Please Visit **www.SixSigmaTutorial.com for the latest eBooks on Six Sigma.**

Table of Contents

Confidence Intervals

Confidence intervals are very important to Six Sigma methodology. To understand Confidence Intervals better, consider this example scenario:

Acme Nelson, a leading market research firm conducts a survey among voters in USA asking them whom would they vote if elections were to be held today. The answer was a big surprise! In addition to Democrats and Republicans, there is this surprise independent candidate, John Doe who is expected to secure 22% of the vote.

We asked Acme, how sure are you? In other words how accurate is this prediction?

Their answer: "Well, we are 95% confident that John Doe will get 22% plus or minus 2% vote"

In the statistical world, they are saying that John Doe will get a vote between 20% and 24% (also known is Confidence Range) with a probability of 95% (Confidence Level).

Definition of Confidence Interval

According to University of Glasgow Department of Statistics, Confidence Interval is defined as:

A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.

In our Acme research example

- The confidence interval is the range 20 to 24
- The confidence level is 95%
- The confidence limits are 20 (lower limit) and 24 (upper limit)
- The unknown population parameter is "What percentage of the total vote will John Doe Get"

The width of the confidence interval, in our case 24-20=4 is a measure that is directly proportional to the precision.

Consider this scenario

What if Acme Nelson's survey predicted that John Doe will get 22% plus or minus 20% of vote? In other words Acme is saying John Doe will get between 2% and 42% of the vote.

How good is this number? Even a monkey can predict that. This is a very wide confidence range and in order to reduce the Confidence Interval, Acme needs to collect more samples.

Confidence Limits

Confidence limits are the lower and upper boundaries of a confidence interval. In our Acme example, the limits were 20 and 24.

Confidence Level

The confidence level is the probability value attached to a given confidence interval. It can be expressed as a percentage (in our example it is 95%) or a number (0.95).

Confidence Interval for a Mean

A confidence interval for a mean is a range of values within which the mean (unknown population parameter) may lie.

Examples of Confidence Interval for a Mean

- A Web master who wishes to estimate her mean daily hits on a certain webpage.
- An environmental health and safety officer who wants to estimate the mean monthly spills.

Confidence Interval for the Difference between Two Means

A confidence interval for the difference between two means specifies a range of values within which the difference between the means of the two populations may lie.

Examples of Confidence Interval for the Difference between Two Means

- A Web master who wishes to estimate her difference in mean daily visitors between two websites.
- An environmental health and safety officer who wants to estimate the difference in mean monthly spills between two production sites.

Confidence Intervals in Six Sigma

When we calculate a statistic for example, a mean, a variance, a proportion, or a correlation coefficient, there is no reason to expect that such *point* estimate would be exactly equal to the true population value, even with increasing sample sizes. There are always sampling inaccuracies, or error.

In most Six Sigma projects, there are at least some descriptive statistics calculated from sample data. In truth, it cannot be said that such data are the same as the population's true mean, variance, or proportion value.

There are many situations in which it is preferable instead to express an interval in which we would expect to find the true population value. This interval is called an *interval* estimate. A *confidence interval* is an interval, calculated from the sample data that is very likely to cover the unknown mean, variance, or proportion.

For example, after a process improvement a sampling has shown that its yield has improved from 78% to 83%. But, what is the interval in which the population's yield lies? If the lower end of the interval is 78% or less, you cannot say with any statistical certainty that there has been a significant improvement to the process.

There is an *error of estimation*, or *margin of error*, or *standard error*, between the sample statistic and the population value of that statistic. The confidence interval defines that margin of error.

