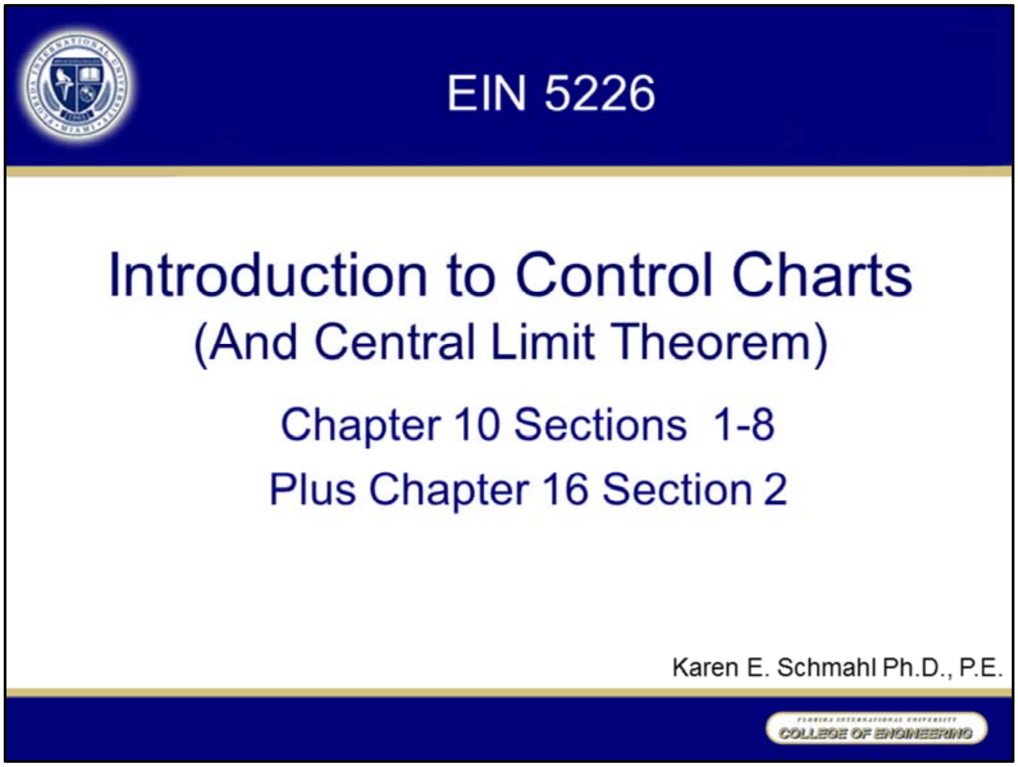
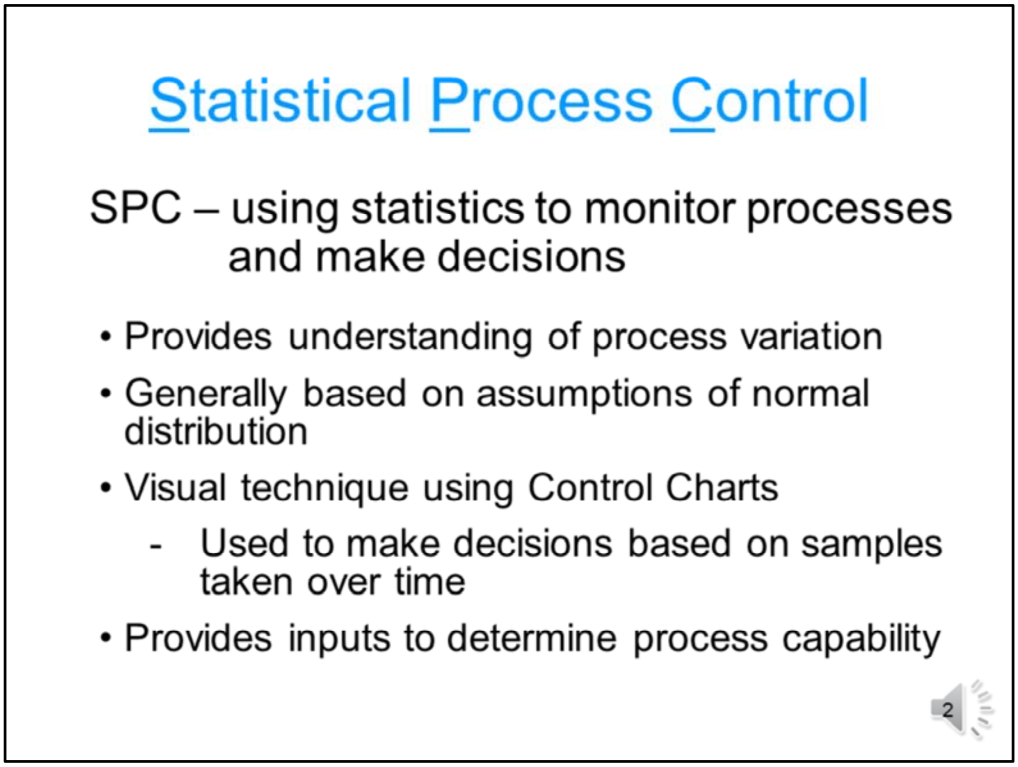


1/16 – major revisions to lecture including

correction to CTL notation, added practice on CTL

moved slides here from next lecture on non-stable processes

moved calc of control limits to next lecture.



Chapter 10 begins looking into the concepts of statistical process control and control charts.

When we use the term statistical process control, we are generally referring to using statistics to monitor processes and make decisions.

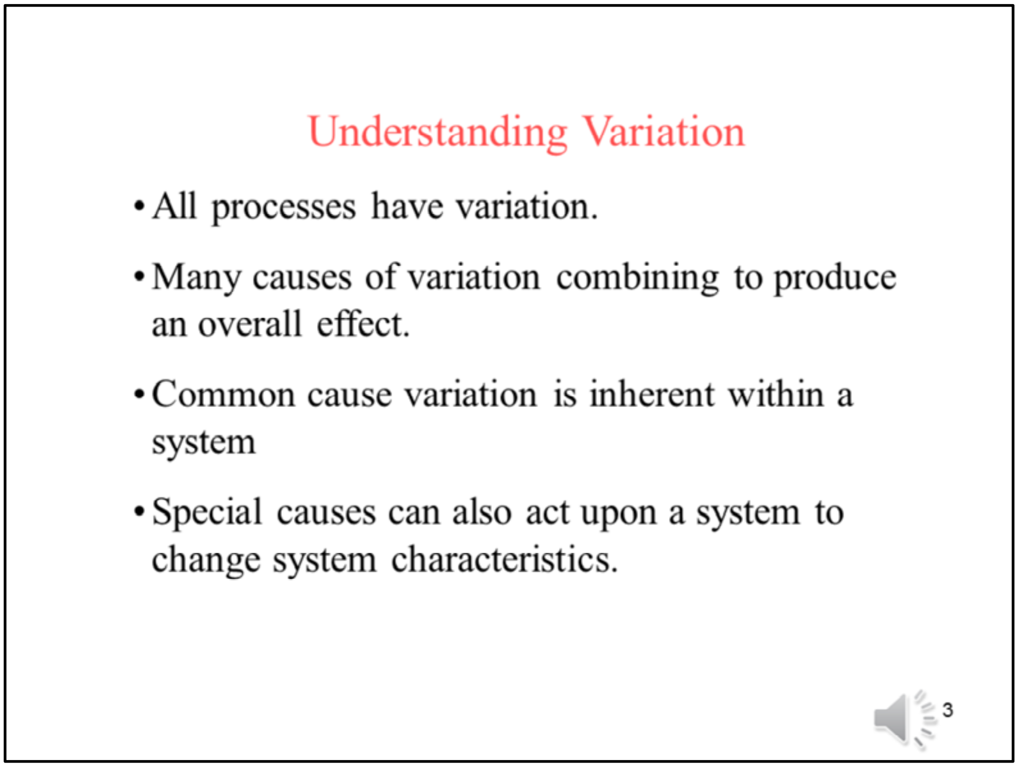
SPC helps us to analyze our processes and to understand the variation that occurs in a process.

