Notes:

Confidence Intervals

Key Learning Points

Describe the importance of establishing confidence intervals.

Explain how to calculate confidence intervals.

Utilize confidence intervals in improvement projects.

What is a Confidence Interval?

Usually population parameters such as the mean or standard deviation are impossible to observe directly. A confidence interval is a range determined by sample data that estimates a true population parameter with a degree of confidence.

Statistical Definition

A (1- α) % confidence interval for a population parameter consists of an estimate bounded by upper and lower confidence limits (CL's) calculated from sample data:

- Probability [Lower CL $\leq \mu \leq$ Upper CL] = $1 - \alpha$
- Probability [Lower CL $\leq \sigma \leq$ Upper CL] = $1 - \alpha$

What is Alpha (α)?

Alpha (α) is the probability that the actual population parameter is located outside the confidence interval (also known as a Type I Error). This probability is always greater than zero, and is frequently established at 5%.

Why Use Confidence Intervals?

Notes:

Sample statistics, such as the sample mean and the sample standard deviation, are only estimates of the true population parameters, μ and σ. Due to inherent sample to sample variability in these estimates, uncertainty is quantified using statistically based Confidence Intervals (CI).

Confidence intervals provide a way to investigate sample to sample variation, and are used to obtain statistical confidence for the population mean, standard deviation, and Cp parameters.

The automotive industry calculates 95% CI for their data.

Some medical applications calculate 99.9% CI for their data.

95% Confidence means that 95 samples out of 100 will yield a CI that contains the true population parameter.

The Confidence Interval Equation

The general confidence interval equation is a parameter estimate plus or minus the standard error for that estimate multiplied times a constant calculated from probability distribution theory. For example, a range of +/- 1.96 standard errors contains 95% of a normal distribution, so the constant for a 95% confidence interval for a population mean is 1.96.

The standard error of a statistic is the standard deviation of the sampling distribution for that statistic. As an example, the standard error for a population mean is attached.

Standard Error (SE Mean)

The standard error of the mean is the standard deviation of the sampling distribution of the mean. The SE Mean is calculated by the sample estimate of the population standard deviation (sample standard deviation) divided by the square root of the sample size:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Where:

- s is the sample standard deviation (i.e., the sample-based estimate of the standard deviation of the population)
- n is the size (number of observations) of the sample

General Form

Confidence Intervals take the general form:

- CI = Statistic ± [K(Sample Standard Deviation)]
- Statistic = Population Parameter Estimate
- K = Constant based on a Statistical Distribution
- For estimating μ K is selected from a family of "t" distributions.
- For estimating σ, K is selected from a family of Chi-Squared distributions.

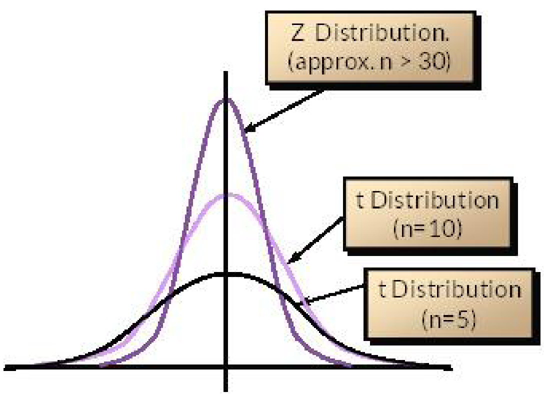
As a Black Belt you will investigate Confidence Intervals for μ, σx, and Cp.

Notes:

Confidence Interval for the Population Mean

In inferential statistics, a random sample of data taken from a population is used to make inferences about the wider population. We will use estimates of the mean and standard deviation calculated from a sample to create a confidence interval for the population mean.

The t Distribution



The t Distribution is a family of bell-shaped (normal-like) distributions whose width is sample size dependent. The smaller the sample size, the wider and flatter the distribution will be. To give an idea of the values of "t" for 95% CIs for different sample sizes, you can look at the table below:

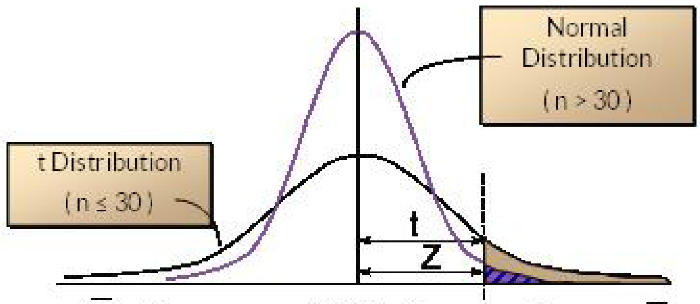
Sample	t Value
5	2.78
10	2.26
20	2.09
30	2.05
100	1.98
1000	1.96

t Distribution vs. Normal Distribution

The distribution you use is dependent on your sample size. When your sample size is less than or equal to 30, you use a t Distribution. When your sample size is above 30, you use the normal distribution.

The difference between the two distributions is when creating a t Distribution, you use a t-Table instead of Z-Table to compute the area in the tails (when sample size is less than or equal to 30).
Note: There is more area under the tail of the "t" as compared to the "Z" distribution.

