equation 4.

Equation 4

$$Z_{USL} = \frac{USL - \left(\bar{x} + t_{\nu}\frac{s}{\sqrt{n}}\right)}{s\sqrt{\frac{n-1}{\chi^2_{\nu}}}}$$

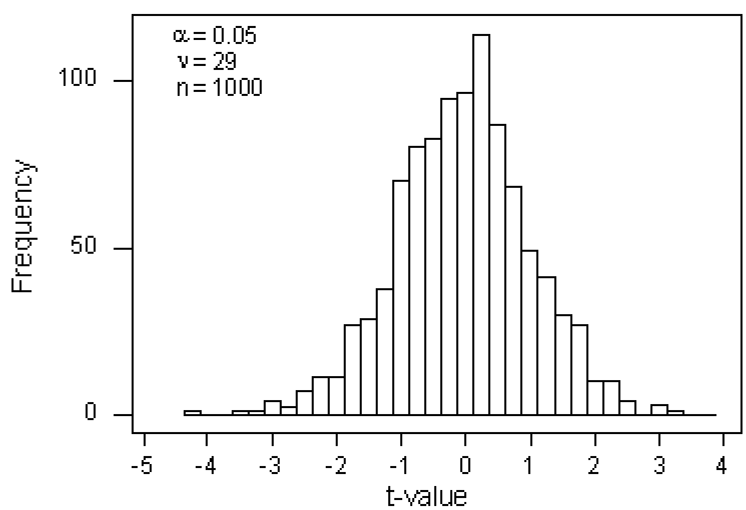
In equation 4, tn is a random observation selected from a t-distribution with n degrees of freedom. ?2 is a random observation selected from a Chi-square distribution with n degrees of freedom. For the Monte Carlo simulation values of t and ?2 are randomly chosen and Z-scores are calculated for each of these trials. A distribution of possible Z-scores results. The resulting Z-scores can be transformed into defect rates in parts per million (ppm). In practice we begin by creating a random sample of 1000 observations from a t-distribution with 29 degrees of freedom (n=30 for this example). The resulting histogram is shown in Figure 3.

Figure 3
t-distribution for 29 degrees of freedom and 1000 observations



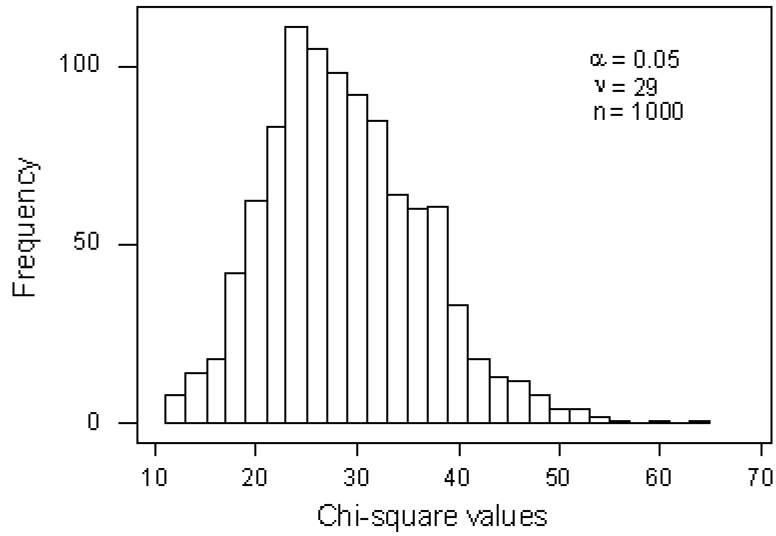
Each observation (an individual t-value) in this distribution is then inserted into the Equation 2 - the confidence interval for the mean. Each result is a member of the distribution of possible population means based on , the standard deviation of the sample and the number of observations (n=30) in the capability study. The resulting histogram is shown in Figure 4.

