Statistical Process Control (SPC) Training Guide

Rev X05, 09/2013

What is data?

- Data is factual information (as measurements or statistics) used as a basic for reasoning, discussion or calculation. (Merriam-Webster Dictionary, m-w.com)
- What does this mean?
  - Data allows us to make educated decisions and take action!

Why is Data Important?

If we don't collect data about a process then what?

- Without data we don't understand the process / product

So what?

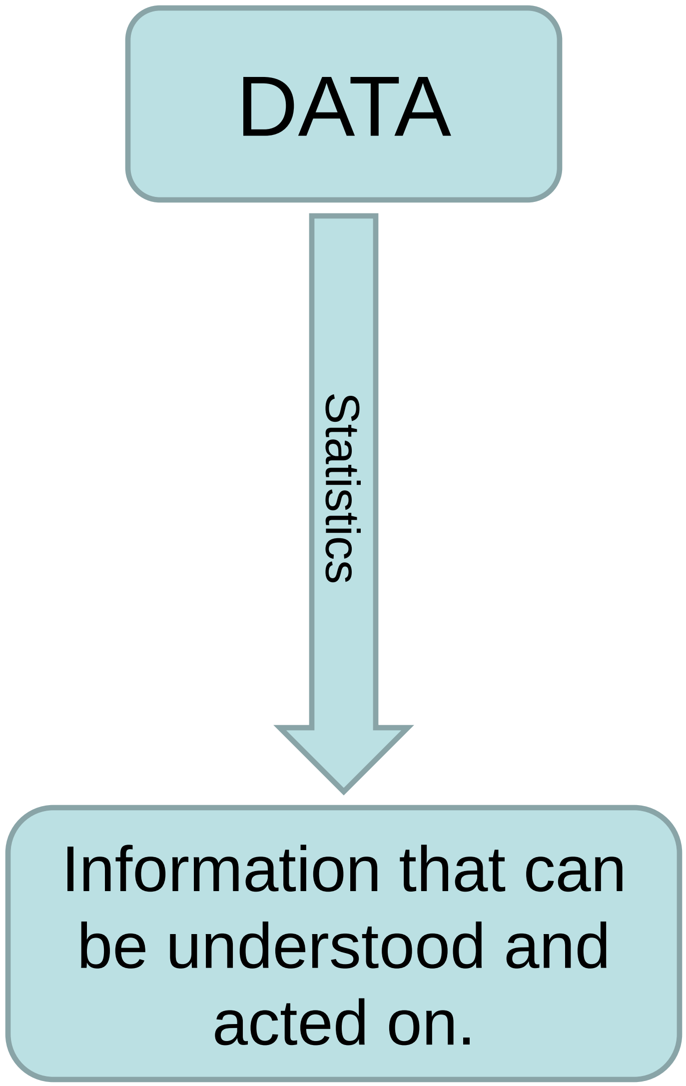
- Without understanding of the process / product we can't control the outcome

Is it important to control the outcome?

- When you can't control the outcome you are dependent on chance.
- You may have a good outcome, you may not.
- Without data collection you may not know either way.

Use of Data

- I’m not collecting data because its non value add.
  - Without data collection there is no way to identify problems, continuously improve or ensure you are meeting the voice of the customer.
- I’m collecting data, but not looking at it. Is that okay?
  - No, the collection of data without analysis is a bigger waste than not collecting data in the first place.
- What should I be doing with the data?

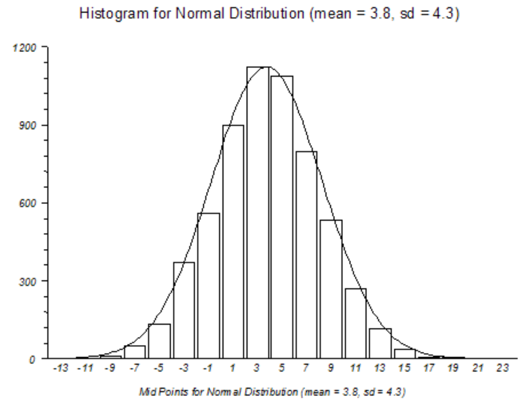

DATA
Statistics
Information that can be understood and acted on.

What are Statistics?

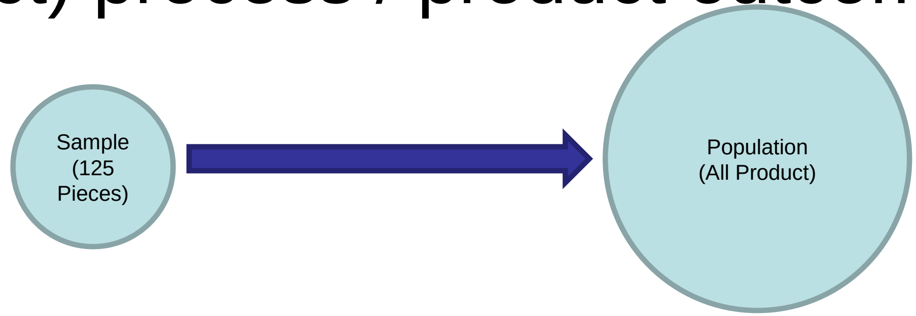
- A branch of mathematics dealing with the collection, analysis, interpretation and presentation of masses of numerical data (Merriam-Webster Dictionary, m-w.com)
- What does this mean?
  - Once data is collected we can use appropriate statistical methods to describe (understand) our process / product and control (predict) the process / product outcome

Statistics Allow us to

Describe our process / product



Control (predict) process / product outcomes



Important Notes

1. Statistical conclusions require useful data
   - We need to measure the right thing
   - Determining what data to collect and how to collect it are important steps in the APQP / continuous improvement process
2. We need to have confidence the data collected is accurate
   - A good measurement system is required to collect data
   - Be sure the measurement system analysis (MSA) is acceptable before collecting data
   - There is a separate training available for MSA
3. Reduce / Eliminate Waste
   - Data that doesn't provide useful information drives waste
   - We want to gain the most useful information from the least amount of data possible!

Types of Data

- Variable Data (Continuous Data)
  - Measurements on a continuous scale
  - Examples
    - Product Dimensions
    - Weight
    - Time
    - Cost
    - Process parameters (cutting speed, injection pressure, etc.)
- Attribute Data (Discrete Data)
  - Data by counting
  - Examples
    - Count of defective parts from production
    - Number of chips on a painted part

Types of Attribute Data

Binomial Distribution

- Pass / Fail (PPM)
- Number of defective bezels
- Number of defective castings

Poisson Distribution

- Number of defects (DPMO)
- Number of defects <u>per</u> bezel
- Number of defects <u>per</u> casting

Which Type of Data is Better?

Variable Data

Pros:

- Provides useful information with smaller sample sizes
- Can identify common cause concerns at low defect rates
- Can be used to predict product / process outcomes (trends)
- Very useful for continuous improvement activities (DOE, Regression analysis, etc.)

Cons:

- Data collection can be more difficult, requiring specific gauges or measurement methods
- Analysis of data requires some knowledge of statistical methods (Control charting, Regression analysis, etc.)

Attribute Data

Pros:

- Very easy to obtain
- Calculations are simple
- Data is usually readily available
- Good for metrics reporting / management review
- Good for baseline performance

Cons:

- Data collection can be more difficult, requiring specific gauges or measurement methods
- Analysis of data requires some knowledge of statistical methods (Control charting, Regression analysis, etc.)

Introduction to Statistics

Goal:

1. Define basic statistical tools
2. Define types of distributions
3. Understand the Central Limit Theorem
4. Understand normal distributions

Definitions

- Population: A group of <u>all</u> possible objects

- Subgroup (Sample): One or more observations used to analyze the performance of a process.

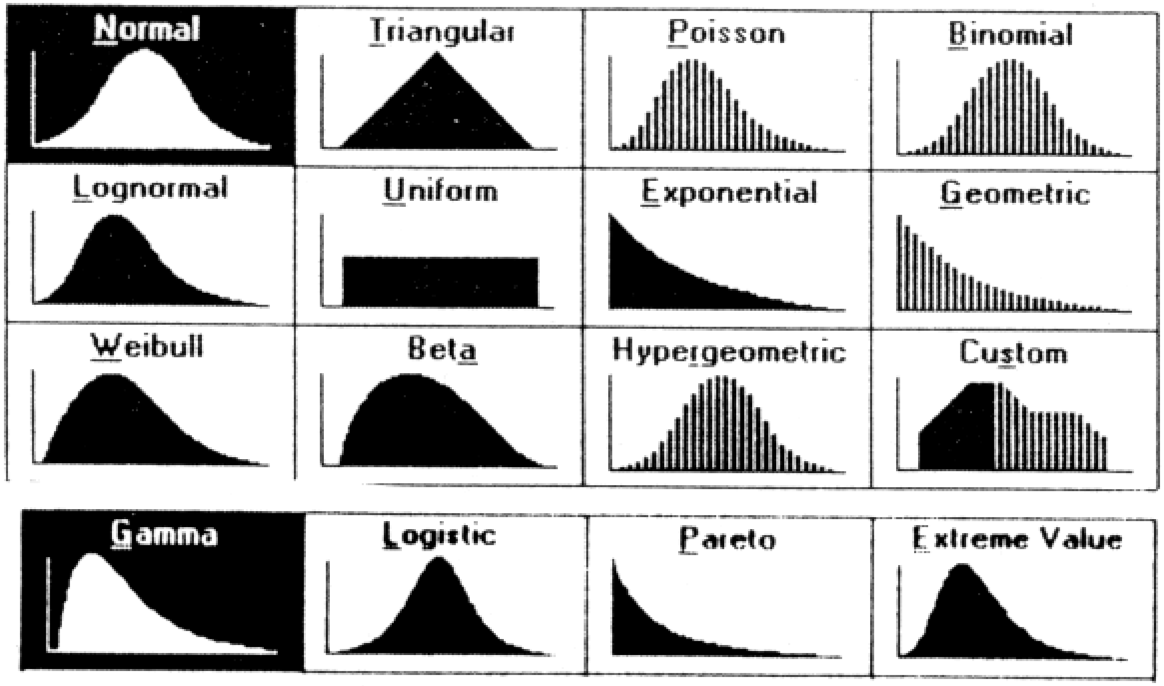
- Distribution: A method of describing the output of a stable source of variation, where individual values as a group form a pattern that can be described in terms of its location, spread and shape.