If σ is unknown, s is the best estimate.



$$t_{n-1} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Example: CI for the Population Mean

Suppose that you want to determine the 95% Confidence Interval for the population mean from 10 samples (n = 10) off a reactor. You sample the reactor and obtain the following sample statistics:
Sample Mean = xbar = 249.6
Sample Standard Deviation = s =14.15

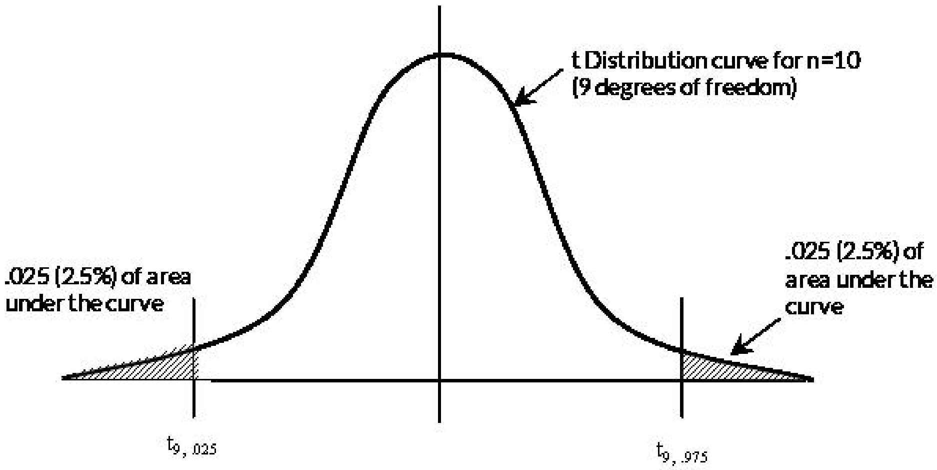
With this sample information, use the general formula for Confidence Intervals to estimate the population mean:

$$\bar{x} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$$

Where:
xbar = sample mean
s = sample standard deviation
n = sample size
tn-1, 1-α/2 = t value for n-1 degrees of freedom and probability 1-α/2

Notes:

Example: Distribution Curve



$t_{9,.025} = -t_{9,.975}$ because the t Distribution is symmetrical

Example: Solution

Using this data, you can be 95% confident that the actual process mean is somewhere between 239.45 and 259.67. Notice that the t value is very close to 2.00.

$$\bar{x} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$$

$$249.56 - 2.262 * \frac{14.15}{\sqrt{10}} \le \mu \le 249.56 + 2.262 * \frac{14.15}{\sqrt{10}}$$

$t_{n-1,1-\alpha/2}$:

df (Degrees of Freedom) = n - 1 = 9

$1-\alpha/2 = 1-.05/2 = .975$

$t_{9,.975} = 2.262$

$249.56 - 10.11 \le \mu \le 249.56 + 10.11$

$239.45 \le \mu \le 259.67$

Confidence Interval for the Population Standard Deviation

Here is the equation for a Confidence Interval for the Population Standard Deviation.

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

Where:
α = 1-% Confidence

Notes:

n = Sample Size

Example: CI for the Population Standard Deviation

Suppose you take a sample of 16 data points and get a standard deviation of 1.66. The Degrees of freedom (n-1) are 16-1 or 15.

With this sample information, use the general formula for Confidence Intervals for Population Standard Deviation to estimate the Interval:

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

Where:
α = 1-% Confidence
n = Sample Size

Example: Solution

Using this data, you can be 95% confident that the actual process mean is somewhere between 1.23 and 2.57.

$$1.66\sqrt{\frac{16-1}{\chi^2_{15,1-.05/2}}} \le \sigma \le 1.66\sqrt{\frac{16-1}{\chi^2_{15,.05/2}}}$$

$$1.66\sqrt{\frac{16-1}{\chi^2_{15,.975}}} \le \sigma \le 1.66\sqrt{\frac{16-1}{\chi^2_{15,.025}}}$$

$$1.66\sqrt{\frac{15}{27.49}} \le \sigma \le 1.66\sqrt{\frac{15}{6.26}}$$

Confidence Interval for the Process Capability (Cp)

You also may need to calculate Confidence Intervals for process capability.

$$\frac{USL - LSL}{6\,s\sqrt{\frac{n-1}{\chi^2_{n-1,\alpha/2}}}} \le C_p \le \frac{USL - LSL}{6\,s\sqrt{\frac{n-1}{\chi^2_{n-1,1-\alpha/2}}}}$$

Where:
X^2_{n-1}, α/2 and X2n-1, 1-α/2 are the lower and upper α/2 percentage points in the X^2 distribution with n-1 degrees of freedom

Notes:

Example: CI for Process Capability

Calculate the 95% CI for Cp as follows:

- USL = 300
- LSL = 200
- s = 14.15
- n = 10
- $\alpha = 0.05$

$$\frac{USL - LSL}{6\,s\sqrt{\dfrac{n-1}{\chi^2_{n-1,\,\alpha/2}}}} \le C_p \le \frac{USL - LSL}{6\,s\sqrt{\dfrac{n-1}{\chi^2_{n-1,\,1-\alpha/2}}}}$$

When Should Confidence Intervals Be Used?

- Confidence intervals are used to put a margin of error on a statistic.
- Recognizing that the statistic measured from the sample is not exactly equal to the overall population parameter, the confidence interval provides an estimate of prediction error.

Pitfalls to Avoid

- Estimates of sample error depend on specific probability distributions.
- The equations used in this module require normally distributed data.

Notes: