AMERICAN MERIDIAN UNIVERSITY

Normality, Stability, Capability

Dr. Bob Gee
Dean Scott Bonney
Professor William G. Journigan
American Meridian University

Learning Objectives

Upon successful completion of this module, the student should be able to:

- Understand tools to verify normality
- Understand tools to verify stability
- Understand the concepts underlying control charts
- Understand process capability terms
- Understand how to assess process capability

Key Questions in the Measure Phase

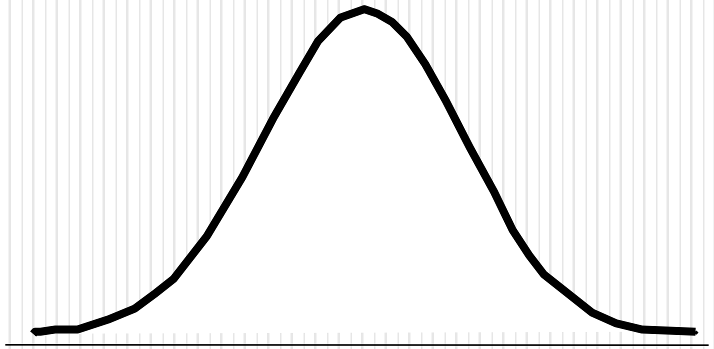
- Does my data exhibit a normal distribution?
- Does my data show that my process is stable?
- Does my data show that my process is capable of meeting the customer's requirements?

Why Is Normality Important?

- Most statistical tests depend on normality.
  - Without bell shaped "tails", tests give skewed results and can be misread.
- Normal data can give projections on what will occur and how often.
- Non-normal data can sometimes be normalized.
  - Everyone's height may not be normal, but the averages of each group (age and gender) are normal.
  - This is an example of the Central Limit Theorem.

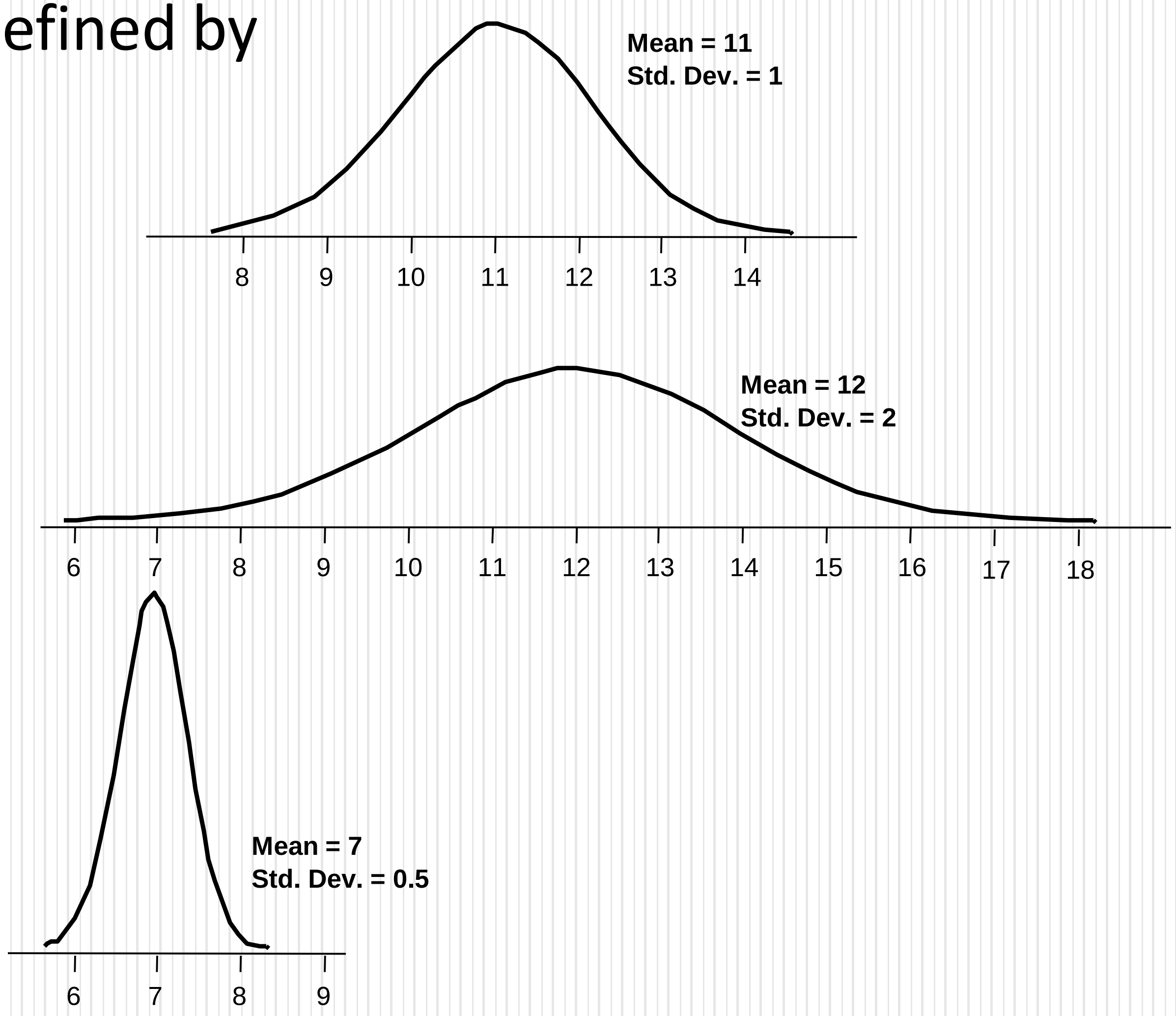
The Normal Curve

- Characteristics
  - Curve theoretically does not reach zero; thus the sum of all finite areas total less than 100%.
  - Curve is symmetric on either side of the most frequently occurring value.
  - The peak of the curve represents the center, or average, of the process.
  - The area under the curve represents virtually 100% of the product the process is capable of producing.



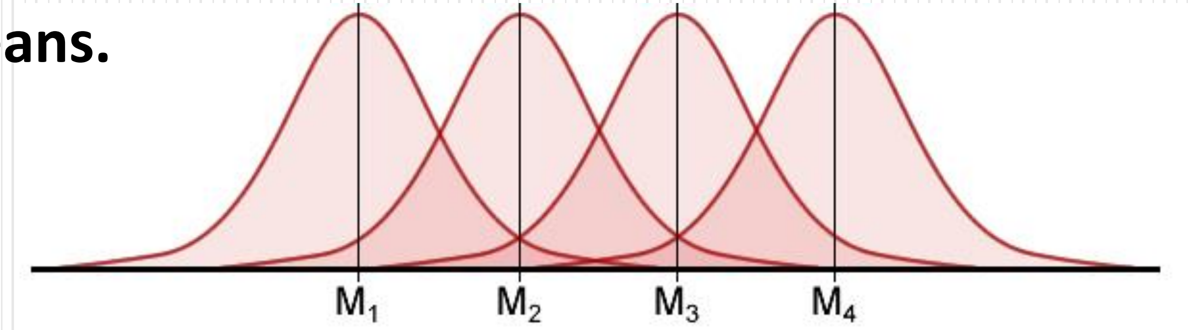
The Normal Curve

- Every Normal Curve is defined by two numbers:
  - Mean
  - Standard Deviation

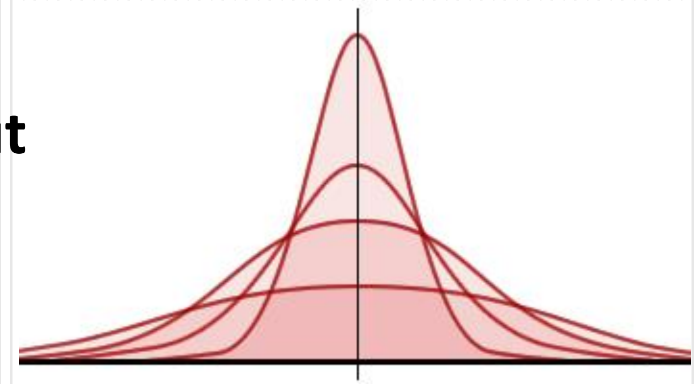


The Normal Distribution

Four Normal distributions that have the same variance but different means.



Four Normal distributions that have the same mean but different variances.

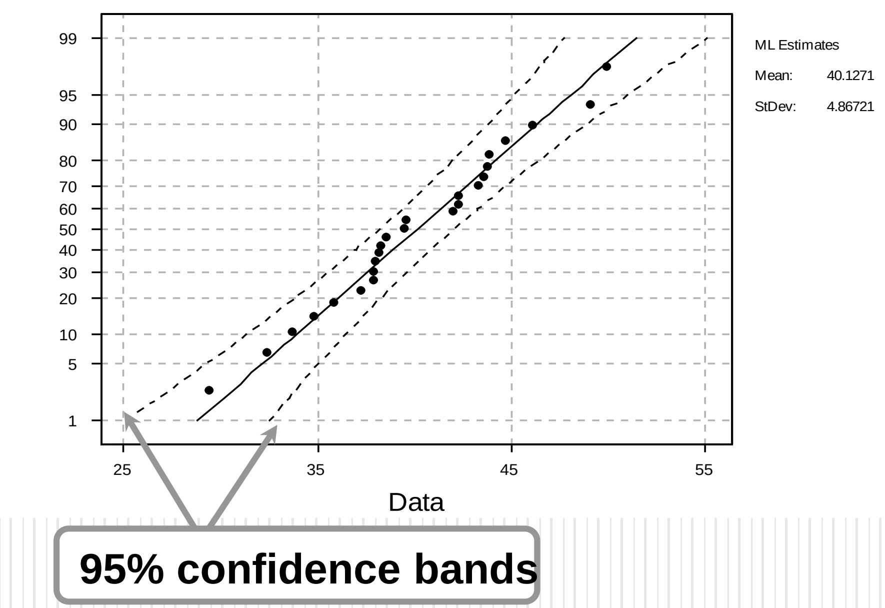


Validating Normality

- A dietician selects a random sample of 13 bottles of cooking oil to determine if the mean percentage of saturated fat is different from the advertised 15%.
- Previous research indicates that the population standard deviation is 2.6%.
- It seems appropriate to use a one-sample Z-test, but the assumption of normality needs to be verified.
- The dietician selects an a-level of 0.10 for the test.

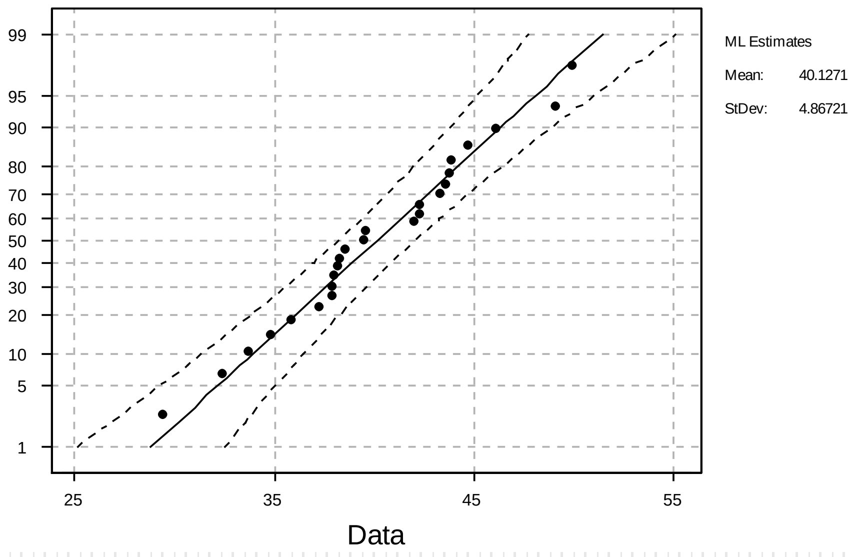
Normal Probability Plot

- Here is a sample Normal probability plot generated in Minitab (n = 25) Graph > Probability Plot
  - If the data are Normal, the points will fall on a "straight" line
  - "Straight" means within the 95% confidence bands
    - You can say the data are Normal if approximately 95% of the data points fall within the confidence bands

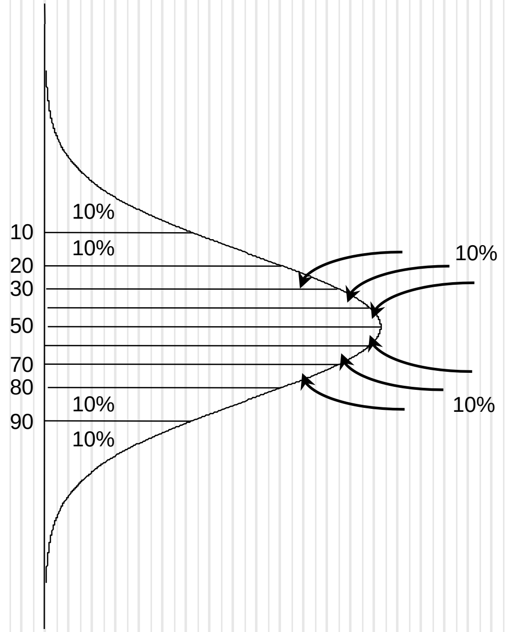


What Is a Normal Probability Plot?

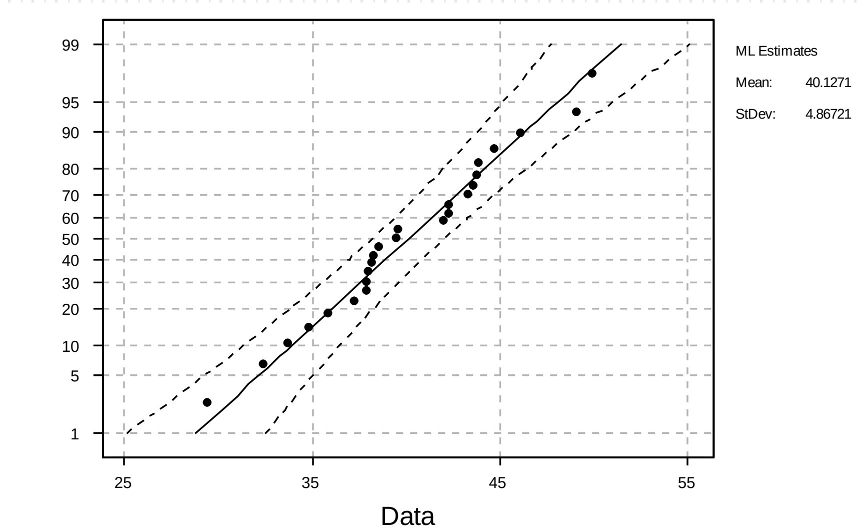
- Normal Probability Plot
  - Data values are on X-axis
  - Percentiles of the Normal distribution are on the Y-axis (unequal spacing of lines is deliberate)

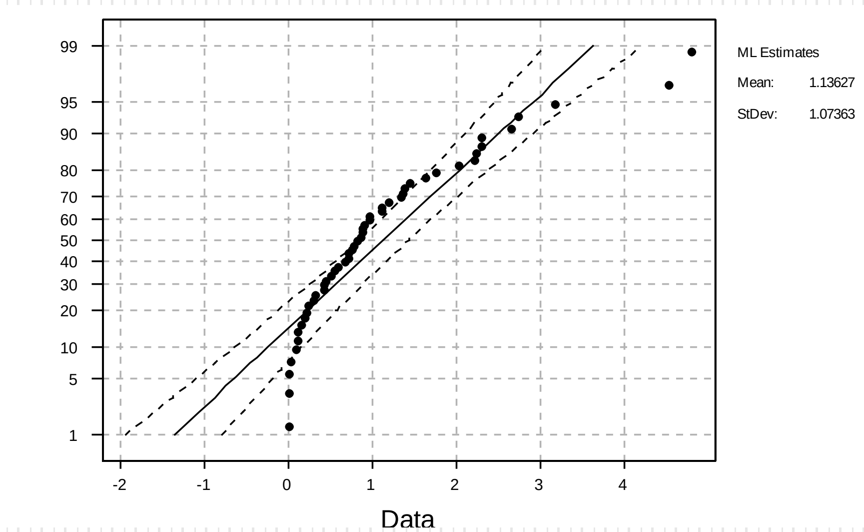


- Equally spaced percentiles divide the Normal curve into equal areas
  - The percentiles match the percent on the vertical axis of the Normal probability plot



Conclusions From Two Normal Probability Plots



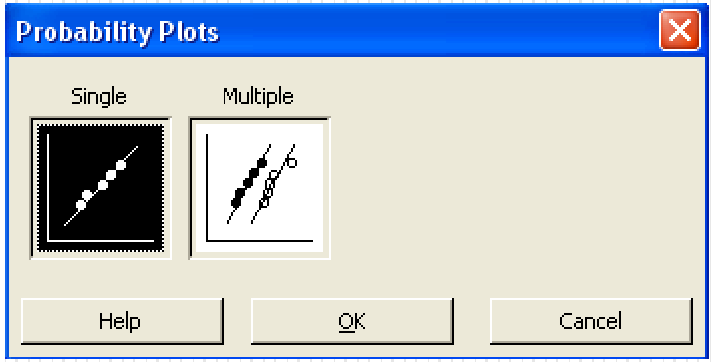


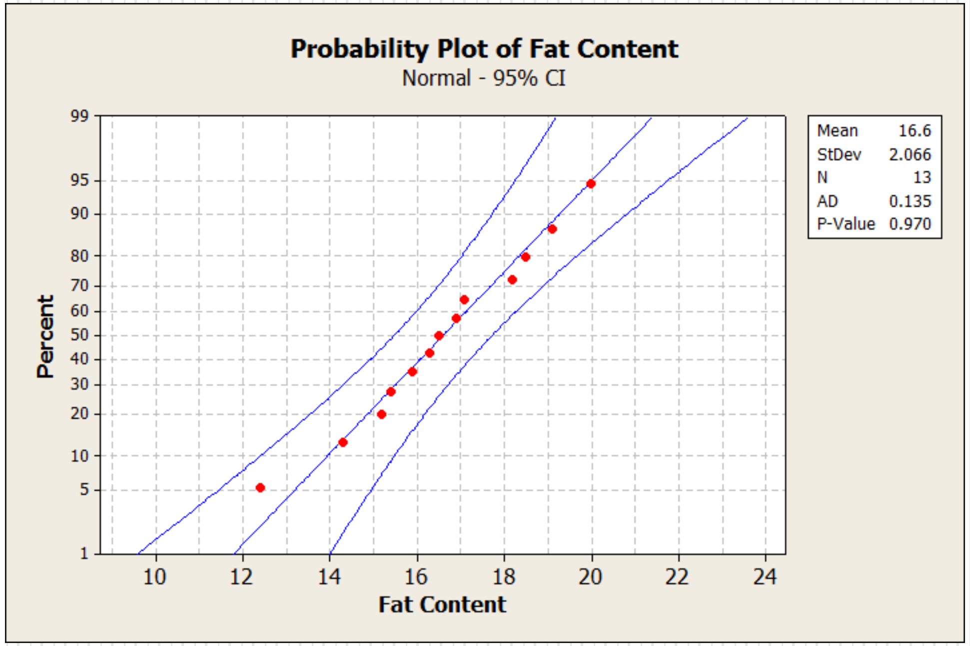
- Not a serious departure from Normality

- There is a serious departure from Normality

Validating Normality in Minitab

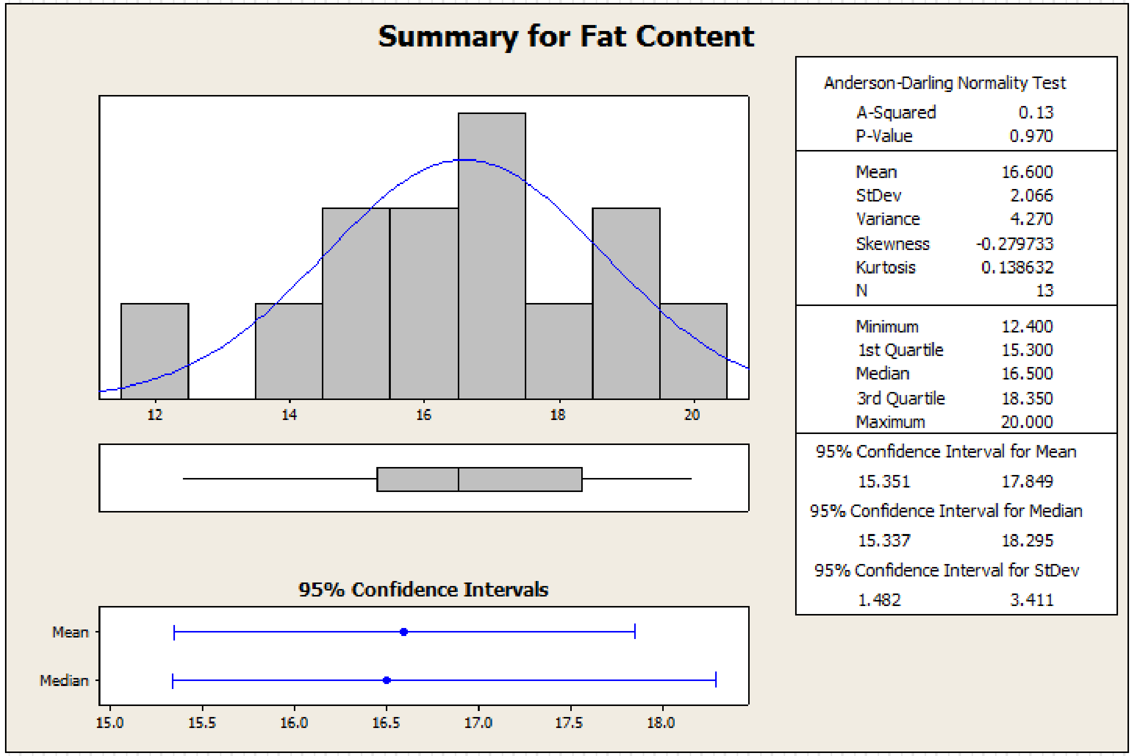
- Open file Fat.mtw
- Select Graph > Probability Plot
- Select Single
- Select OK
- Double click Fat Content
- Select OK

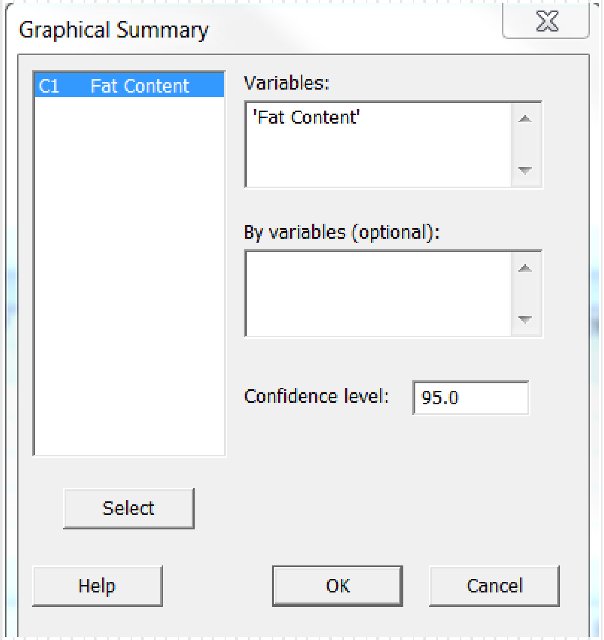




Validating Normality in Minitab

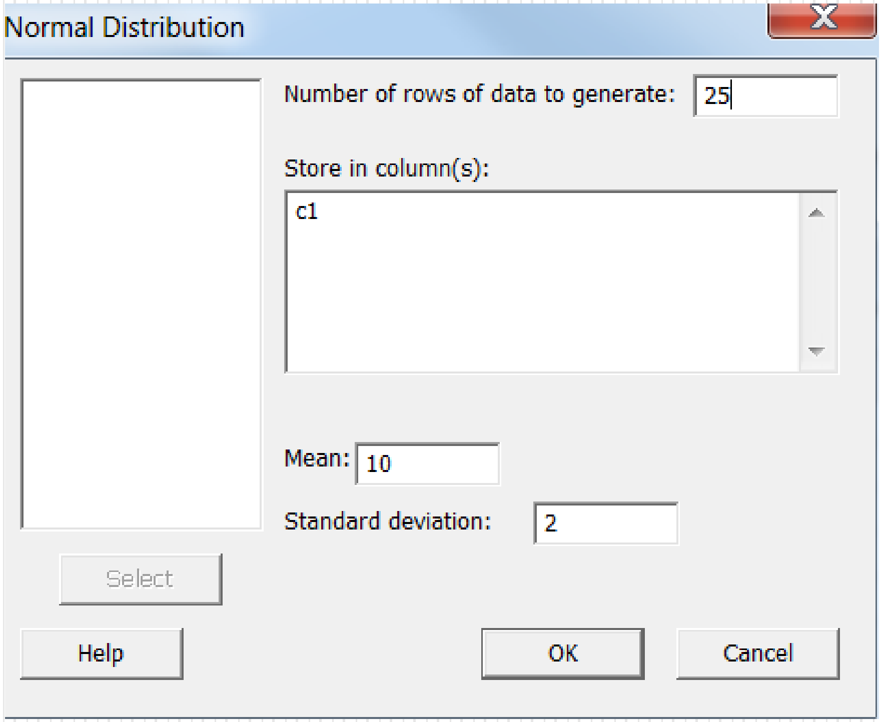
- Stat > Basic Statistics > Graphical Summary
- Enter Fat Content in the Variables Box
- Select OK





Validating Normality in Minitab

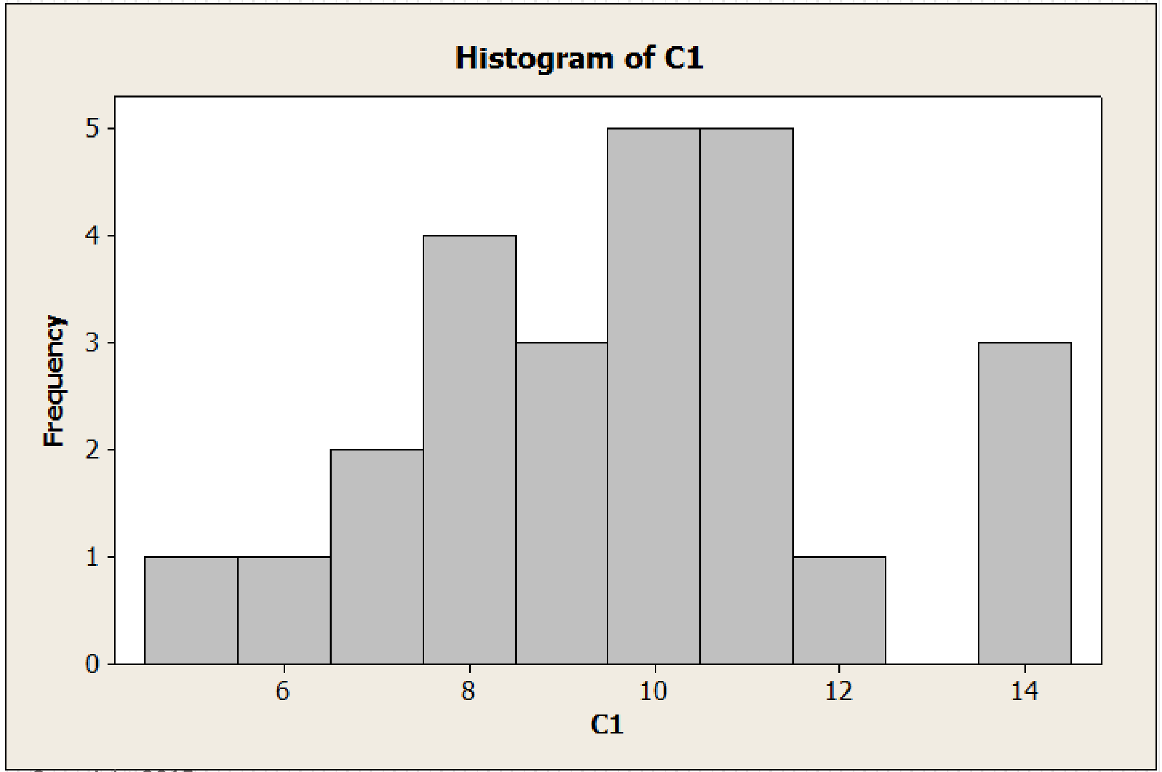
- Open a new (blank) worksheet in Minitab.
- Calc > Random Data > Normal
- Generate a random sample of 25 data points, mean = 10, st. dev. = 2, and store it in C1
- Select OK



Create a Histogram

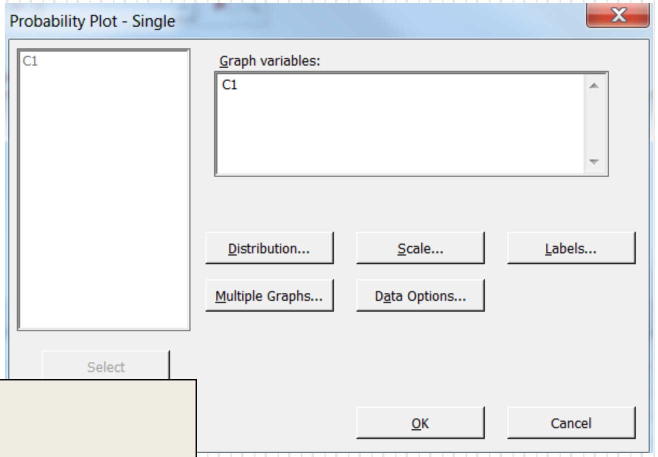
- Create a histogram of the random data you just stored in C1
- Graph > Histogram > Simple > Select OK
- Enter C1 in Graph Variables > Select OK

Yours will be different

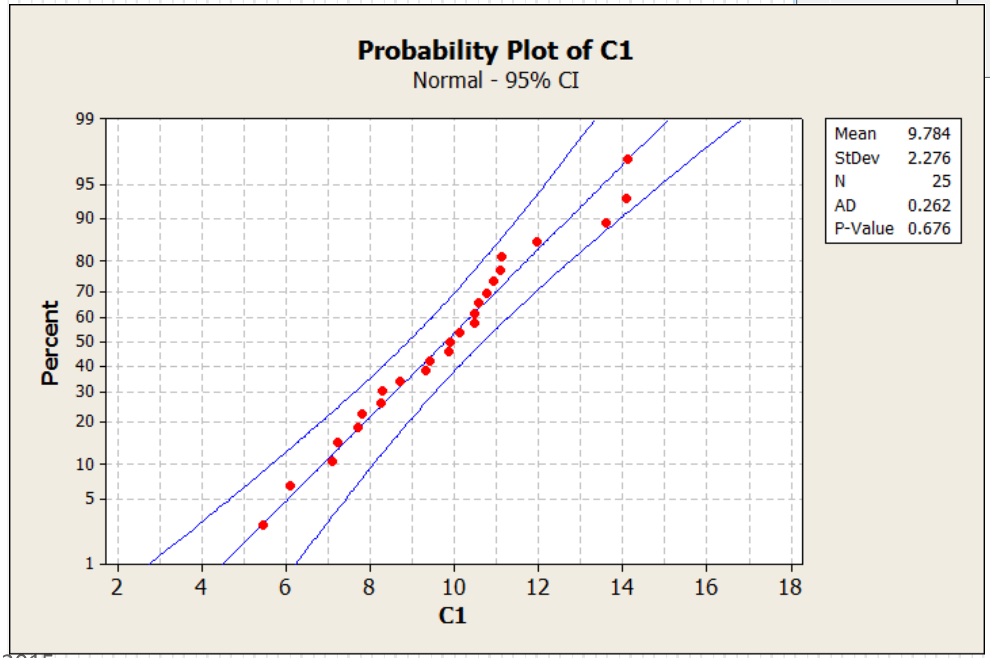


Validating Normality with Minitab

- Select Graph > Probability Plot > Single
- Select OK
- Enter C1 data in the Graph Variables Box
- Select OK



What are your conclusions?



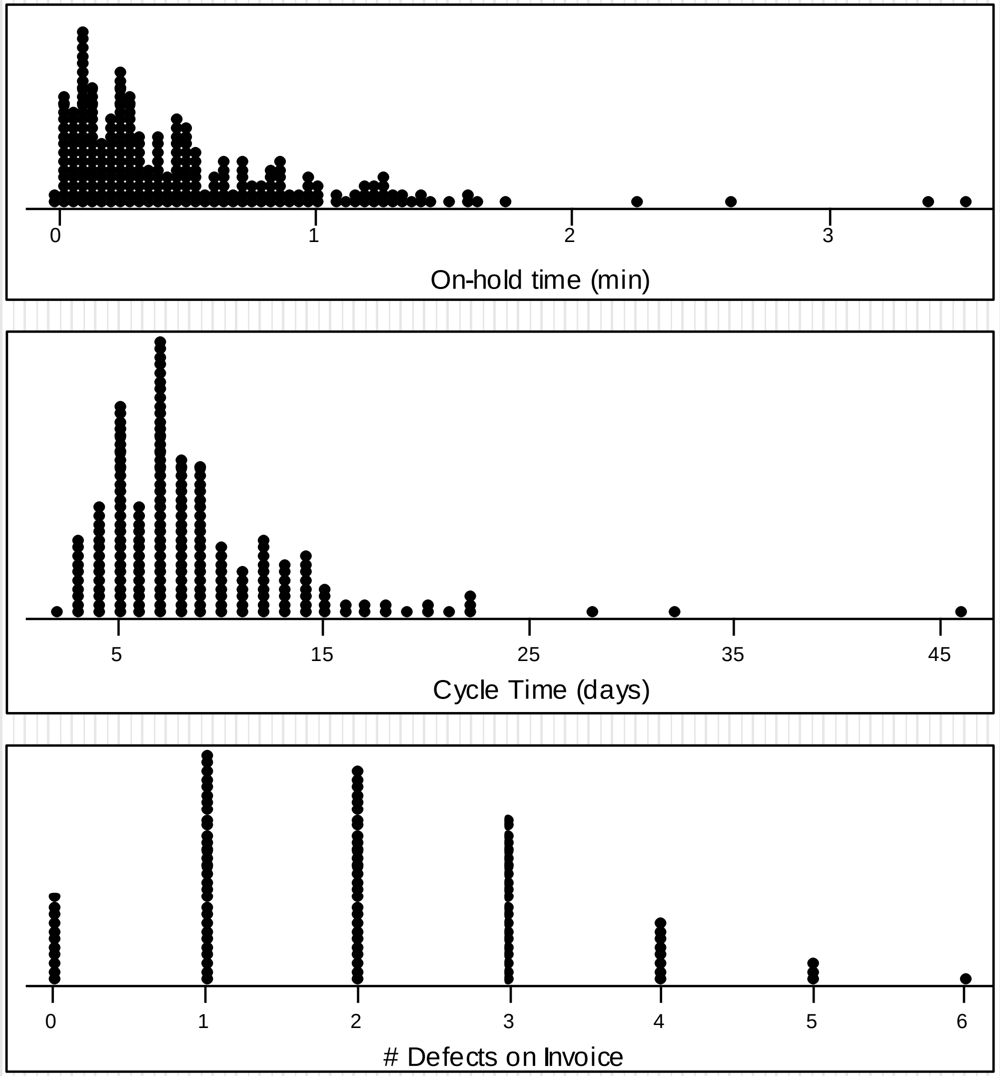
Normality Tests

- The Anderson-Darling Normality test can be used as an indicator of goodness-of-fit.
- It produces a p-value, which is a probability that is compared to the decision criteria, alpha (α) risk.
- Assume $\alpha = 0.05$, meaning there is a 5% risk of rejecting the null when it is true.
- The hypothesis test for this example is:
- Null (H0) = The data is normally distributed
- Alternate (Ha) = The data is not normally distributed

Normality Tests

- If the p-value < α, there is evidence that the data does not follow a normal distribution.
- If the p-value > α, there is evidence that the data is normally distributed.

Not All Data Are Normal



Distributions that have a long tail in only one direction are said to be **skewed**

Understanding Normality: As Easy as 1-2-3!

1) **Check to see if the data is normal**

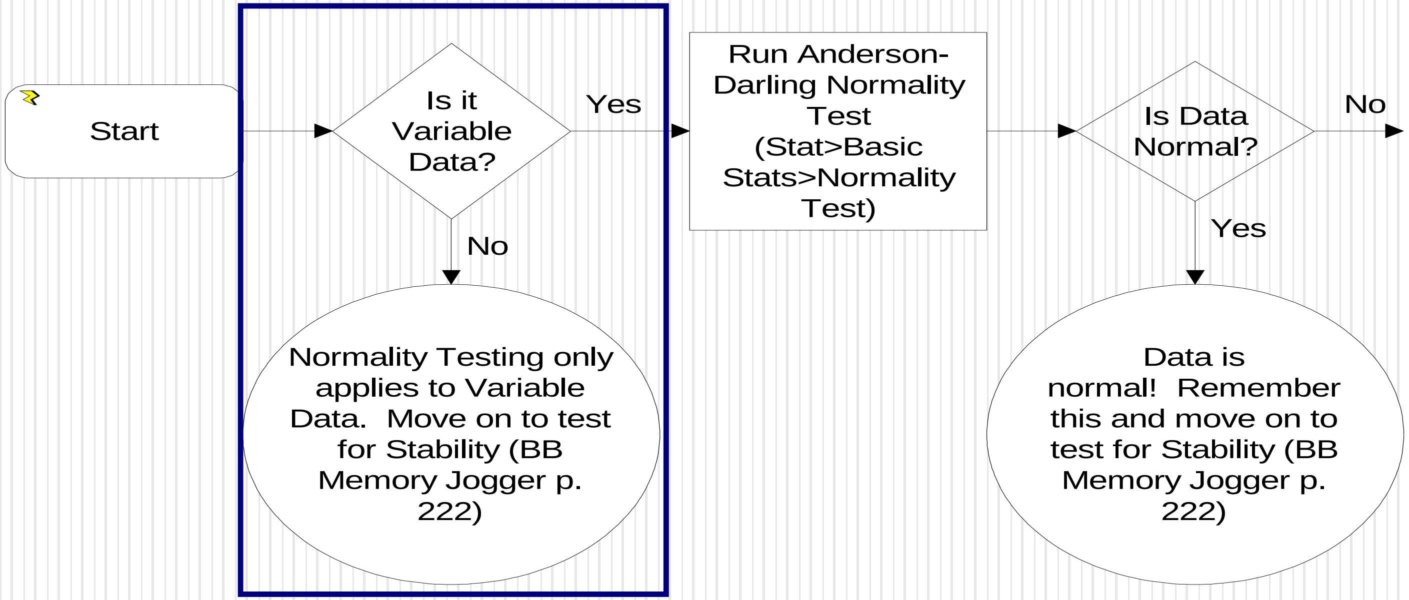
If not…

2) **Check to see if the data can be transformed to become normal**

If not…

3) **Check to see if the data comes from some known non-normal distribution**

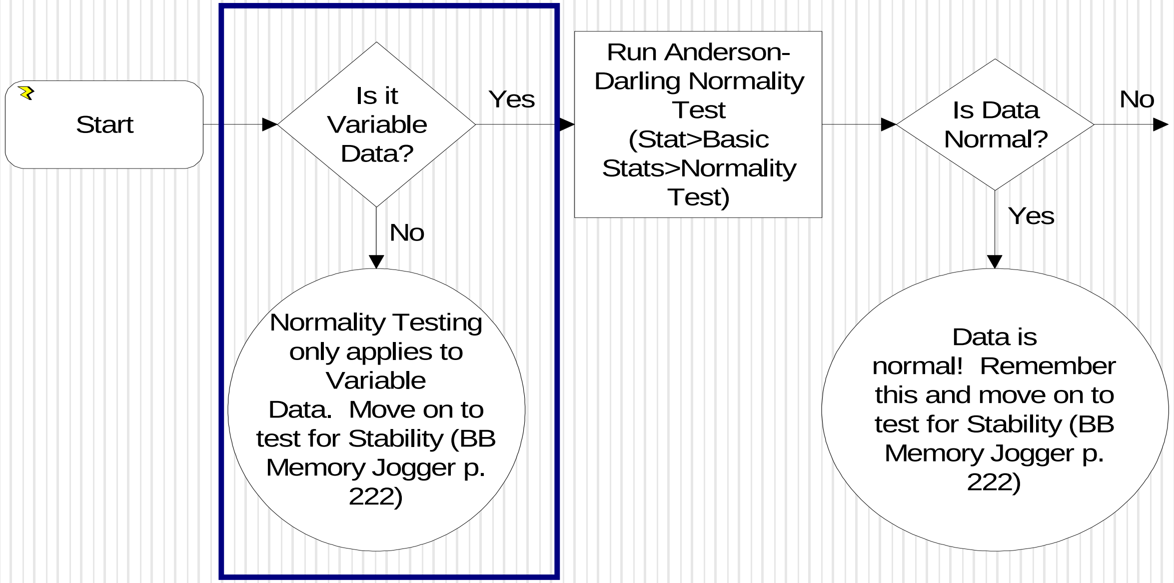
Step 1: Is the Data Normal?



Note that normality is a characteristic of continuous data (variable data).

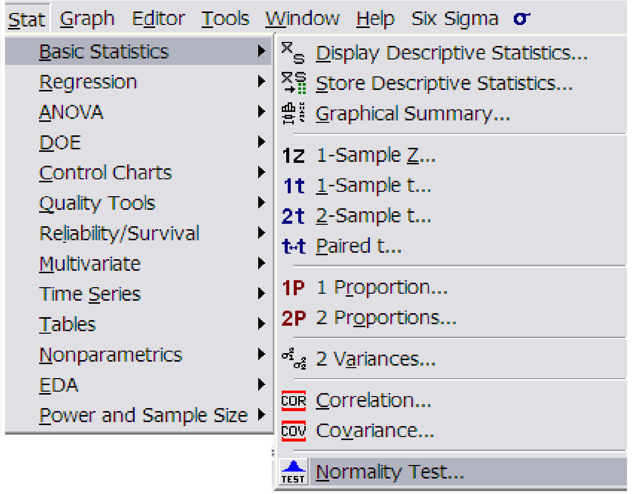
To understand Attribute data, skip Normality and move directly to Stability!

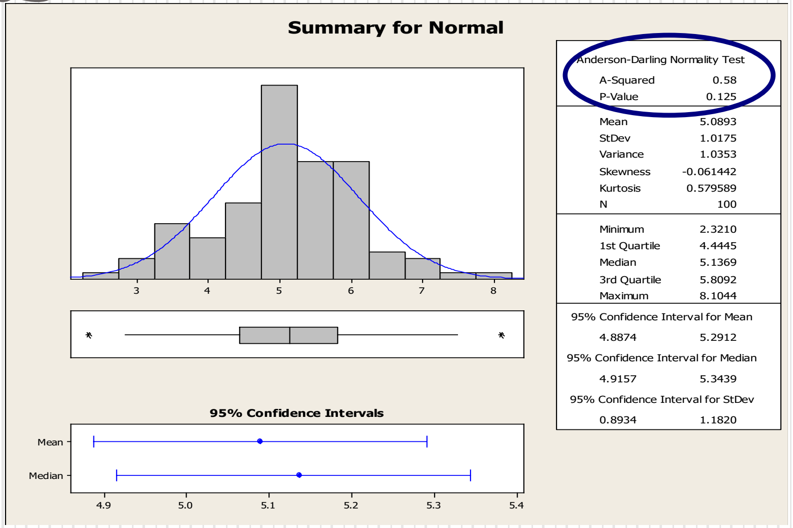
Step 1: Is the Data Normal?

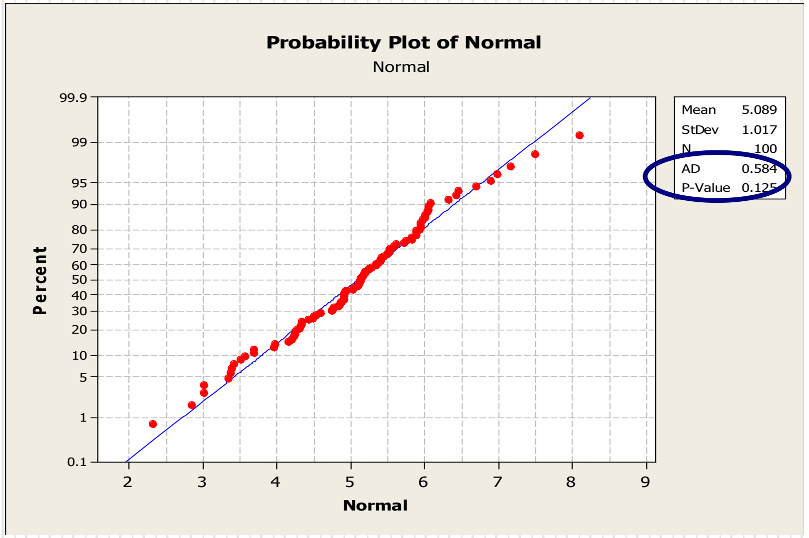


The Anderson-Darling test is the most common test for normal data. We run the test to determine the probability that your data was pulled at random from a normally distributed population.

Example: Normality Tests







Stat >Basic Statistics > Graphical Summary

or

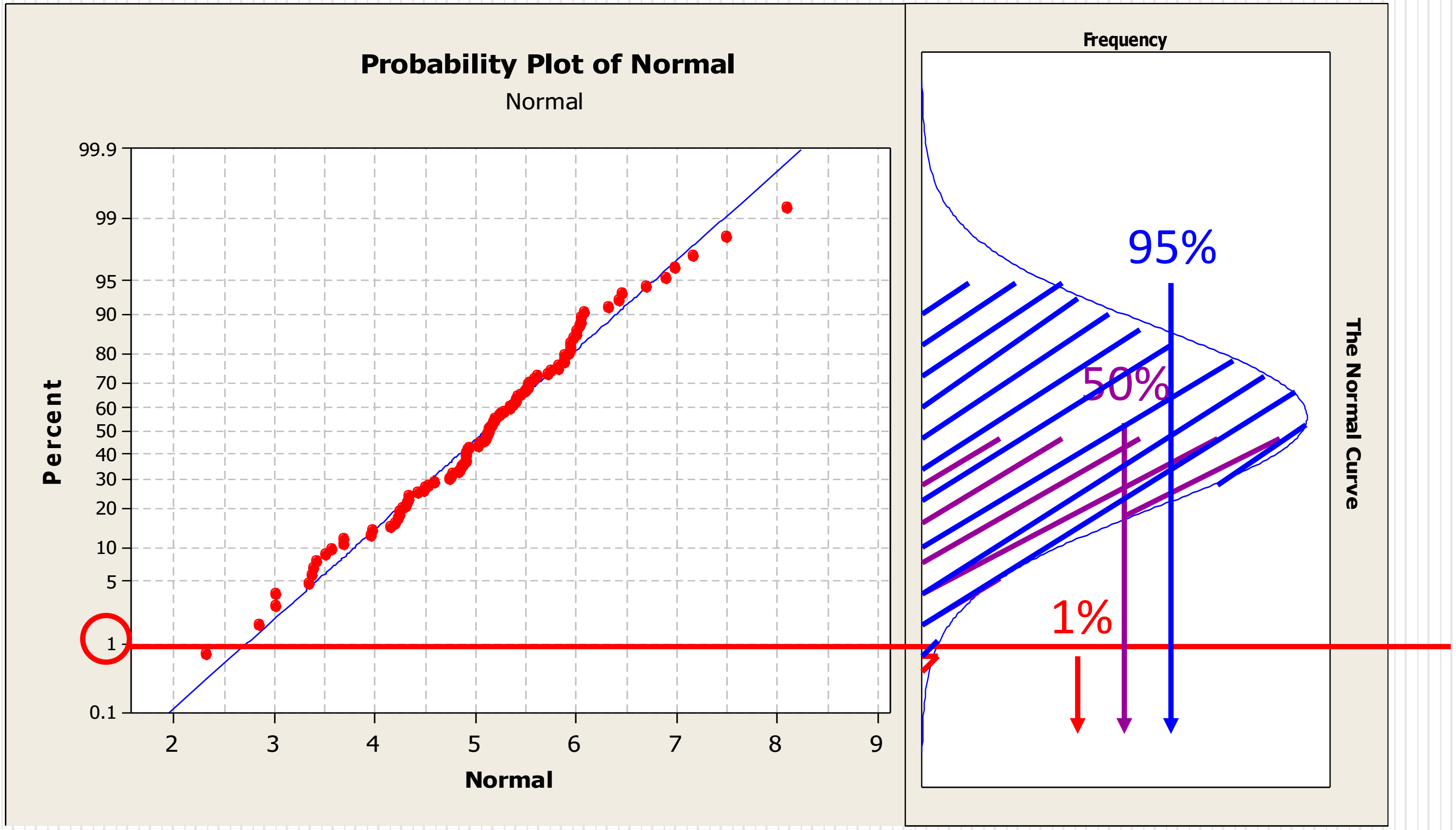
Stat > Basic Statistics > Normality Test

p-value equals the probability that the data was pulled at random from a normal population.

If p>0.05, we assume normality

Interpreting Probability Plots

Probability of finding that value
Read on Y-Axis



Actual measured data value
Read on X-Axis

Stability

Why Measure Stability?

- A process that is out of control is a process with immeasurable variation.
- An out of control process is a process that cannot use statistical tools to improve the process.
- An uncontrollable process has outputs that cannot be predicted.
- An uncontrollable process cannot provide good data for decision making.
- Technically, measuring process variation measures the level of control in a process.

Data in the Measure Phase

- Does my data exhibit a normal distribution?
- Does my data show that my process is stable?
- Does my data show that my process is capable of meeting the customer's requirements?

Types of Variation

- Common Cause
  - Normal variation of a process
  - Over time the occurrence of this type of variation is predictable
  - Variation sources are usually hard to define
  - Described statistically as random or noise
- Special Cause
  - Unusual variation in a process
  - Occurrence of this type of variation is unpredictable
  - Variation sources are easily defined
  - Described statistically as pattern- or trend-based, or a signal

Control Charts

- A control chart is a graphical method for determining if a process is in a state of statistical control.
- The decision is made by comparing control limits with values of some statistical measure calculated from the data.

The Value of Process Control

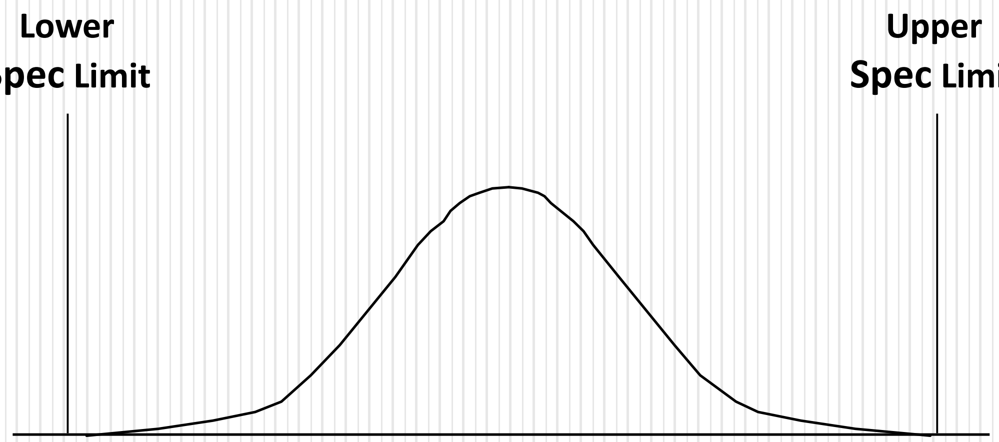
- When a process is in control:
  - You can predict what it will do in the future in terms of its average performance and its variation.
  - You can estimate the capability of the process to meet specifications.
  - It reduces process variation and process cost.

CAUTION!
When a process is not stable, we cannot draw valid conclusions about the process' ability to meet specifications!

Control Limits

- Control limits are bounds on a control chart that serve as a basis for judging if a process is in a state of statistical control.
- These limits are calculated from process data.

Control Limits vs. Specification Limits



Distribution of Individual Data Values from the process

Specification Limits **are set by the Customer**

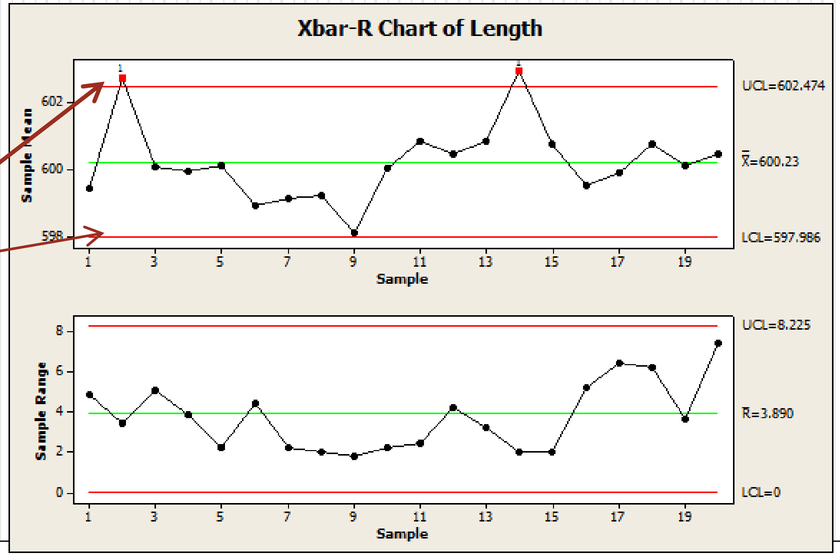
Vs.

Control Limits

are set by the Process

Distribution of Sample Averages (n=3) plotted on X/R control chart

(narrower distribution than individual values due to Central Limit Theorem)



Control Chart Formulas

- Control limits for X-bar and R charts are based on formulas that estimate the standard deviation (s) of the sample averages (X-bar) using the average of the sample ranges (R-bar).

Control Chart Constants

Formulas:

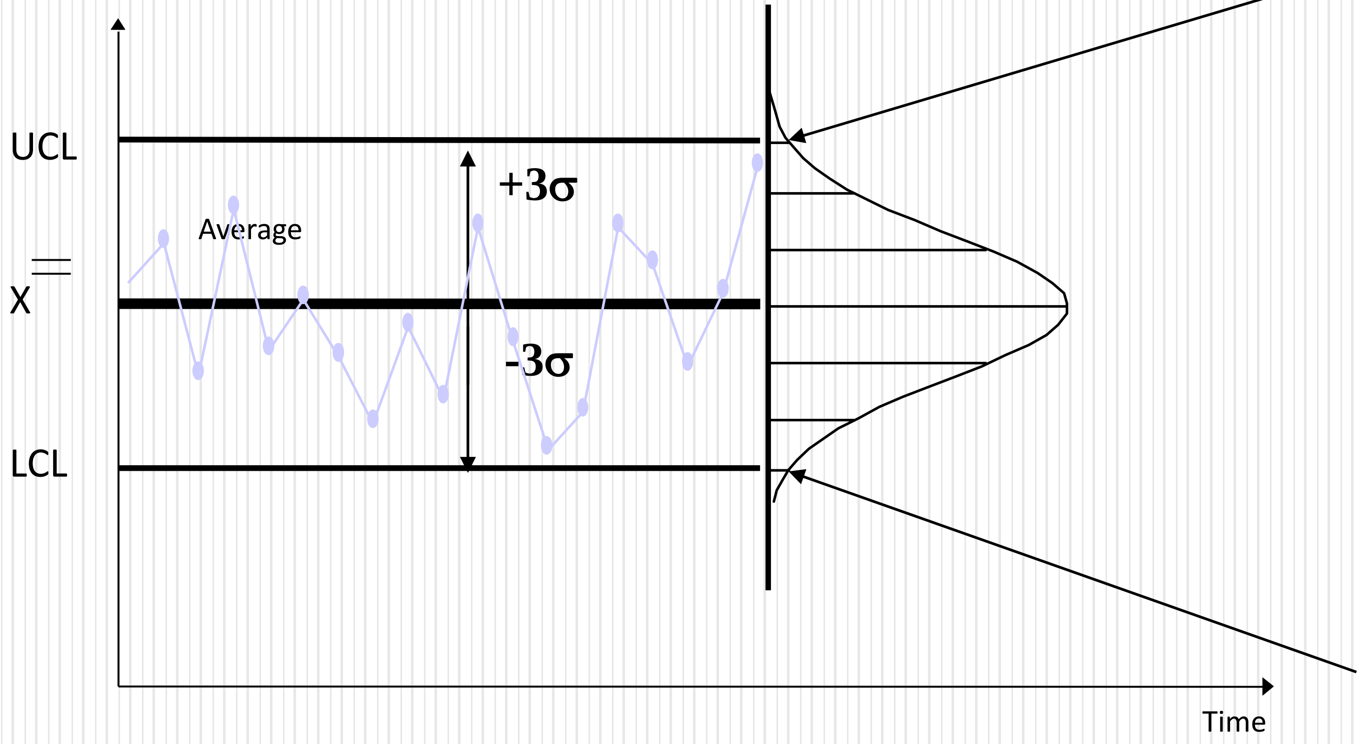
$$UCL_{\bar{X}} = \bar{\bar{X}} + A_2\bar{R}$$
$$LCL_{\bar{X}} = \bar{\bar{X}} - A_2\bar{R}$$
$$UCL_R = D_4\bar{R}$$
$$LCL_R = D_3\bar{R}$$

$$\hat{\sigma} = \left(\frac{\bar{R}}{d_2}\right)$$ **(If the range chart is in control)**

Sample Size	A_2	D_3	D_4	d_2
2	1.880	-	3.267	1.128
3	1.023	-	2.574	1.693
4	0.729	-	2.282	2.059
5	0.577	-	2.114	2.326
6	0.483	-	2.004	2.534
7	0.419	0.076	1.924	2.704
8	0.373	0.136	1.864	2.847
9	0.337	0.184	1.816	2.970
10	0.308	0.223	1.777	3.078

Transitioning to Control Charts

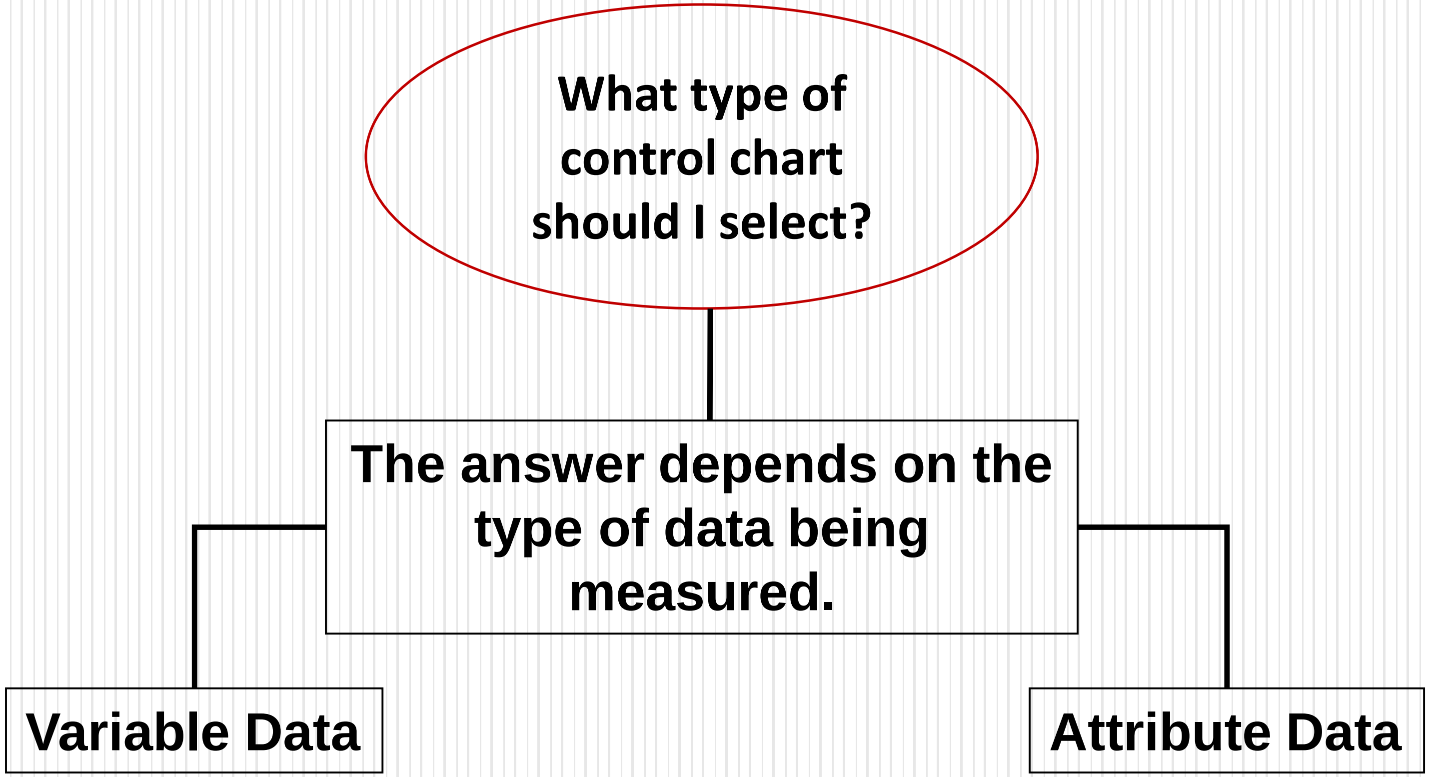
- The normal curve is found within the limits of the control chart.
- Control limits are the boundaries of where we expect the process to operate in the future.



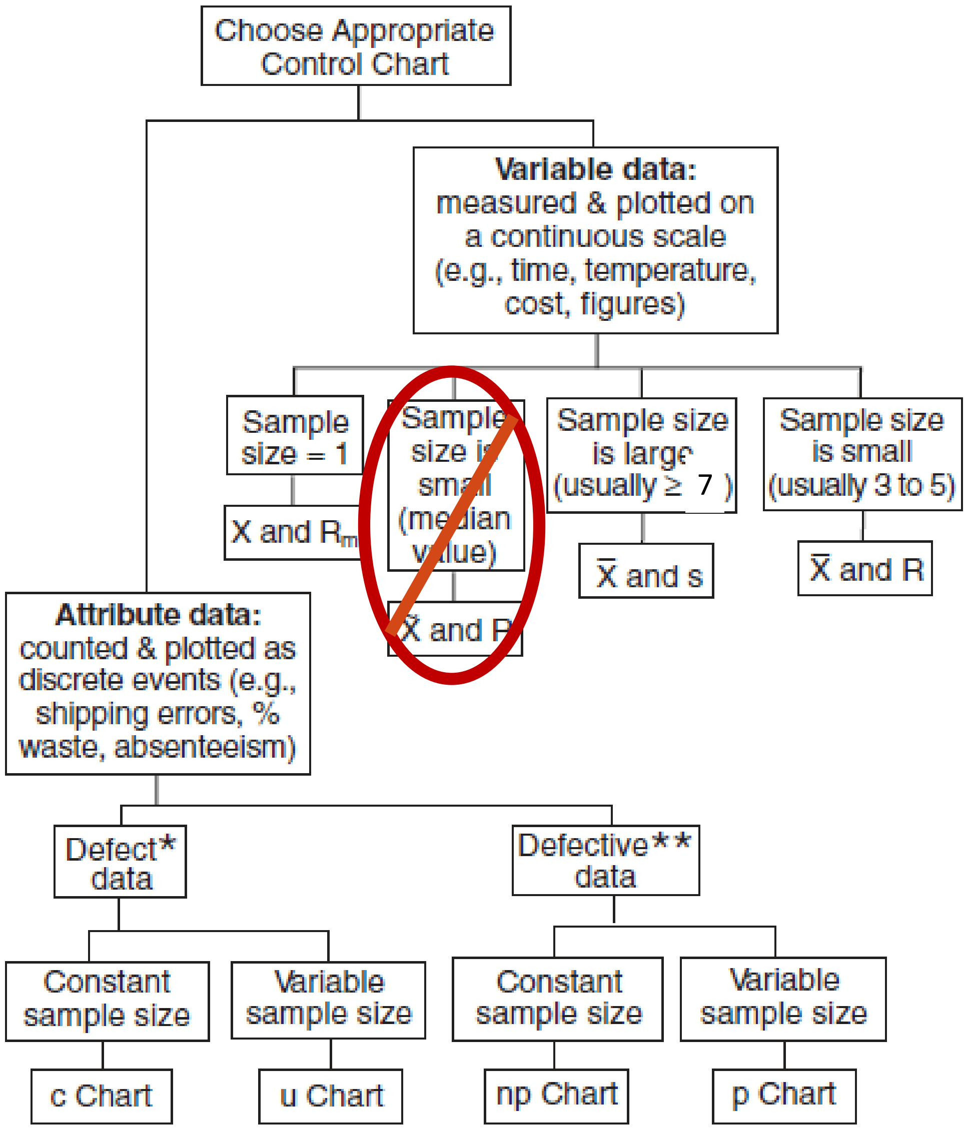
Control limits are determined using a mathematical approximation of the standard deviation of the data plotted.

Here, *averages* are plotted, so we use the standard deviation of the averages.

Selecting the Correct Control Chart



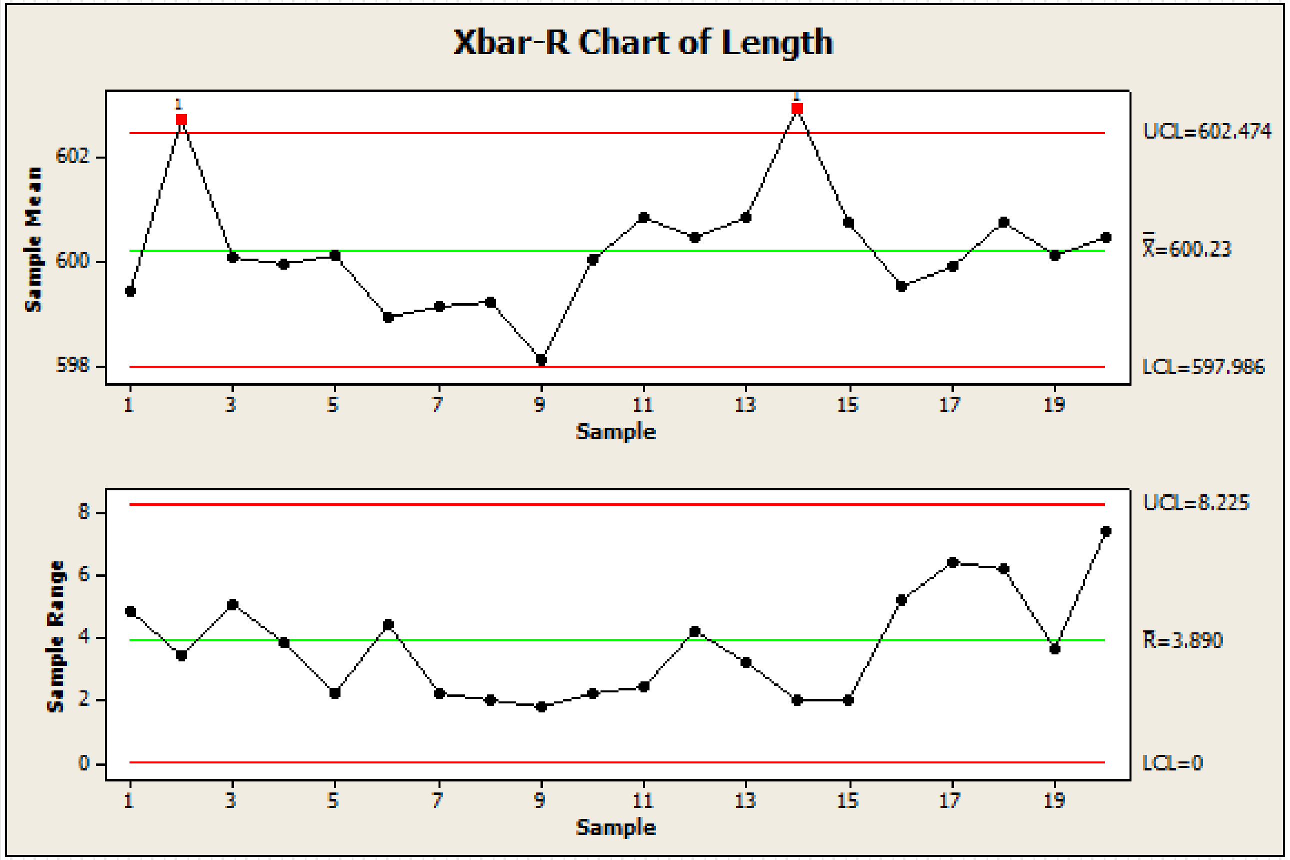
Variables Control Charts



Common Chart Types

- X-bar and R charts for variable data use the average to monitor the center of the distribution and the range to monitor the spread.

Synonyms:
Average
"X-bar"
$\overline{X}$



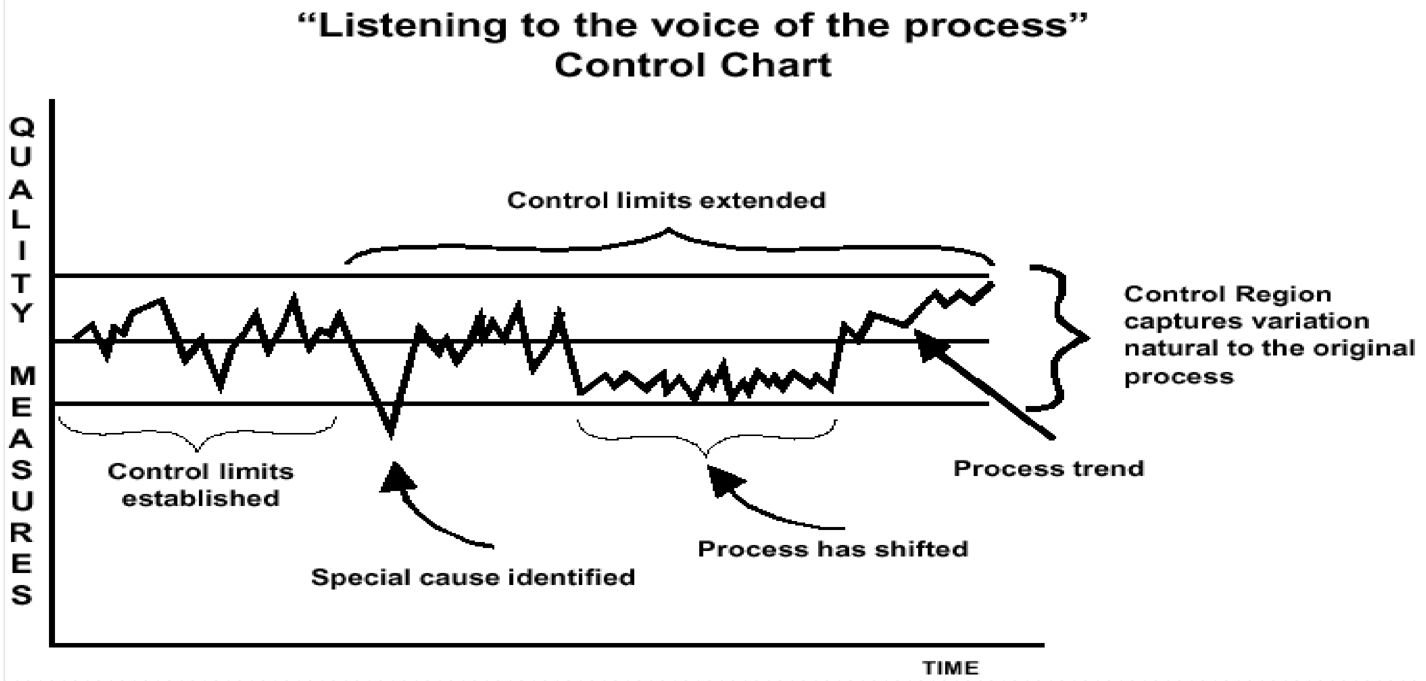
Interpreting Control Charts

- Several guides exist:
  - One point plots outside control limits (outlier)
  - Two out of three consecutive points plot on the same side of the centerline in zone A or beyond (shift)
  - Four out of five consecutive points plot on the same side of the centerline in zone B or beyond (shift)
  - Nine consecutive points plot on one side of the centerline (shift)
  - Six consecutive points increasing or decreasing (trend)
  - Fourteen consecutive points that alternate up and down
  - Fifteen consecutive points within Zone C (above and below the average.
- Refer to p. 230-231 in your BB Memory Jogger, or…
- Use Minitab to help identify patterns.

Causes for Patterns

Outliers

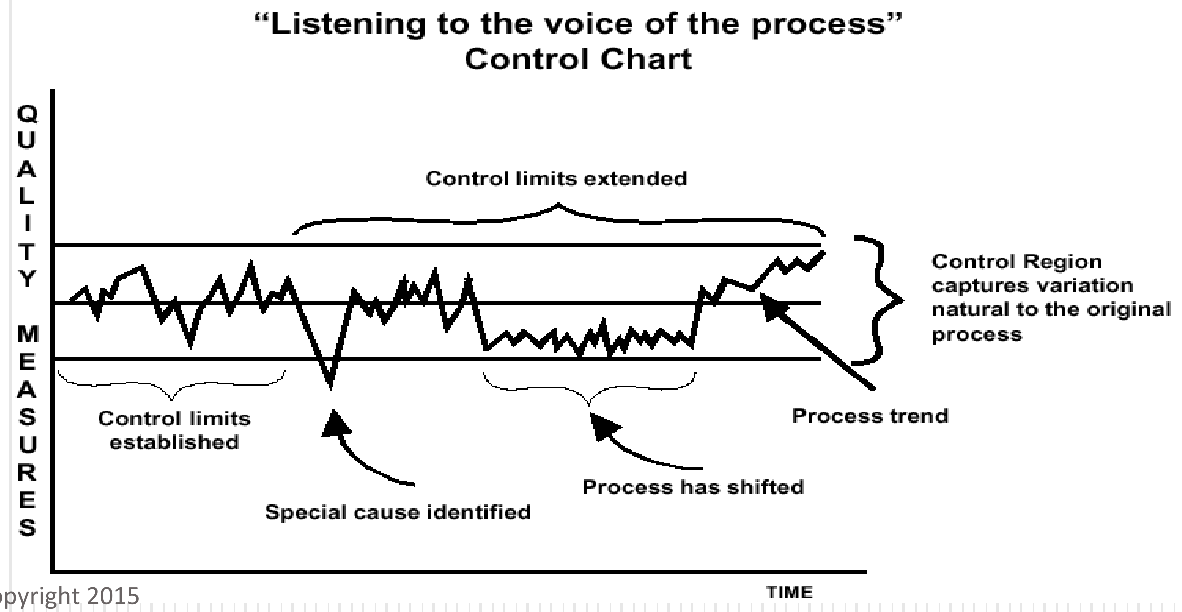
- Incorrect process settings, error in measurement, sub-grouping or plotting, incomplete operation, machine and tool breakdowns, power surge.



Causes for Patterns

Shifts

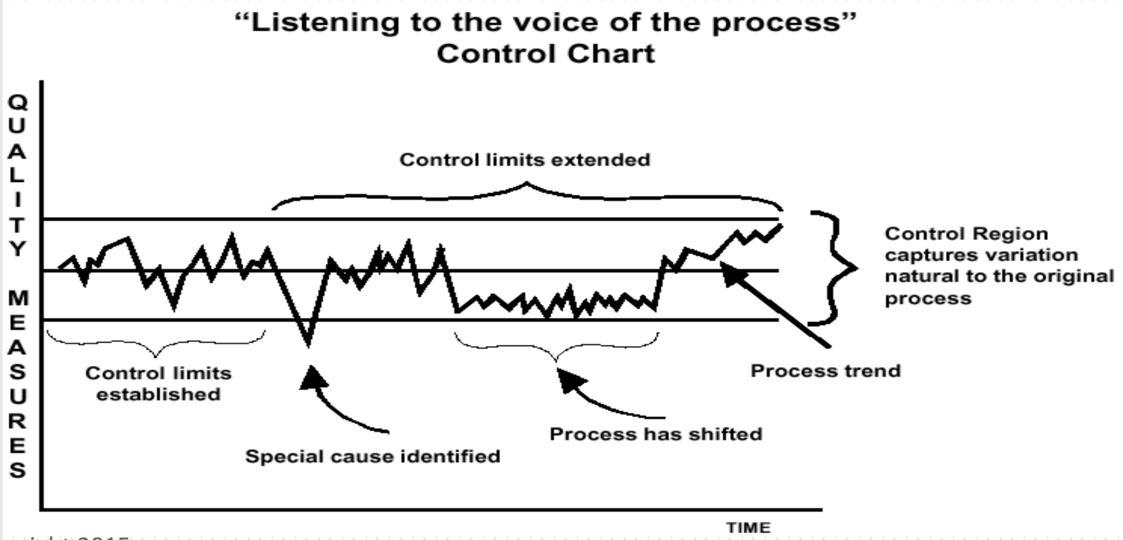
- Introduction of a new material, machine, operator, inspector or test set, new process controls, maintenance, process changes, change in proportion of materials from different sources.



Causes for Patterns

Trends

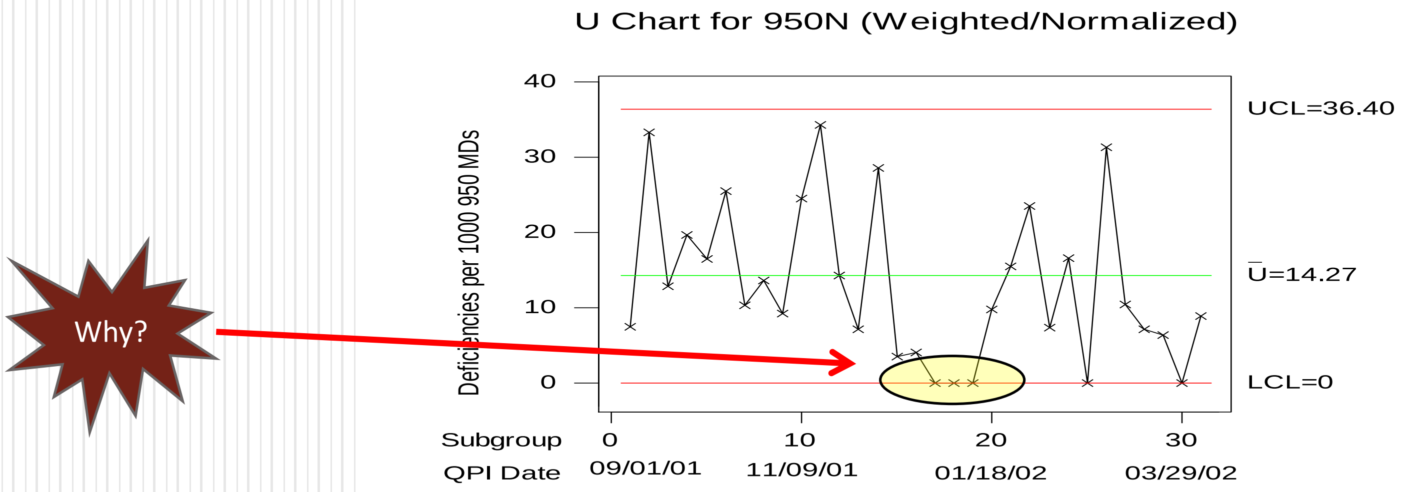
- Tool or fixture wear, deterioration of materials, aging, changes in maintenance or calibration, environmental factors, human factors, production schedules, gradual changes in materials or process, accumulation of waste products, or machine warm-up.



Causes for Patterns

Cycles

- Does the time period of the cycle suggest a cause?
- Environmental factors, worn locations on tools or fixtures, human factors, gage changes, voltage fluctuations, shift changes, systematic rotation of equipment or materials, merging of subassemblies or processes.



Causes for Patterns

Stratification

- Could each subgroup be a mixture of data from more than one source?
- Non-random sampling, miscalculation, incorrect chart type, non-rational sub-grouping, reduction in process variability, or changes in inspection process.

Causes for Patterns

Mixtures

- Could subgroups be coming from two sources alternately?
- Two or more different materials, operators, designs, or testers mixed in the process, over-adjustments of the process, poor sampling procedures, or control of two or more processes on the same chart.

Control Chart Dangers

- "In control" does not necessarily mean "capable"
- It takes time to identify and resolve problems
- Control charts signal problems but they still don't indicate the reasons for the problems
- Data gathering and setting limits is not easy; use of past data that "happens to be available" may not be good enough

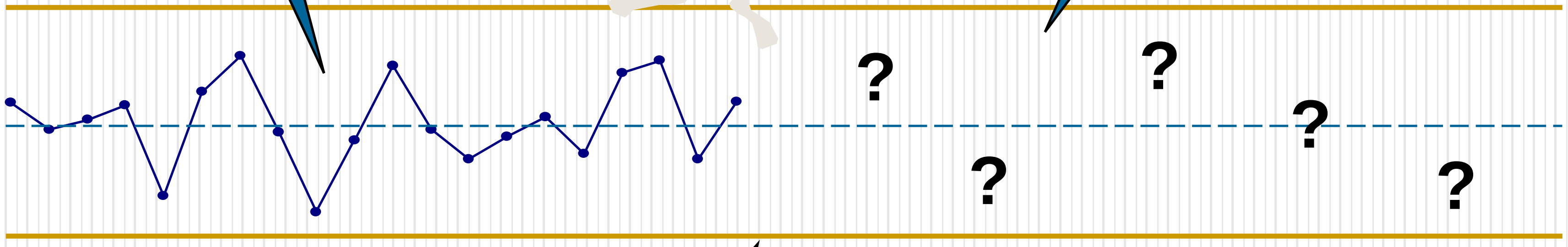
Control Chart Advantages

- Systematic and efficient method for turning data into actionable information
- Lets people make decisions from FACTS
- Highlights special cause impacts to a process
- Provides warning of degradation before making defect products / services
- Establishes controls for continuous improvement and shows evidence of improvements
- Involves everyone and builds worker knowledge of the process

A Process Is In Control When...

1 "Through the use of past experience . . .

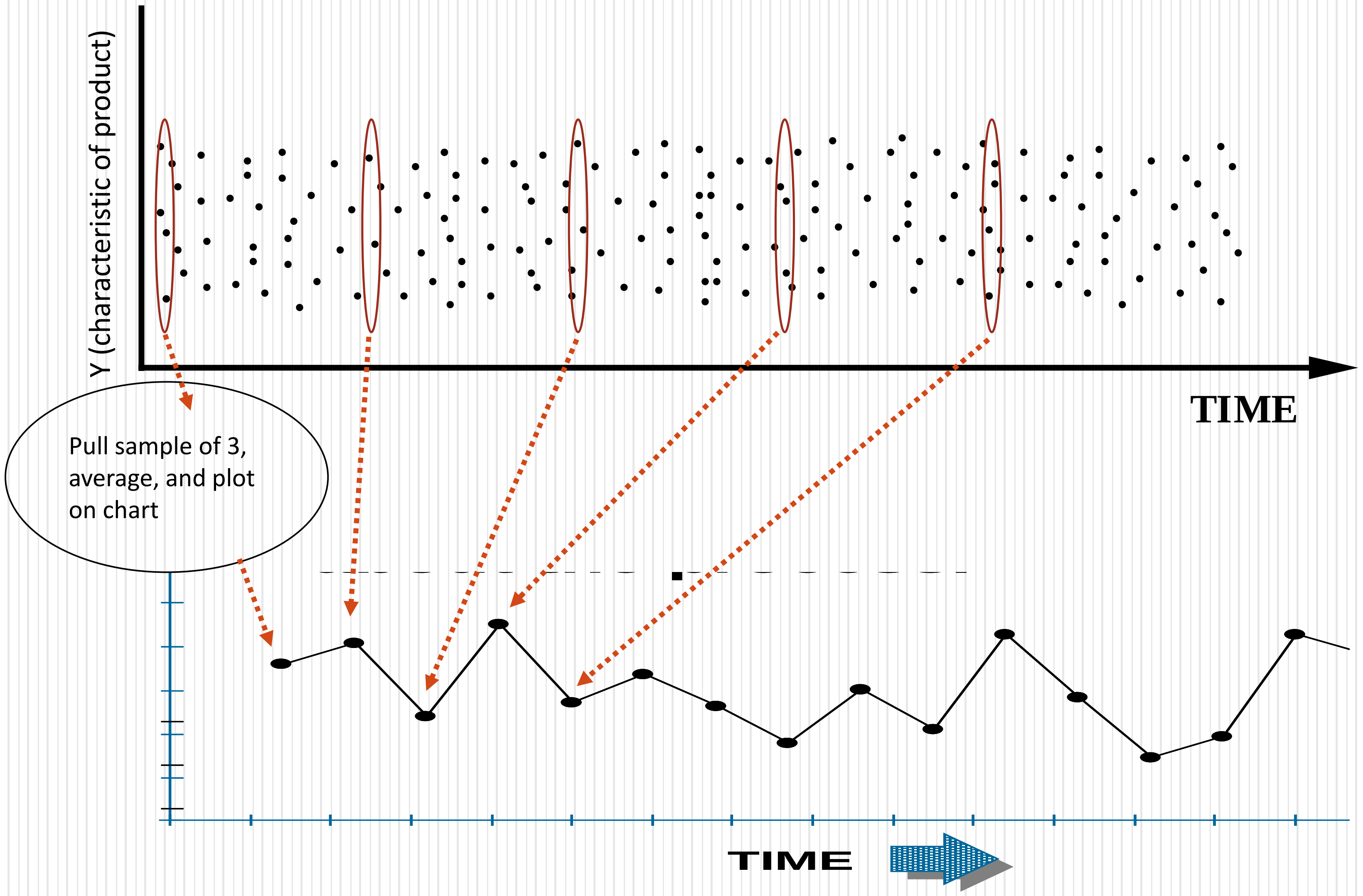
3 "How the phenomenon may be expected to vary in the future."



2 "We can predict, at least within limits . . .

Walter A. Shewhart
Economic Control of Quality of Manufactured Product
Published in 1931

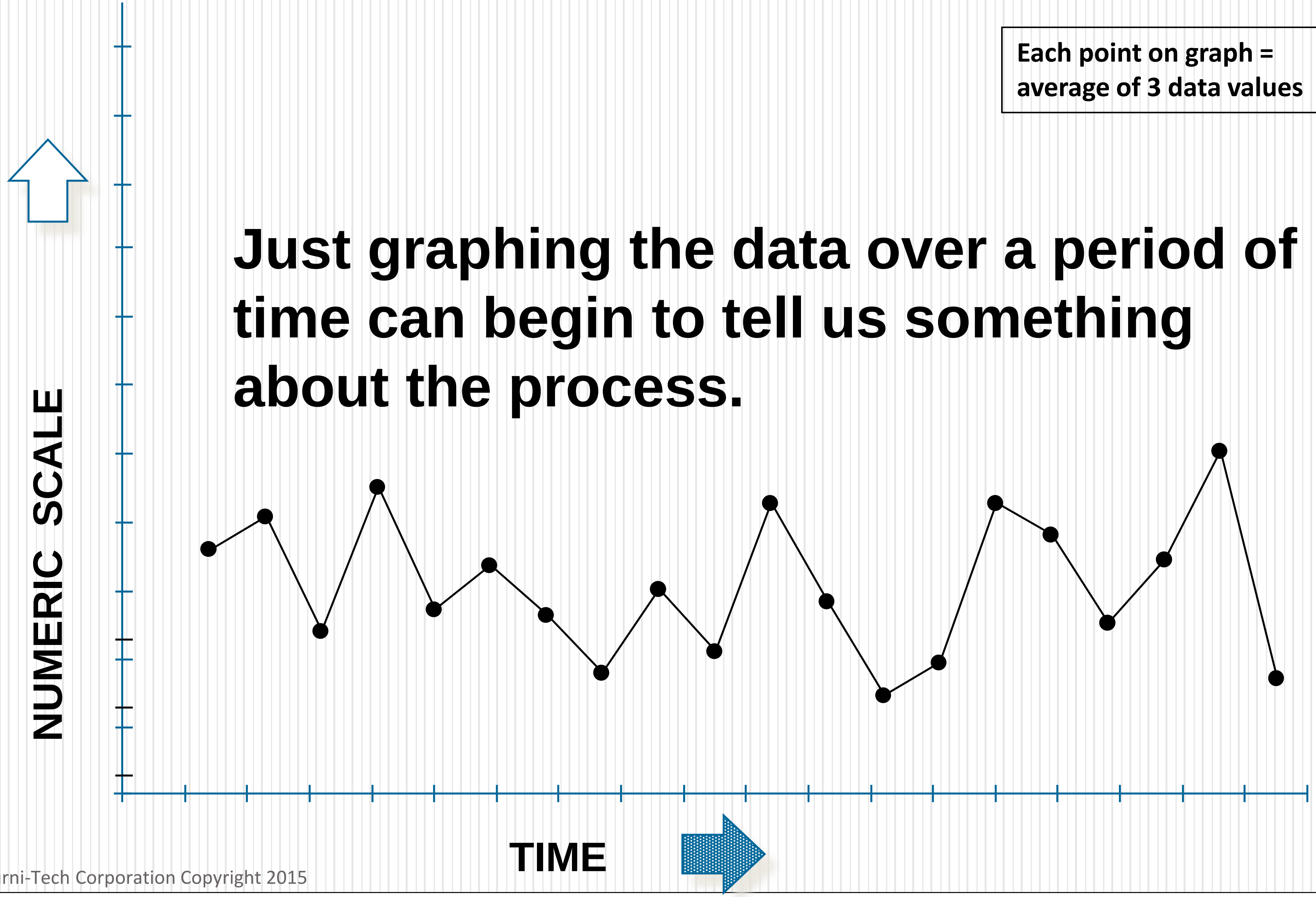
Pull Samples From The Process Output



Plot the Data in Time Order

Each point on graph = average of 3 data values

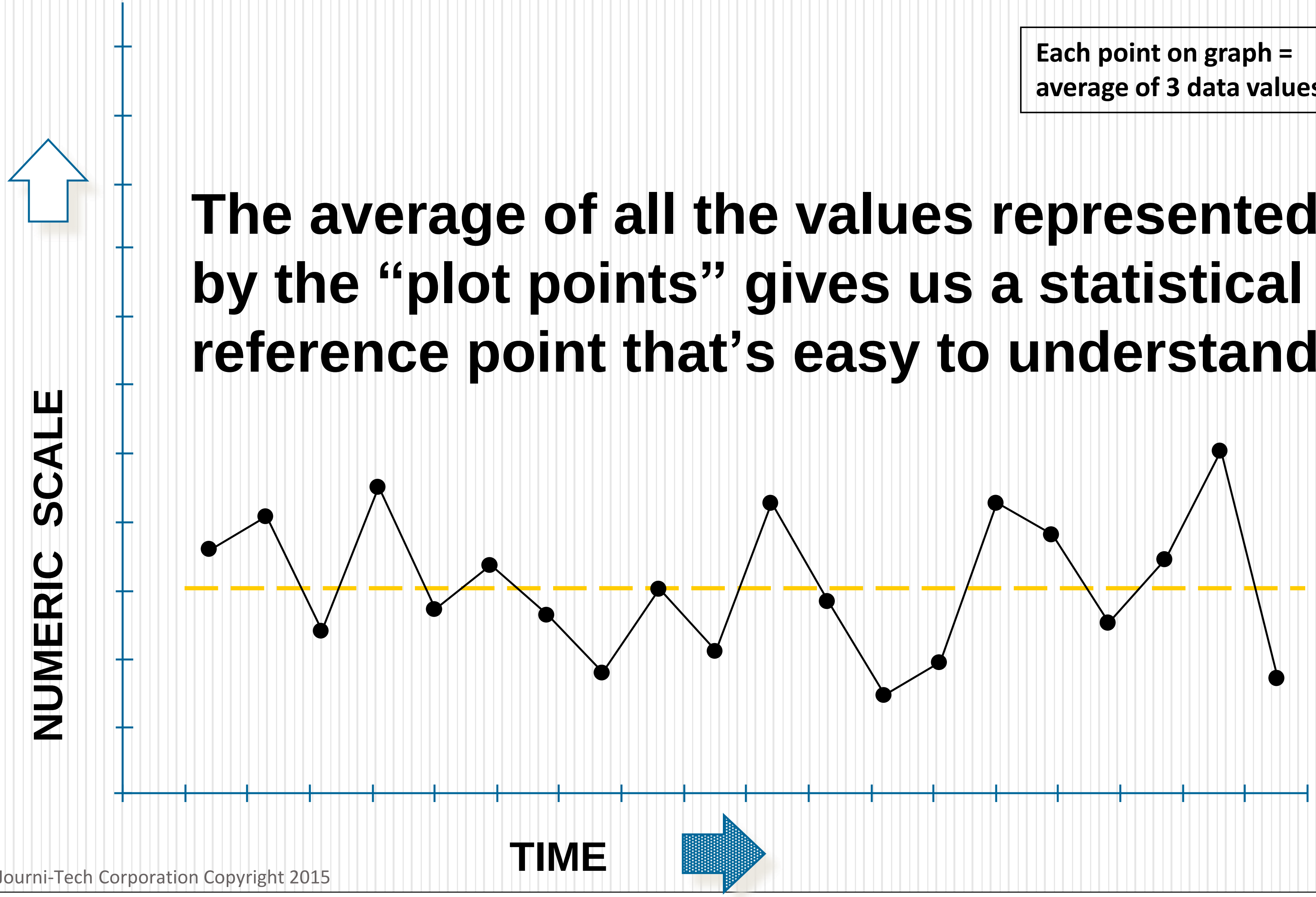
Just graphing the data over a period of time can begin to tell us something about the process.



Draw the Centerline

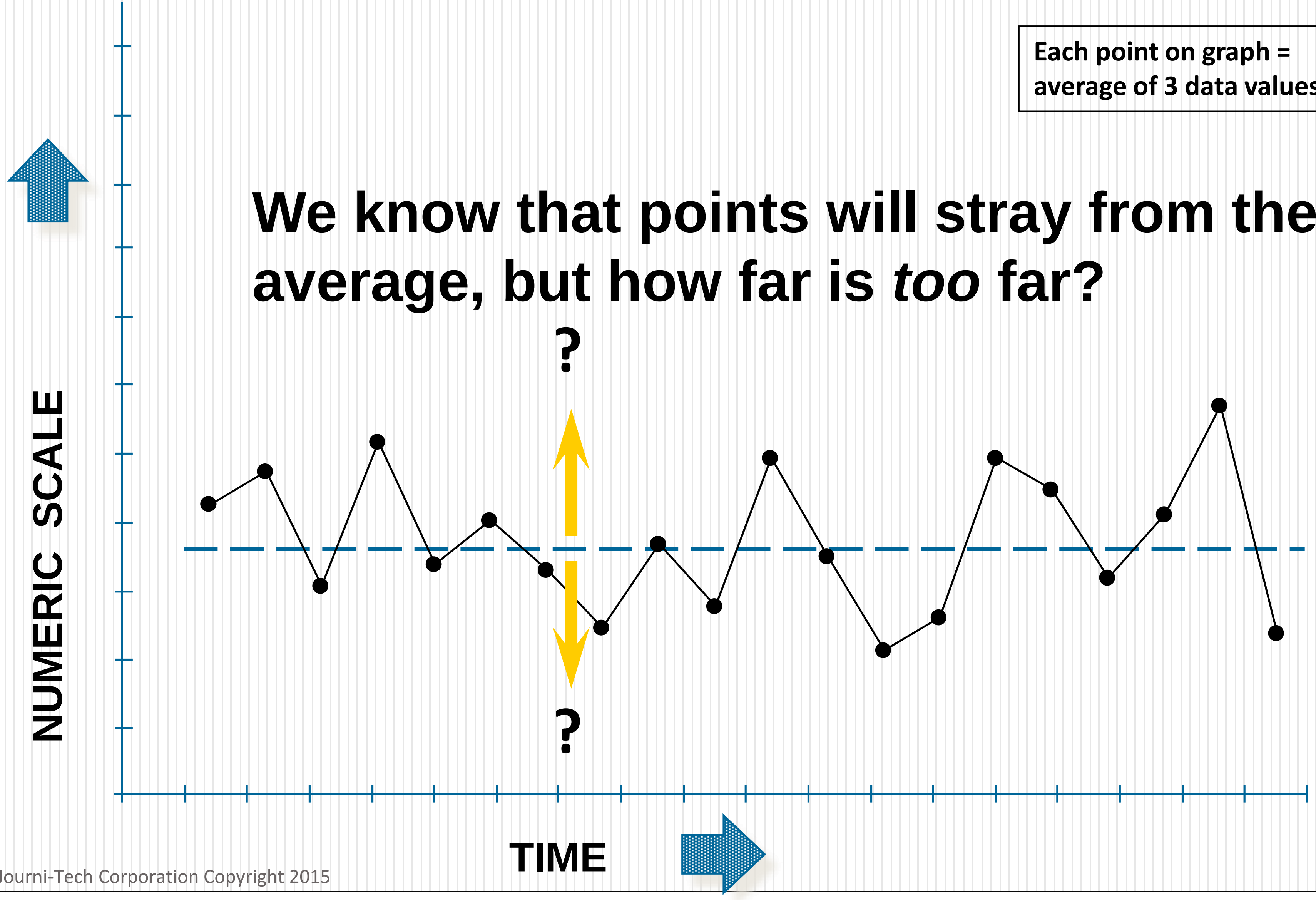
Each point on graph = average of 3 data values

The average of all the values represented by the "plot points" gives us a statistical reference point that's easy to understand.