Measures of Data Distribution

- Measures of Location
  - Mean (Average)
    - The sum of all data values divided by the number of data points
  - Mode
    - The most frequently occurring number in a data set
  - Median (Midpoint)
- Measures of Spread
  - Range
    - The difference between the largest and smallest values of a data set
  - Standard Deviation
    - The square root of the squared distances of each data point from the mean, divided by the sample size. Also known as Sigma (σ)

Distribution Types

- Discrete
  - Binomial
  - Poisson
- Continuous
  - Normal
  - Exponential
  - Weibull
  - Uniform



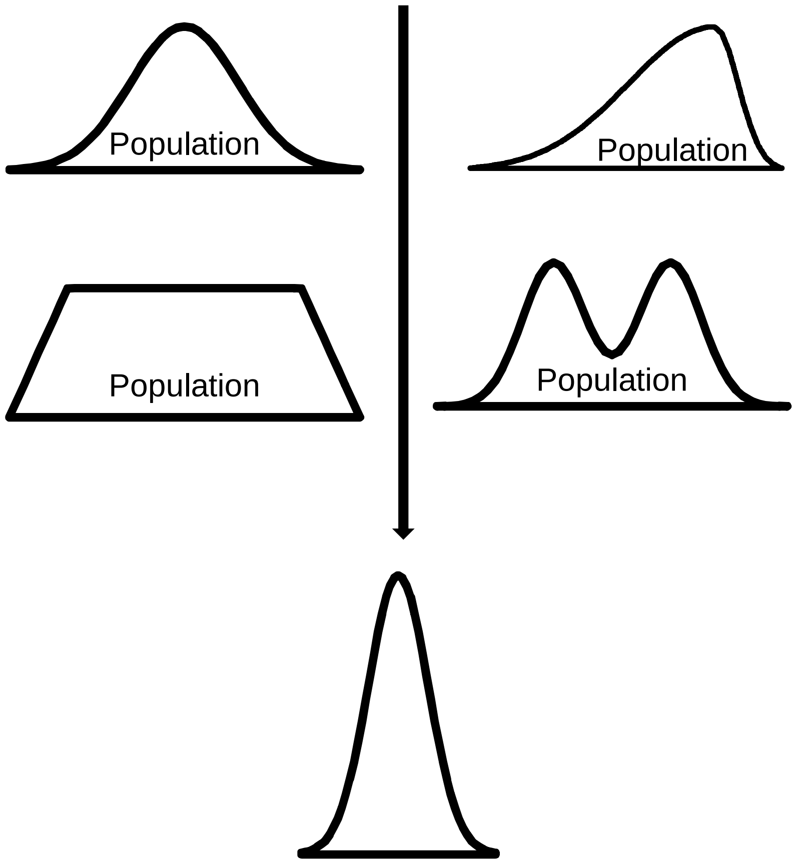
Central Limit Theorem

The Central Limit Theorem is the basis for sampling and control charting (of averages).

There are 3 properties associated with the CLT;

1. The distribution of the sample means will approximate a normal distribution as the sample size increases, even if the population is non-normal
2. The average of the sample means will be the same as the population mean
3. The distribution of the sample means will be narrower than the distribution of the individuals by a factor of $\frac{1}{\sqrt{n}}$, where n is the sample size.

CLT Property 1



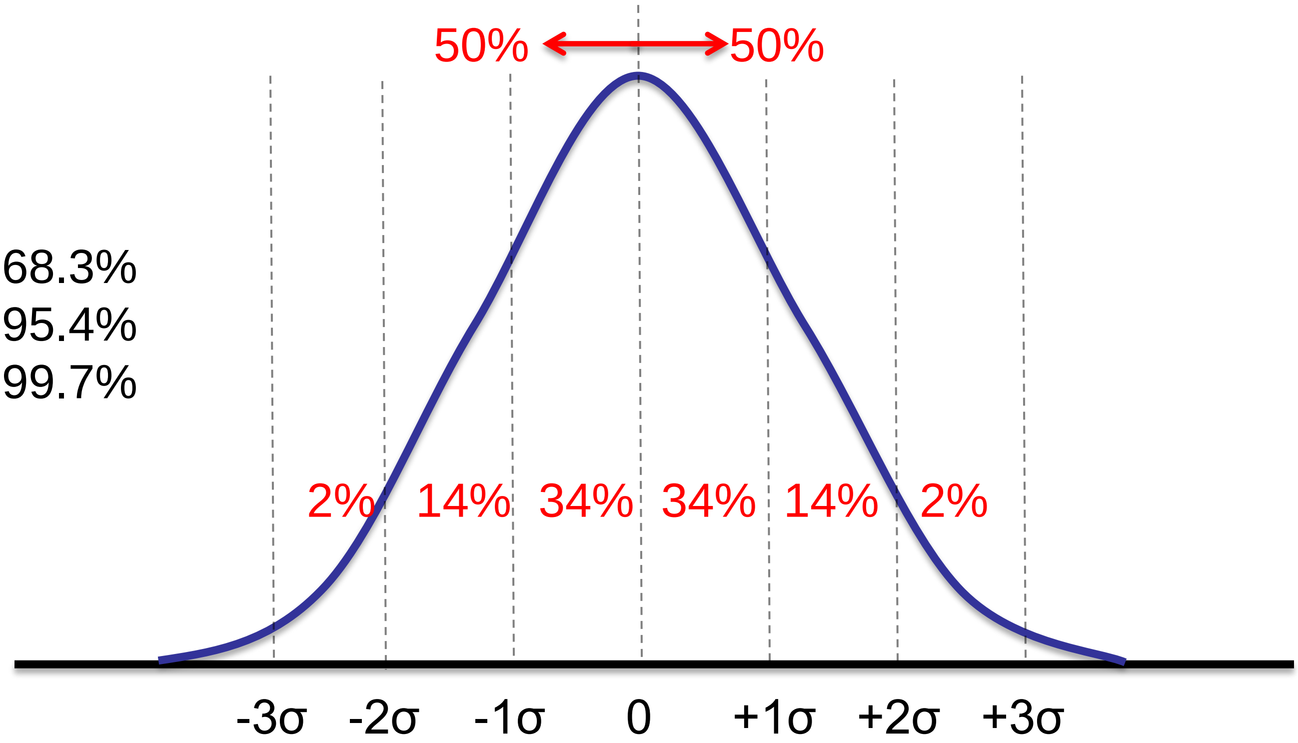
Distribution of sample means from any population

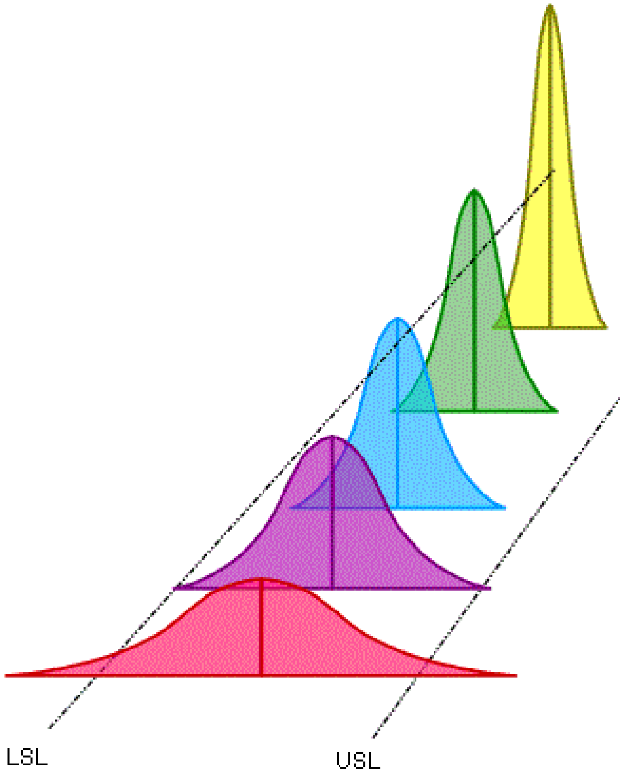
The distribution of the sample means of any population will approximate a normal distribution as the sample size increases, even if the population is non-normal.

Because of this property control charts (for averages) is based on the normal distribution.

The Normal Distribution

$\mu \pm 1\sigma$ = 68.3%
$\mu \pm 2\sigma$ = 95.4%
$\mu \pm 3\sigma$ = 99.7%





Cp	Sigma	ppm
2.00	6.0	3.4
1.67	5.0	233
1.33	4.0	6,210
1.00	3.0	66,800
0.67	2.0	308,540

PROCESS CAPABILITY

Process Capability

Goal

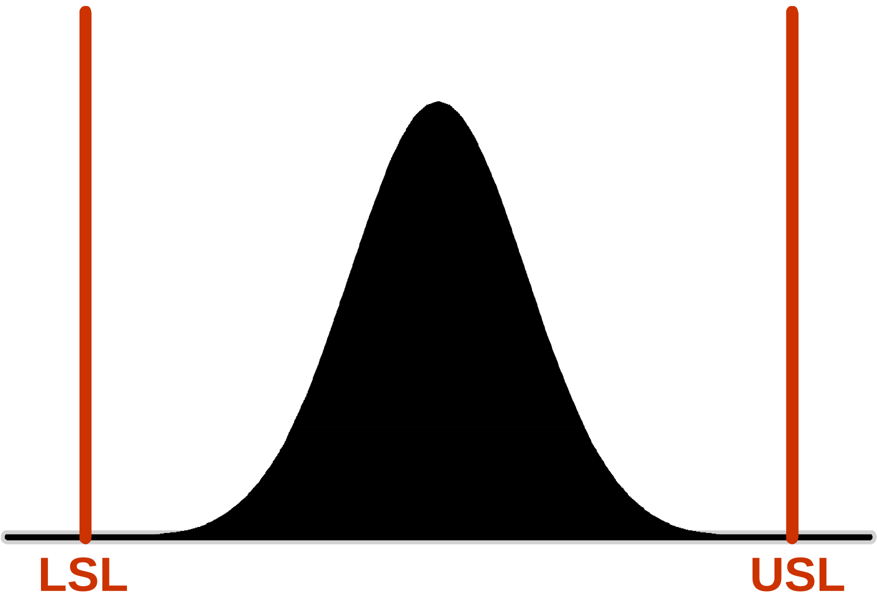
- Understand process capability and specification limits
- Understand the procedure of calculating process capability
- Understand C_p, C_{pk}, P_p and P_{pk} indices
- Estimating percentage of process beyond specification limits
- Understanding non-normal data
- Example capability calculation

Specification Limits

Individual features (dimensions) on a product are assigned specification limits.

