Control Chart Basics
Primary Knowledge Unit
Instructor Guide

Note to Instructor

This Primary Knowledge (PK) unit provides an overview of control charts, one of the primary tools of Statistical Process Control. Control charts are used to ensure that a process stays within its inherent variation. This unit covers how to construct control charts and how to interpret them.

This Primary Knowledge unit is part of the *Statistical Process Control Learning Module*, which covers the following:

- Statistical Process Control Knowledge Probe (KP) Pre-test
- Introduction to Statistical Process Control (PK)
- **Control Chart Basics (PK)**
- SPC Resistance Activity
- Activity (Advanced) – An MEMS Process Problem (Found in the SCME Systematic Approach to Problem Solving Learning Problem)
- Final Assessment

This companion Instructor Guide (IG) contains all of the information in the Participant Guide (PG) as well as answers to the coaching and review questions at the end of the unit. A PowerPoint presentation is provided for a classroom presentation. The PowerPoint is a summary of the PG.

Description and Estimated Time to Complete

Control Charts are one of the primary tools used in Statistical Process Control or SPC. Control charts help to graphically represent different aspects of a process and to assist engineers and technicians with producing a quality product. In this unit you learn the basics of control charts, how they are constructed, and how they are interpreted.

<u>Estimated Time to Complete</u>
Allow approximately 30 minutes to read through this unit.

Unit and Related Activity Objective(s) / Outcomes

Objectives

- Explain normal distribution and how it relates to $\bar{X}$*-charts*.
- Given a set of data, construct an $\bar{X}$*-chart*.
- Interpret control charts by applying the Shewhart rules to identify processes that are Out of Control (OOC).

Outcomes

You should be able to construct a basic $\bar{X}$*-chart* and apply Shewhart rules to interpret a control chart. You should also be able to explain what is meant by an out of control process.

Terminology

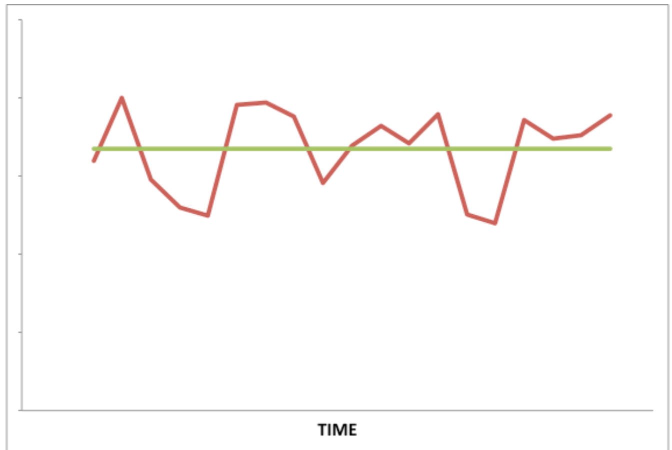
Attribute Data
Control Chart
Mean
Median
Normal Distribution
Standard Deviation
Target
Variable Data
Variance

Introduction to Control Charts

A Bell Laboratories engineer named Dr. Walter A. Shewhart introduced the basic concept of control charts in the 1920's. Shewhart was able to identify the importance of the reduction of variation in a process. In order to achieve continuous improvement he stressed the necessity of understanding and reducing this variation. He introduced a control chart as a tool to bring a process into a state of statistical control. These charts are sometimes referred to as Shewhart Charts.

Control charts characterize process variation over time and distinguish between *common or inherent cause variation* and *special cause variation*. Common cause variation is usually due to inherent elements of the process that is often referred to as background noise. It is variation that is expected and predictable in a process. Special cause variation is different. It is usually due to some special or assignable cause, something that is not expected or predictable. Control charts show what the process is doing and, when the process changes, show very quickly that a change has occurred.

Shewhart defined two types of variation, *controlled* and *uncontrolled*. *Controlled* variation is predictable in a probabilistic sense and can be represented by a consistent pattern that can be attributed to common causes or that background noise (e.g., small fluctuation in temperatures, different operators/technicians, different equipment running the same process, process that don't repeat "exactly") . If this variation is controlled, then the process can be said to be in control. A process, which is in control, has a predictable or expected variation with no abnormal or deviant data values, nor any consistent trends away from the normal variation. A process, which is in control, has no surprise data values or sets of data values (trends or patterns). Based on what we know at this point, the process on the right, looks in-control. There are no trends, patterns or extreme highs or lows.



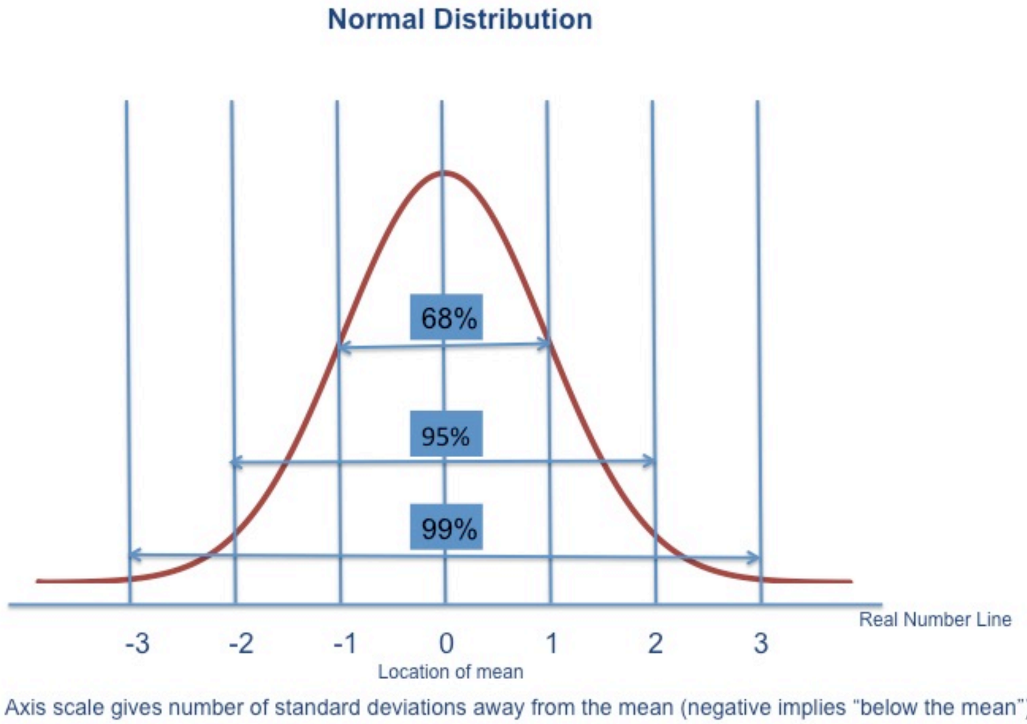
Uncontrolled variation is when the variation within a process becomes unpredictable, due to a special cause. The special cause of such variation needs to be identified and corrected in order to remove this uncontrolled variation from a process. A control chart is a helpful tool in characterizing the variation for a process that is already in control, and for identifying when a process is no longer in control, or in other words, out of control (OOC).

Understanding the basic mathematical concepts involved with SPC is important in developing and utilizing control charts correctly. Some basic statistical concepts that you should be familiar with are *sample median, sample mean (μ or $\bar{X}$), sample range*, *sample variance*, (σ^2) and the square root of the sample variance - *sample standard deviation* (σ). Using these statistical concepts will assist you in creating a control chart to monitor process variation. For a detailed description of these statistical concepts and calculations, please review *Introduction to Statistical Process Control.*

Normal Distribution

When considering process data, it is important to understand the nature of your data, how it behaves and how it is distributed. In statistical terms, this is called the *probability distribution* of the data set. There are many different types of probability distributions, but the scope of this lesson focuses on a continuous *normal distribution*. A normal distribution is most often assumed because most process data is distributed normally, or at the very least, a normal distribution can be a good approximation.