When we use statistical process control we are assuming the that characteristics we are monitoring follow a normal distribution.

We are going to be creating control charts – these graphs let us see what is happening with our process and to be able to react appropriately to variation.

The data used in creating control charts also provides information required to help analyze how capable the process is in meeting requirements.

Before we start talking about control charts, we will review some of the concepts on variation. We will also gain an understanding of the central limit theorem in working with the normal distribution.

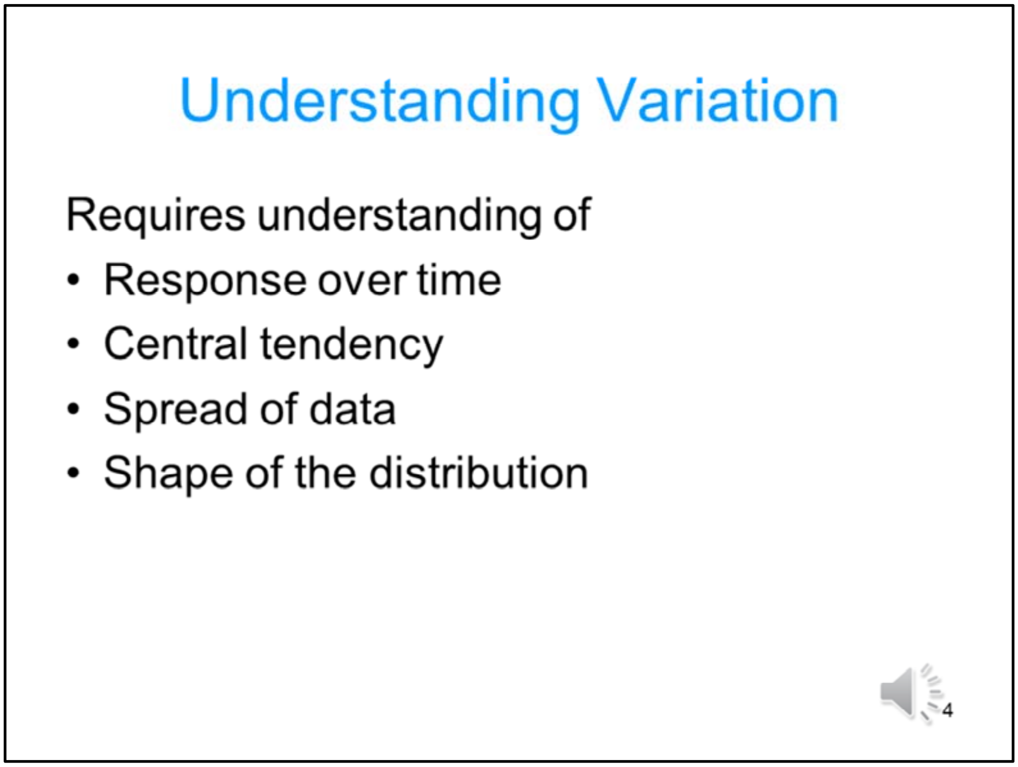


All processes have variation. No two outputs will ever be exactly the same. Sometimes the variation is very evident. Sometimes the variation is so small that you would think there is not any – but variation is present.

In processes, there are many causes of variation combining to produce an overall effect. Equipment, methods of operation, personnel, measuring instruments, the material, even the environment can cause variation.

Variation can be inherent within a system. That is common cause variation.

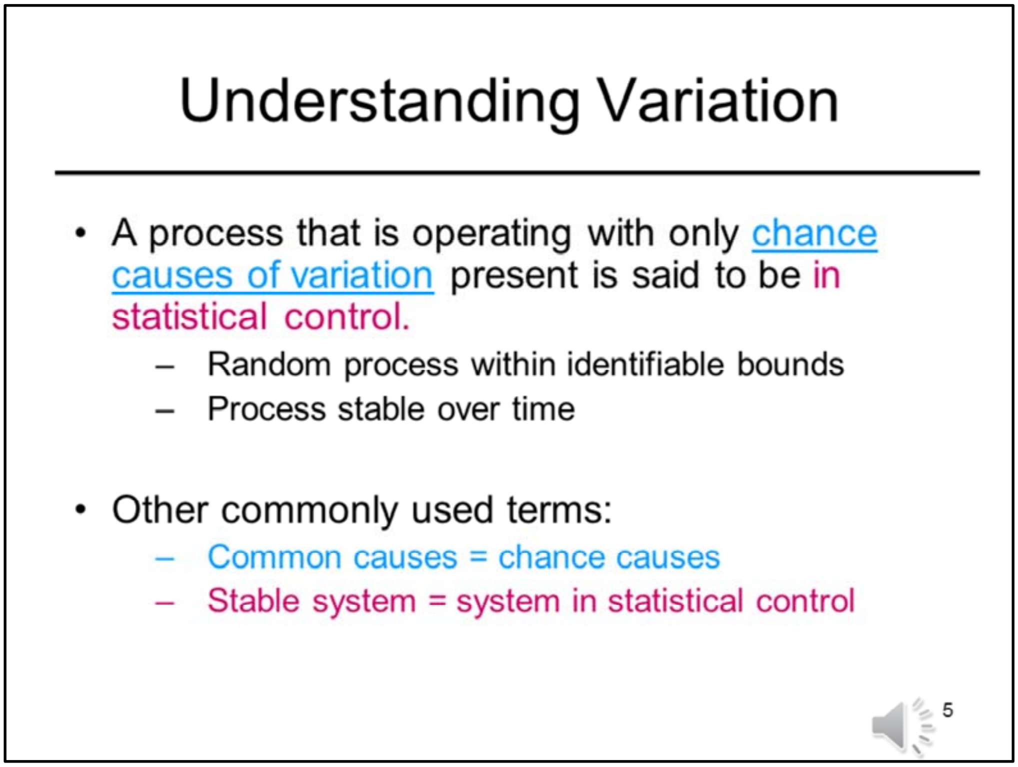
If something unusual happens to disrupt the system or change the system characteristics that would be a special cause of variation.



When we are looking at our data we want to know if there are just common causes of variation present or if special causes are there as well. If the data was taken over time, we need to look at it over time and we need to look at the spread of the data.