Figure 4
Distribution of delivery time averages



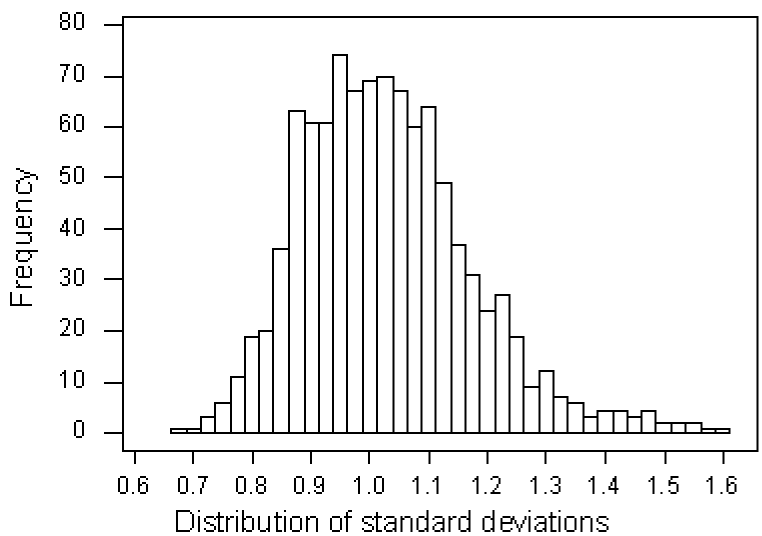
Likewise, we can create a distribution of standard deviations by first generating a random sample of 1000 observations from a Chi-square distribution with 29 degrees of freedom (n=30 for this example). See Figure 5 on the next page.

Figure 5
Chi-square distribution for 29 degrees of freedom and 1000 observations



Each observation (an individual Chi-square value) in this distribution is then inserted into the Equation 3 - the confidence interval for the standard deviation. Each result is a member of the distribution of possible population standard deviations based on the standard deviation of the sample and the number of observations (n=30) in the capability study. The resulting histogram is shown in Figure 6.

Figure 6
Distribution of delivery time standard deviations



We now have 1000 pairs of mean and standard deviation observations. Z-scores for each of these pairs can be calculated using Equation 1 (USL=12). The resulting histogram of potential Z-scores and defect levels in parts per million (ppm) are shown below in Figures 7 and 8. Defect levels in parts per millions are the proportion of observations times 1X106 that exceed the USL based on a given Z-score.[3]