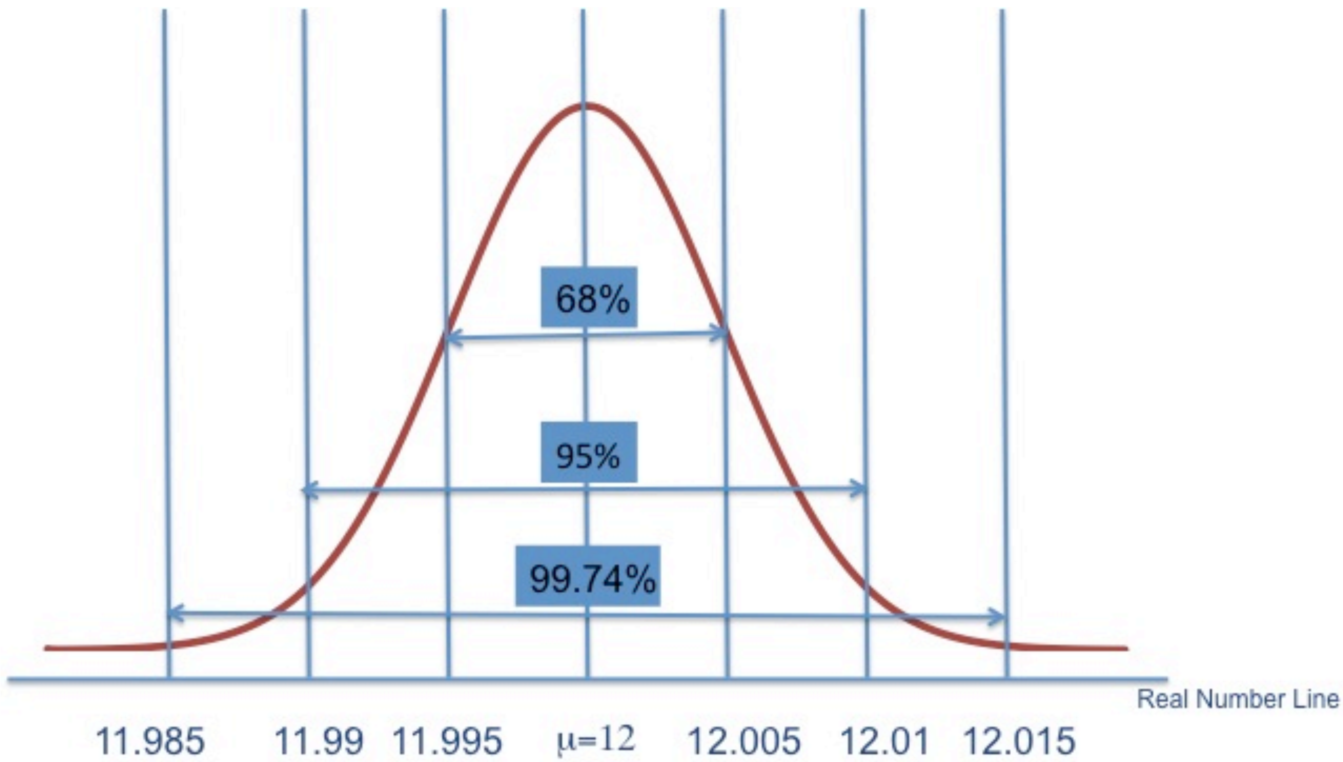
A normal distribution is also referred to as a Gaussian Distribution and it follows a bell-shape pattern. A normal distribution is shown below with the vertical lines signifying numerical values at one (1), two (2), and three (3) standard deviations (σ) *above and below* the mean.



The well-known bell-shaped curve has the following important properties:

- It is symmetrical about its mean. The mean is central in the distribution of data values. It is at the apex of the curve.
- The width of the bell-shaped curve increases as the standard deviation increases and decreases as the standard deviation decreases.
- Approximately 68% (68.26%) of the data will be within one standard deviation on either side of the mean
- Approximately 95% (95.44%) of the data will be within two standard deviations on either side of the mean
- Approximately 99% (99.74%) of the data will be within three standard deviations on either side of the mean.

Let's look at an example. Suppose the voltage measurements produced by a process represent a normal distribution that has a mean of 12.00 volts (μ =12V) and a standard deviation of 0.005 volts (5mV), ($\sigma = 0.005$V). These values were calculated using a very large sample of 10,000 voltage measurements taken from the process. After looking at the data, you find about 99.74% of the data values fall within the range of 11.985 to 12.015 volts ($\mu \pm 3\sigma$), 95% of the measurements are found within the range from 11.99 to 12.01 volts ($\mu \pm 2\sigma$), and about 68% of the measurements are within the range of 11.995 to 12.005 volts ($\mu \pm 1\sigma$). Unless the distribution changes, this is what you should expect subsequent samples from the distribution to demonstrate. Take a look at the plot below that illustrates the normal distribution of the voltage measurements.



How can we tell from the distribution if the variation is due to common causes or special causes (controlled or uncontrolled variation)? Later we shall see that control charts enable us to detect variation that is due to special causes by detecting a sizeable shift in the sample mean or sample standard deviation.

Characteristics of Control Charts

In the following discussions, we assume that the plotted statistics represent independent samples from a normal distribution.

Variable and Attribute Data

Two major types of data are involved in Statistical Process Control (SPC) – variable data and attribute data. Data based on measurements is called *variable data*. Data based on counts is called *attribute data*.

- Variable Data – Data based on measurements (continuous)
- Attribute Data – Data based upon counts (discrete)

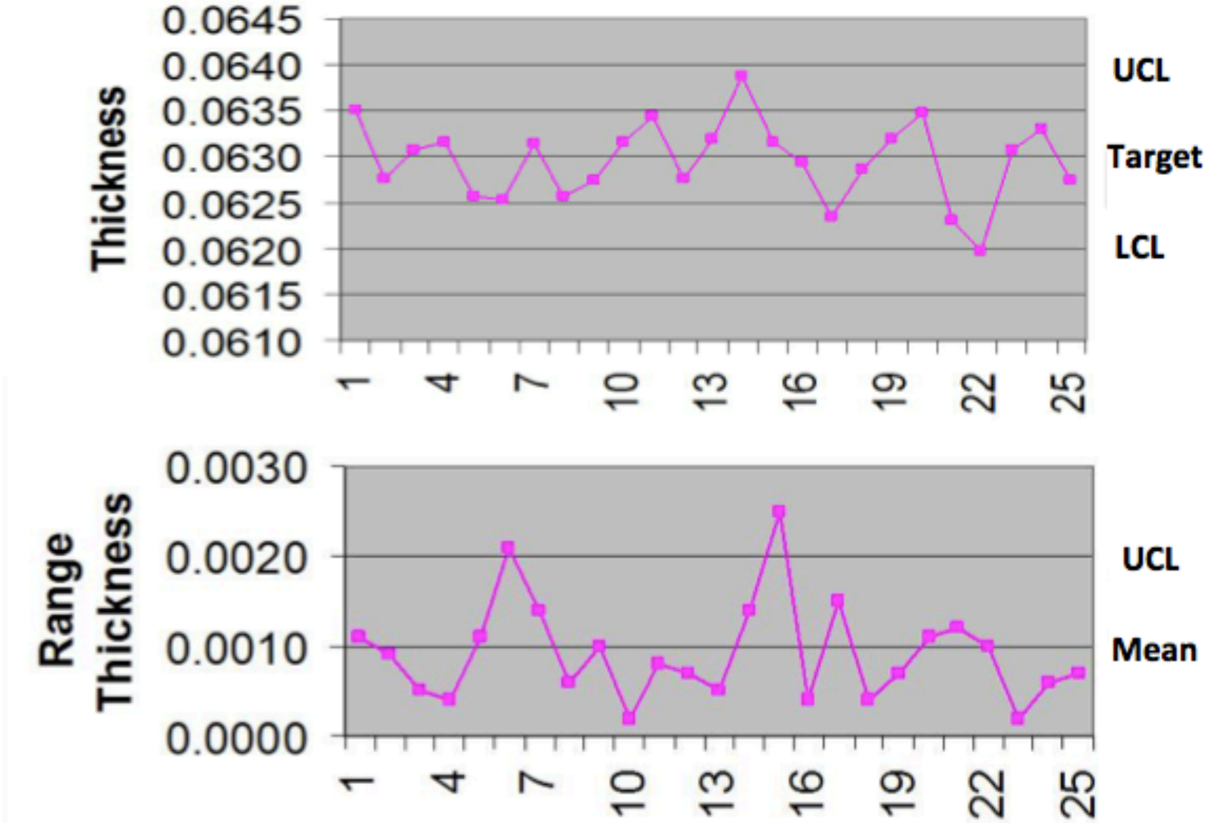
Variable data is more commonly used at detecting out of control (OOC) conditions in a process situation than attribute data. Variables data is continuous and quantitative, meaning that a variable measure can conceptually take on any value in any given interval. Common variables include length, yield, time, weight, temperature, pressure, line width, and film thickness.