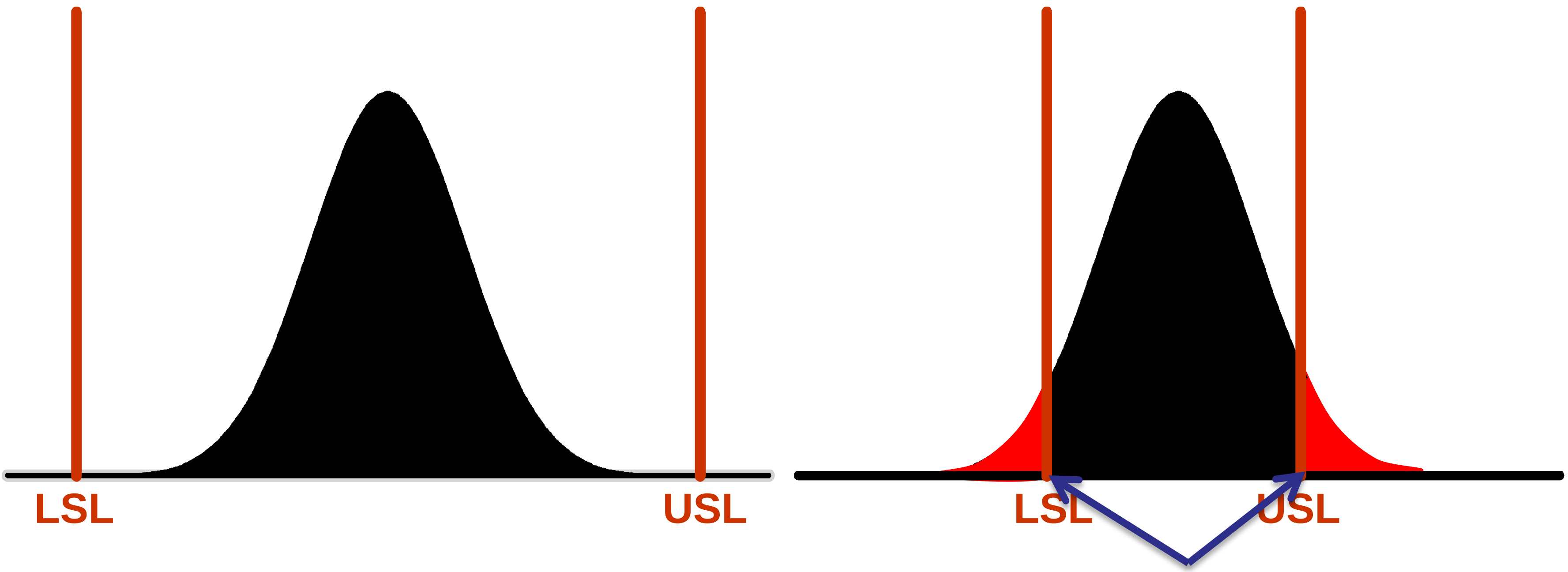
How do we determine a process is able to produce a part that meets specification limits?

Process Capability!



Process Capability and Specification Limits

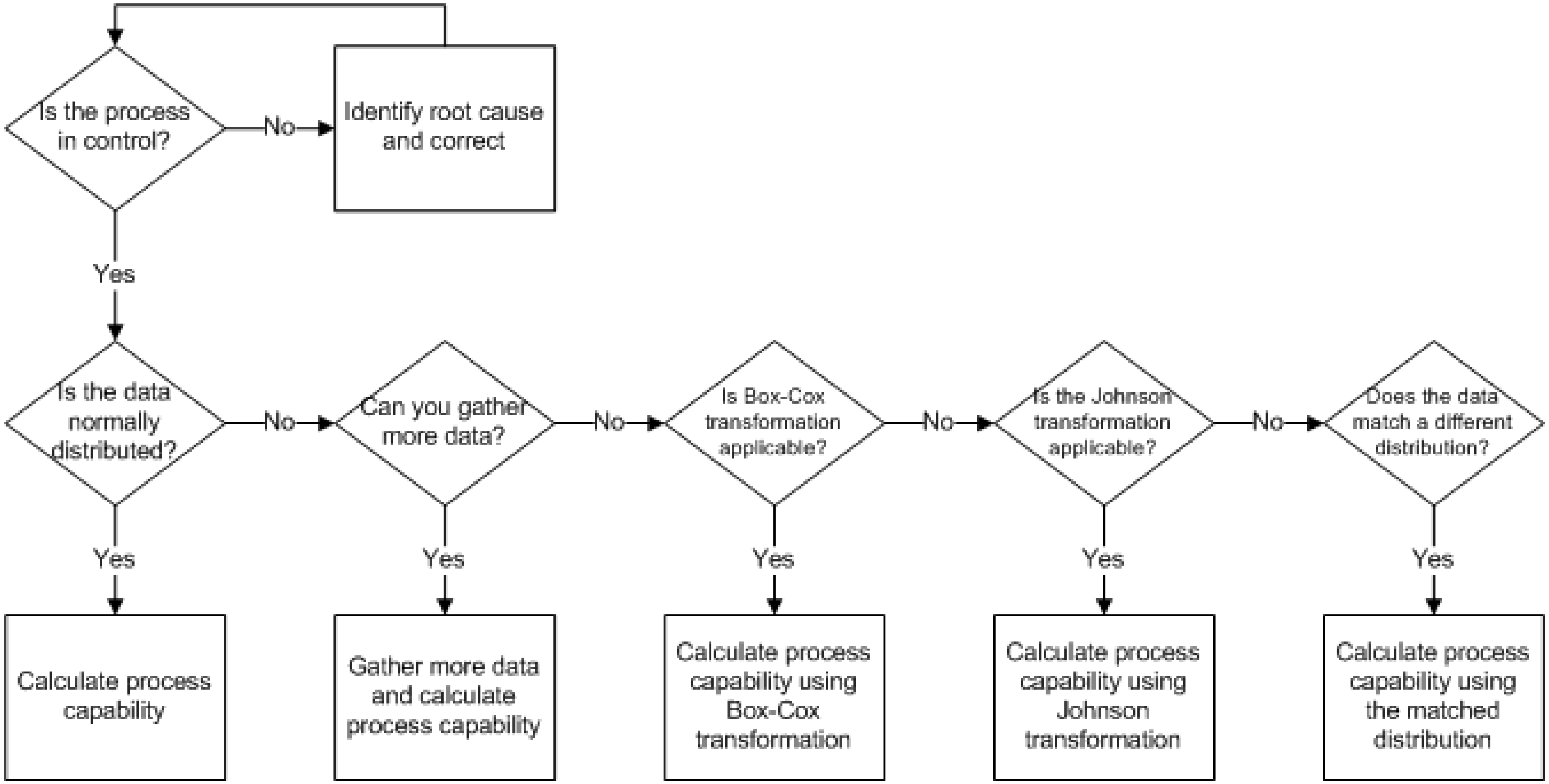
Process capability is the ability of a process to meet customer requirements.