Figure 7
Possible Z-scores for example

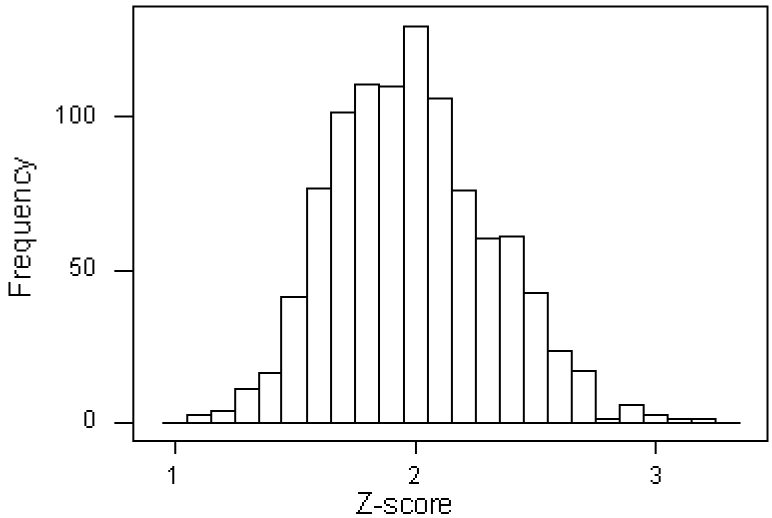
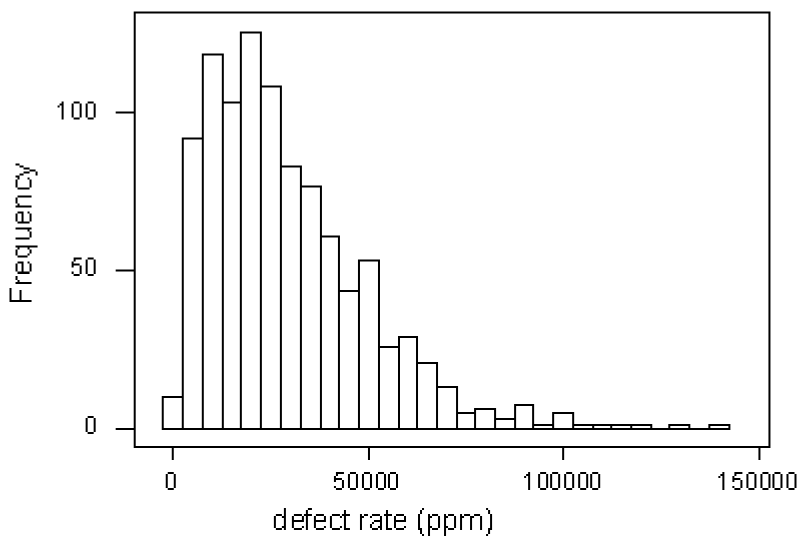
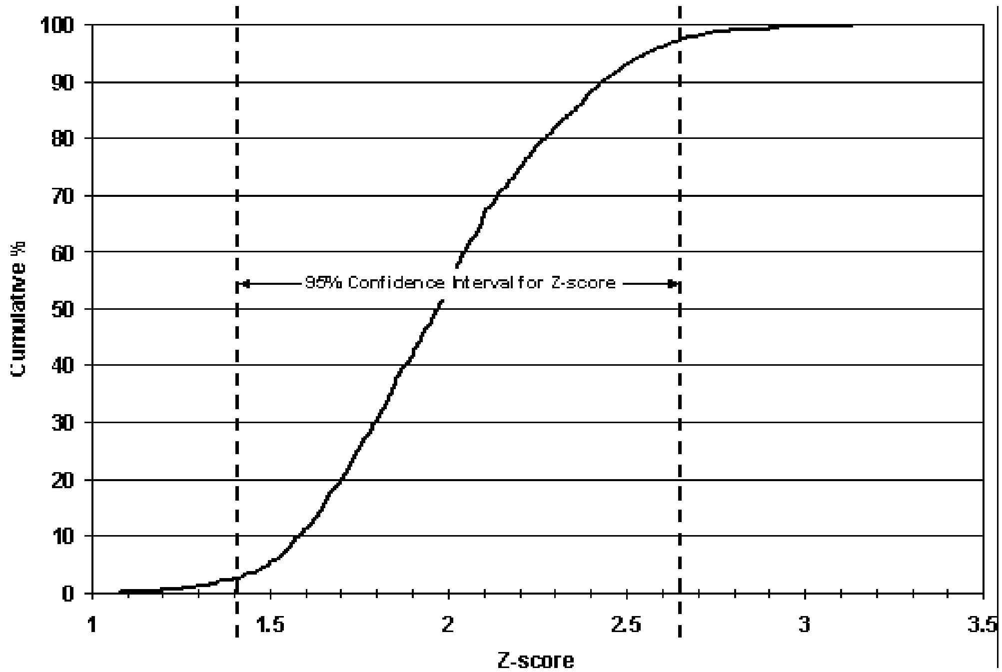


Figure 8
Defect Rate (ppm) based on Z-score



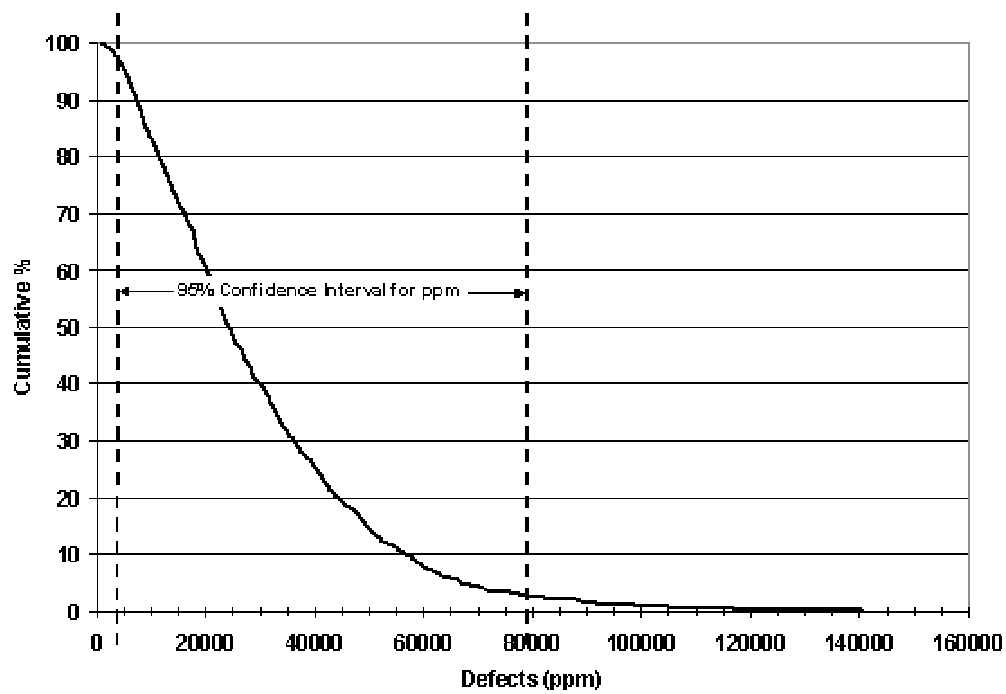
We then plot the cumulative distribution function (CDF) of the resulting Z-scores and identify the Z-scores where the CDF equals 2.5% and 97.5% - see Figure 9. This essentially identifies the "95% Confidence Interval" for the Z-score.