How Much Variation is Normal?

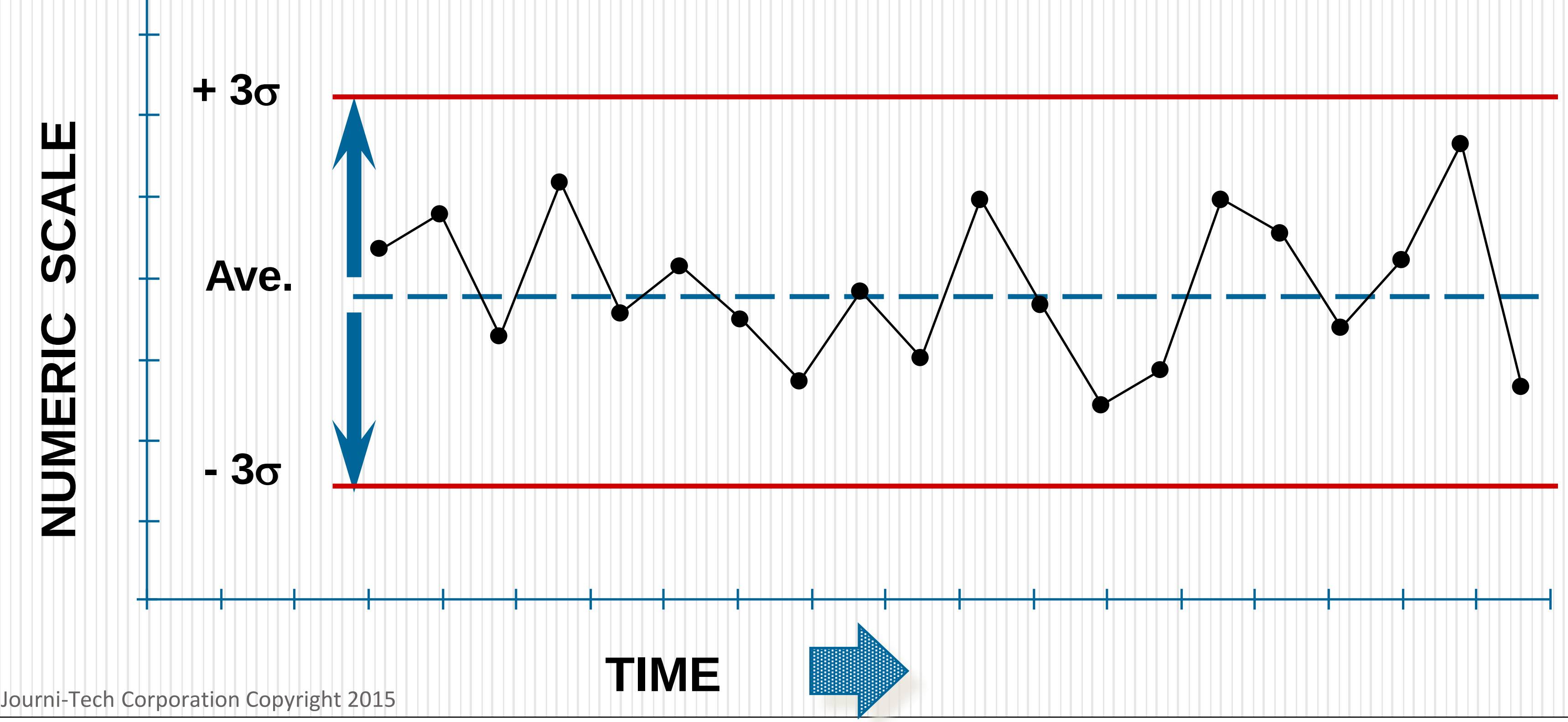
Each point on graph = average of 3 data values



Establish Control Limits

Each point on graph = average of 3 data values, therefore this is an X-bar chart

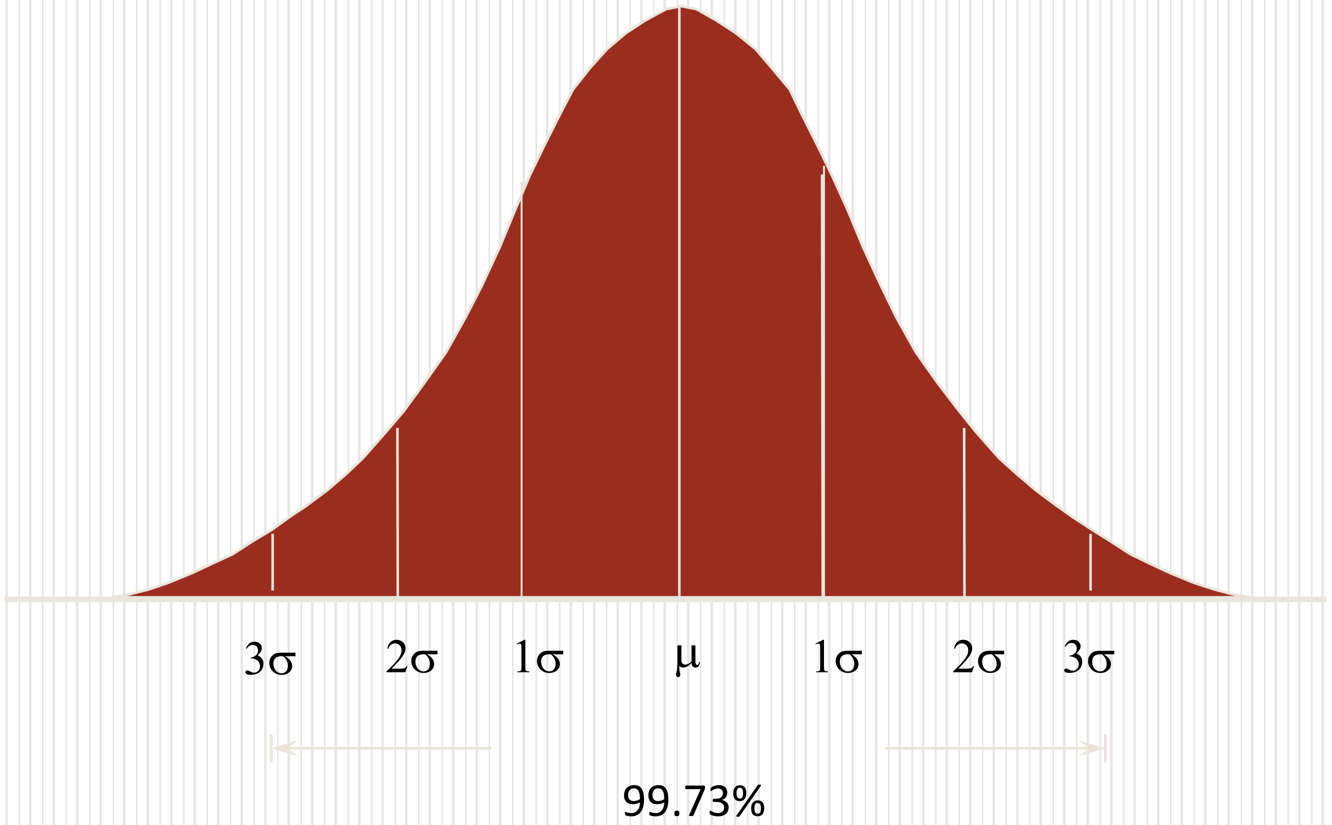
Control limits are near $\pm 3\sigma$ from the average



Question:

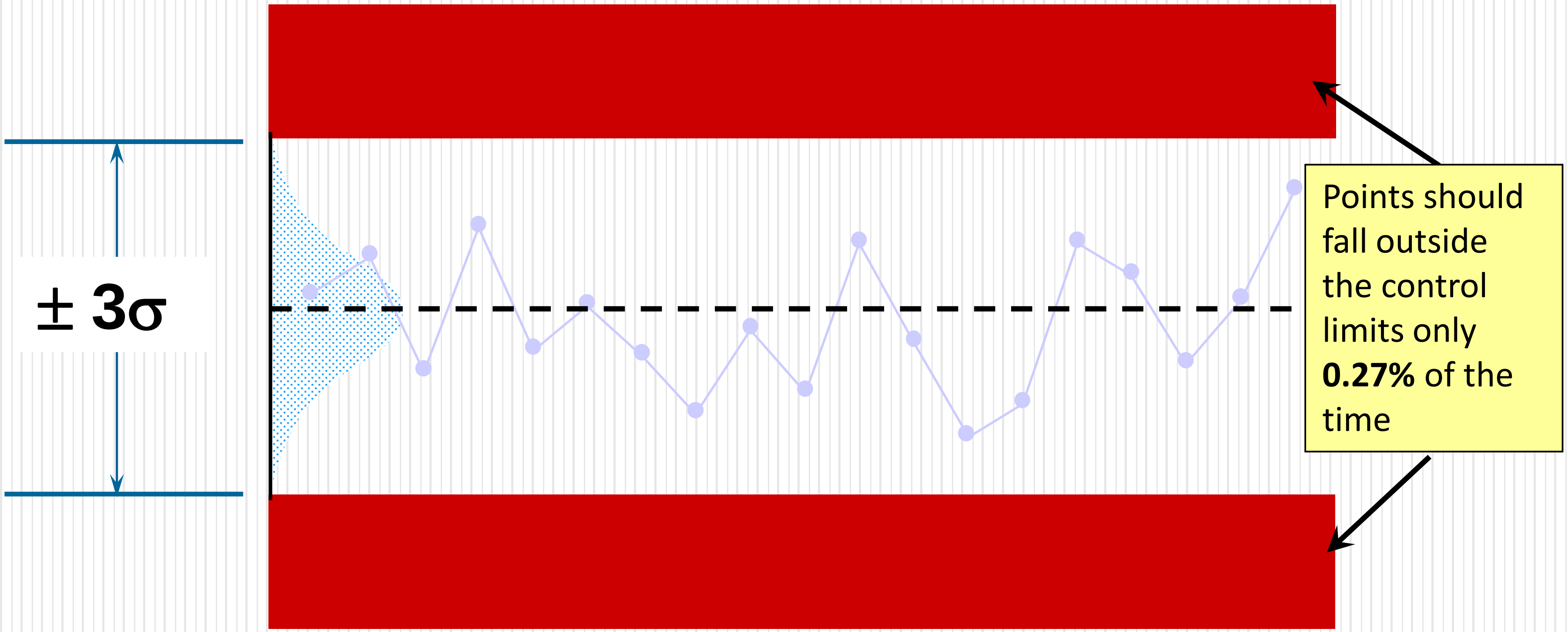
What is the relationship between the control limits shown here and the Specification Limits required by the customer?

The Normal Distribution



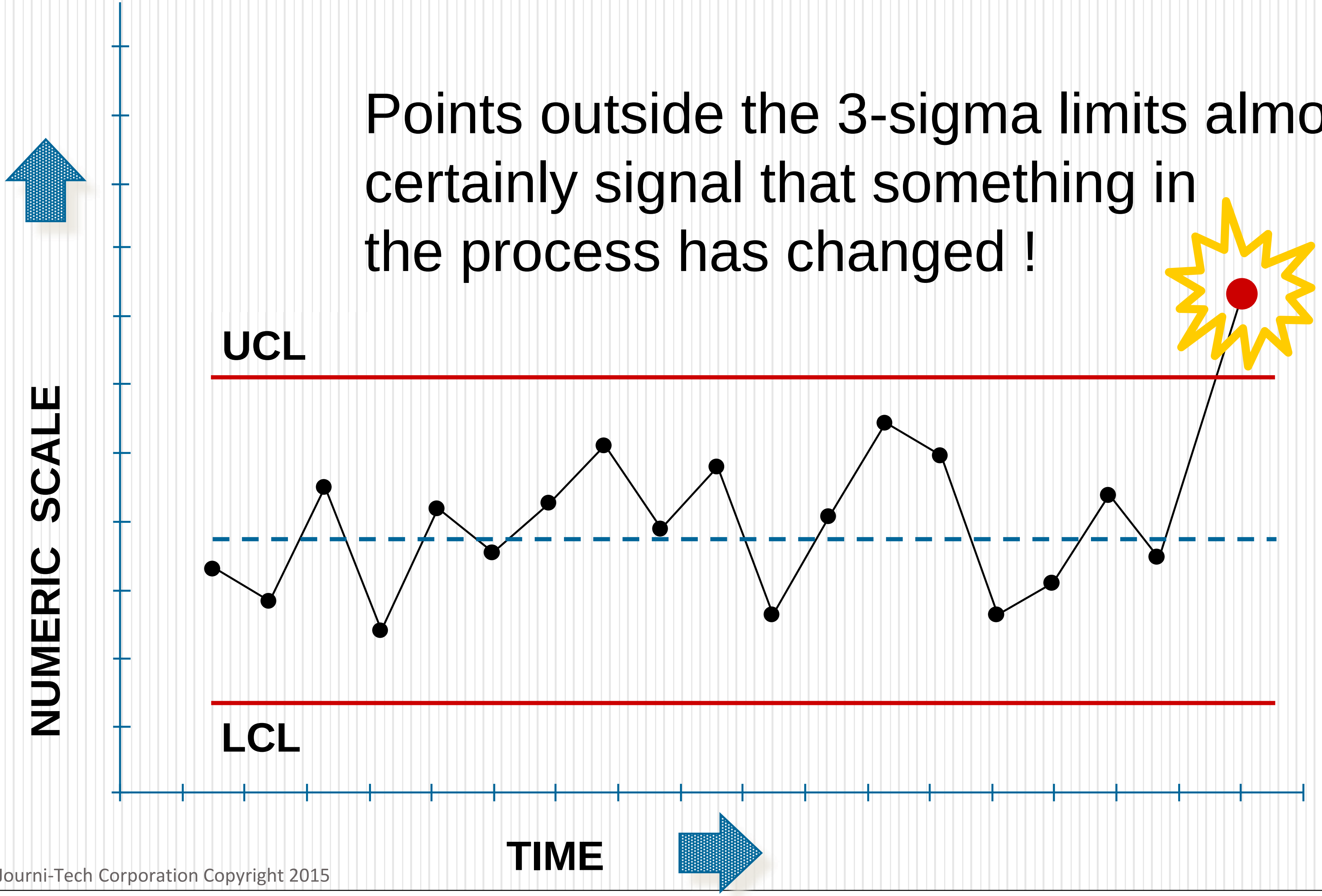
Areas under the normal distribution curve

Why Control Limits Are Set near +/- 3 Sigma



When a process is stable, >99% of the output falls within ± 3σ of the average

Out of Control Points



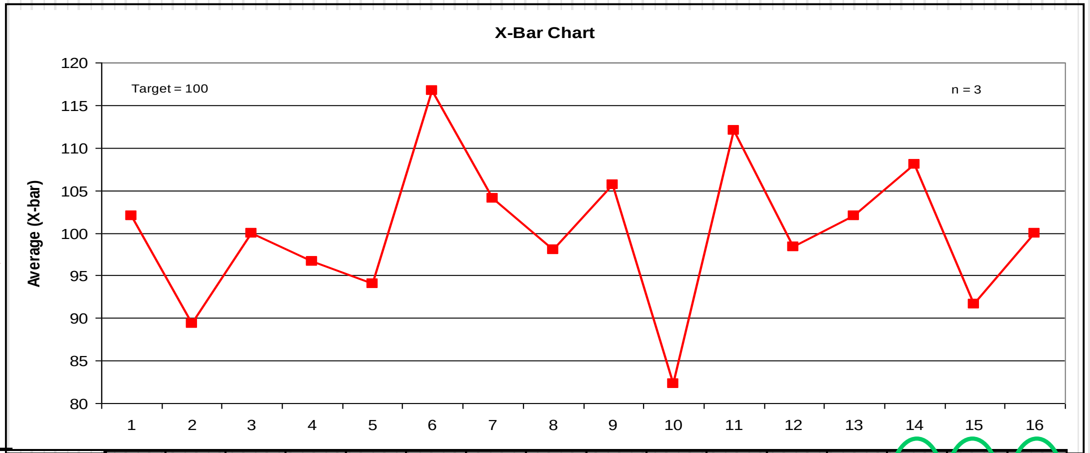
What Is a Subgroup?

- Population - the whole set of possible outcomes, i.e. every part produced or all letters mailed, etc.
- Subgroup - a small portion of a population used to help determine characteristics about the whole body.
- A sub-grouping strategy targets specific characteristics within the population.
- A sampling plan refers to how individual measurements are gathered in order to compose a subgroup.

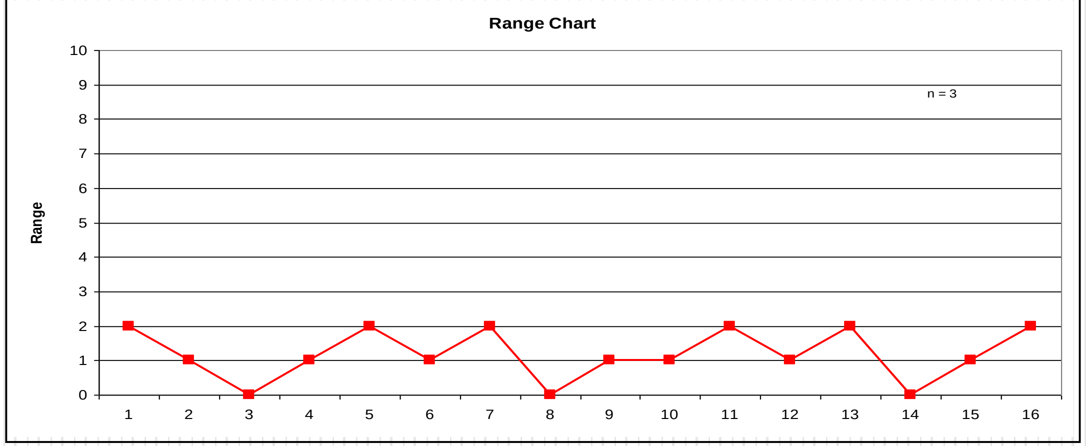
Steps to Draw an X-bar and R Chart

- Determine an appropriate sampling plan.
  - What do you want to know? What potential sources of variation are captured within/between subgroup?
- Collect the Sample Data.
- Calculate the average (x-bar) and range (R) for each subgroup.
- Plot the data.
- Calculate the control limits for the range chart.
  - If the range chart is not in control, take appropriate action.
  - If the range chart is in control, calculate limits for the X-bar chart.
  - If the X-bar chart is not in control, take appropriate action.
  - If both charts are in control, take appropriate action.