Attributes data is discrete and qualitative or categorical, meaning that the measure can take only a countable number of values in an interval. An attribute is a characteristic that is either there, or not. An attribute or characteristic itself cannot necessarily be measured, but those characteristics can be accounted for. Common attributes measured include counts, number of defects, fraction defective, number of values outside of specs, failure of success, number of defective die, number of particles on a wafer (i.e., acceptable or unacceptable results).

An example of *variable* data is the oxide thickness measurements of wafers in microchip fabrication. An example of *attribute* data is the number of rejected wafers due to microcontamination.

Control charts use both variable and attribute data and for each type of data, there are different types of control charts. In the manufacturing industry, the most commonly used data to monitor a process is variable data. The most widely used data to monitor process parameters or variables are the mean that is plotted on an $\bar{X}$*-chart*. Sometimes the range of the data is used along side the $\bar{X}$*-chart*. The range of data is plotted on a *R-chart*. The $\bar{X}$*-chart* uses the sample mean (μ or $\bar{X}$) of a large, initial sample of data as the centerline, and the mean plus or minus 3 times the sample standard deviation ($\mu \pm 3\sigma$), as the upper and lower control limits. Once the $\bar{X}$*-chart* is applied to the process, it shows how the mean of small process samples fluctuates over time.

The R-chart uses the mean range of a large initial sample as the centerline and the range plus or minus 3 times the sample standard deviation of the mean range ($R \pm 3\sigma_R$) as the upper and lower control limits. For any given process, the $\bar{X}$*-chart* can be used alone or together with the *R-Chart*. Sometimes, the mean of the measurements can appear to be in control, but the process is not behaving as it should. The range chart of the same data can provide more insight. Let's take a look at an example.



The top chart is an X-bar chart of photoresist thickness. The bottom chart is a range chart for the photoresist thickness. Notice that the range variability can be out-of-control while the mean of the measurements is in control. In the range chart, data points 6 and 15 show an out-of-control range, but an in-control-mean. Whereas data point 22 shows an out-of-control mean and in-control range. Differences such as this help us to better analyze the cause of such problems, because we have even more data. An in-control-range, but out-of-control mean, could indicate a shift in the process mean, where as an out of control (OOC) range and an in control mean could mean an increase in variability. Whether one or both of these charts is used depends on the product and the process.

This lesson discusses how to create a $\bar{X}$*-chart* in detail. The range charts are created in a similar fashion.

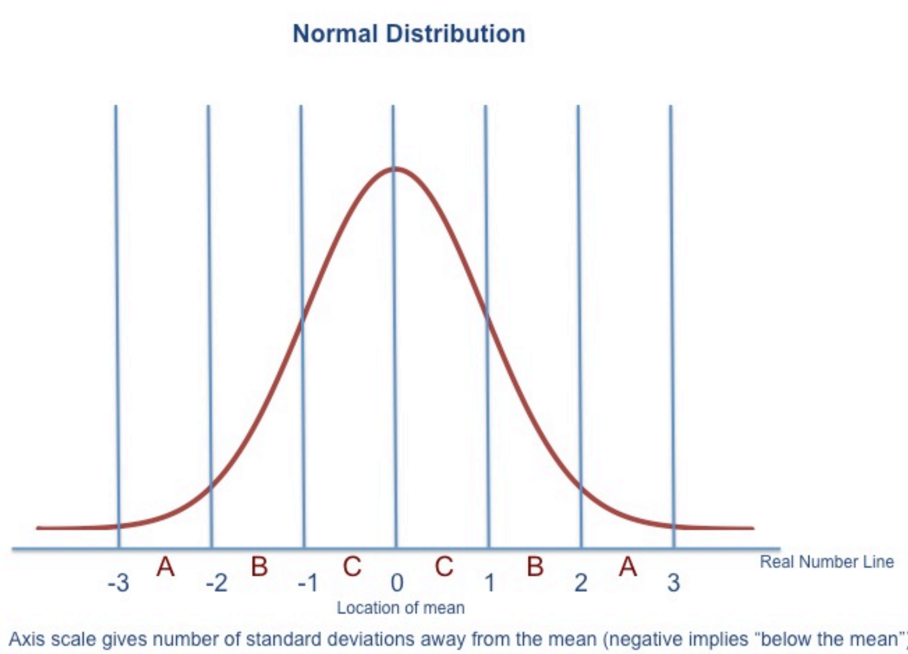
Mean or $\overline{X}$-chart

Variables data can conceptually take on any value in an interval (e.g., process temperature can be any value between 800° C and 1200°C). Attributes data is discrete and the measure can take only a countable number of possible values in interval process (e.g., the number of defective die on wafer with 200 die). As we discussed earlier, a major assumption is that the data being charted exhibits a normal distribution.

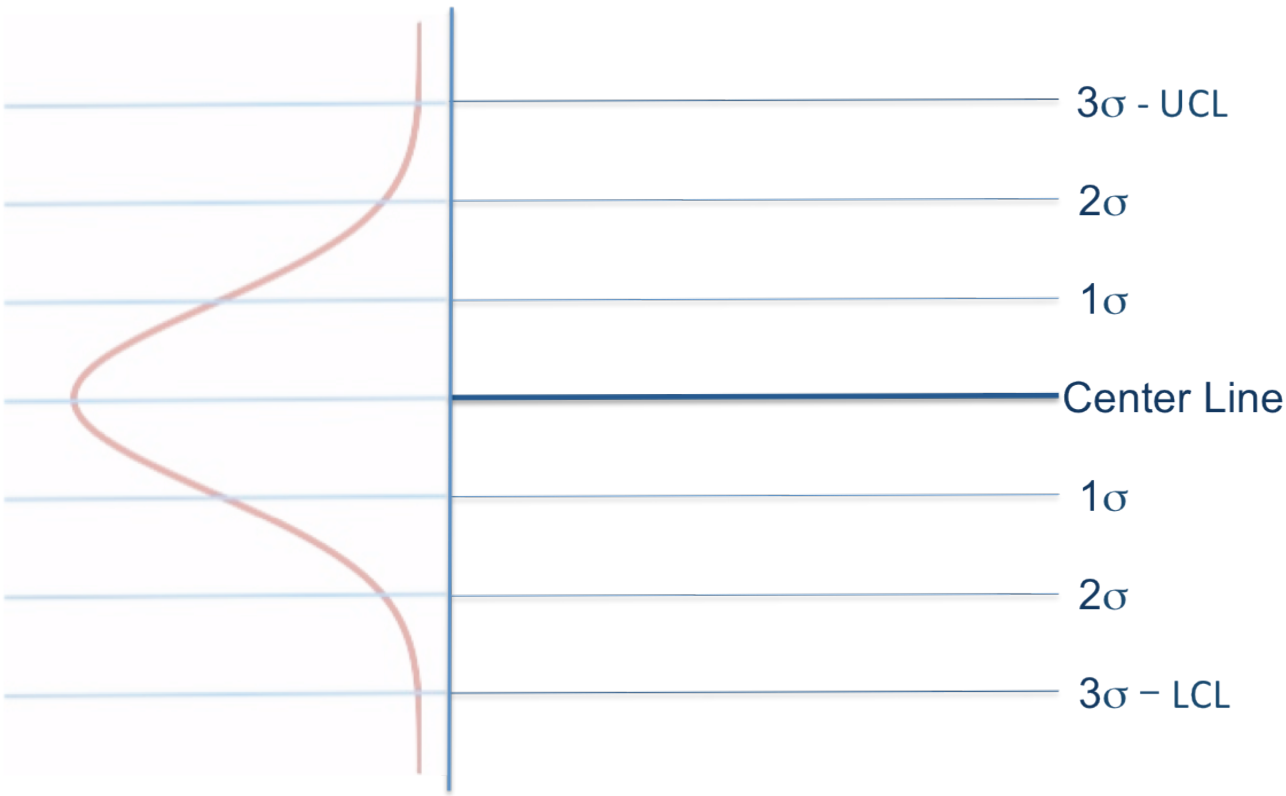
Regardless of the type of data, the basic form of the control chart is the same. The chart has a centerline or target. As stated earlier, the $\bar{X}$*-chart* uses the sample mean (μ or $\bar{X}$) as the centerline. There are one or more control lines that are drawn above and below the centerline. The control lines are determined by a certain number of standard deviations away from the centerline or target. It is important that both the centerline and the standard deviation be correctly calculated and defined because the decision as to whether the process is in or out of control is determined by the data within or outside of these values.

Centerlines and Control Limits for $\bar{X}$-chart

Recall our discussion of the normal distribution. Let us define zones based on 1, 2 and 3 standard deviations above and below the mean (target) of the normally distributed quantity. The zones are labeled C, B, and A below. The mean (μ) of the data is the centerline which is the target. As previously discussed, in a normal distribution a certain amount of the collected data fall within specific zones. The C zone contains the 68% of the data that is $\pm 1\sigma$ from the mean. The B zone contains the 95% of the data that is $\pm 2\sigma$ from the mean, and the A zone contains the 99% of the data that is $\pm 3\sigma$ from the mean. (Recall that σ is the symbol for standard deviation.) In control charts, the center line represents the mean values of a large sample of data for a process that is in control.



Now let's flip this normal distribution curve on its side and develop the $\bar{X}$-*chart*. Again, the centerline is the mean, μ, and the $\mu \pm 3\sigma$ values are referred to as the Upper Control Limit (UCL) and the Lower Control Limit (LCL). The $\mu \pm 2\sigma$ lines are sometimes indicated and used as warning limits.

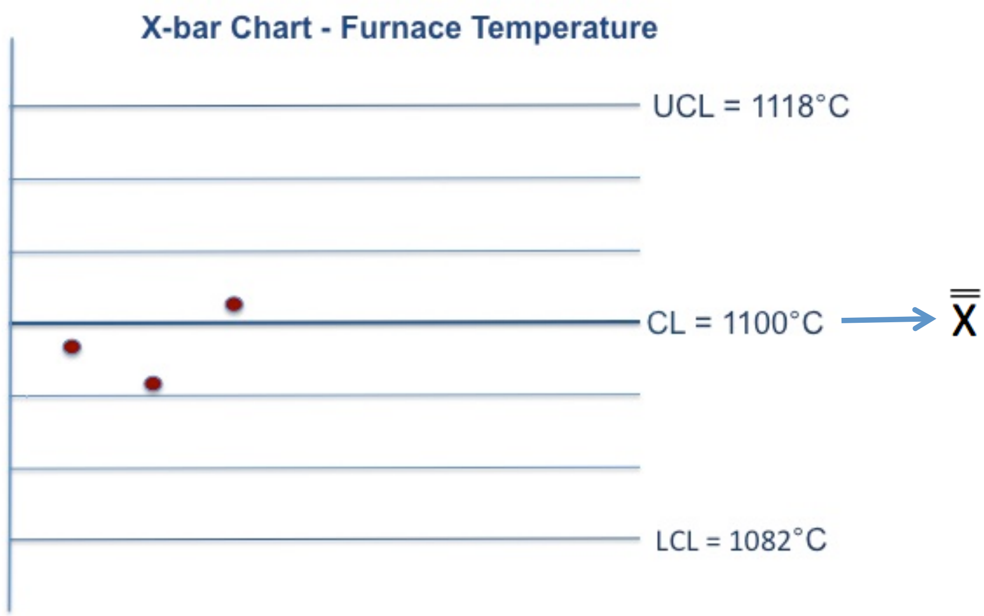


Remember that the centerline and the upper and lower control limits on this graph were calculated using a large sample of collected data points. It is better to have as much data in the sample as you can. This is because this data sample is used to calculate the parameters within which to determine whether or not the process is in control. As in the normal distribution curve, on this $\bar{X}$-*chart,* the centerline is the mean of the sample data and the three lines above and below the centerline are based on the standard deviation of the data and represent 1, 2, and 3 standard deviations away from the mean. Let's put some numbers on this graph representing the process temperature of a diffusion furnace used to fabricate MEMS.

Over several thousands of temperatures collected from the process diffusion furnaces, it was found that the mean temperature was 1100° C. The standard deviation for the collected temperatures was 6° C. Therefore, the upper and lower control limits are the mean ±3 σ or 1118° C and 1082°C, respectively. That tells us, that when oven temperatures are randomly collected during the process, the mean of those collected temperatures should fall, 99% of the time, between 1118° C and 1082°C.

So let's collect some data. Part of your job as a technician is to record an average run temperature for each process run. So this chart shows three average run temperatures for 3 different runs. For example, let's say that your three average run temperatures were 1098 °C, 1095 °C, 1101 °C. Take those data points and plot them on the graph. Your plot should look like the one below. What would you say about the process temperatures that you just plotted? Your temperatures definitely fell within the upper and lower control limits; actually, they were within $\pm 1\sigma$ from the mean. Therefore, the furnace temperatures are behaving the way you would expect them to.

You may sometimes see X-bar, $\bar{X}$ and X-bar bar, $\bar{\bar{X}}$ represented in control charts. Let's talk briefly about the difference between the two. It may not always be reasonable to plot the measurement data for each and every data point. Let's say you took several temperature measurements during the furnace process. The temperature measurements may vary during each of the furnace runs. You wouldn't want to plot each and every temperature measurement. You would want to plot the *mean* temperature measurement for the run. The mean temperature measurement for each run would be represented by $\bar{X}$. Thus, the centerline, target, or average of the mean of the temperature measurements would be represented by $\bar{\bar{X}}$. So on the chart below, each data point is actually an $\bar{X}$ value, or the mean of the temperature measurements for a single run. And again, the centerline is the average of all of the means of many historical furnace runs.



Once the mean and the control limits are calculated for a process, they stay that way as long as the process doesn't change. However, if the product changes, or the process for a product changes, then the mean and the control limits are revised. The mean and the control limits may also be changed as part of a process improvement plan. Let's say that the oven temperatures stay within $\mu \pm 1\sigma$ for several weeks showing that the furnace temperature variation has improved to where oven temperature are between $\mu \pm 1\sigma$, 100% of the time rather than the predictable 68% of the time (based on a normal distribution). Management may decide to tighten up the control limits by recalculating the mean and standard deviation of the current variation of temperatures. One should NEVER change the control limits to accommodate an unstable process.

Notice that the x-axis of the control chart is time based. The process is monitored over time in order to evaluate if the process is experiencing common cause variations or special cause variations over time. Common cause variations are attributed to things that we would expect (e.g., change in manufacturing room temperature or line personnel). Special cause variations are variations that are caused by changes, extremes or unexpected events (e.g., change in vendors of a product ingredient, low level coolant, leaks in a vacuum line). Again, the Upper Control Limit (UCL) and the Lower Control Limit (LCL) are the $\mu \pm 3\sigma$ values. If the process is experiencing only common cause variation, the data points should fall within this range ~99% of the time and exhibit the percent criteria of a normal distribution (68%, 95%, 99%).

Why are $\mu \pm 3\sigma$ values chosen? Shewhart basically set the "3 sigma" limits based upon several statistical concepts, such as Chebyshev's inequality[3] and the Vysochanskii-Petunin inequality[3]. This lesson does not discuss these inequalities in detail but further discussion can be found in the *Basics of Statistics PK*. Also, recall from our discussion of normal distributions, 99.74% of the data occur within 3 standard deviations of the mean.

Control Limits are not Specification Limits

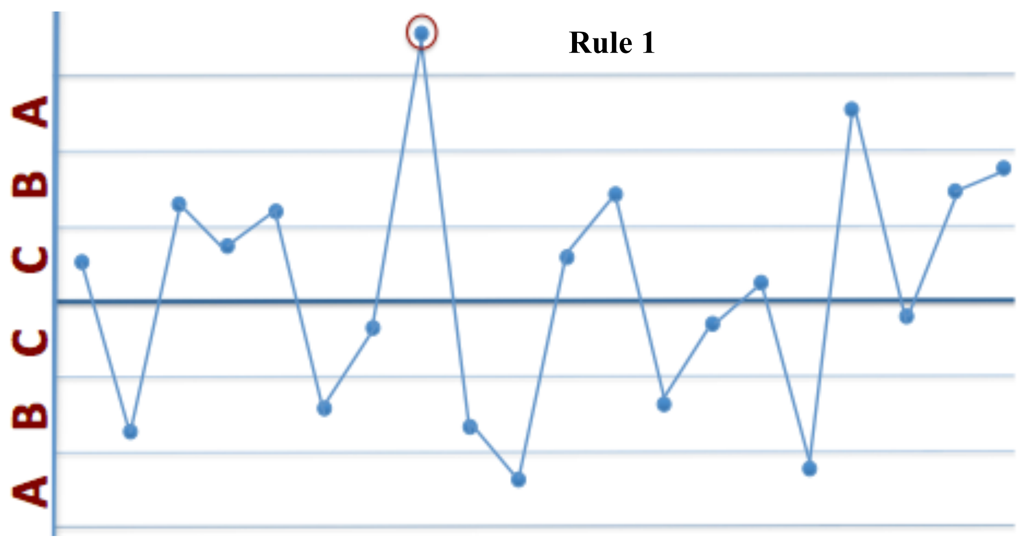
As we just discussed, the control chart centerline (mean) and control limits (standard deviation) are derived from real-time process data. Specifications, however, are the boundaries at which a product is said to be acceptable or not acceptable, not defective or defective based on the customer's criteria. A process can be in statistical control but may be producing a product that is not within specifications. The opposite can also happen. A process may be statistically out of control, but producing products that are within specifications. It is the role of the process and product design teams to setup a process that is "in control" *and* producing a product "within specifications". Generally, it is the role of the manufacturing engineer and technician to keep the process "in control", and the role of the process design team to ensure the when the process goes on-line that it "in control" and "within specifications".