Acceptable capability

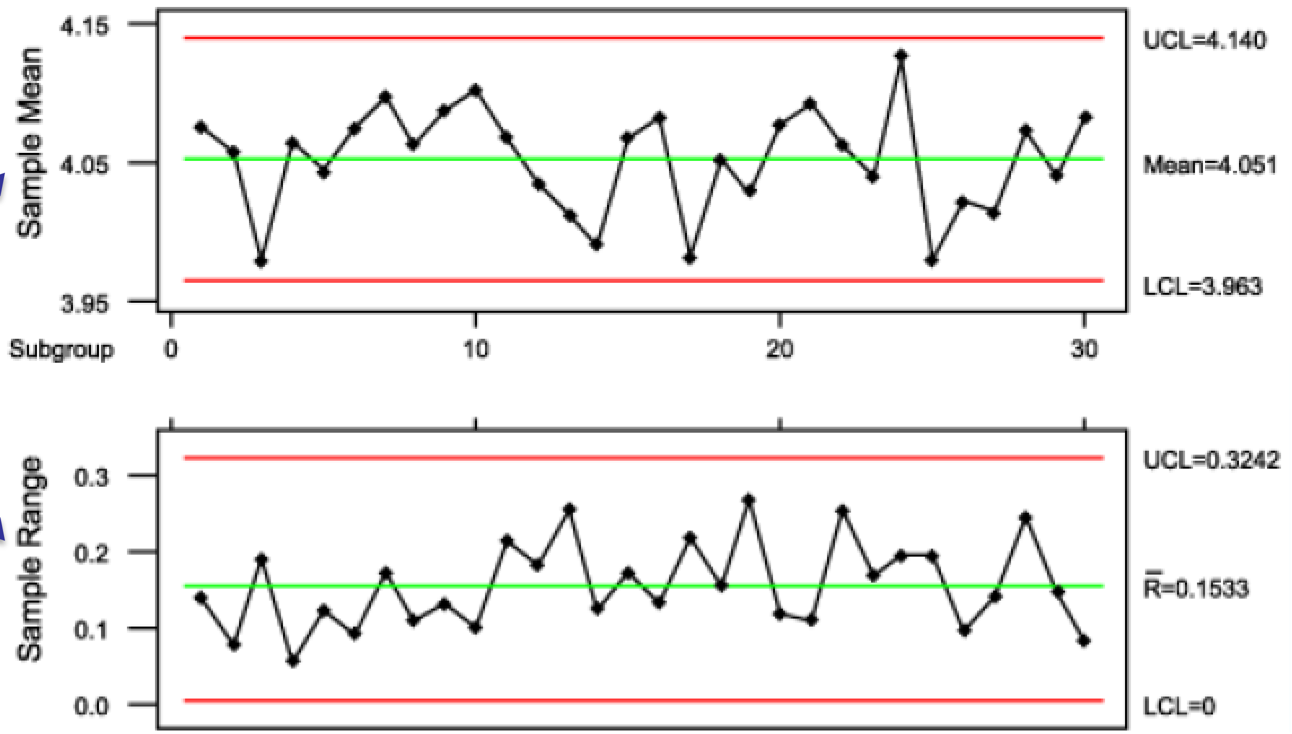
Product outside of spec limits

Calculating Process Capability



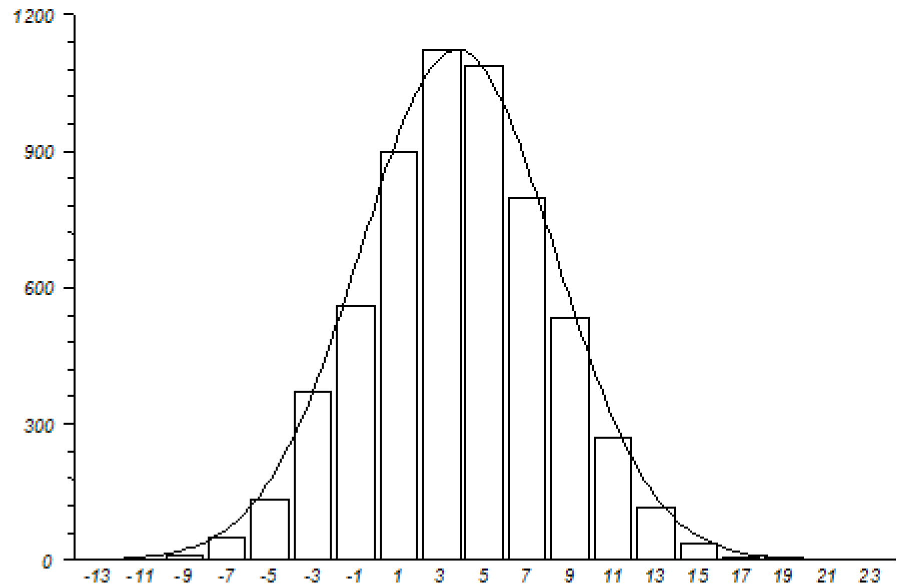
Determine Stability

All sample means and ranges and in control and do not indicate obvious trends



Determine Normality

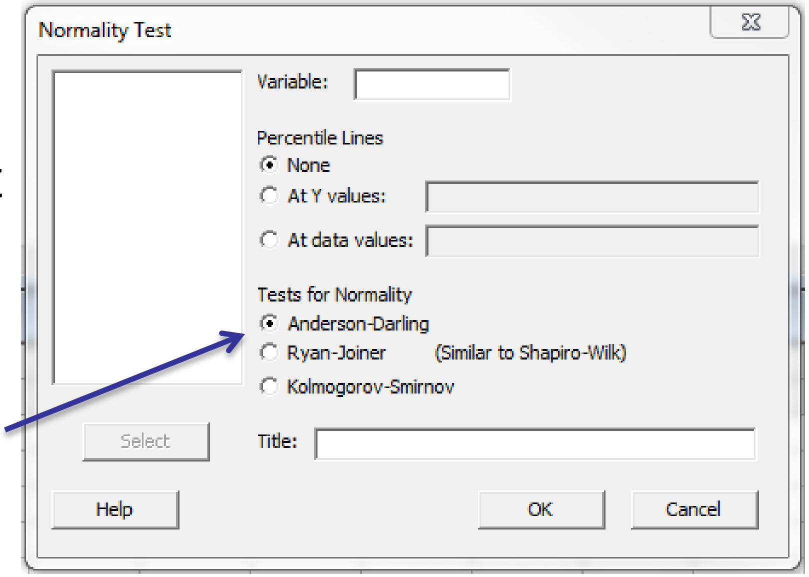
- Placing all data in a histogram may be used to help determine normality. If the data represents a normal curve.



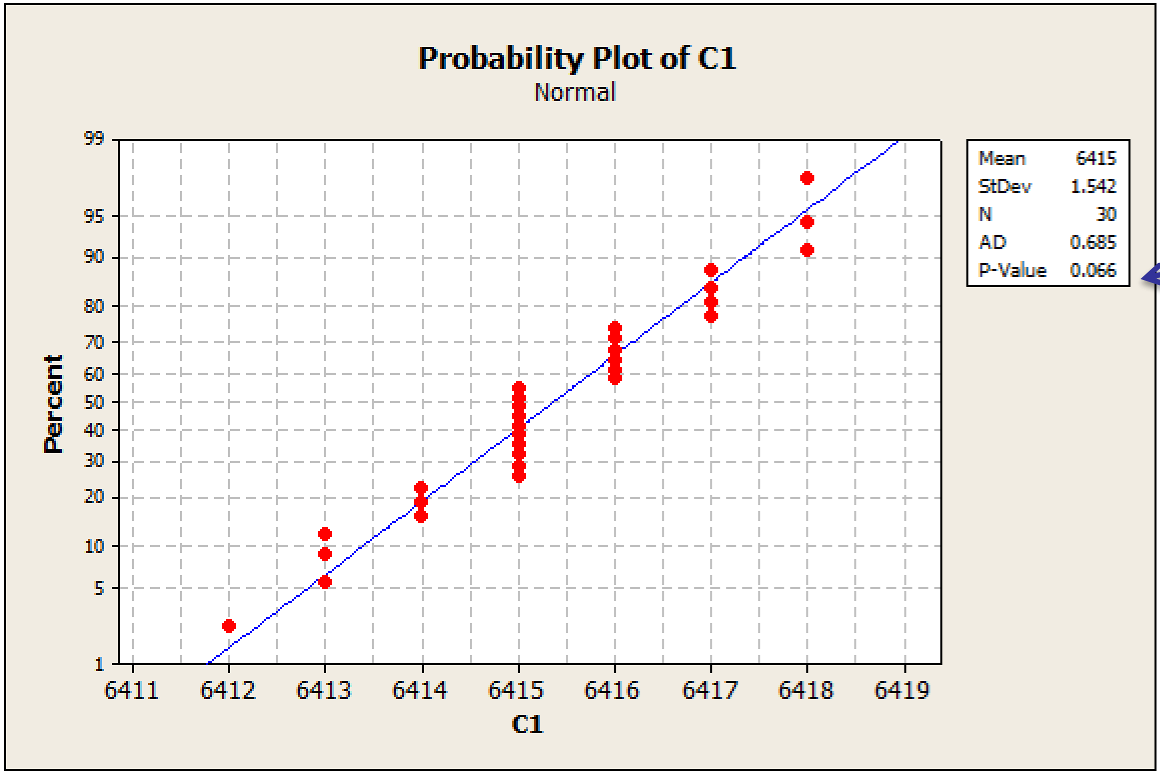
Determine Normality

- A more statistical method is to use the Anderson Darling test for normality
- In Minitab go to:

  Stat > Basic Statistics > Normality Test
- Select Anderson Darling and click ok



Interpreting Normality



The p-Value must be greater than or equal to 0.05 to match the normal distribution.

Process Indices

- Indices of process variation only, in regard to specification;
  - C_p and P_p
- Indices of process variation and centering combined, in regard to specification;
  - C_{pk} and P_{pk}

NOTE

Before calculating capability or performance indices we need to make sure of a couple of things!

1. The process needs to be stable (in control)
2. The process needs to be normal
3. A completed MSA needs to prove the measurement system is acceptable

If the above items are not met understand the results of the capability studies may be inaccurate. Also, per the AIAG PPAP manual (4th Edition) if the above items are not met corrective actions may be required prior to PPAP submittal.

C_p Overview

C_p = Potential Process Capability = $\frac{Tolerance}{Width\ of\ Distribution} = \frac{Voice\ of\ Customer}{Voice\ of\ Process}$

This index indicates ***potential*** process capability.
Cp is not impacted by the process location and can only be calculated for bilateral tolerance.

C_p Calculation

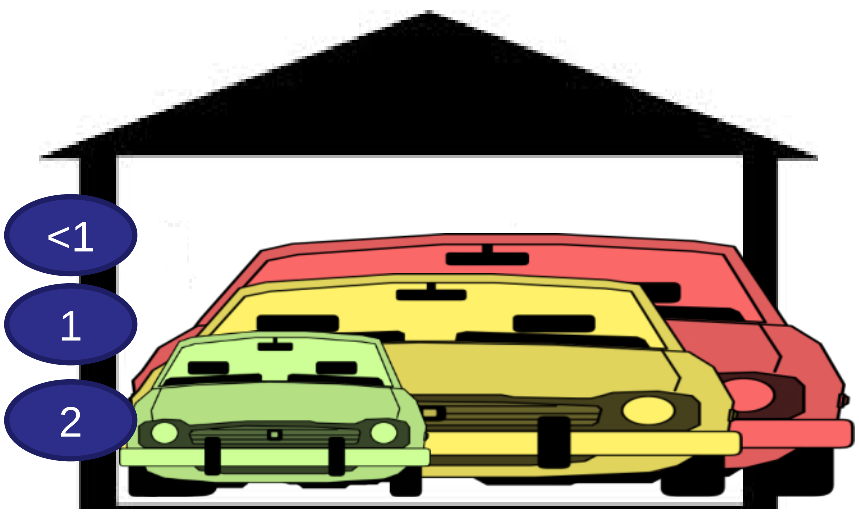
$$C_p = \frac{USL - LSL}{6\sigma_c} = \frac{USL - LSL}{6\left(\frac{\bar{R}}{d_2}\right)}$$

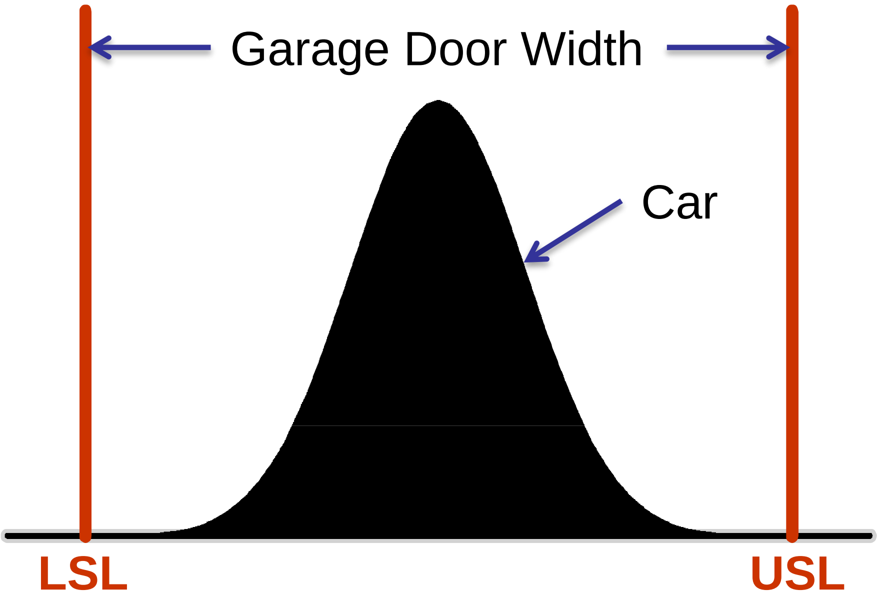
Where:

USL = Upper specification limit
LSL = Lower specification limit
$\bar{R}$ = Average Range
d_2 = a constant value based on subgroup sample size

Subgroup Size	d_2
2	1.128
3	1.693
4	2.059
5	2.326
6	2.534
7	2.704
8	2.847
9	2.970
10	3.078

Interpreting C_p





Cp = Number of times the car (distribution) fits in the garage (specification limit)

C_{pk} Overview

C_{pk} is a capability index. It takes the process location <u>and</u> the capability into account.