X-Bar and R Chart



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Item #1	103	89	100	97	93	117	104	98	106	82	112	99	103	108	91	99
Item #2	101	89	100	96	95	117	103	98	106	83	113	98	101	108	92	101
Item #3	102	90	100	97	94	116	105	98	105	82	111	98	102	108	92	100
X-Bar (avg)	102	89	100	97	94	117	104	98	106	82	112	98	102	108	92	100
Range	2	1	0	1	2	1	2	0	1	1	2	1	2	0	1	2

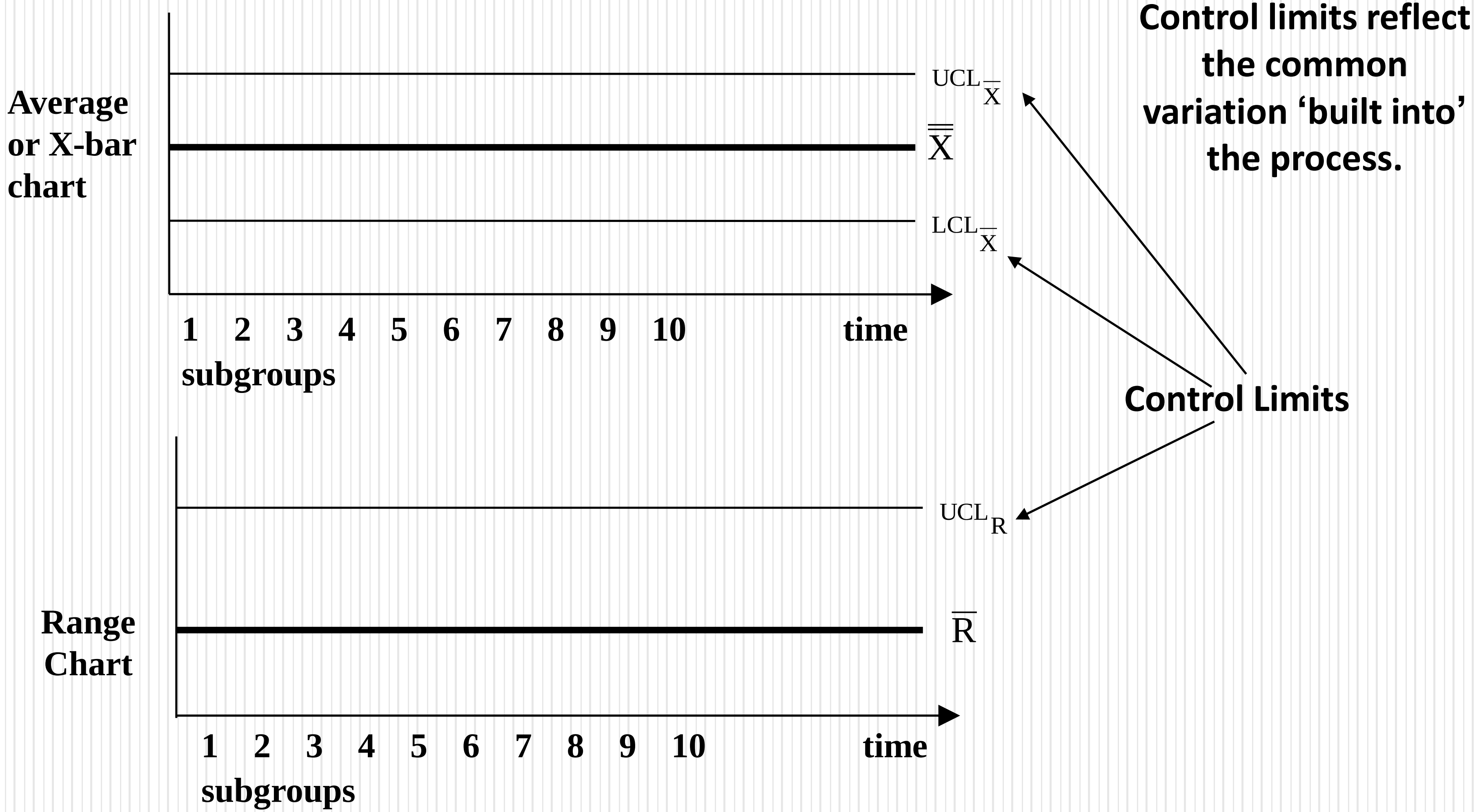


***Between*-**
subgroup
variation

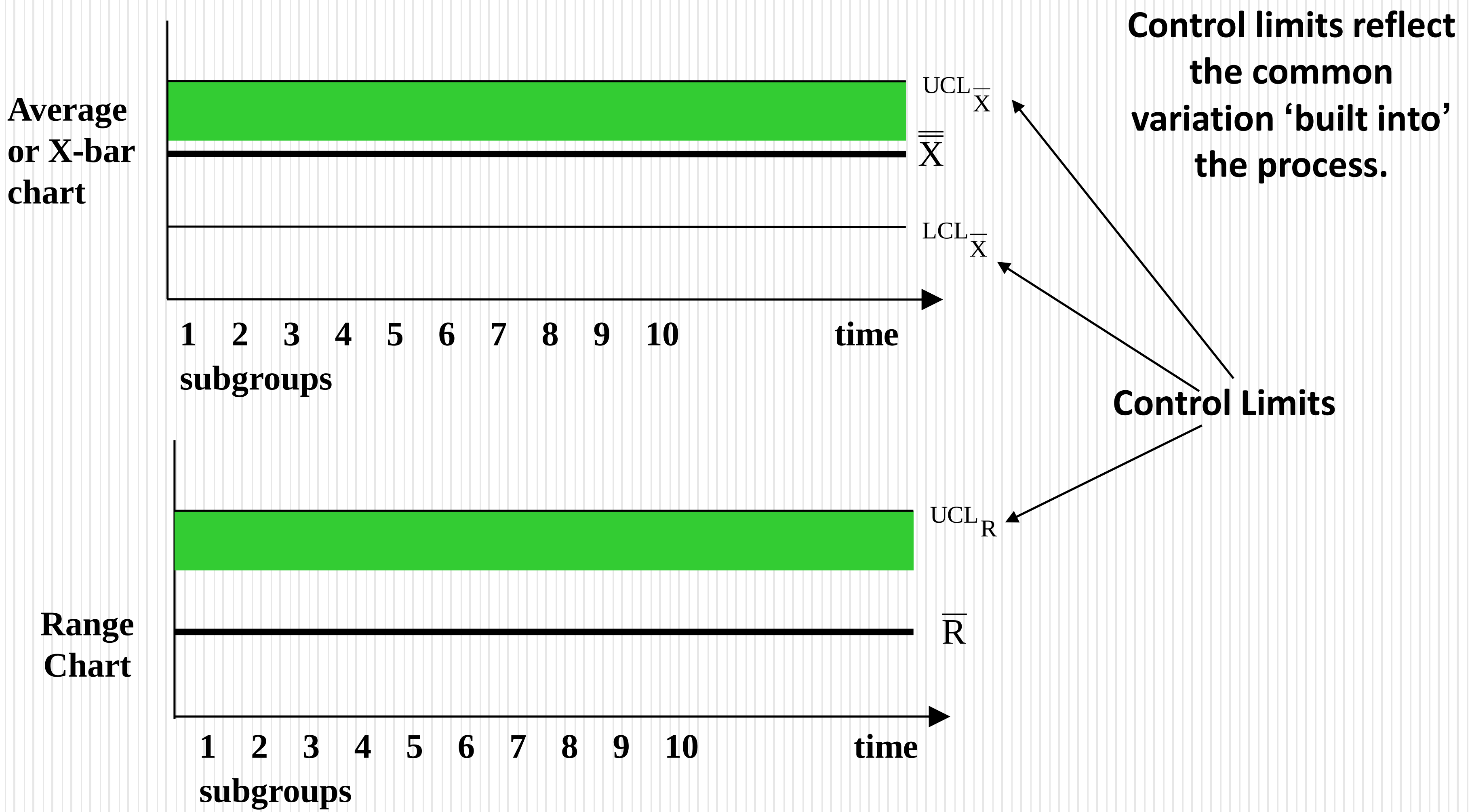
Provides a basis for estimating typical variation within the process

***Within*-**
subgroup
variation

X-Bar and R Chart



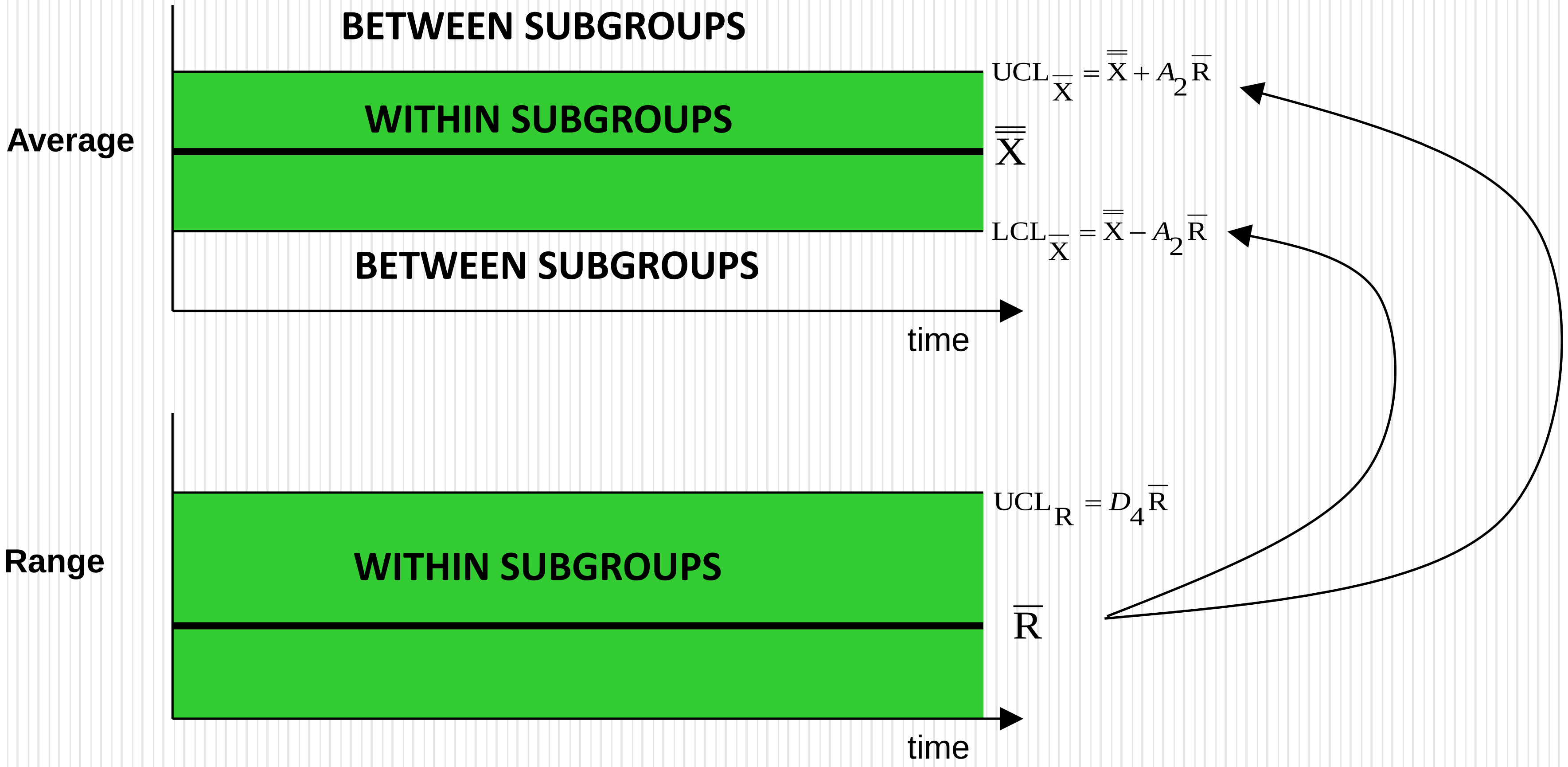
X-Bar and R Chart



X-Bar and R Chart Definitions

- X-bar & R charts are a way of displaying variables data
  - Examples of variables data: Width, Diameter, Time, etc…
- R (Range) Chart
  - Displays within subgroup variation of the process.
  - Is the variation of the measurements within subgroups consistent?
- X-bar Chart
  - Displays between subgroup variation of the process.
  - Is the variation between the averages of the subgroups more than that predicted by the variation within subgroups?

Relationship of X-Bar and R Chart



Capability

Process Baseline

- One of the team's tasks in the Measure phase is to report the current process baseline.
- The team will continue to collect data on the project metric throughout the project and compare it to the baseline.
- Process capability is one way to report the process baseline.
- Graphically, the baseline may be reported with a run chart or even a Pareto plot.

Process Capability – What Is It?

Quantifiable comparison of Voice of Customer (spec limits) to Voice of the Process (control limits)

- Most measures have some target value and acceptable limits of variation around the target
- The extent to which the “expected” values fall within these limits determines how capable the process is of meeting its requirements
- Consider key measures of process performance in:
  - Help Desk Responsiveness
  - Customer Queue Time
  - Service Cost/Order
  - Revenue/Employee
  - Job Acceptance Rate
  - Service Treatment (complaints)
  - On-Time Delivery

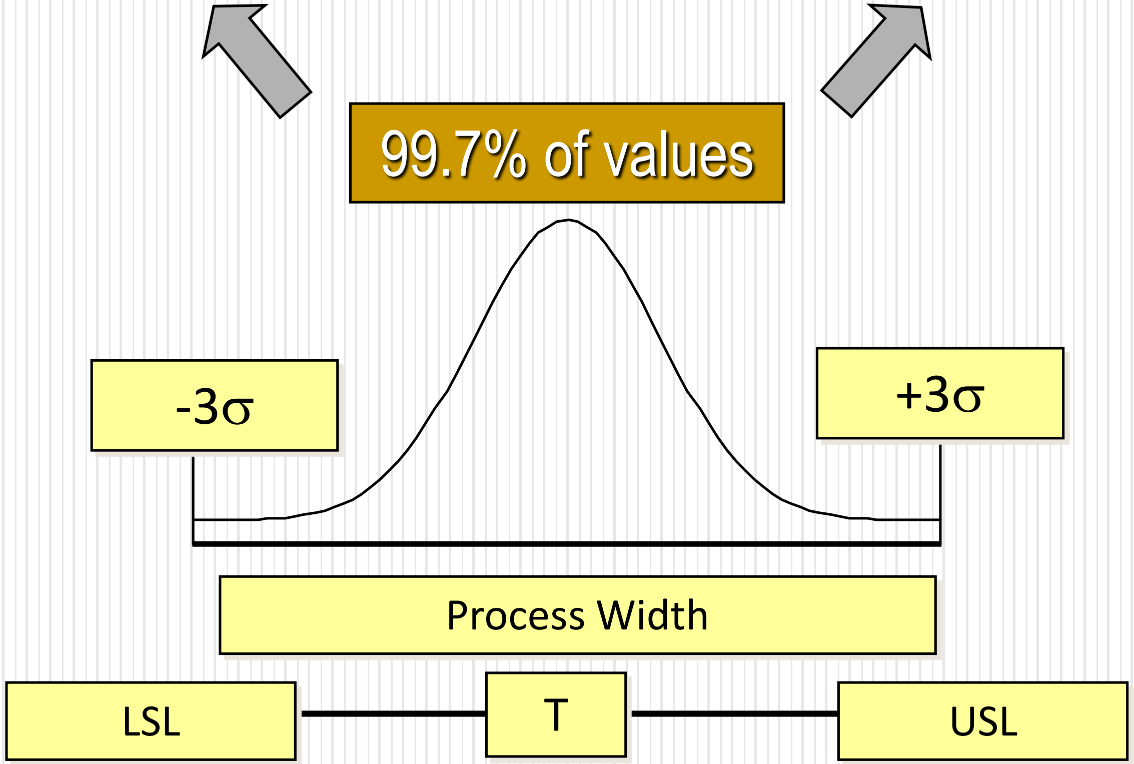
Process Capability – C_p

- Ratio of total variation allowed by the specification to the total variation actually measured from the process
- Use C_p when the mean can easily be adjusted (i.e., transactional processes where resources can easily be added with no or minor impact on quality) AND the mean is monitored (so process owner will know when adjustment is necessary – doing control charting is one way of monitoring)
- Typical goals for C_p are greater than 1.33 (or 1.67 for safety items)

If C_p < 1, then the variability of the process is greater than the specification limits

Process Capability – C_p

$$C_p = \frac{\textit{Allowed variation (spec.)}}{\textit{Normal variation of the process}} \quad \text{or} \quad C_p = \frac{|USL - LSL|}{6\sigma}$$



Process Capability Terminology

- **Specification Limits**
  - Boundaries, usually set by management, engineering or customers, within which a process must operate
  - Voice of the customer
  - Is the service or product meeting the customer's expectations?
- **Process Capability**
  - Compares the output of an in-control process to specification limits using capability indices
  - Answers the question of how well a process meets a customer's expectations

Why Assess Process Capability?

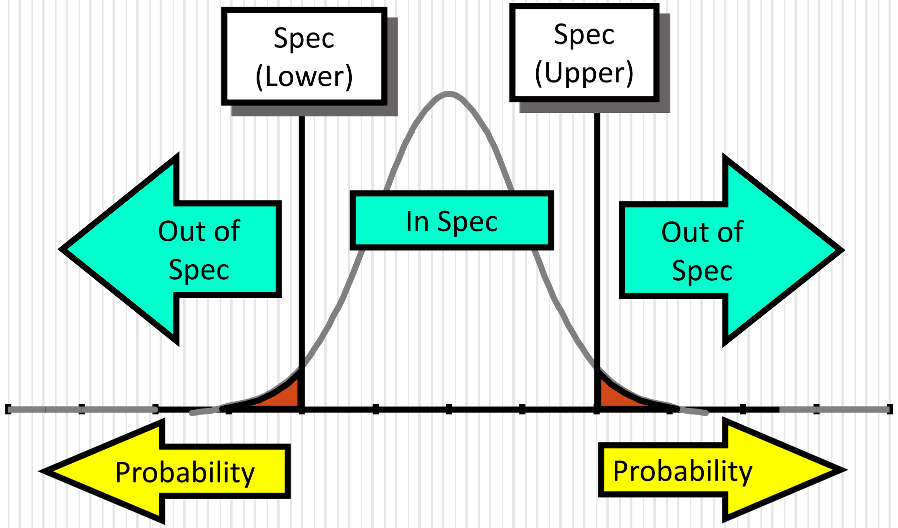
- Allows us to quantify the nature of the problem:
  - Are the specifications correct for the parameter (Y) of interest?
  - Is the location of the central tendency of the parameter (Y) centered within the appropriate specifications?
  - Is the process variation in the parameter greater than allowed by the specifications?
  - Is the measurement system affecting our ability to assess true process capability?
- Allows the organization to predict defect levels.