Statistical process control has to do with process "predictability". Process "capability" has to do with consistency and meeting specifications. A process that has a small enough variation to meet specifications consistently, exhibits capability.

As a general rule, specification limits are NOT put on a control chart! The reason is that there is a tendency for people to put all of their attention only on how well the process is doing relative to specifications and to ignore the control limits. Purposes of the control chart are to detect special cause variation, which destroys process predictability, and to detect and prevent process problems before they occur. If all the emphasis is placed on the specification limits, there little attention is paid to the chart, the control limits, and variation in processes. The "voice of the process" is ignored.

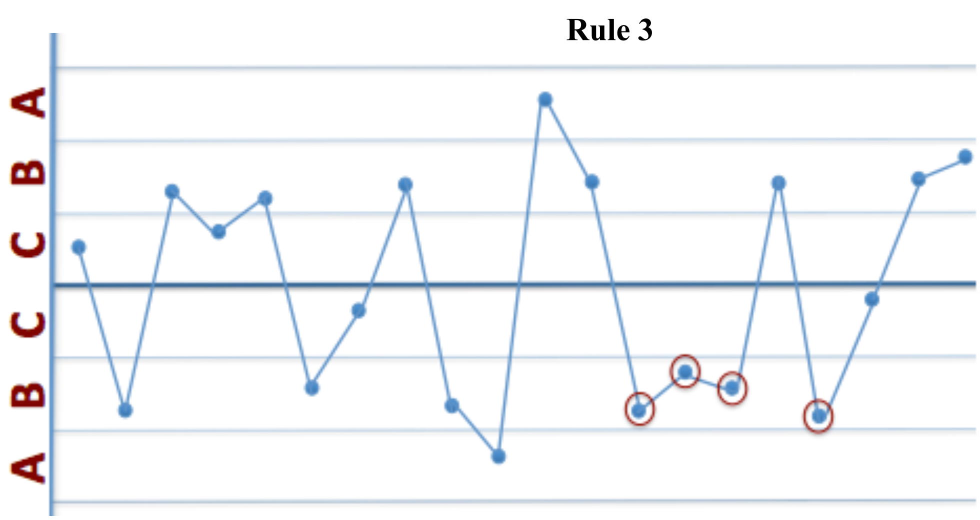
Shewhart Rules (Western Electric Rules)

How can we determine when a process is out of control or not behaving the way it should? Walter Shewhart came up with eight rules that are commonly used in industry today. The eight *Shewhart rules* (also referred to as Western Electric Rules) are used as signals that suggest that the data we are collecting is ***not*** what would be expected if only common causes of variation were present. In other words, the rules are criteria for identifying that special cause variation is currently present in the process. Below is a description of each of the eight (8) rules. Different companies use different rules for their processes, so it's good to understand each rule and why each rule indicates the presence of special cause variation.



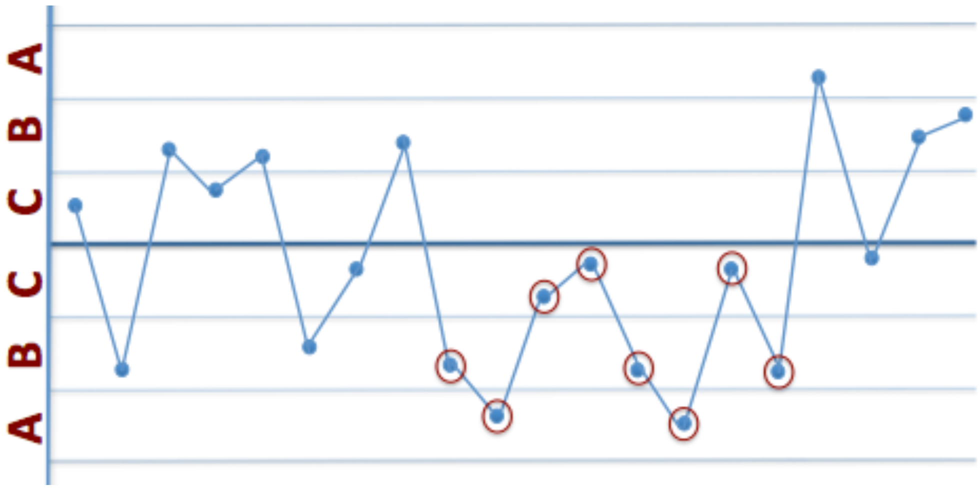
Rule 1: The existence of a number that is not in any of the zones labeled A, B, and C. (See special, encircled point above.) This would be a single point outside the $\mu \pm 3\sigma$ zone.

Rule 2: Two out of three successive numbers in a zone A or beyond (by beyond we mean away from the mean). This would be two out of three successive points outside $\mu \pm 2\sigma$ zone.



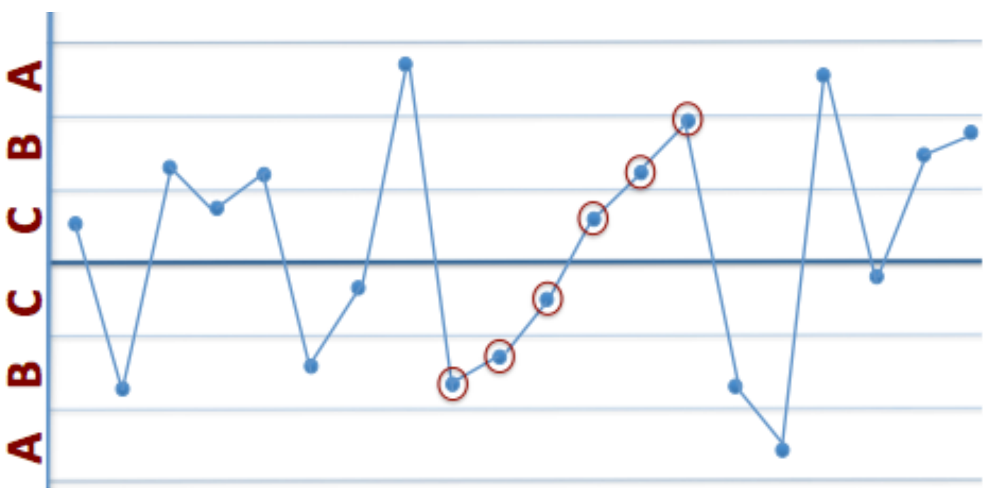
Rule 3: Four out of five successive numbers in a zone B or beyond. This would be four out of five successive points outside $\mu \pm 1\sigma$ zone.

Rule 4



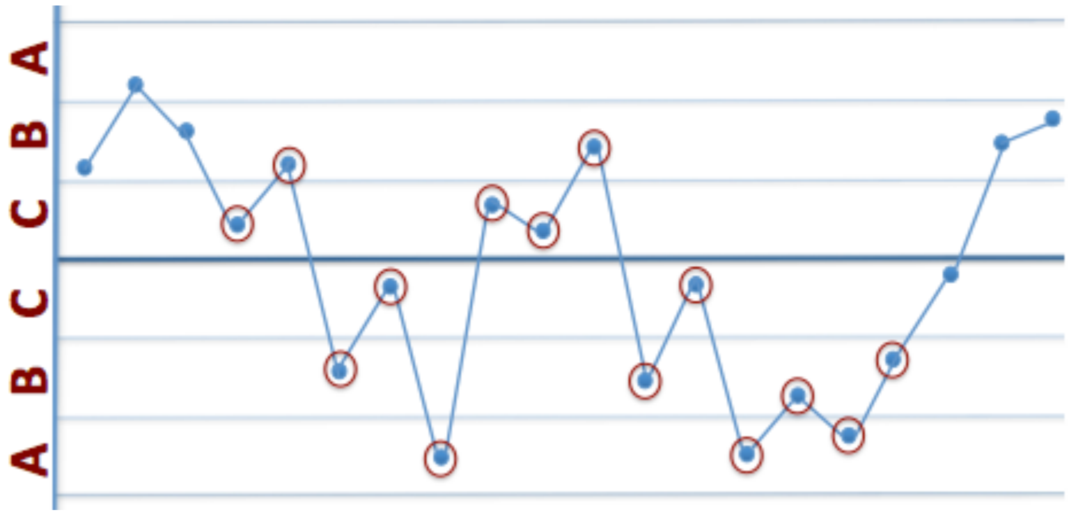
Rule 4: Eight or more successive numbers either strictly above or strictly below the mean (the centerline).