C_{pk} can be calculated for both single sided (unilateral) and two sided (bilateral) tolerances.

For bilateral tolerances C_{pk} is the minimum of CPU and CPL where:

$$\text{CPU} = \frac{USL - \bar{\bar{X}}}{3\sigma_c} = \frac{USL - \bar{\bar{X}}}{3\left(\frac{\bar{R}}{d_2}\right)} \quad \text{and} \quad \text{CPL} = \frac{\bar{\bar{X}} - LSL}{3\sigma_c} = \frac{\bar{\bar{X}} - LSL}{3\left(\frac{\bar{R}}{d_2}\right)}$$

Where:

$\bar{\bar{X}}$ = Process average

USL = Upper specification limit

LSL = Lower specification limit

$\bar{R}$ = Average Range

d_2 = a constant value based on subgroup sample size

Note that $\frac{\bar{R}}{d_2}$ is an estimate of the standard deviation

C_{pk} Example

Subgroup Size	d_2
2	1.128
3	1.693
4	2.059
5	2.326
6	2.534
7	2.704
8	2.847
9	2.970
10	3.078

GHSP is provided a machined casting with a hole diameter of

16.5 ± 1.0mm.

The supplier has collected 25 subgroups of 5 measurements and wants to determine the process capability.

With a subgroup size of 5 d2 is 2.326 (table to right).

If the process average ($\bar{\bar{X}}$) is 16.507 and the average range ($\bar{R}$) is 0.561, then the C_{pk} is calculated as follows:

$$C_{pk} = \text{Minimum} \left\langle \text{CPU} = \frac{USL - \bar{\bar{X}}}{3\sigma_c} = \frac{USL - \bar{\bar{X}}}{3\left(\frac{\bar{R}}{d_2}\right)}, \text{CPL} = \frac{\bar{\bar{X}} - LSL}{3\sigma_c} = \frac{\bar{\bar{X}} - LSL}{3\left(\frac{\bar{R}}{d_2}\right)} \right\rangle$$

$$= \text{Minimum} \left\langle \text{CPU} = \frac{17.5 - 16.507}{3\left(\frac{0.561}{2.326}\right)}, \text{CPL} = \frac{16.507 - 15.5}{3\left(\frac{0.561}{2.326}\right)} \right\rangle$$

$$= \text{Minimum} \langle \text{CPU} = 1.372 , \text{CPL} = 1.392 \rangle$$

$$= 1.372$$

C_p / C_{pk} Review

C_p indicates how many process distribution widths can fit within specification limits. It does <u>not</u> consider process location. Because of this it only indicates ***potential*** process capability, not ***actual*** process capability.

C_{pk} indicates ***actual*** process capability, taking into account both process location and the width of the process distribution.

P_p Overview

This index indicates ***potential*** process performance. It compares the maximum allowable variation as indicated by tolerance to the process performance

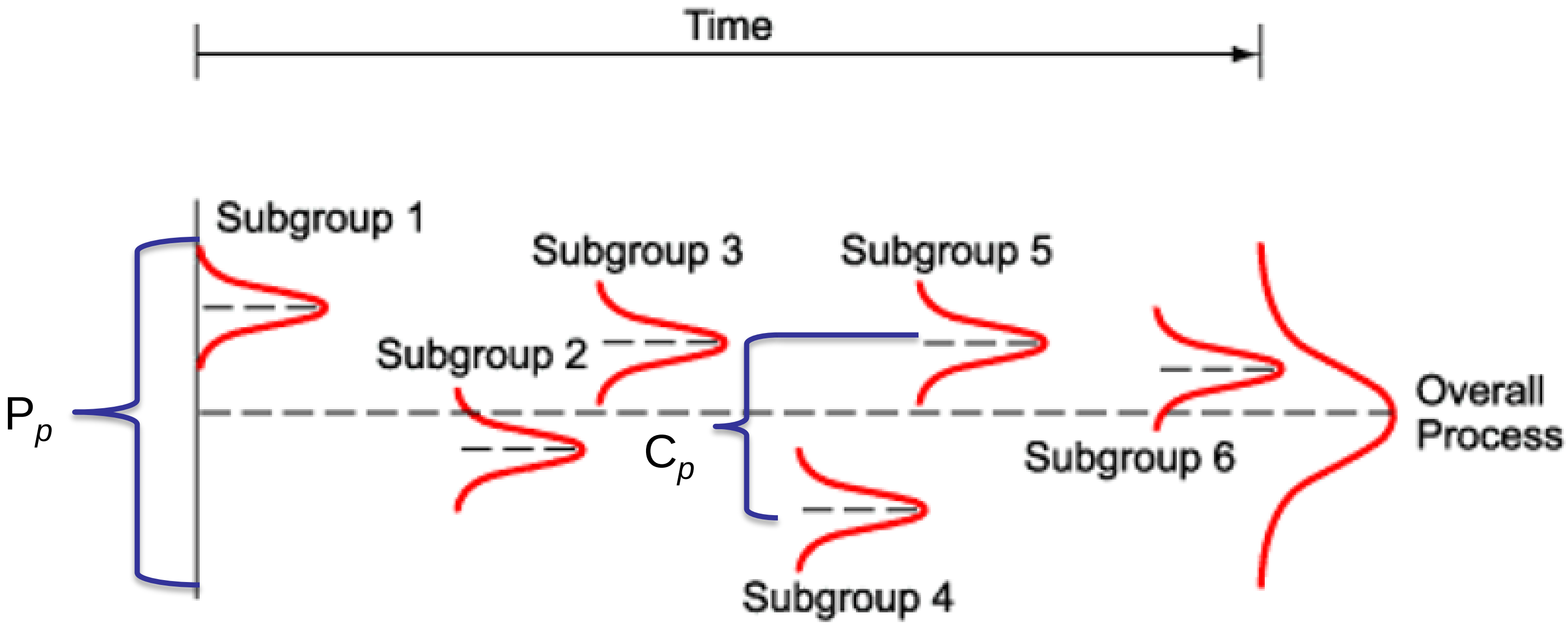
P_p is not impacted by the process location and can only be calculated for bilateral tolerance.

P_p must be used when reporting capability for PPAP

$$P_p = \frac{USL - LSL}{6\sigma_p}$$

Where:
σ = Standard deviation
USL = Upper specification limit
LSL = Lower specification limit

P_p and C_p



C_p takes into account within subgroup variation (average range)
P_p takes into account between subgroup variation

P_{pk} Overview

- P_{pk} is a performance index. It indicates actual process performance, taking into account both process location and overall process variation.
- P_{pk} shows if a process is actually meeting customer requirements.
- P_{pk} can be used for both unilateral and bilateral tolerances.
- P_{pk} must be used when reporting capability for PPAP

P_{pk} Calculation

Ppk is the minimum of PPU and PPL where:

$$\text{PPU} = \frac{USL - \bar{\bar{X}}}{3\sigma_p} \quad \text{and} \quad \text{PPL} = \frac{\bar{\bar{X}} - LSL}{3\sigma_p}$$

Where:
$\bar{\bar{X}}$ = Process average
USL = Upper specification limit
LSL = Lower specification limit
σ = Standard deviation

P_p / P_{pk} Review

P_p indicates how many process distribution widths can fit within specification limits. It does <u>not</u> consider process location. Because of this it only indicates ***potential*** process performance, not ***actual*** process performance.

P_{pk} indicates ***actual*** process performance, taking into account both process location and the width of the process distribution.

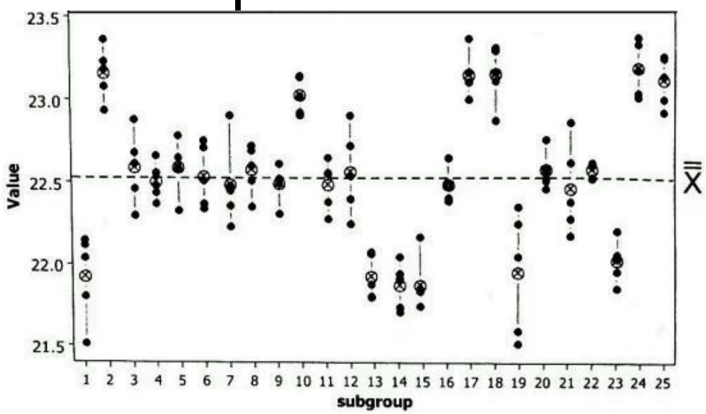
A Few Notes

- C_{pk} will always be smaller than C_p, unless the process is centered. If the process is centered the two value will be the same.
- P_{pk} will always be smaller than P_p, unless the process is centered. If the process is centered the two values will be the same.
- C_p and P_p cannot be calculated for unilateral tolerances.
- C_{pk} and P_{pk} can be calculated for both unilateral and bilateral tolerances.

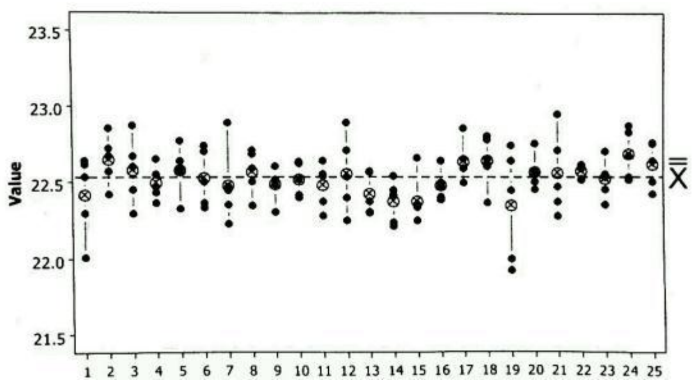
- Often, short term capability will be calculated using C_p and C_{pk}. This is because these indices use $\frac{\bar{R}}{d_2}$ to estimate standard deviation, which is a calculation of within subgroup variation. This calculation tells us how good a process could ***potentially*** be at its best performance.
- Pp and Ppk uses the actual (overall or between subgroup) standard deviation to calculate performance. As such, when reporting capability use P_p and P_{pk}.

C_{pk} and P_{pk} Compared

Ppk = 0.71
Cpk = 1.80



Ppk = 1.71
Cpk = 1.80



These two data sets contain the same data.

The top data set is from an immature process that contains special cause variation.

The bottom data set has the same within group variation, but has between group variation removed.

The bottom data set shows a process that is in statistical control.