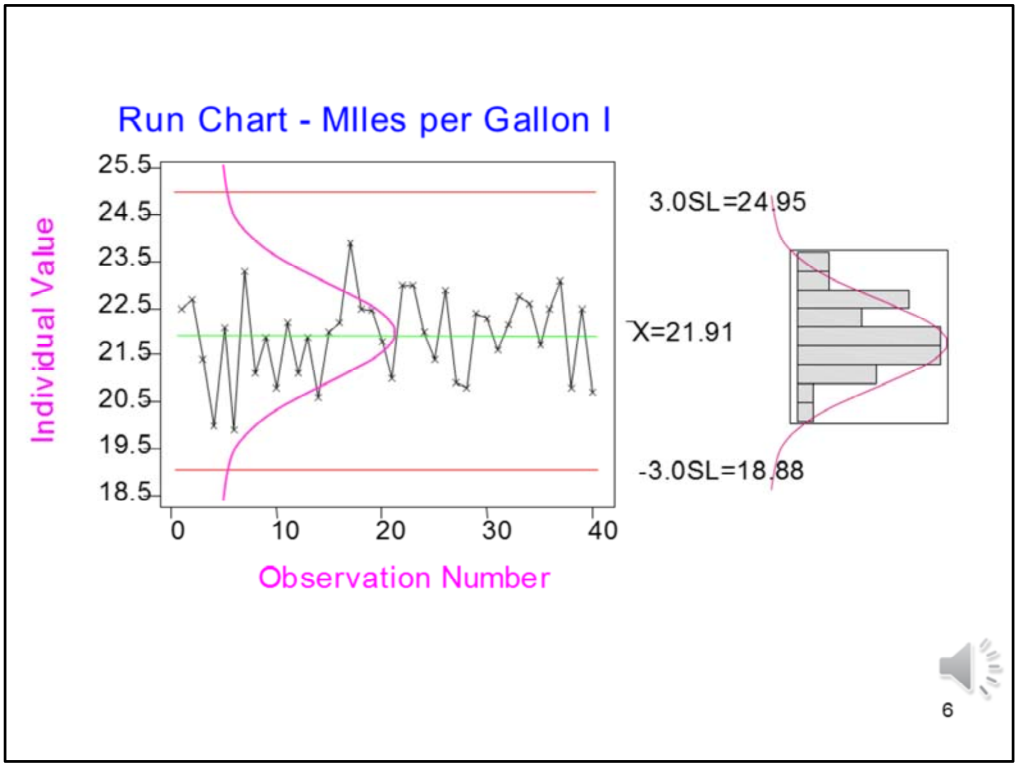
So that we can analyze the data we also want to know where the data is located, what is its central tendency. We are usually, but not always looking, at the mean of a sample when we refer to central tendency.

We said that when we use SPC techniques we are assuming a normal population or that the distribution can reasonably be approximated with the normal distribution.



With SPC, you will be answering the question " is this process in statistical control? Or "Is the system stable?" What that means is that there is only common cause variation present.

When there is only common cause variation the data will be random within identifiable bounds, and the process will be stable over time.



We looked at this miles per gallon data previously. It looks here to be random over time, with approximately equal numbers of observations above the mean and below.

Also to be normal we know that more points should be closer to the mean rather than further away. Remember 68% of the values would fall within plus or minus 1 standard deviation of the mean, and 95% within 2 standard deviations. We can also verify this by looking at the histogram of the data or running a normal probability plot.

I have drawn 3 sigma bounds on the run chart. Recall that when we first started talking about the normal distribution, I said that plus or minus 3 sigma is considered to be the natural process limits when referring to data that is normally distributed.

Some normal review

If an observation is taken from a normal population, what is the probability that the value will be is further away from the mean than 3σ?

Hint

A. 0.0013 B. 0.3821 C. .0027 D. 0.7642

7

Hint: Since Z is the number of standard deviations away from the mean a value is, to work this problem, just look in the standard normal table and find the area in the tail less than Z=-3 and multiply by 2.

A incorrect this is for just one tail.
B incorrect
C correct. It is .0027 rather than .0026 due to some rounding issues with the.0013
D incorrect. This is what was in the tails for +/-0.3 rather than +/-3.

Some normal review

If an observation is taken from a normal population, what is the probability that it is above the mean?

A. 0.75 B. 0.50 C. 0.67 D. 0.97

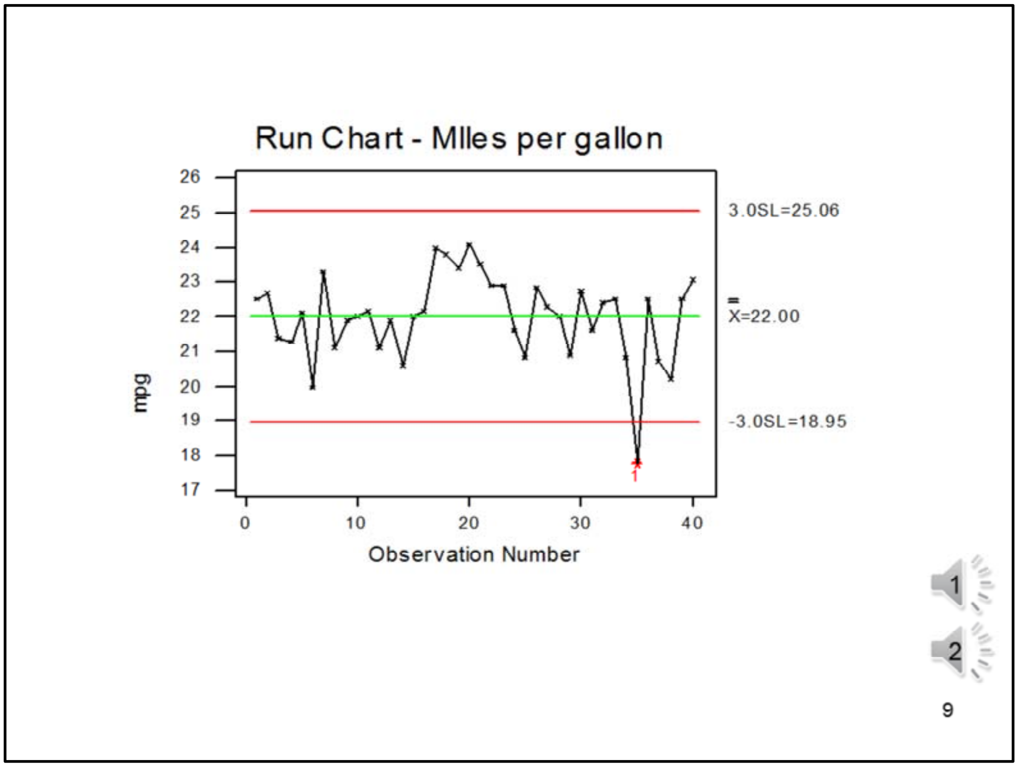
If 7 observations are taken in a row from a normal population, what is the probability that all 7 observations are above the mean?

A. 0.9922 B. 0.50 C. 0.0027 D. 0.0078

8

B correct. Hopefully you did not even have to think about this. Since the normal distribution is symmetric, 50% of the observations will be above the mean.

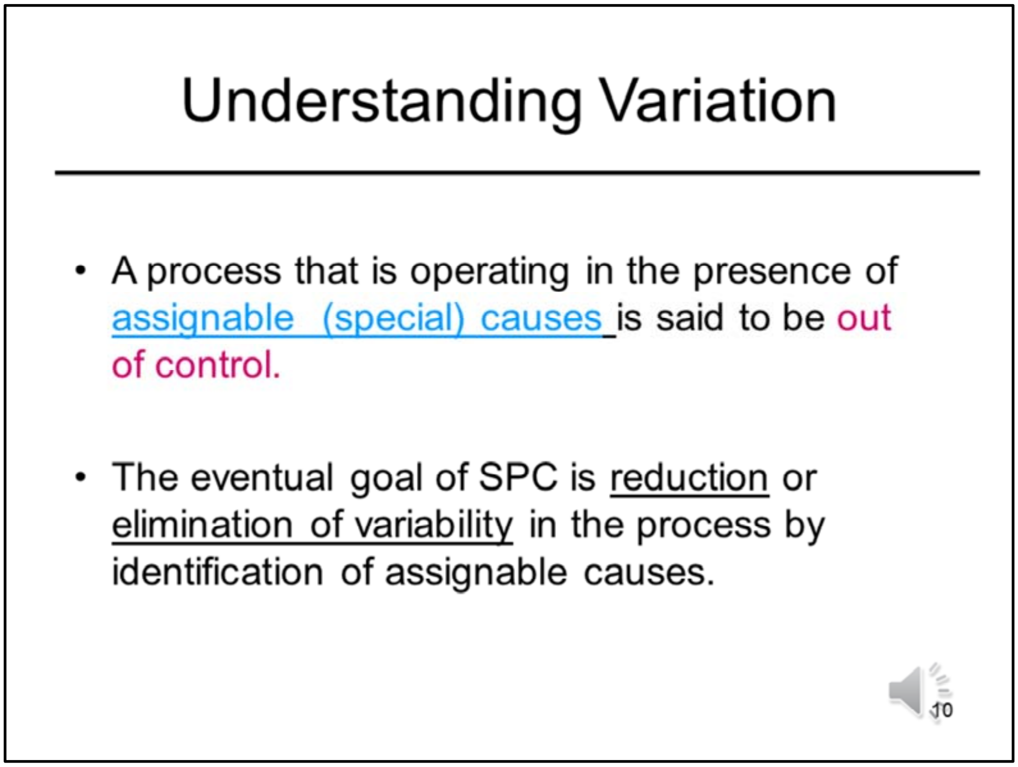
D correct. We are taking independent observations from the population. So we multiply the individual probabilities. So .5 raised to the seventh is .0078. - less than 1% chance of this happening.