Rule 5



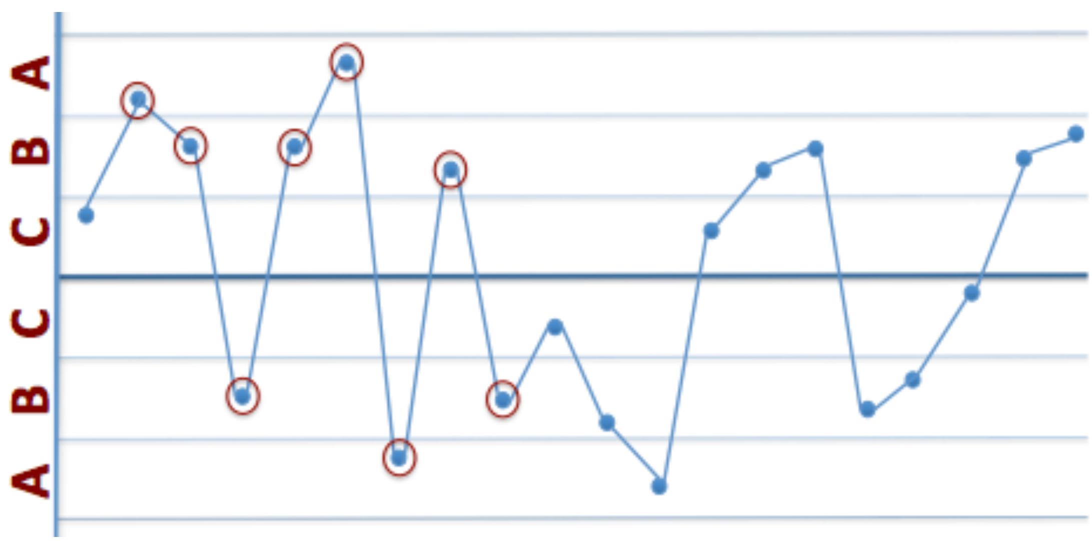
Rule 5: Six or more successive numbers showing a continuous increase or continuous decrease.

Rule 6



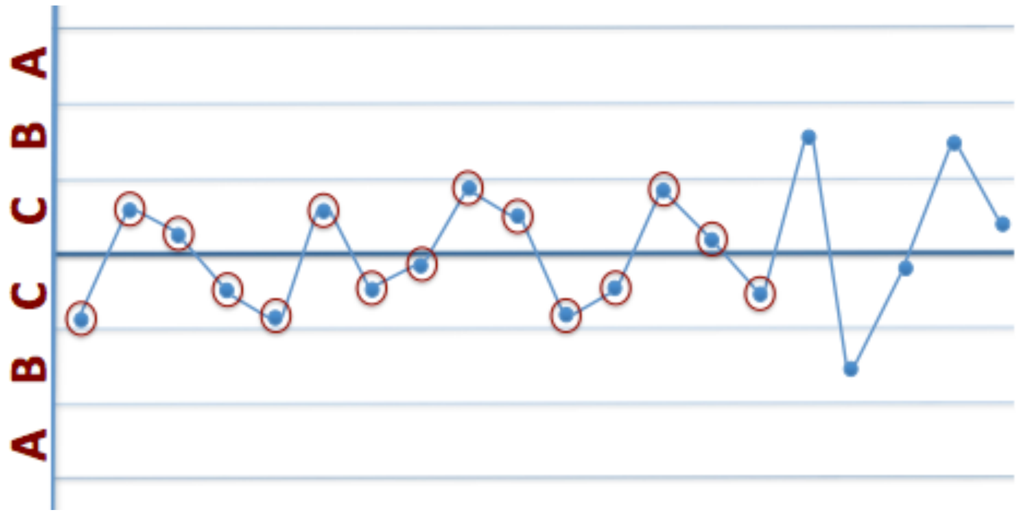
Rule 6: Fourteen or more successive numbers that oscillate in size (i.e. smaller, larger, smaller, larger)

Rule 7



Rule 7: Eight or more successive numbers that avoid zone C.

Rule 8



Rule 8: Fifteen successive points fall into zone C only, to either side of the centerline.

The process is said to be in statistical control if and only if the variation is only due to common causes. If one or more of the above rules is broken, this indicates a special cause variation and the process is not exhibiting statistical control or is said to be Out of Control (OOC). If the process is out of control, the next step is to look for the root cause of the special cause variation and then remove this cause to bring the process back into statistical control.

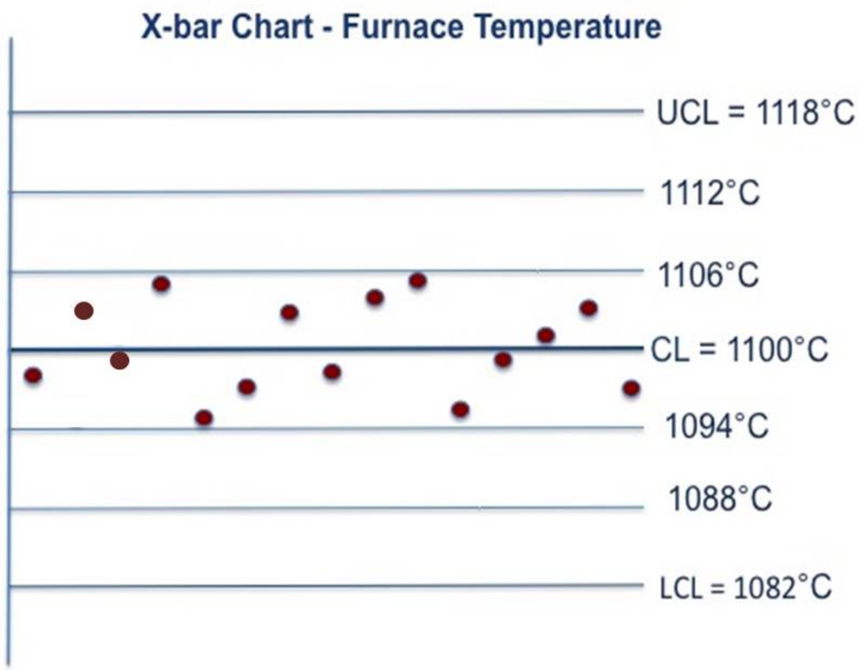
Type I and Type II errors

In the context of control charts, there are two types of errors: Type I error and Type II error.

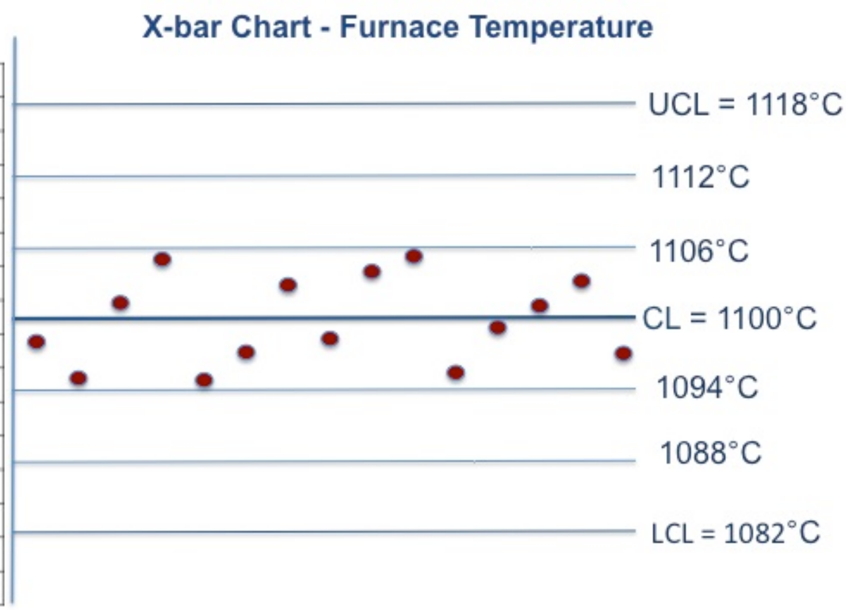
- Type I error occurs when the decision rules (e.g. Shewhart rules) lead you to decide that special cause variation is present when in fact it is *not* present. This would be a false alarm.
- Type II error occurs when the decision rules lead you *not* to decide that special cause variation is present when in fact it is present. This would be a miss.