This example may be found in the AIAG SPC (Second Edition) Manual on page 136

Process Indices Review

Process Indices

1. Cp = 4.01, Cpk = 4.01
   Pp = 4.00 , Ppk = 3.99

2. Cp = 0.27, Cpk = 0.26
   Pp = 0.25, Ppk = 0.24

3. Cp = 4.01, Cpk = -2.00
   Pp = 3.99, Ppk = -2.01

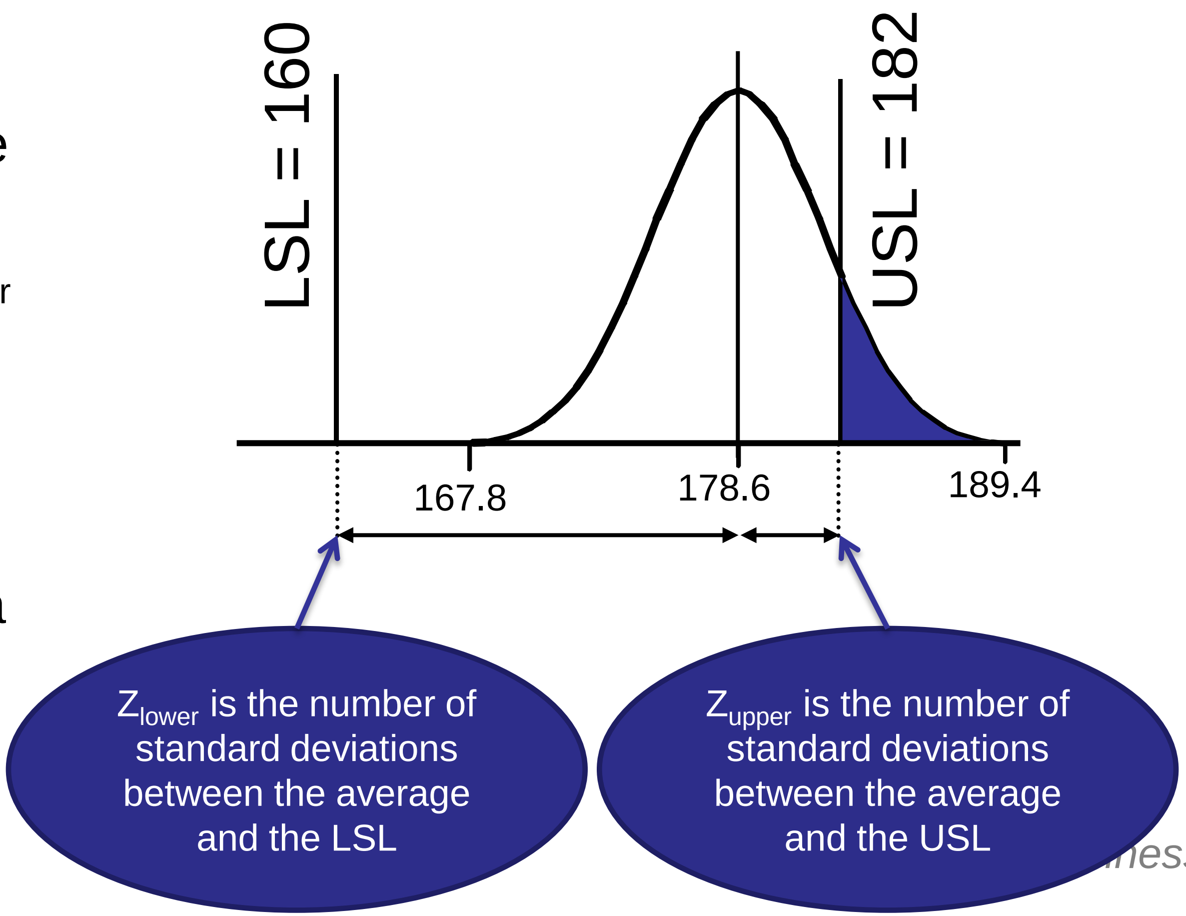
4. Cp = 4.01, Cpk = 4.00
   Pp = 2.01, Ppk = 2.00

Interpretation

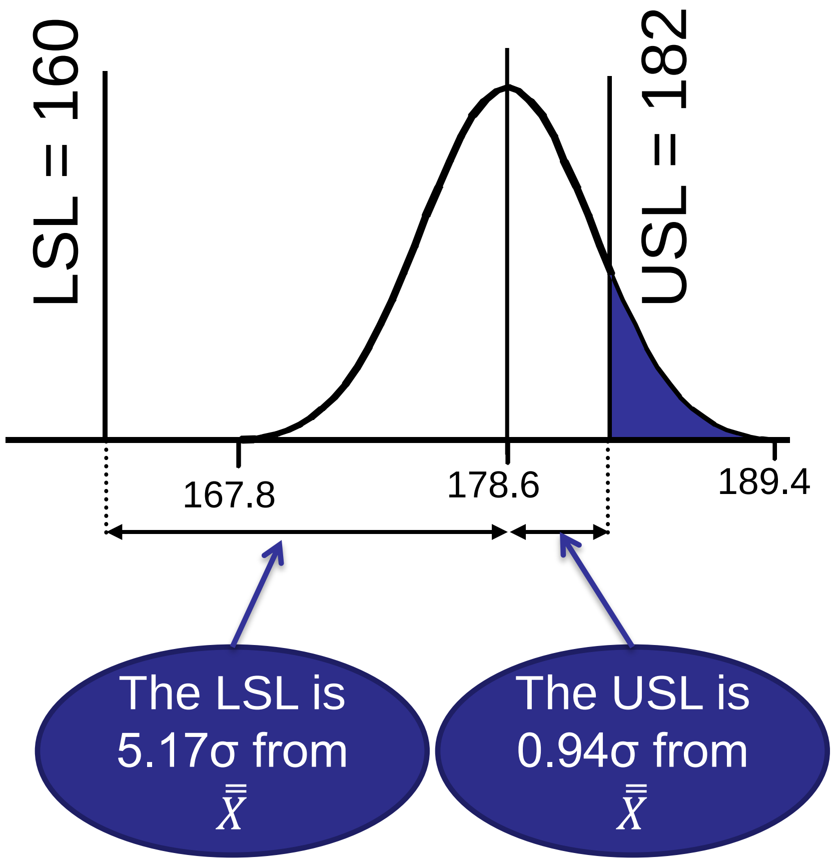
1. This process is stable and produces almost all parts within specification
2. Significant common cause variation exists
3. Significant special cause variation exists
4. Improvement can be made by centering the process

Estimating Percent Out of Specification

- To estimate the percentage of product that falls outside of the specification limits we need to compute Z_{upper} and Z_{lower}
- For this example assume an average range of 8.4 from a stable process using a sample size of 5



Estimating Percent Out of Specification



$$\sigma = \frac{\bar{R}}{d_2} = \frac{8.4}{2.326} = 3.6$$

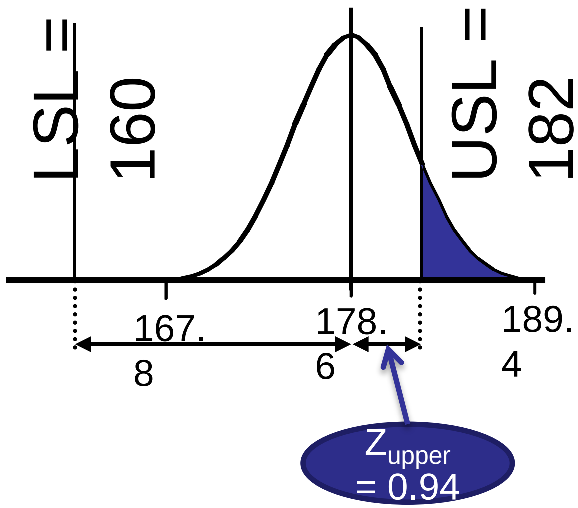
$$Z_{upper} = \frac{USL - \bar{\bar{X}}}{\sigma}$$

$$Z_{upper} = \frac{182.0 - 178.6}{3.6} = 0.94$$

$$Z_{lower} = \frac{\bar{\bar{X}} - LSL}{\sigma}$$

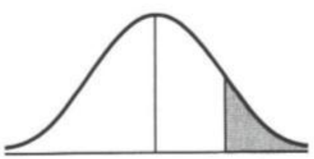
$$Z_{lower} = \frac{178.6 - 160.0}{3.6} = 5.17$$

Estimating Percent Out of Specification



Next, we need to reference a z table. From the z table we find 0.94, which corresponds to a proportion of 0.1736

This convert to 17.36% defective or 173,600 PPM



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Understanding Non-Normal Data

What happens when data used for capability is not normally distributed?

- C_p and P_p indices are robust in their accuracy with regards to non-normal data.
- C_{pk} and P_{pk} indices are not robust with regards to non-normal data.
- Calculating (and making decisions based on) P_{pk} and C_{pk} indices assuming normality with non-normal data can be misleading.

Understanding Non-Normal Data

What do I do if my data is not normal?

- Gather more data
  - This may or may not be an option given timing and cost.
  - The Central Limit Theorem (covered earlier) states that <u>all</u> population means resemble the normal distribution with larger sample size
- Transform the data
  - Using either the Box-Cox or Johnson transformations we can transform non-normal data to a near normal form
  - This allows use to accurately calculate Cpk and Ppk indices and the proportion nonconforming
- Calculate capabilities based on different distributions

Box-Cox Transformation

- The goal of the Box-Cox transformation is to identify a Lambda value (λ).
- The Lambda value will then be used in the function X^{λ} to transform the data (X) from a non-normal set into a normal data set.
- The formula for this transformation is $W=X^{\lambda}$ where:

$$-5 \leq \lambda \geq 5$$

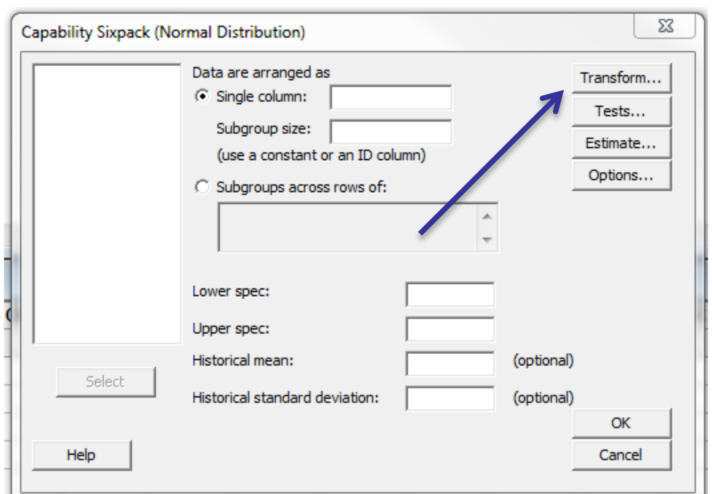
and $\lambda = 0$ for the natural log transformation