1)
So continuing to take data on my gas mileage, I see that I had seven points in a row above the mean. We just determined that the probability of 7 points in a row above the mean if the distribution is from a stable normal population is .0078. So this is likely not data from a stable normal population. This would likely be special cause variation. In fact, this was data from a road trip, so my gas mileage was better than my typical driving habits. We can see this in the run chart.

2)
We later have a point outside of the 3 sigma bounds. What is the probability that a point will be below that 3 sigma limit if the process is in statistical control? Remember that 99.73% of observations would be between the 3 sigma bounds, so only .0027 is outside. So if you split that between the two sides only .13 % of the time will you have a point below the lower bound? So likely this is not common cause variation.

This is what we are going to be doing with the control charts. We want to see if a process is stable and to identify special causes as they occur. Except with a control chart, we are monitoring the process with sample statistics rather than individual observations. By using sample statistics, we can actually identify the potential special causes more quickly.



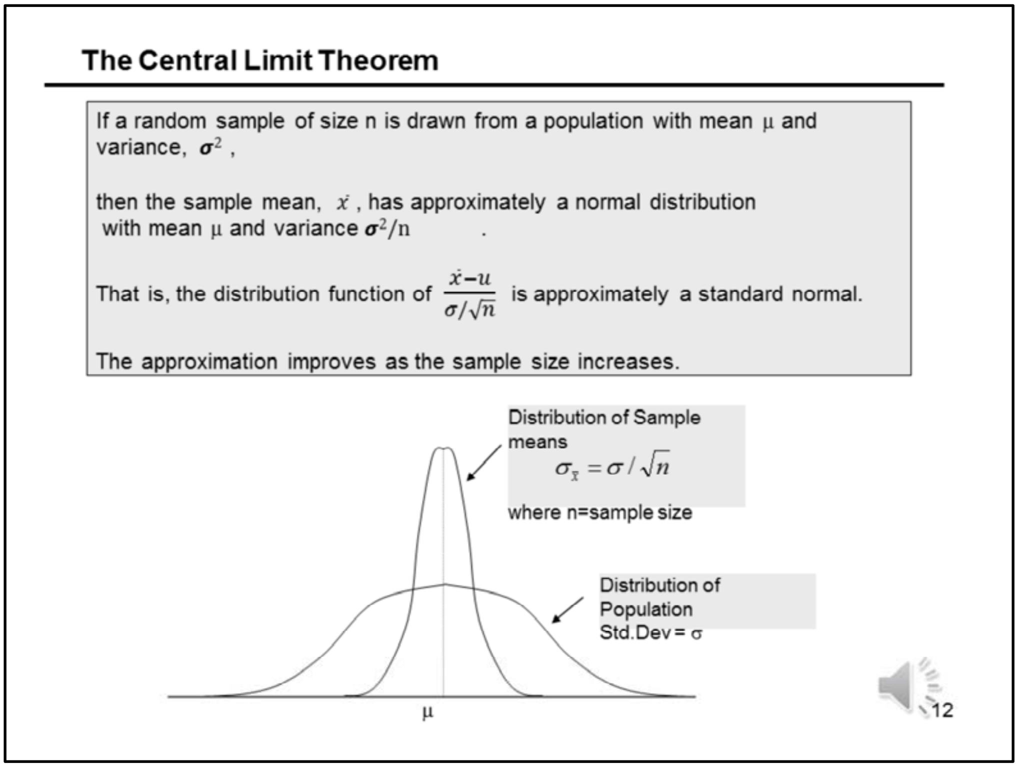
So it looks like on the last slide we had a couple of assignable causes. And from a statistical point of view the process is said to be out of control.

Our objective with SPC is to be able to identify those assignable causes and differentiate them from common cause variation. Once we know what is what, we can work to eliminate variability due to the assignable causes. It will also highlight the magnitude of the common cause variation which may also need improvement.

Decision Making with Sample Data

- Statistics - used to make decisions about processes
- Using sample statistics to make decisions as we monitor a process with control charts
- Need to understand how taking samples, and the distribution of the sample statistic, relates to the population

In order to make decisions about populations with sample data, we need to understand how our sample statistics relate to the population parameters. In the first control chart we are going to be working with we will be plotting sample means. So we will take a sample from a process, calculate its average, and plot the averages. We need to understand how the distribution of sample means relates to the population distribution. This will be explained with the Central Limit Theorem.



Section 16.2 of your text covers the Central Limit Theorem prior to discussing confidence intervals. But the central limit theorem also forms the statistical basis for control charts, so I will be covering it now.

Look at the figure. There are two normal curves, both with the same mean. The bottom curve has a higher standard deviation so it is spread out more. The tall narrow curve has a lower standard deviation. The bottom curve represents the distribution of a normal population with standard deviation of sigma.

From that population, if we were to take some samples of size 2 and average the two observation values, the numbers that we come up with would generally be closer to the mean, than if we took one at a time. If we took samples of size n=ten, and calculated the sample mean, that number would be even closer to the mean. That is what the central limit theorem says – as your sample size grows, the distribution of the average of the sample will be closer to the true mean.

We know the mathematical relationship between these - the standard deviation of the sample means, sigma xbar is equal to the standard deviation of the population divided by the square of the sample size.

And although my figure shows the population as having a normal distribution, it does not matter what the population distribution is, the distribution of sample means will follow the normal distribution.

You will have an exercise to do which will help you understand this as well. But lets look at it phrased one more way, and make sure we understand the math before you do the exercise.

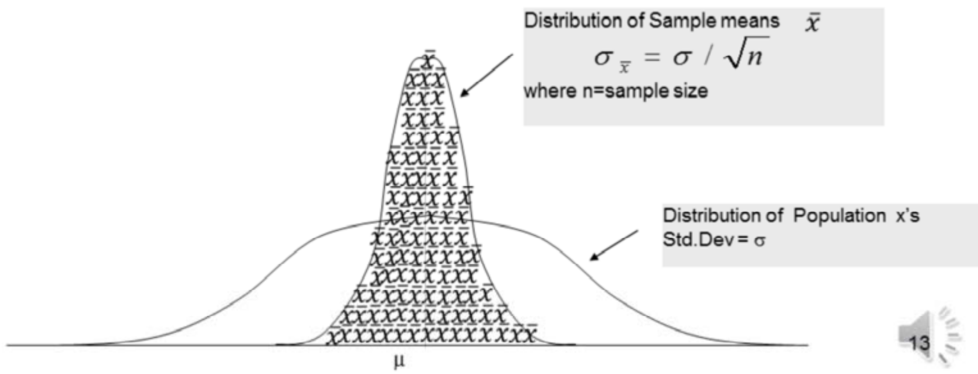
The Central Limit Theorem – Standard Error of the Mean

A population has a mean of $\boldsymbol{\mu}$ and standard deviation, $\boldsymbol{\sigma}$.