Figure 9
95% Confidence Interval for Z-score



Based on this 95% confidence interval the true Z-score for the process could be as high as 2.653 or as low as 1.412. These Z-values correspond to defect levels of 3989 ppm (Z=2.653) and 78975 ppm (Z=1.412). Recalling that we calculated a Z-score of 2.0 (22750 ppm) we see that we may be overestimating the capability by 247% or underestimating the capability by 82%. Remembering that the process capability before the Black Belt project had a Z-score of 1.5 we cannot say that the improvement had a statistically significant impact since a Z-score of 1.5 is greater than the lower confidence limit for the Z-score of 1.412. Figure 10 illustrates the 95% confidence interval for the defect rate in ppm.

Figure 10
95% Confidence Interval for Z-score



Clearly, when a capability analysis is based on only 30 observations the uncertainty in the results can be quite large. Depending on the frequency with which your processes generate data and the ease with which process measurements can be collected 30 observations may be all that you can afford to collect for a capability study. On the other hand, if sample size is not an issue, how many samples should be collected to minimize the uncertainty in a capability study?

To answer this question, we follow the procedure described above in which we create random t and Chi-square distributions based on the sample size of interest and then calculate potential Z-scores based on Equation 4. Results of this procedure are illustrated in Figures 11-13. Figures 11 and 12 illustrate the upper and lower confidence limits for the Z-score and defect rate for different sample sizes. Figure 13 illustrates the width of the confidence interval. The width of the confidence interval is calculated by subtracting the lower confidence level from the upper confidence level. Table 1 lists 95% confidence interval values for different sample sizes.

It is worth noting that these results are consistent to those of Bissell [5] although the approaches are different. Bissell calculates confidence intervals for Cpk. Defining Z=3Cpk the results presented here match those in Reference [5] for large sample sizes. For small sample sizes the confidence level widths presented here are wider. The differences result from the use of the t-distribution PDF here vs. the Z-distribution PDF in [5].