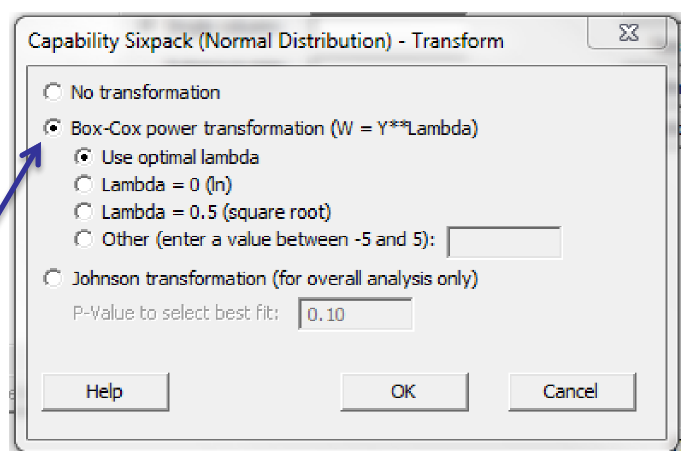
$\lambda = 0.5$ for the square root transformation

Box-Cox in Practice

- The Box-Cox transformation can be used in Minitab's capability analysis of normal data.
- When running a capability study in Minitab select:

  Stat > Quality Tools > Capability Sixpack > Normal
- Then select "Transform" and check "Box-Cox power Transformation"





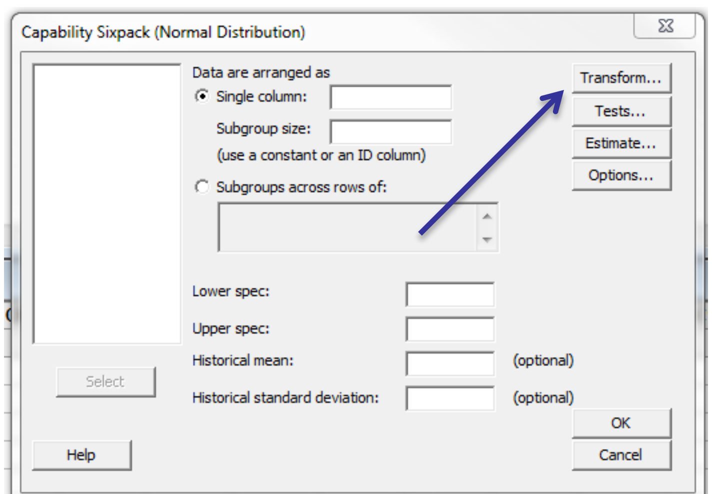
Johnson Transformation

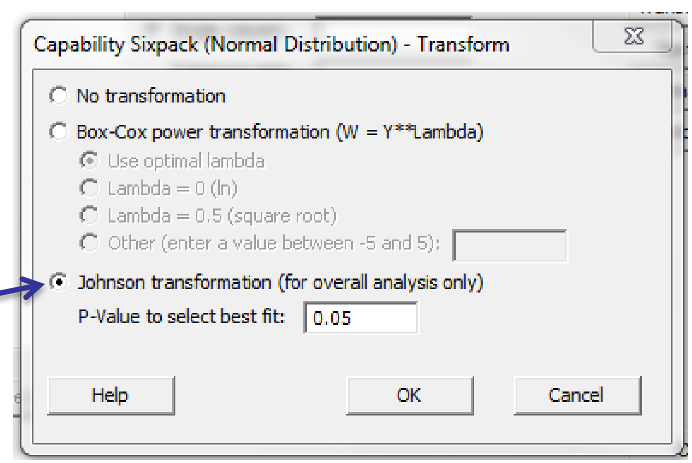
- The Johnson Transformation uses a system of transformations that yield approximate normality.
- The Johnson Transformation covers:
  - Bounded data
  - Log normal data
  - Unbounded data
- This system of transformations cover all possible unimodal distribution forms (skewness-kurtosis)
- This transformation can also be ran in Minitab

Johnson Transformation in Practice

- The Johnson Transformation can be used in Minitab’s capability analysis of normal data.
- When running a capability study in Minitab select:

  Stat > Quality Tools > Capability Sixpack > Normal
- Then select “Transform” and check “Box-Cox power Transformation”





Example Process Capability

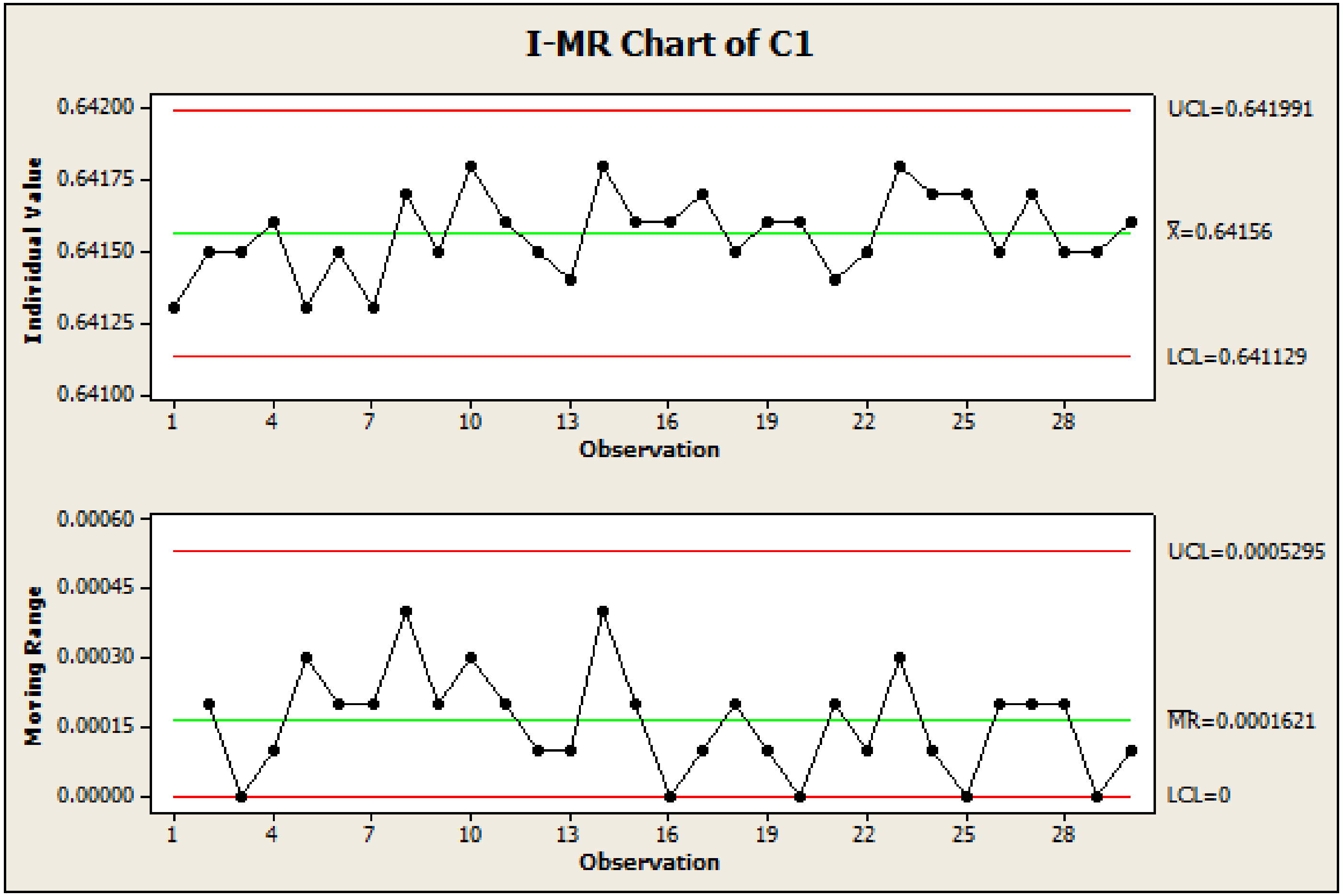
Using the data below captured from an early pre-production run calculate initial process capability.

0.6413	0.6416
0.6415	0.6417
0.6415	0.6415
0.6412	0.6416
0.6413	0.6416
0.6415	0.6414
0.6413	0.6415
0.6414	0.6418
0.6415	0.6417
0.6418	0.6417
0.6416	0.6415
0.6415	0.6417
0.6414	0.6415
0.6418	0.6415
0.6416	0.6416

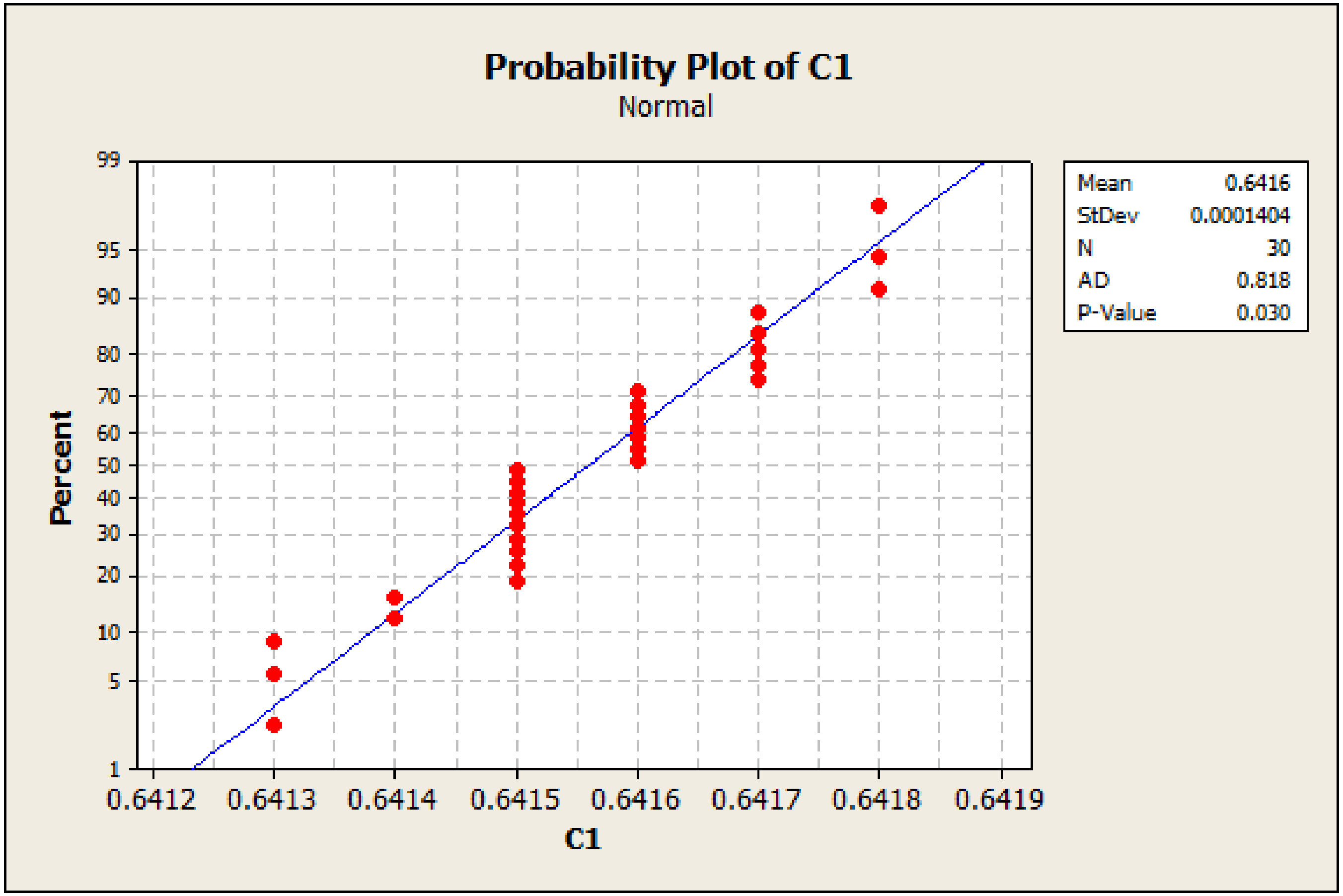
The process specification is 0.642 +0.001 / -0.002