If multiple samples of size n are taken from that population, the sample means, $\bar{x}$, are normally distributed with the mean of $\boldsymbol{\mu}$ and standard deviation

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$

This value is called the **Standard Error of the Mean.**



Let me try to state this a little bit differently.

A population of x’s has a mean of u and a standard deviation of sigma. In this case, the population is normally distributed as shown in the bottom curve.

We take multiple samples of the same size, n and calculate the mean for each one. If we were to build a histogram of these xbars, they would also be normally distributed. They would have a smaller spread.

There is a mathematical relationship for the sigma of the population and the standard deviation of the sample means. The standard deviation of the x bars is equal to sigma over the square root of n. This value is know as the standard error of the mean. Lets do a calculation on the next slide, then I would like for you to do the simulation exercise before continuing with the lecture.

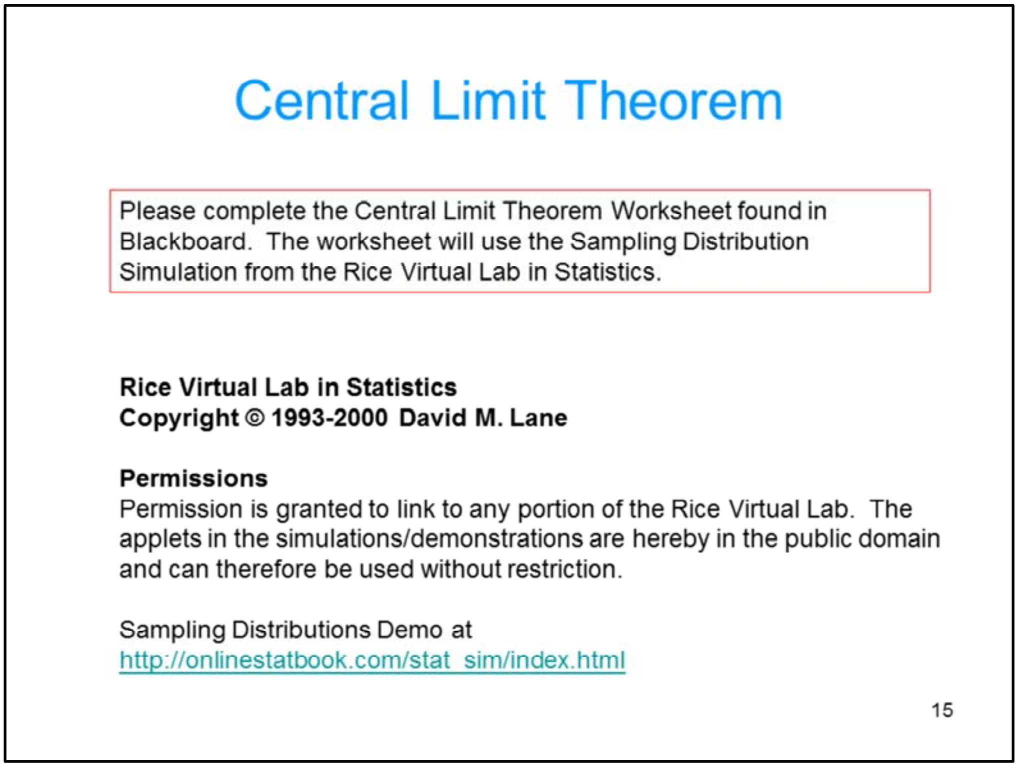
$\sigma_{\bar{x}} = \sigma / \sqrt{n}$

What is theoretical standard deviation, $\sigma_{\bar{x}}$, of the distribution of sample averages of size n=3 where for the population μ= 37 and σ= 1.2?

A) 0.400 B) 0.604 C) 0.693 D) 0.750

14

$\boldsymbol{\sigma}$ xbar = $\boldsymbol{\sigma}$/square root of n = 1.2/ square root of 3 = .693

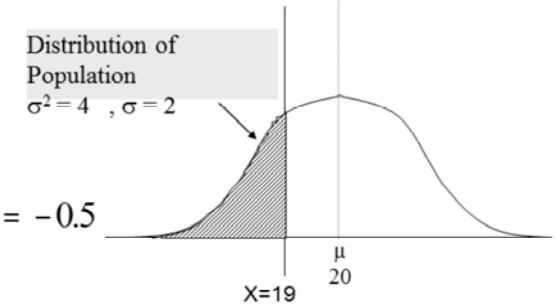


Rice Virtual Lab in Statistics
Copyright © 1993-2000 David M. Lane

Permissions
Permission is granted to link to any portion of the Rice Virtual Lab.

The applets in the simulations/demonstrations are hereby in the public domain and can therefore be used without restriction.

Historical data has shown the battery life to have an average value of 20 hours and a variance of 4. What is the probability that a single battery will have a life less than 19 hours?



$$Z = \frac{x - u}{\sigma} = \frac{19 - 20}{2} = -0.5$$

From normal table:

P(X<-.5) = 0.3085

For the population, we can determine that 30.85% will have a life less than 19 hours.

16

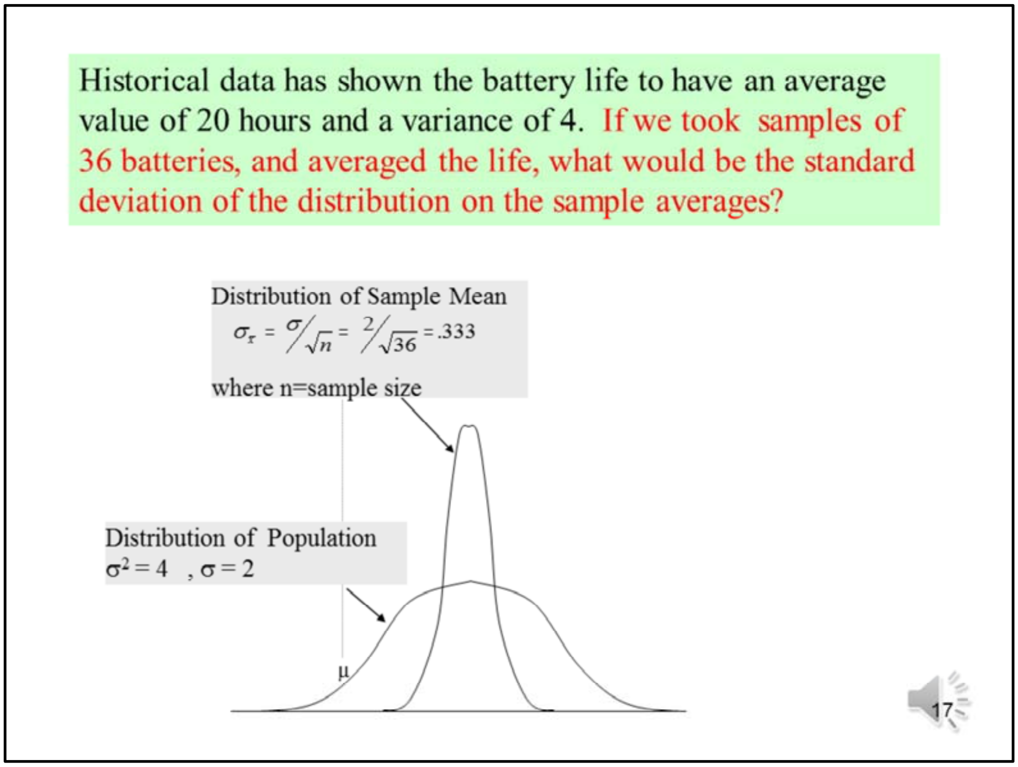
So now that we know about the Central Limit Theorem and the standard error of the means, we need to start learning what we can do with it.