Let's take our previous example of furnace temperatures and see if you can use the Shewhart rules to determine if the process is in control. You have collected more data from 12 more process runs, calculated the mean of each run, and plotted them it on the SPC chart. Is this process in control? If not, which Shewhart rule did the data violate?

Mean Furnace Temperatures °C
1098
1103
1099
1105
1095
1097
1102
1098
1103
1105
1096
1099
1101
1104
1096



Furnace Temp °C
1098
1095
1101
1105
1095
1097
1102
1098
1103
1105
1096
1099
1101
1104
1096

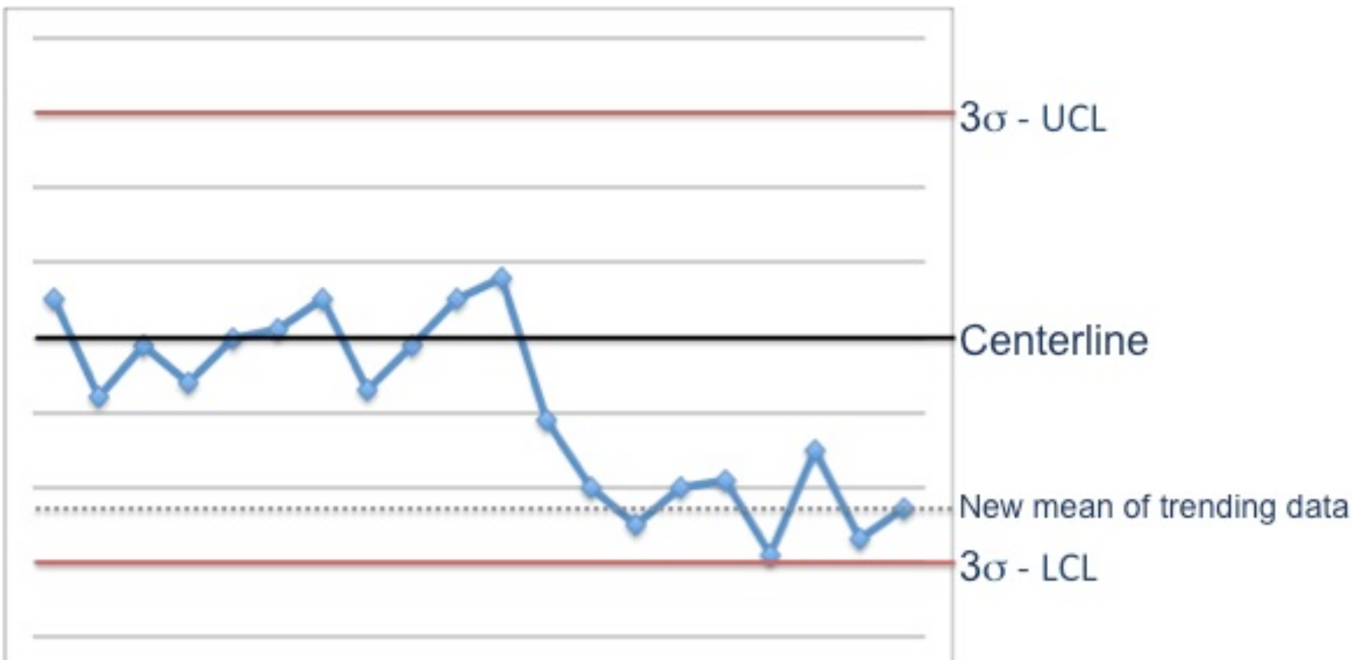


If you said that the process was out of control, then you are correct. The fifteen data points were are all in zone C, or within $\pm 1\sigma$ from the centerline. This is a violation of Rule 8. If any ONE of these points had been outside of $\pm 1\sigma$ and within $\pm 3\sigma$ from the CL, then this process would be "in control".

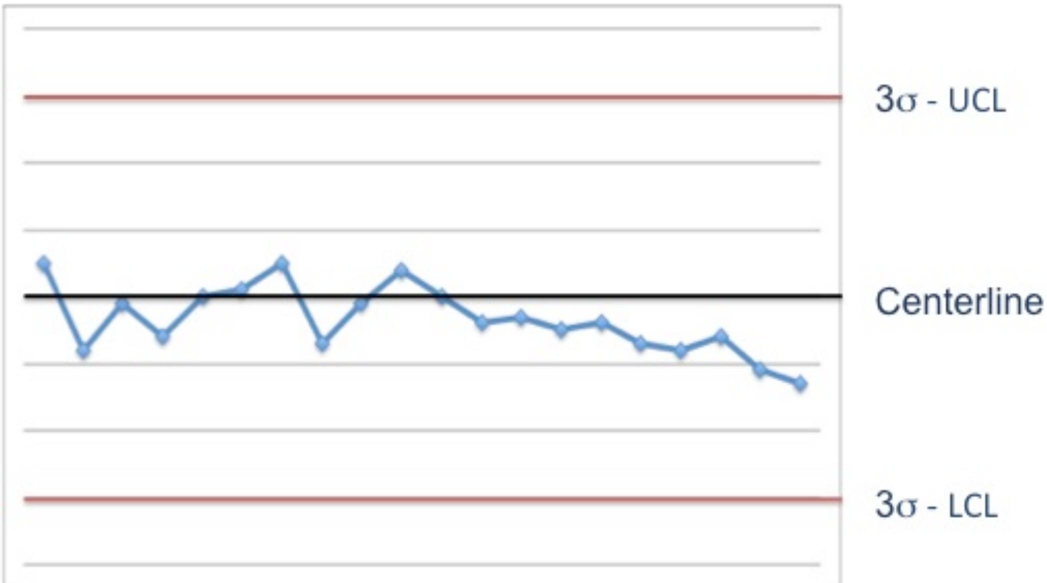
Process Changes

There are many instances when your control chart may not produce an out of control (OOC) signal, but the process may be trending or changing within the control limits. It is important to be aware of these types of changes regardless of the rules you are using to declare your process out of control. There are three different patterns that could indicate a change in the process.

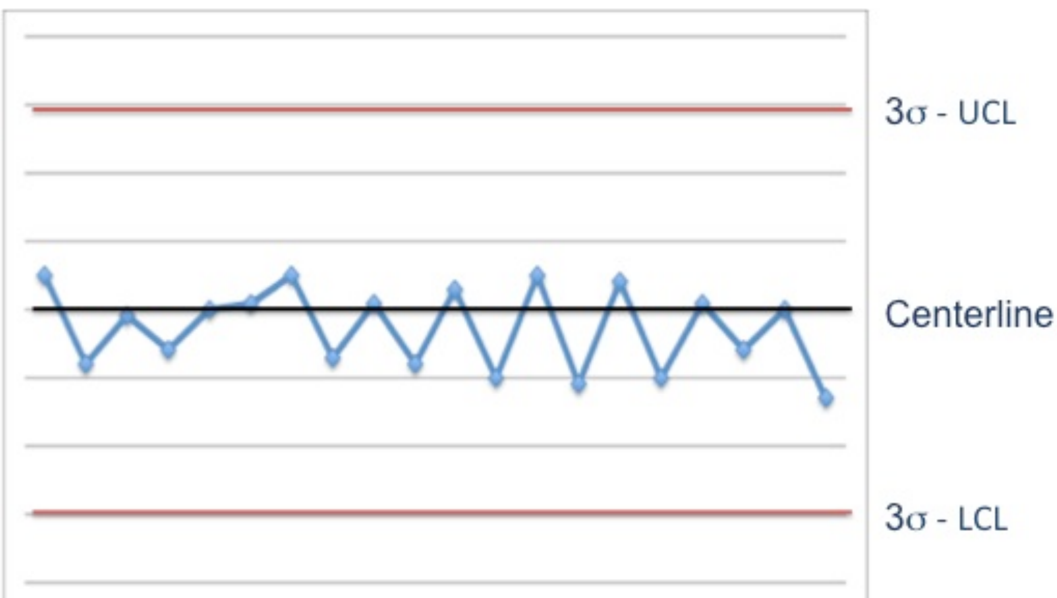
1. Shift – When the data starts to center around a different mean or center line.