What is Process Capability?

Process capability is simply a measure of how good a metric is performing against an established standard. Assuming we have a stable process generating the metric, it also allows us to predict the probability of the metric value being outside of the established standard(s).

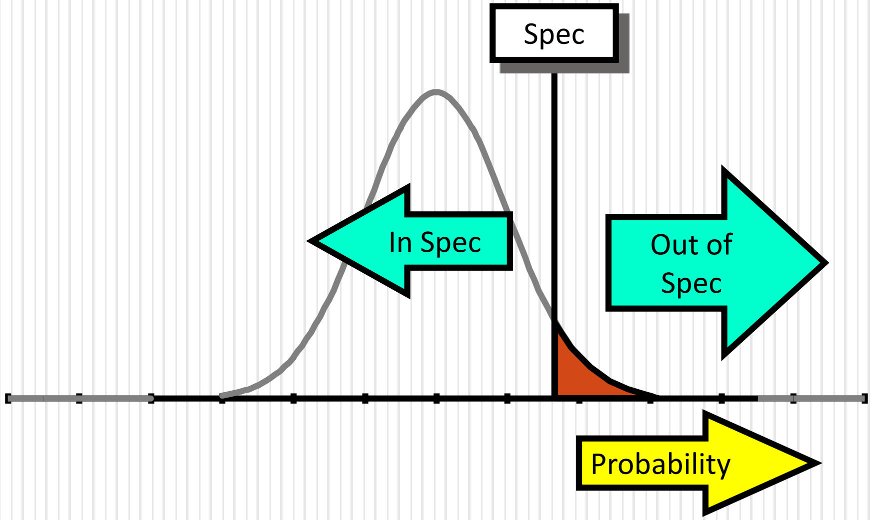
Upper and Lower Standards (Specifications)

Example: Diameter, length, etc.



Single Standard (Specification)

Example: Time, Roundness, etc.

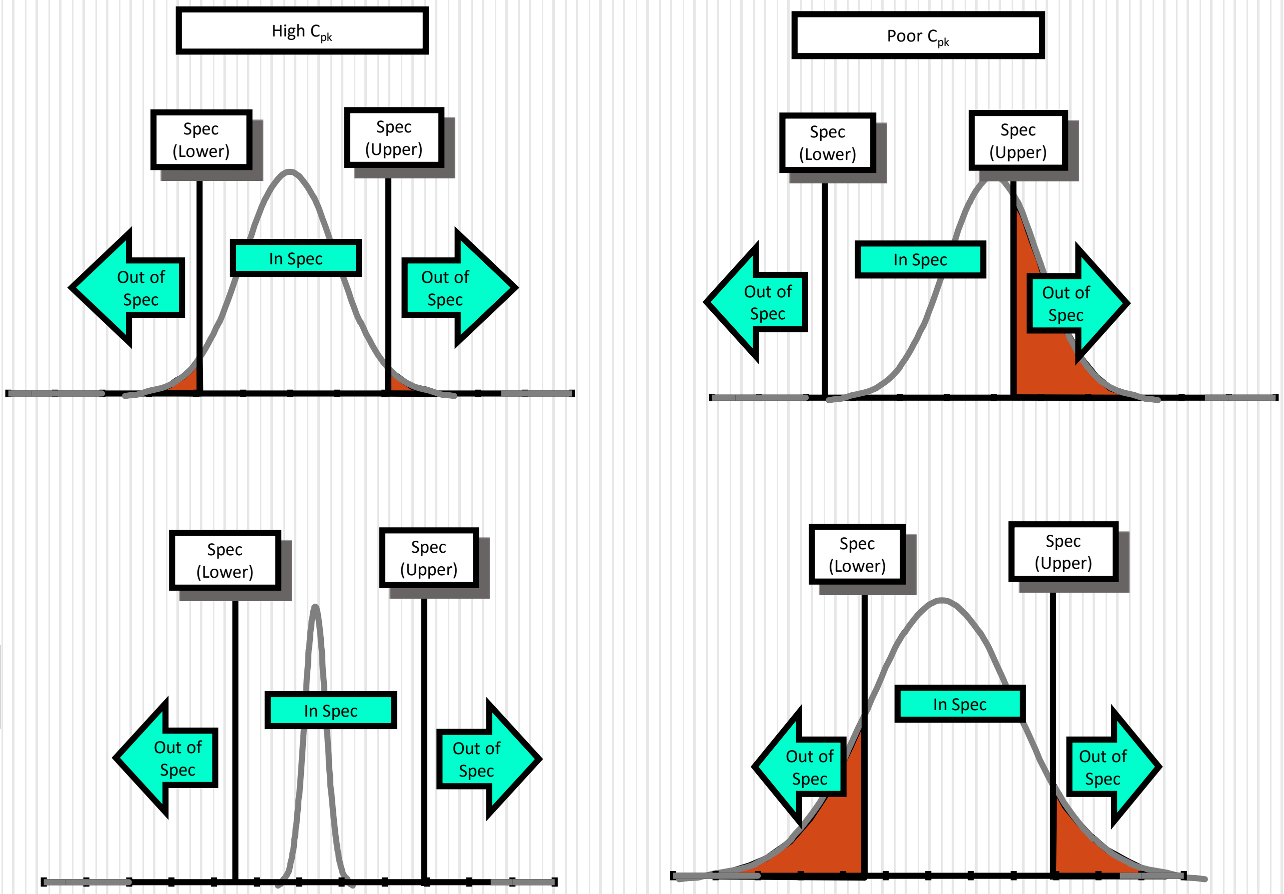


What is Process Capability?

Process capability (C_{pk}) is a function of how the population is centered *and* the population spread.

Process Center

Process Spread



Process Capability Study

- Process capability is composed of variation and centering
  - The POTENTIAL process capability index is a measure of the variation or spread only, and is expressed as C_p
    - $C_p = (USL-LSL)/(6*\sigma)$
  - The OVERALL process capability index combines the effect of variation with how well the process is centered, and is called C_{pk}
    - $C_{pk} = MIN(\mu\text{-}LSL, USL\text{-} \mu)/(3*\sigma)$, where μ = population mean

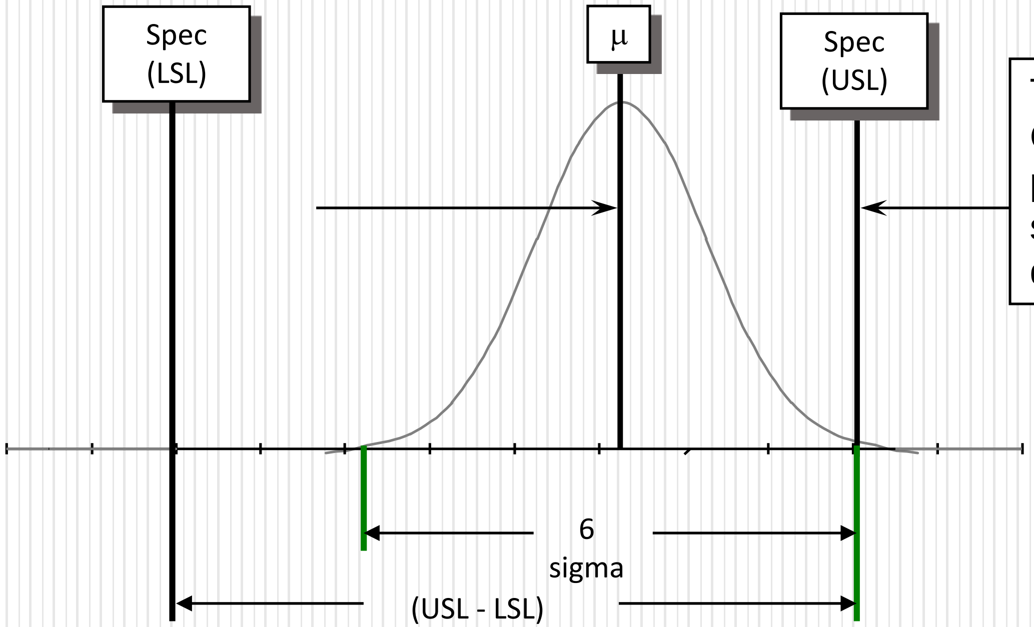
Benefits of a Process Capability Study

- The three biggest benefits of the capability study are:
  - Saving money and improving customer satisfaction by allowing us to identify and eliminate causes of scrap
  - Saving money by allowing us to quickly reduce material usage without generating scrap
  - Saving money and improving customer satisfaction by showing how reduction of process variations can reduce material usage AND scrap

Process Capability Study

- Keys to success:
  - Distinguish carefully between the variation (spread) and the average (center)
  - Use your team or process experts to help with the analysis
  - Look for slow changes over time (drift) and sudden changes in the center (shift)
  - Observe how changes in CTP’s impact the capability
  - Note how improvements in capability can lead to productivity

Process Capability Study

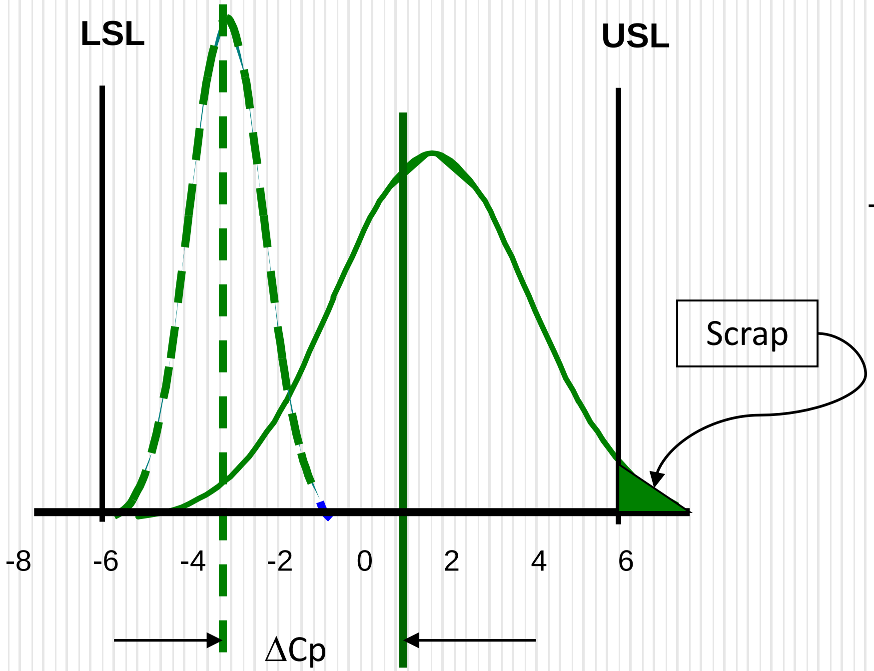


To determine **overall capability**, C_{pk}, find the distance between the population mean and the nearest spec limit ($|\mu$-USL; LSL-$\mu|$). This distance divided by 3σ is C_{pk},

The **potential capability** (C_p) of this process is determined by [(USL-LSL)/6σ].

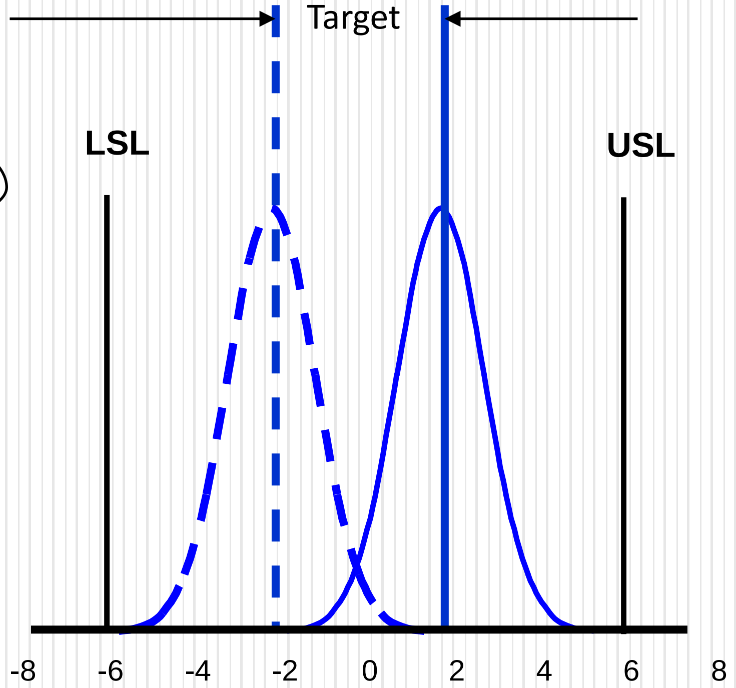
In a capable process, C_{pk} is 1.0 or greater

Process Capability Study



Savings from reducing variation, lowering target, and from eliminating scrap

Savings from simply lowering the target without changing process capability



Process Capability

- C_p relates tolerance spread to process capability
- Process capability = 6σ
- FIND Cp: USL = 12, LSL = 4, $\sigma = 1$
- Cp = USL - LSL = 12 - 4 = 1.33
- 6σ 6
- When Cp = 1.33, the process spread is 3/4 of the tolerance spread

Process Capability

* If the tolerance is 0.5" +/- 0.006", what must 6σ be to yield a $C_p = 1.33$?

$$\frac{USL - LSL}{6\sigma} = \frac{0.506 - 0.494}{6\sigma} = 1.33\ (C_p)$$

$$6\sigma = \frac{0.012}{1.33} = 0.009$$

* Using the calculated process capability above, find σ.

$\sigma = 0.009/6 = 0.0015$

* Given $\bar{X} = 0.503$, find C_{pk}.

$$C_{pk} = \text{MIN}\{\bar{X} - LSL/3\sigma\ ;\ USL - \bar{X}/3\sigma\}$$
$$\text{MIN}\{0.503 - 0.494/3\,(0.0015)\ ;\ 0.506 - 0.503/3\,(0.0015)\}$$
$$\text{MIN}\{0.009/0.0045\ ;\ 0.003/0.0045\}$$
$$\text{MIN}\{2\ ;\ 0.67\} \quad C_{pk} = 0.67$$

Process Capability

* The center of the tolerance spread is 0.500 inches. What is C_{pk} when $\bar{X}$ = 0.500 inches?

$$\frac{0.500 - 0.494}{0.0045} \; ; \; \frac{0.506 - 0.500}{0.0045}$$

1.33 ; 1.33

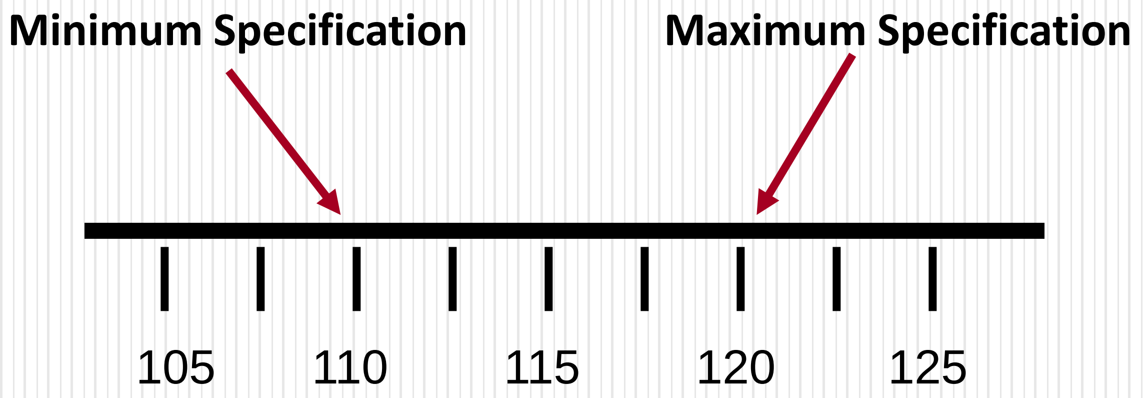
What does the calculated C_{pk} tell you?

A Practical Illustration

A popular coffee bar has started receiving complaints from customers about the coffee temperature.

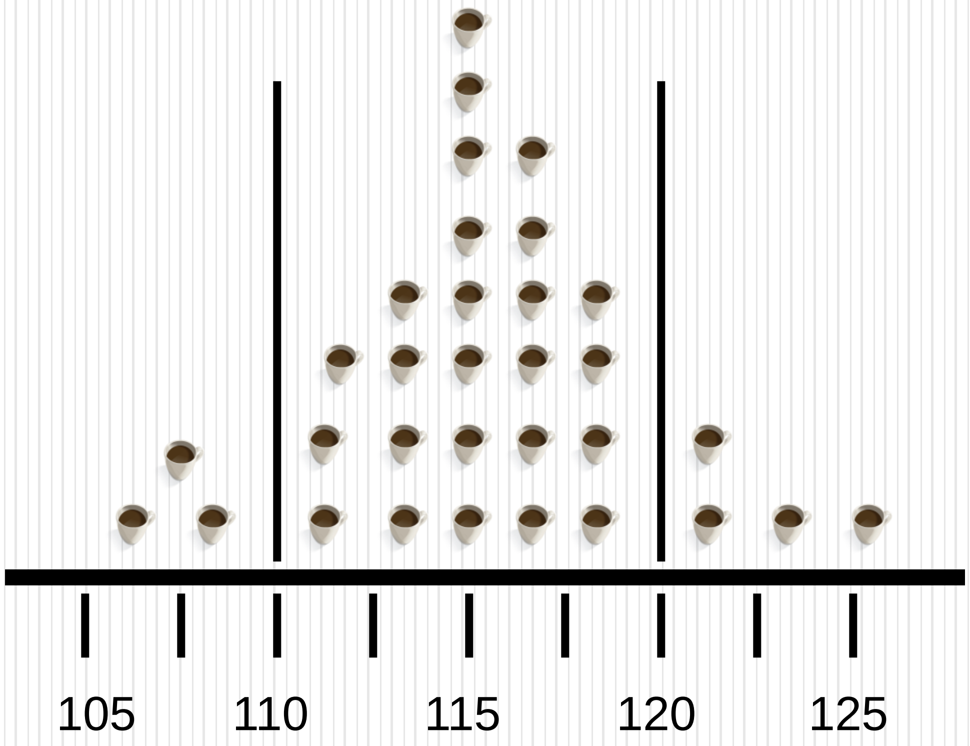
Some customers say the coffee is too hot, but others have complained it is too cold.

A Practical Illustration



Customer surveys and industry benchmarking indicate that customers will be satisfied if the temperature is between 110 and 120 degrees F.

A Practical Illustration



Capability Indices

A capability index uses both the process variability and the process specification limits to quantify whether or not a process is capable of meeting customer expectations.

C_p	C_{pk}
Indicates how well the process distribution fits within its specification limits. This index reflects how capable the process would be if centered	Incorporates information on spread as well as the process mean so it is an indicator of how well the process is actually performing.

If $C_p = C_{pk}$, the process is centered. If $C_p > C_{pk}$ the process is not centered.

Summary

In this module you have learned about:

- Tools to verify normality
- Tools to verify stability
- The concepts underlying control charts
- Process capability terms
- How to assess process capability