This first problem is referring to the distribution of the population.

Read problem.

We work this as a straight forward normal probability problem.

We identify what area under the curve we want. In this case x less than19. We calculate the associated Z value which is -.5. The we look up the probability in the Z table. And we find .3085. So for the population, we can determine that 30.85% will have a life less than 19 hours.



Now we have our same historical data with the average battery life of 20 hours and the variance of 4. This time the question is ….

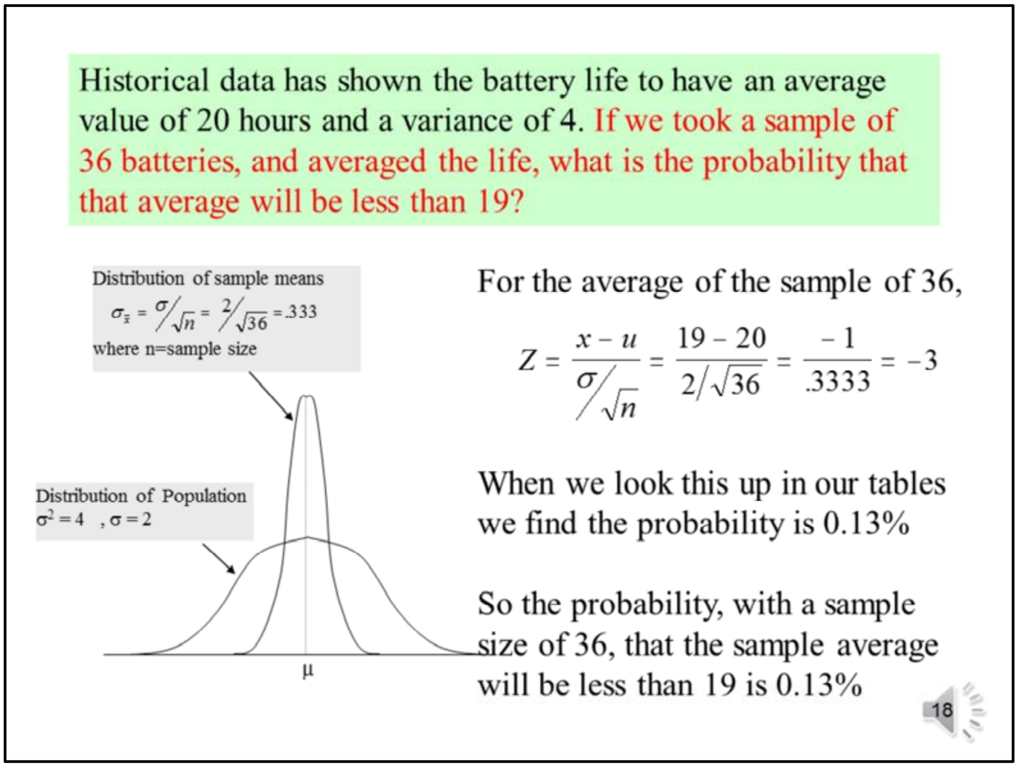
If we took samples of 36 batteries, and averaged the life, what would be the standard deviation of the distribution on the sample averages?

So here we are talking about sample averages

We must use the standard error of the means formula to determine the standard deviation of the distribution of the sample means.

Our sample size, n is 36. the population standard deviation is 2.

Plugging into the equation we get Sigma xbar = sigma / square root of n = 2 over the square root of 36 which is 2 over 6 or .333.



We are looking at the same population again.

Now we want the probability that the average of a sample of size 36 will be less than 19.

So we will use the curve for the distribution of the sample means. This curve has the same mu as the population, 4. We know the x we are concerned with is 19.

We need to calculate Z. In the numerator for Z, we have x-mu or 19 minus 20 which gives us -1

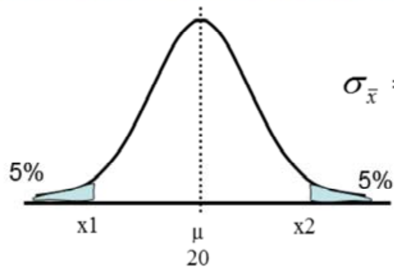
In the denominator, we have to have the standard deviation of the sample means because that is the curve we are working with.

So rather than sigma, we have to use sigma over the square root of n. So in the denominator we have 2 over the square root of 36.

Working the problem through we find Z is -3. Looking up the Z value in the table, we find the associated probability is 0.13.

So the probability, with a sample size of 36, that the sample average will be less than 19 is 0.13%

Taking samples n=36, from a population where the known mean is 20 and the standard deviation is 2. Between what values should 90% of the sample averages fall?



$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 2/\sqrt{36} = .333$$

$$Z = \frac{x - u}{\sigma/\sqrt{n}}$$

From the tables, for 5% : Z = ±1.645

$$x = u + Z\frac{\sigma}{\sqrt{n}}$$

-1.645 = (x1 - 20)/ .333 x1 = 19.45
+1.645 = (x2 - 20)/ .333 x2 = 20.55

Therefore 90% of our sample averages should fall between 19.45 and 20.55

This is again working with the distribution of the sample means. This time I do not show the population distribution as well.

You have worked this type of problem before. See if you can follow it through.

$\sigma_{\bar{x}} = \sigma / \sqrt{n}$

What is theoretical standard deviation, $\sigma_{\bar{x}}$, of the distribution of sample averages of size n=5 where for the population μ= 82 and σ= 3.1?

A) 0.408 B) 1.521 C) 0.620 D) 1.386

If a sample of size 5 is taken from the above distribution, what is the probability that the average of the sample will be less than 84?

A) 0.433 B) 0.925 C) 0.783 D) .877

20

σ xbar = ***σ***/square root of n = 3.1/ square root of 5 = 1.386. D correct

B correct. We are looking at the distribution of the sample means. So when we work the normal probability problem, we need to use the standard error of the means, sigma xbar, to determine Z. Z = (x-mu)/sigma where sigma is sigma xbar. So Z = (84-82)/1.386 = 1.443. Looking that up in the table we find the probability that z is less than 1.443 is .925.

Z=(84-82)/1.386 = 1.443

From Table P(Z<1.443) = .925

A process has a mean, $\mu = 10$, and standard deviation, $\sigma = 1.5$. What is theoretical standard deviation, $\sigma_{\bar{x}}$, of the distribution of sample averages where samples of size $n=5$ are taken?

A) 4.472 B) 2.00 C) 0.671 D) 0.300

For the distribution of the sample means, what percent of the sample means would be within +/- 3 standard deviations?