2. Trend – When the process mean begins to gradually move in one direction.



3. Cycle – When the data begins to increase or decrease in a cyclical or repetitive manner.

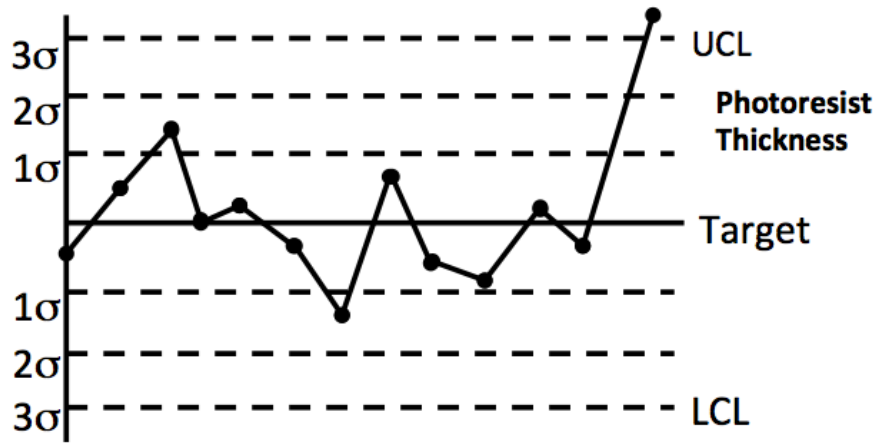


It is always important to keep an eye on the process and notice any changes that are out of the ordinary. A shift in the mean, trending data, and data cycling in a repetitive manner are only a few examples of how data can indicate a process changing. Remember, understanding process variability is a key to a quality product!

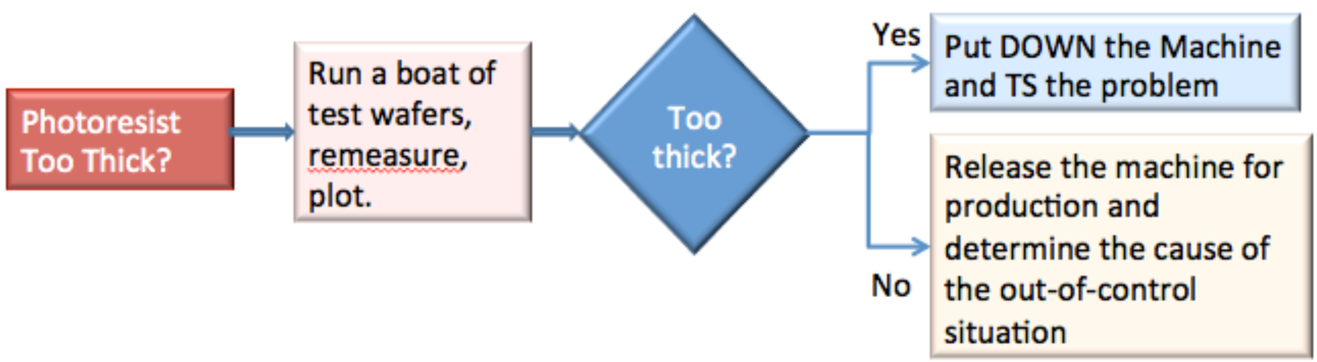
Out of Control Action Plan (OCAP)

So what do you do when you've identified an out of control process? Now it is time to employ some problem solving techniques. A specific problem solving strategy that is used in industry, specific to a SPC chart detecting a process out of control, is an Out of Control Action Plan or OCAP

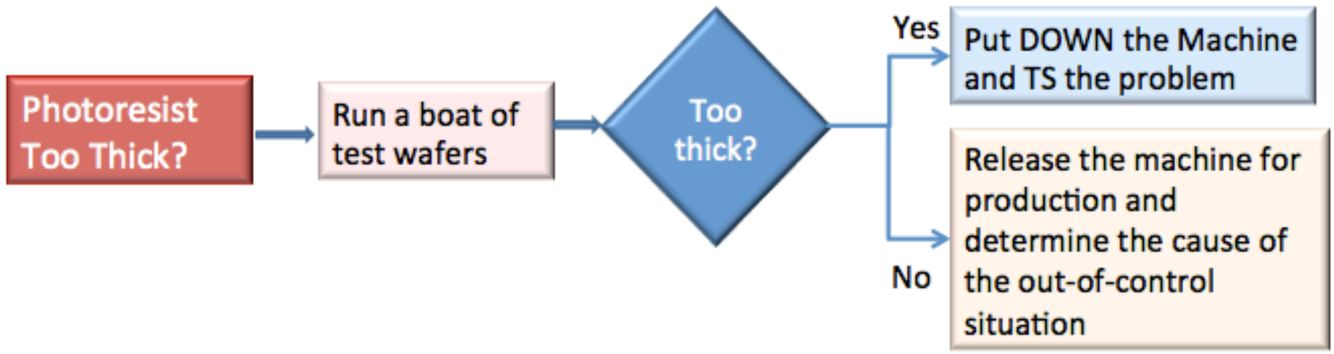
Let's look at an example. You are a technician in the photolithography aisle of a local MEMS fabrication facility. After randomly testing several wafers from the last processing batch and plotting the data on a control chart, you identify an out-of-control situation with resist thickness.

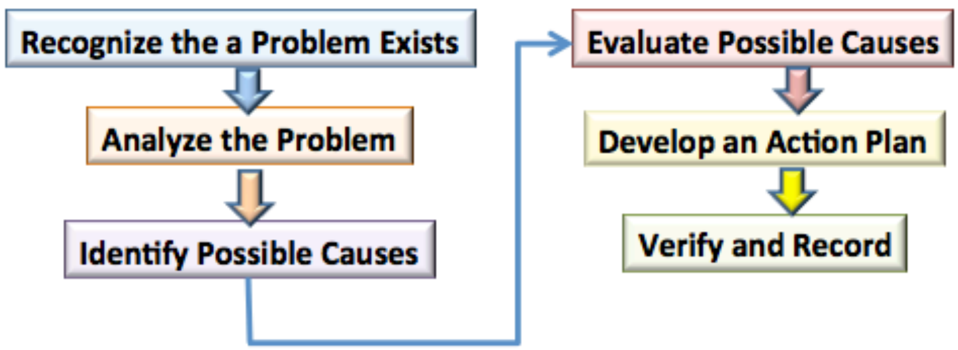


Normally in an out-of control situation like this, the technician would first complete the Out of Control Action Plan (OCAP). Here is a short example of such a procedure. In industry, it would be more extensive than this. The control chart has identified that the resist thickness is out-of-control by having a resist measurement above the upper control limit. The OCAP says to run a boat of test wafer, re-measure, and plot. If the resist thickness is no longer too thick, then release the machines for production and determine the cause of the out-of-control situation (because apparently it has corrected itself, but you don't want a repeat the problem).



So what IF the test wafers do come out with resist thickness that is too thick? At this point the technician would put down the machine and troubleshoot the problem using a systematic approach to problem-solving. For a detailed troubleshooting example, read the Problem Solving PK available at scme-nm.org.





Summary

Statistical Process Control (SPC) is a scientific method that can provide much information about a process and how specific process parameters are varying with time. In order to produce a quality product, the amount of variation should be understood and controlled. SPC is a common set of tools used in industry to help manage this variation. Control charts help to identify variation and identify process issues before they begin to affect the product.

Most process data that is monitored in industry follows a normal distribution. With this assumption, it is easy to employ a basic $\bar{X}$*-chart*. With knowledge of basic statistical concepts, such as mean, variance, and standard deviation, it is easy to create a SPC chart with upper and lower control limits with which to monitor a specific variable or attribute of a process and the variation of the variable or attribute over time. You can then use Shewhart rules, which are standard, rules used in industry to help with the identification process.

Statistical Process Control is meant to be used as a tool. It should be used in conjunction with the knowledge and expertise of the technicians and engineers who run a process. The SPC charts can provide valuable information, and when used along with problem solving strategies, they can help to establish and maintain a quality product.

References

1. Dr. Michael Leeming, University of Arizona
2. Dr. Richard Prairie, Statistical Professor, University of New Mexico
3. Montgomery, D.C. (1985). Introduction to Statistical Quality Control (2nd edition). New York: Wiley.
4. Devore, J.L. (1995). Probability and Statistics for Engineering and the Sciences (4th edition). New York: Duxbury Press.

Disclaimer

The information contained herein is considered to be true and accurate; however the Southwest Center for Microsystems Education (SCME) makes no guarantees concerning the authenticity of any statement. SCME accepts no liability for the content of this unit, or for the consequences of any actions taken on the basis of the information provided.

Support for this work was provided by the National Science Foundation's Advanced Technological Education (ATE) Program through Grants. For more learning modules related to microtechnology, visit the SCME website (http://scme-nm.org).