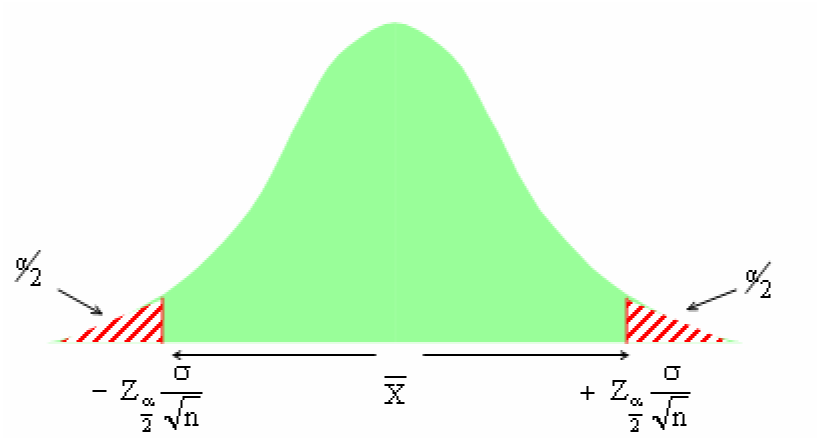
The next page shows a decision tree for selecting which formula to use for each situation. For example, if you are dealing with a sample mean and you do not know the population's true variance (standard deviation squared) or the sample size is less than 30, than you use the t Distribution confidence interval. Each of these applications will be shown in turn.

Decision Tree for selecting What Formula to use:

<u>For a Mean</u> with known variance & $n \geq 30$	Use Z Distn Confidence Interval	$C.I. = \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
with unknown variance or $n < 30$	Use t Distn Confidence Interval	$C.I. = \bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$
<u>For a Variance</u>	Use the χ^2 Distn Confidence Interval	$C.I. = \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}$
<u>For a Proportion</u>	Use Z Distn Confidence Interval if $n \times \hat{p}$ and $n \times \hat{q}$ are each ≥ 5	$C.I. = \hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Z Confidence Interval for Means

This Z Confidence Interval for Means applies to a mean from a normal distribution of variable data. Use the normal distribution for the confidence interval for a mean if the sample size n is relatively large (≥ 30), and σ is known.



The confidence interval (C.I.) includes the shaded area under the curve in between the critical values, excluding the tail areas (the α risk). The entire curve represents the most likely distribution of population means, given the sample's size, mean, and the population's standard deviation.

$$\text{C.I.} = \overline{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

where $\overline{X}$ = the sample mean

σ = the population standard deviation

$Z_{\frac{\alpha}{2}}$ = the Z value for the desired confidence level α (obtained from an Area Under the Normal Curve table)

Here we are making an assumption that the underlying data we are working with is distributed like the bell curve shown.

The most common confidence interval used in industry is probably the 95% confidence interval. If we were to use its formula on many sets of data from the population, then 95% of the intervals would contain the unknown population mean that we are trying to estimate. And 5% of the intervals would not contain the population mean. 2.5% of the time, the interval would be low, and 2.5% of the time, the interval would be too high.

The probability is 95% that the interval contains the population parameter. The 95% value is the *confidence coefficient*, or the *degree of confidence*. The end points of the interval are called the *confidence limits*. In the graphic on the previous page, the endpoints are defined by

$$\overline{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Example - Z Confidence Interval for Means

Calculate a 95% C.I. on the mean for a sample (n = 35) with an x-bar of 15.6" and a known σ of 2.3":

$$C.I. = \overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

where $\overline{X}$ = the sample mean
σ = the population standard deviation
$Z_{\frac{\alpha}{2}}$ = the Z value for the desired confidence level α