A) 99.73% B) 99.99% C) 23.58% D) 62.57%

For the distribution of the sample means, the x values at plus and minus three sigma from the mean is

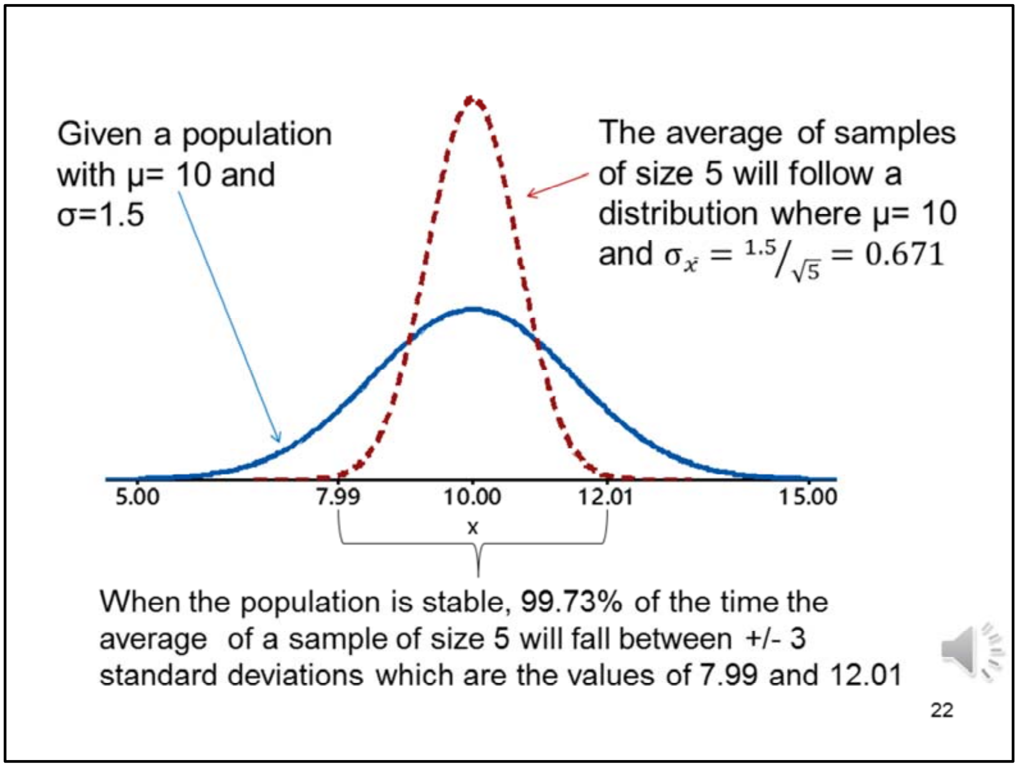
A) 7.45, 12.68 B) 7.00, 13.00 C) 7.99, 12.01

21

A incorrect, did you divide the mean rather than the process standard deviation by the square root of 5?
B incorrect
C correct. 1.5 divided by the square root of 5 = 0.671.
D incorrect. Did you forget to take the square root of the 5?

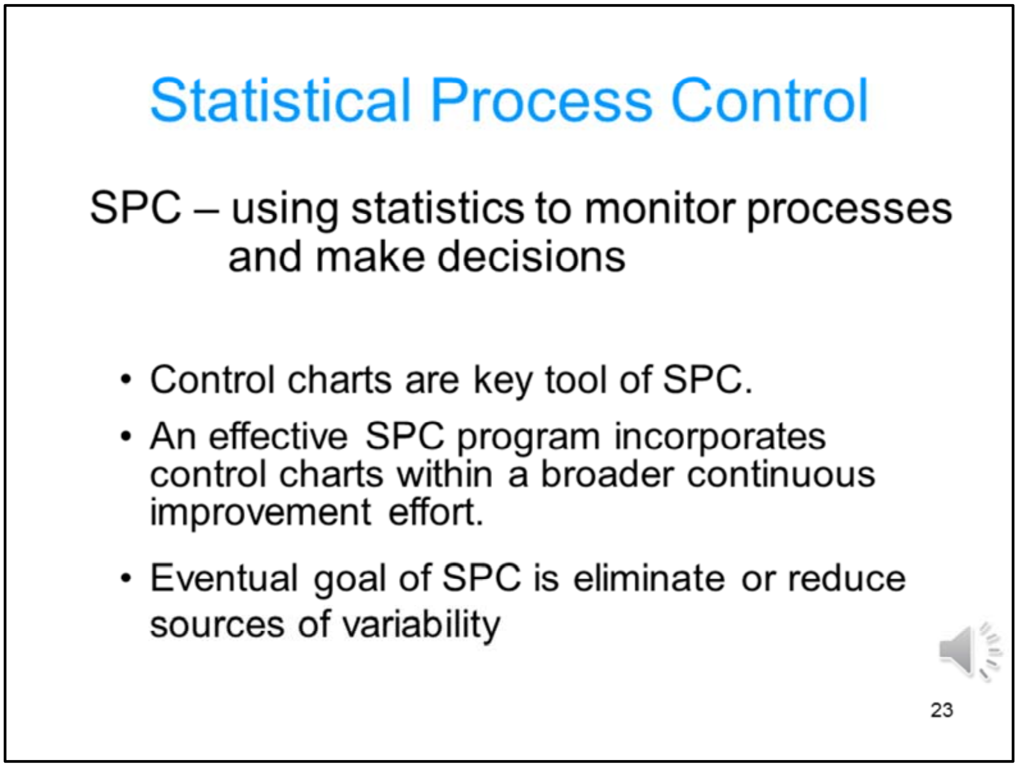
A correct. Remember Z is the number of standard deviations away from the mean you are in a distribution. When we look up Z=-3 in the standard normal table we find the amount in the tail is .0013. Since the distribution is symmetric, the same amount would be in the other tail. Actually those numbers are rounded so the total in the two tails is .0027. The questions asks “within” +/- 3 standard deviations so 1 - .0027 = .9973.

C correct. This is 10 plus or minus .671 times three.



This is a depiction of the problem from the previous slide. A process has an output where the mean is 10 and standard deviation of 1.5. We decide to take samples of size 5. We determined that the distribution of the sample means would have a standard deviation of 0.671.

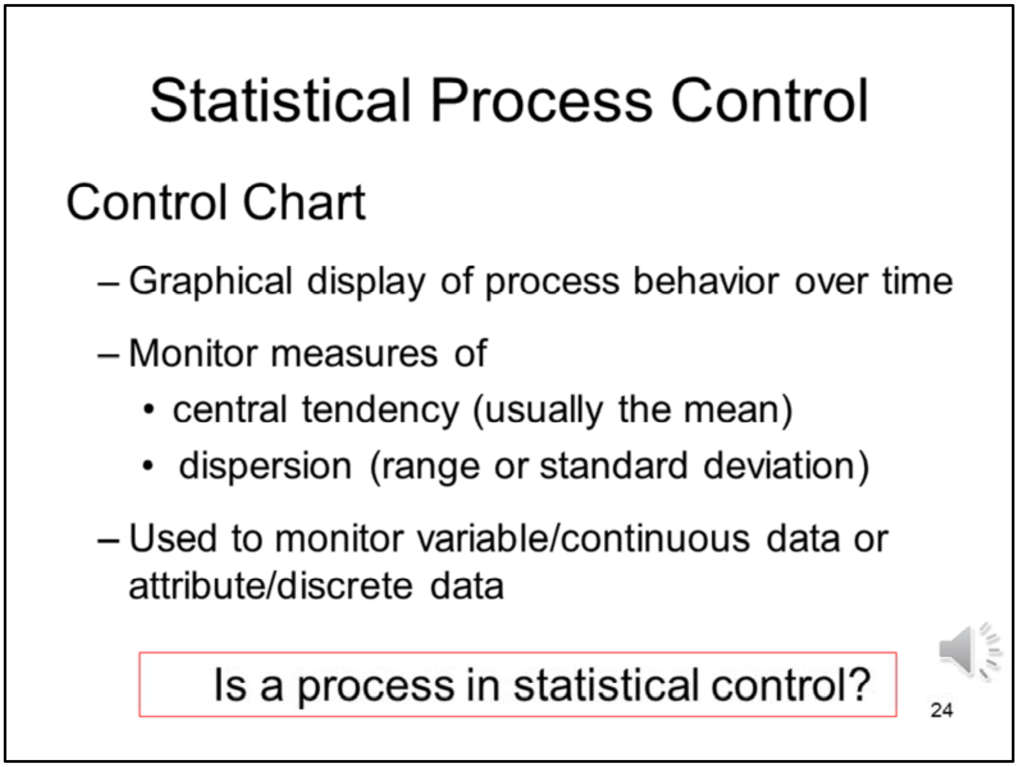
We then determined the values between which the 99.73% of the sample means would be expected to fall. These values are associated with the +/- 3 sigma values for the normal curve for the distribution of sample means.