30 data points were collected (subgroup of 1)

Is the data stable?

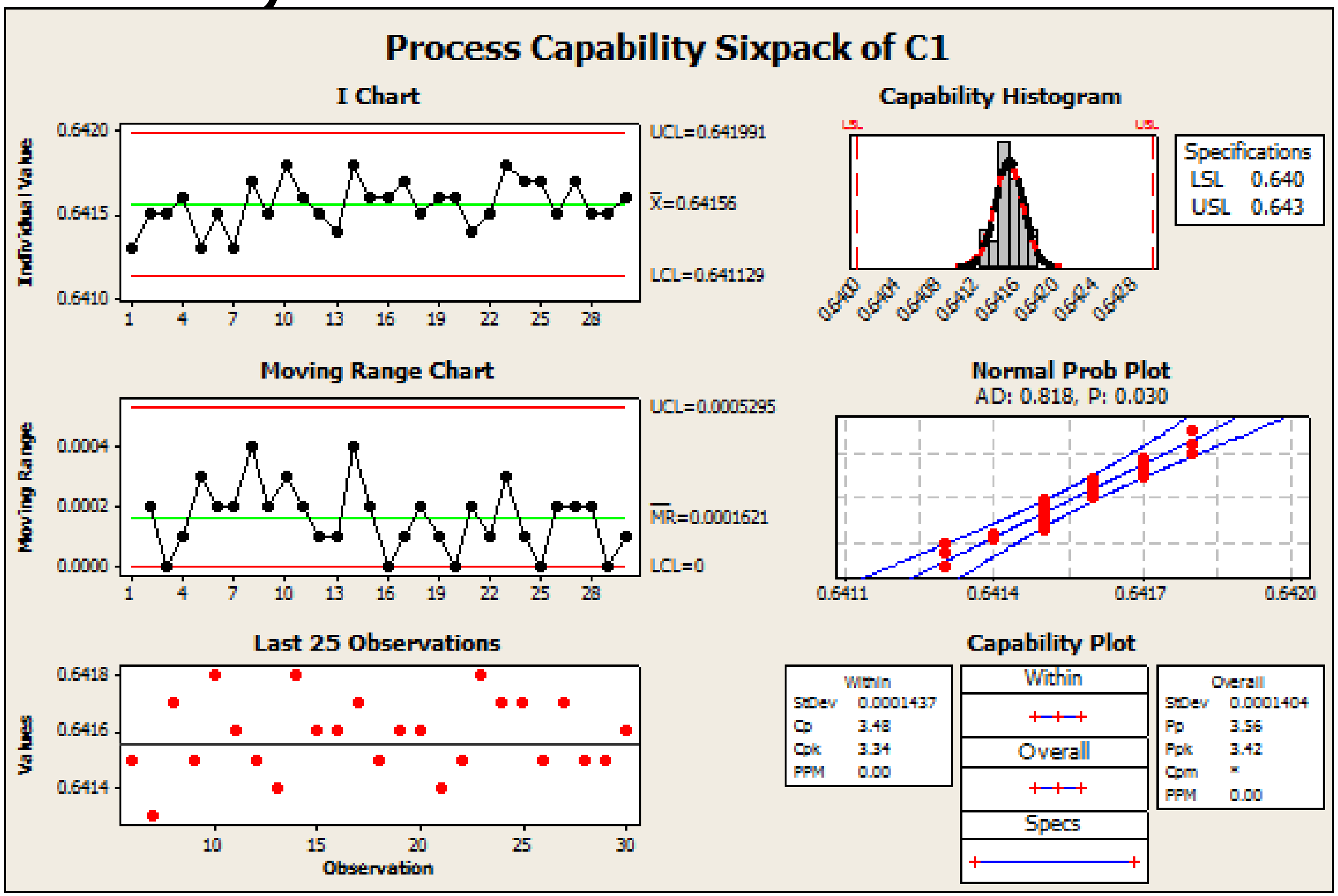


Is the data normal?



Calculate Capability…

Analyze Results



Process Stable?
Yes!

Data set normal?
Yes!

Capability Acceptable?
Yes!

SPC

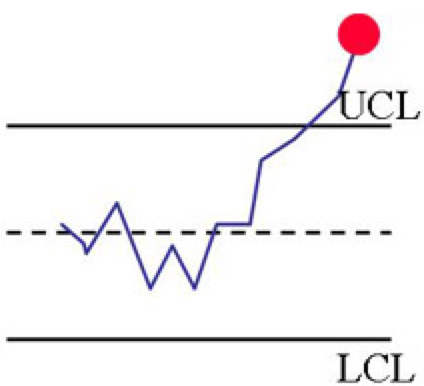
STATISTICAL PROCESS CONTROL

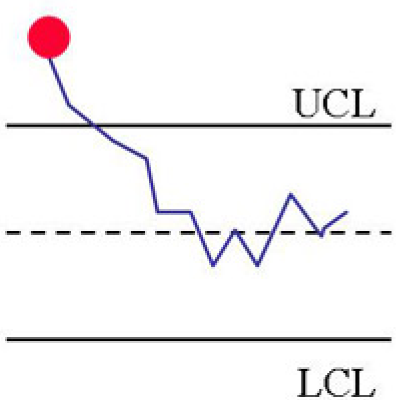
The process talks through the control chart

The organization Listens and reacts

-Root cause
-Learning (Understanding)
-Corrective Action
-Share knowledge

The process responds





Why do we need process control?

Why prevention instead of detection?

Detection tolerates waste, Prevention avoids it.

Variation

There are two types of variation

1. Common Cause

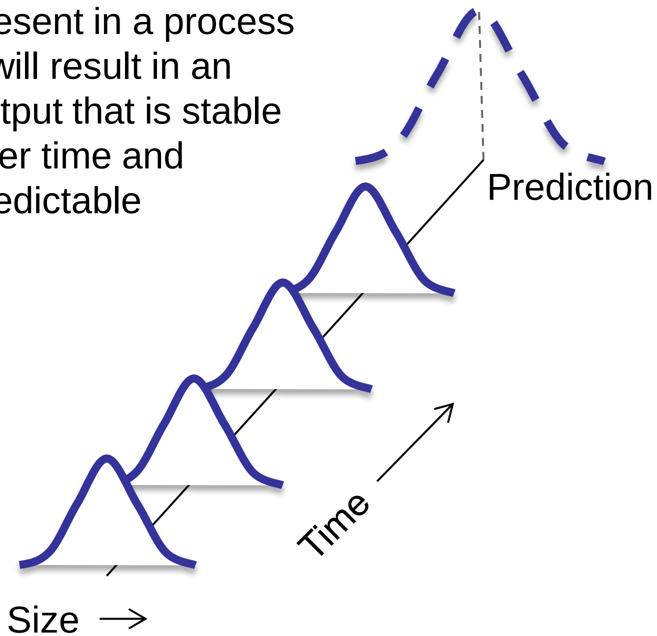
The many factors that result in variation that is consistently acting on a process. If only common causes are present in a process it is considered stable and in control. If only common cause variation is present in a process the process outcome is predictable.

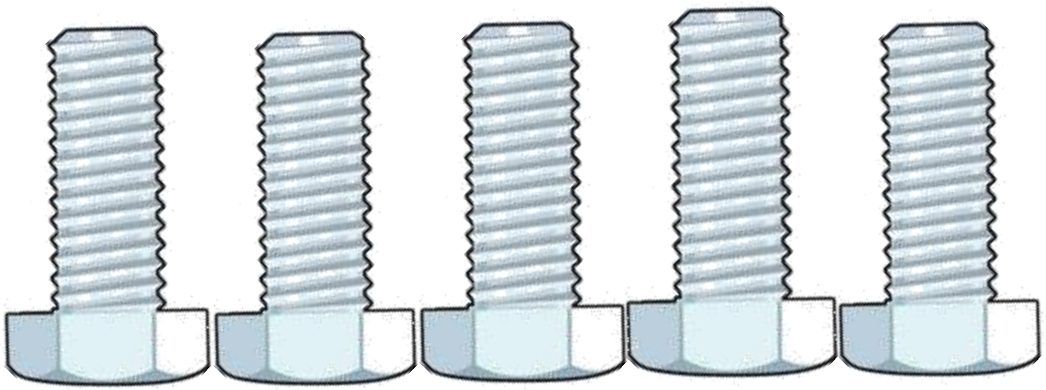
2. Special Cause

Also known as assignable causes, special causes are factors that result in variation that only affect some of the process output. Special cause variation results in one or more points outside of controls limits and / or non-random patterns of points within control limits. Special cause variation is not predictable and, if present in a process, results in instability and out of control conditions.

Common Cause Variation

If only common cause variation is present in a process it will result in an output that is stable over time and predictable





Common cause variation can result in, for example, slight variation in the size of a bolt (exaggerated visual)