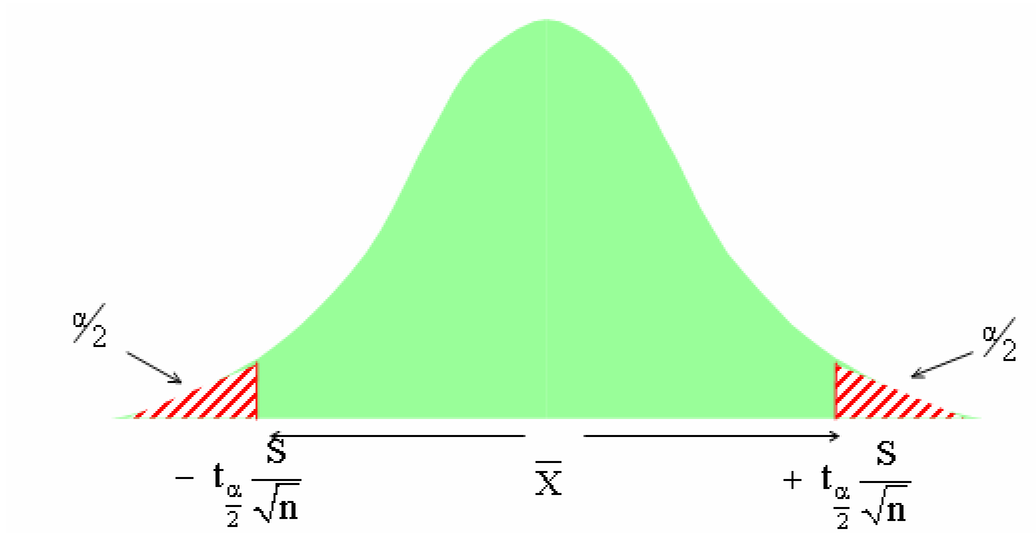
$$C.I. = 15.6 \pm 1.96 \frac{2.3}{\sqrt{35}}$$

$$= 15.6" \pm .8" = 14.8 \text{ to } 16.4"$$

This interval represents the most likely distribution of population means, given the sample's size, mean, and the population's standard deviation. 95% of the time, the population's mean will fall in this interval.

t Confidence Interval for a Mean

Use the t distribution for the confidence interval for a mean if the sample size n is relatively small (< 30), and/or σ is not known.



The confidence interval (C.I.) includes the shaded area under the curve in between the critical values, excluding the tail areas (the α risk). The entire curve represents the most likely distribution of population means, given the sample's size, mean, and standard deviation.

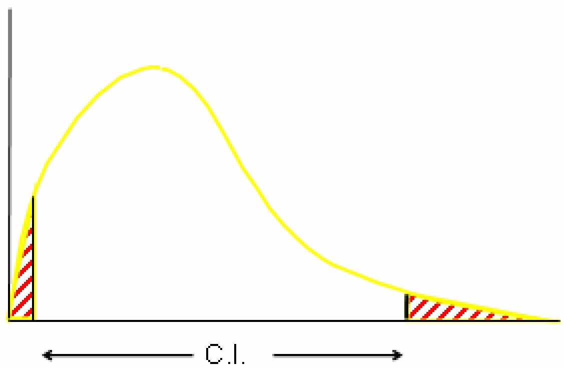
$$\text{C.I.} = \overline{X} \pm t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$

where $\overline{X}$ = the sample mean
S = the sample standard deviation
$t_{\frac{\alpha}{2}}$ = the t distribution value for the desired confidence level α

t Confidence Interval for a Variance

Use the $\chi 2$ (chi-squared) distribution for the confidence interval for the variance

The confidence interval (C.I.) includes the area under the curve in between the critical values, excluding the tail areas (the α risk). The entire curve represents the most likely distribution of population variances (sigma squared), given the sample's size and variation.



$$\text{C.I.} = \frac{(n-1)\,S^2}{\chi^2_{\alpha/2,\,n-1}} \leq \sigma^2 \leq \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,\,n-1}}$$

where:

n = the sample size

S^2 = the sample variance

$\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,\,n-1}}$ & $\frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,\,n-1}}$ = the χ^2 distribution values for the desired confidence level α and for n-1

Example - t Confidence Interval for a Variance

Calculate a 95% C.I. on variance for a sample (n = 35) with an S of 2.3”

$$C.I. = \frac{(n-1)\,S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,n-1}}$$

$$C.I. = \frac{(34)\,2.3^2}{51.966} \leq \sigma^2 \leq \frac{(34)\,2.3^2}{19.806}$$