Figure 11
95% Confidence limits for Z-score vs. sample size

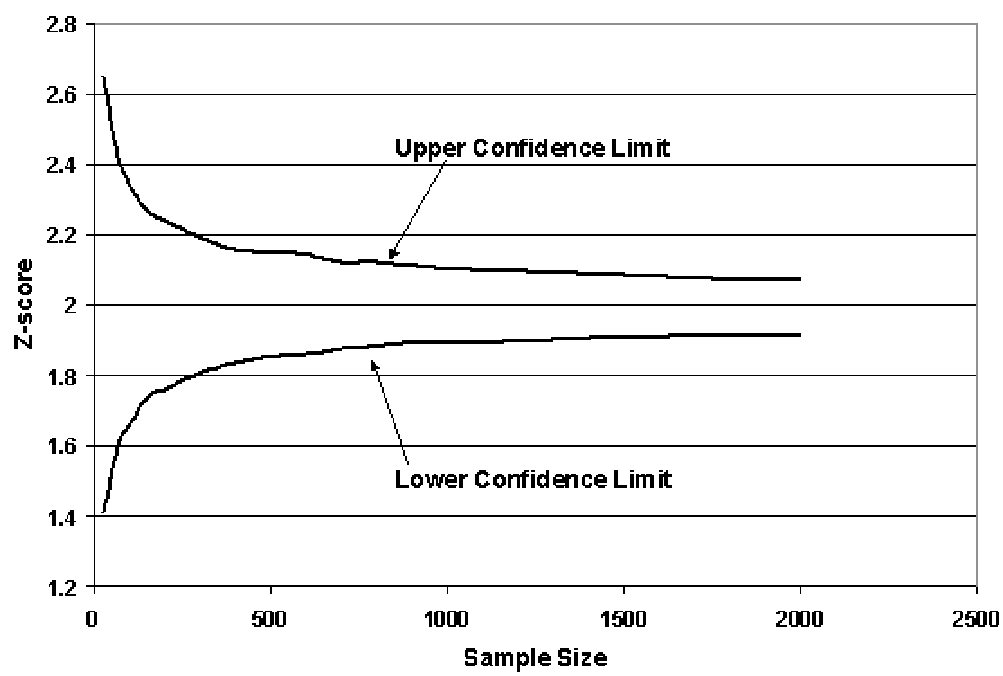


Figure 12
95% Confidence limits for defect rate vs. sample size

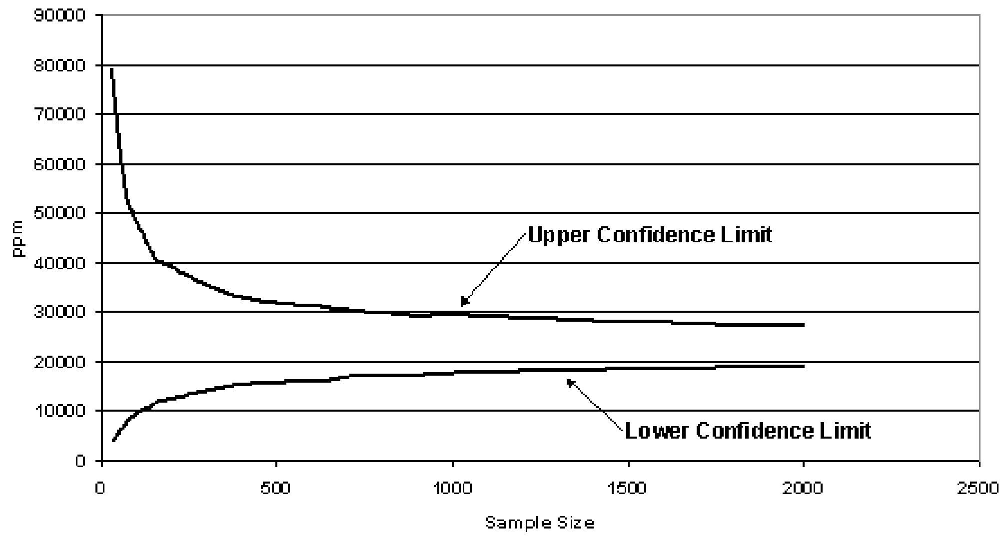


Figure 13
Confidence interval width vs. Sample Size

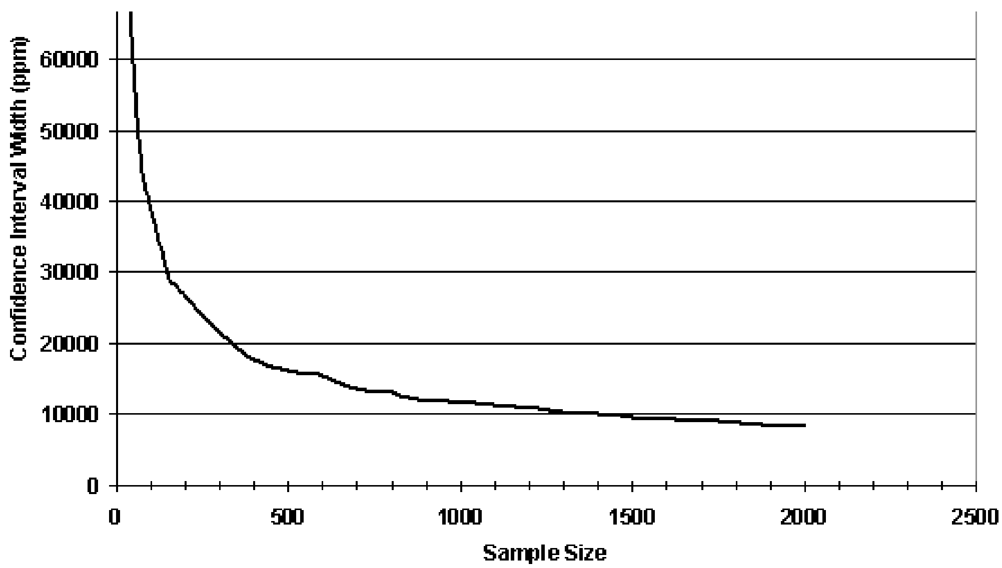


Table 1
95% Confidence levels for Z and ppm vs. sample size

Sample Size	Z (lower)	Z (upper)	ppm (lower)	ppm (upper)	Confidence Interval Width (ppm)
30	1.41	2.65	78974	3991	74983
50	1.51	2.53	65695	5665	60031
60	1.56	2.47	59691	6704	52988
75	1.62	2.41	52937	8011	44926
100	1.66	2.35	48681	9406	39275
150	1.74	2.27	41271	11485	29786
200	1.76	2.24	39013	12457	26556
300	1.81	2.20	35469	13989	21481
400	1.84	2.16	32986	15370	17616
500	1.86	2.15	31774	15750	16024
600	1.86	2.15	31276	15915	15361
700	1.88	2.12	30318	16871	13447
800	1.88	2.12	29808	16884	12924
900	1.89	2.12	29053	17189	11864
1000	1.89	2.11	29250	17619	11631
1500	1.91	2.09	27993	18404	9589
2000	1.92	2.07	27324	19043	8280

From Figure 13 we observe that for the width of the confidence interval to narrow to less than 10000 ppm (1%) a sample size of approximately 1300 is required. For many processes a sample size of 1300 may simply be too expensive or take to long to collect to be practical. So what is a Black Belt to do? Recognizing that capability metrics like the Z-score are not exact is an important first step. Quantifying the uncertainty and defining acceptable levels of uncertainty are the next.

Conclusion

Through the use of standard definitions of the confidence intervals for the mean and standard deviation coupled with Monte Carlo simulation we can determine confidence intervals for the Z-score. The analysis highlights that significant uncertainty may exist in our capability measurements and that caution should be exercised when reporting