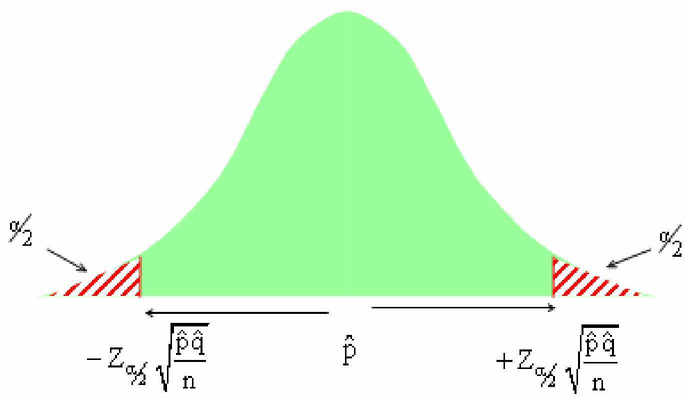
$$= 3.5 \text{ to } 9.1''$$

This interval represents the most likely distribution of population variances, given the sample’s size and variance. 95% of the time, the population’s variance will fall in this interval.

Z Confidence Interval for Proportions

This Z Confidence Interval for Proportions applies to an average proportion (which is from a binomial distribution).

Use the normal distribution for the confidence interval for a *proportion*.
If $n \times \hat{p}$ and $n \times \hat{q}$ are each ≥ 5, p is approximately normally distributed.



The confidence interval (C.I.) includes the shaded area under the curve in between the critical values, excluding the tail areas (the α risk). The entire curve represents the most likely distribution of population proportions, given the sample's size and $\hat{p}$ value.

$$\text{C.I.} = \hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where $\hat{p}$ = the sample proportion of "successful" outcomes

$\hat{q} = 1 - \hat{p}$

$Z_{\alpha/2}$ = the Z distribution value for the desired confidence level α

Example - Z Confidence Interval for Proportions

Calculate a 95% C.I. on the proportion for a sample (n = 35) with an proportion (such as fraction nonconforming) $\hat{p}$ of .65.:

$$C.I. = \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

where $\hat{p}$ = the sample proportion of "successful" outcomes

$\hat{q} = 1 - \hat{p}$

$Z_{\alpha/2}$ = the Z distribution value for the desired confidence level α

The Z distribution will work for this application since n x $\hat{p}$ and n x $\hat{q}$ are both greater than 5.

$$C.I. = .65 \pm 1.96\sqrt{\frac{.65(.35)}{35}}$$

$$= .65 \pm .08 = .57 \text{ to } .73$$

This interval represents the most likely distribution of population proportions, given the sample's size and proportion. 95% of the time, the population's true proportion will fall in this interval.

Confidence Intervals Summary

Confidence intervals are very crucial to Six Sigma. Confidence intervals provide important information as they give us a range of possible values and attach a confidence level to the interval.

Please Visit www.SixSigmaTutorial.com for the latest eBooks on Six Sigma.

Six Sigma Training & Consulting

If you like the information in this paper, and you need training, consider Qi2 for your needs. We specialize in on-site work-shops. If you have two or more people for training, you are probably at the break-even point (vs. sending them off somewhere for an off-the-shelf, public course).

Qi2 offers both Black Belt and Green Belt workshops, in addition to Executive Overviews, Champion training, and Lean Six Sigma. Obtain your certification from Qi2. Typically, the Black Belt curriculum takes 4 weeks, and the Green Belt 2 weeks. The latter does not include Design of Experiments or Statistical Process Control. You can see typical workshop outlines at **www.SixSigmaQi2.com**.

We will customize the workshops (at no charge) to include examples in your product/process terminology. Each attendee receives a comprehensive set of training/ reference materials for each week of training. The courses are designed to be practical and flexible. The Black and Green Belt candidates will be given time in each workshop week to apply the 6σ steps and tools to their own projects just after the material is presented.

The course and the materials are organized by the 12 or so steps of DMAIC, so the attendees will have a roadmap of 6σ implementation that they experience in the workshop and then further apply in between the separated training weeks.

Contact Qi2 at **qi2@qualityi2.com** for a quote for on-site training at your location, or for an instructor bio and a list of references.