such data to management and making business decisions. At the same time we must remember that Six Sigma is much more than a metric. Six Sigma is a set of tools, a problem solving process and a company culture changing initiative that focuses on data based decision-making.

References

1] Anderson, D.R., Sweeney, D. J. and Williams, T. A., Statistics for Business and Economics, Seventh Edition, (Cincinnati, OH: South-Western College Publishing, 1999).[2] Pyzdek, T., The Complete Guide to Six Sigma (Tuscon, AZ: Quality Publishing, 1999).[3] Ryan, B. F. and Joiner, B. L. MINITAB Handbook (Pacific Grove, CA, Duxbury Press, 2000).[4] Morgan, B. J. T., Elements of Simulation, (New York, NY: CRC Press, 1984).[5] Bissell, A. F., “How Reliable is your Capability Index?” Applied Statistics, 30, 1990, pp. 331-340.

Six Sigma **Qualtec**

Your Strategic Partner

Six Sigma Qualtec is a premier provider of process management and performance improvement consulting, training, and technology solutions that drive breakthrough growth, productivity and value for our clients.

We are unique in our ability to customize the integration of management disciplines to meet the industry-specific requirements of global leaders in financial services, natural resources, manufacturing, process and service industries.

Six Sigma Qualtec
821 Alexander Road
Suite 130
Princeton, NJ 08540 • USA
toll free (800) 247-9871
phone (609) 925-9458
fax (609) 419-9855
email info@ssqi.com
website www.ssqi.com

Six Sigma Qualtec
1295 W. Washington Street
Suite 208
Tempe, AZ 85281 • USA
toll free (800) 247-9871
phone (480) 586-2600
fax (480) 586-2586
email info@ssqi.com
website www.ssqi.com

Six Sigma Qualtec
P.O. Box 2959
Kenilworth
CV8 1XR
United Kingdom
tel +44 (0) 1926 859555
fax +44 (0) 8701 400023
email info@ssqi.co.uk
website www.ssqi.co.uk