Statistical process control is using statistics to monitor processes and make decisions..

The primary tool within SPC is the control chart. Unless the control chart is used within a broader concept of continuous improvement it is really not very meaningful. We have to take actions to actually effect positive changes.

The ultimate goal of SPC is to eliminate or reduce sources of variability. That is why it was important that we review the types of variability before we learn how to implement SPC.



A control chart is a line graph which displays characteristics of a processes behavior over time.

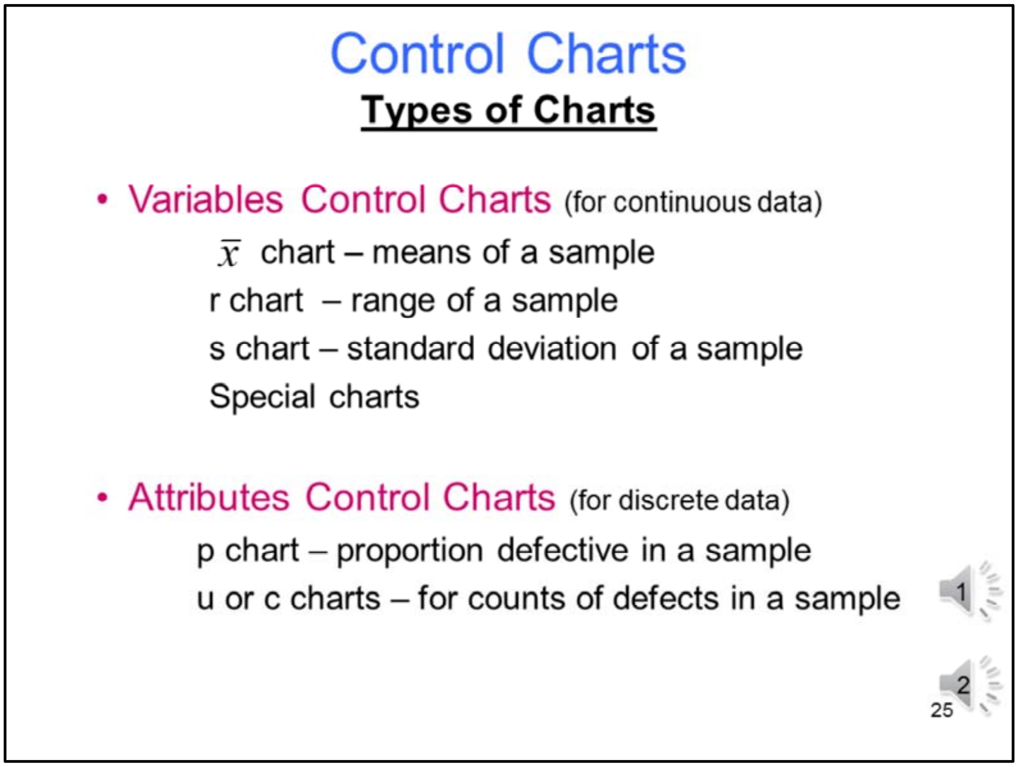
We can monitor the location of the data using a central tendency measure. This is usually the mean, although I have seen charts using the median.

We also want to monitor the dispersion. We will do this using the range or standard deviation.

There are charts for both continuous data and discrete data.

The main question the charts are answering is " Is the process in statistical control?"

If the chart indicates it is not in statistical control, action needs to be taken.



Control charts for continuous data are called variables control charts. And control charts using discrete data are attribute control charts.

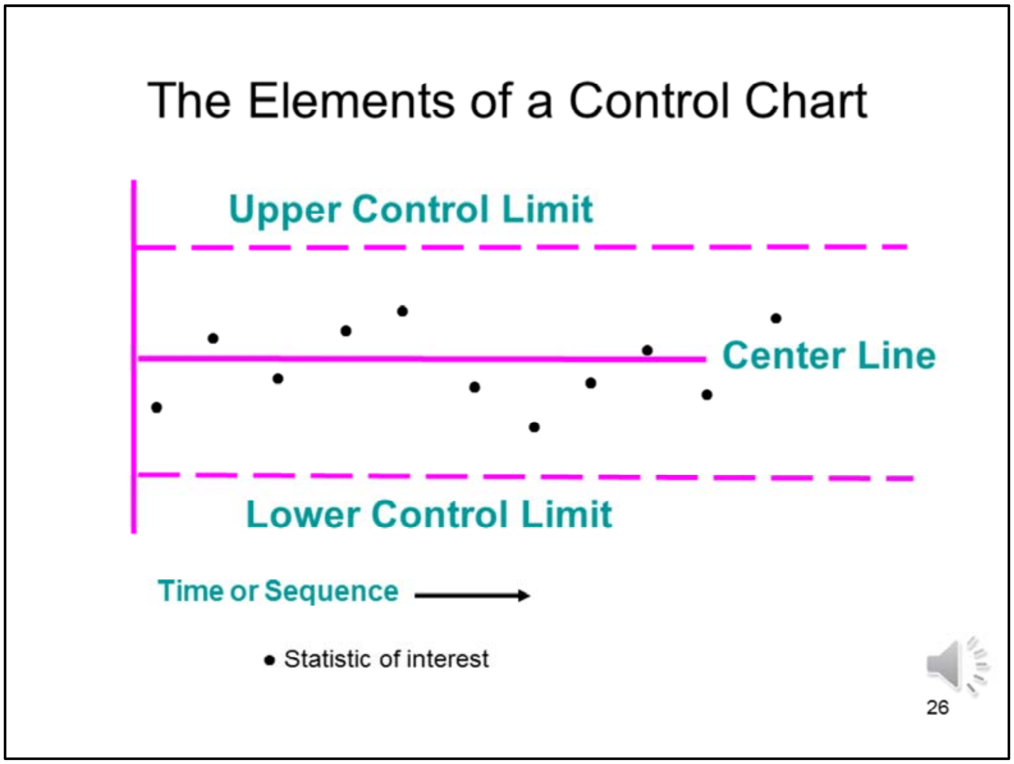
On the variables control chart, we are taking samples then plotting the mean of the sample on an xbar chart. For the same sample, we would also plot a measure of dispersion, either the range or the standard deviation of the sample.

I mentioned that I had seen median charts before. There are also some other special types of variables charts that we will take a look at. The most common variables charts use xbar and r. That is what we will be focusing on in this lecture.

For attribute control charts, the most common is a p chart which plots the proportion defective in a sample. For counts of defects in a sample, there are two types of charts u and c which we will take a look at.

------Clip 2---------

In this course we will not be covering attribute control charts. But I did want to make you aware that they exist. If you are taking, or will be taking EIN 5332 Quality Engineering, we will cover the attribute control charts, as well as the special charts for variables data in that course.



All control charts have the same basic format.

The center line is the average of the statistic of interest.

Then the bounds in which the data will fall are put on the chart as the upper and lower control limits.

The center lines and control limits are established during an initial study of a stable process.

Once established, the chart is placed into use. Samples are taken over time and plotted on the chart to monitor the process.

General Model of a Control chart

$$UCL = \mu_W + 3\sigma_W$$
$$Center \quad Line = \mu_W$$
$$LCL = \mu_W - 3\sigma_W$$

μ_W = mean of the sample statistic, w.

σ_W = standard deviation of the statistic, w.

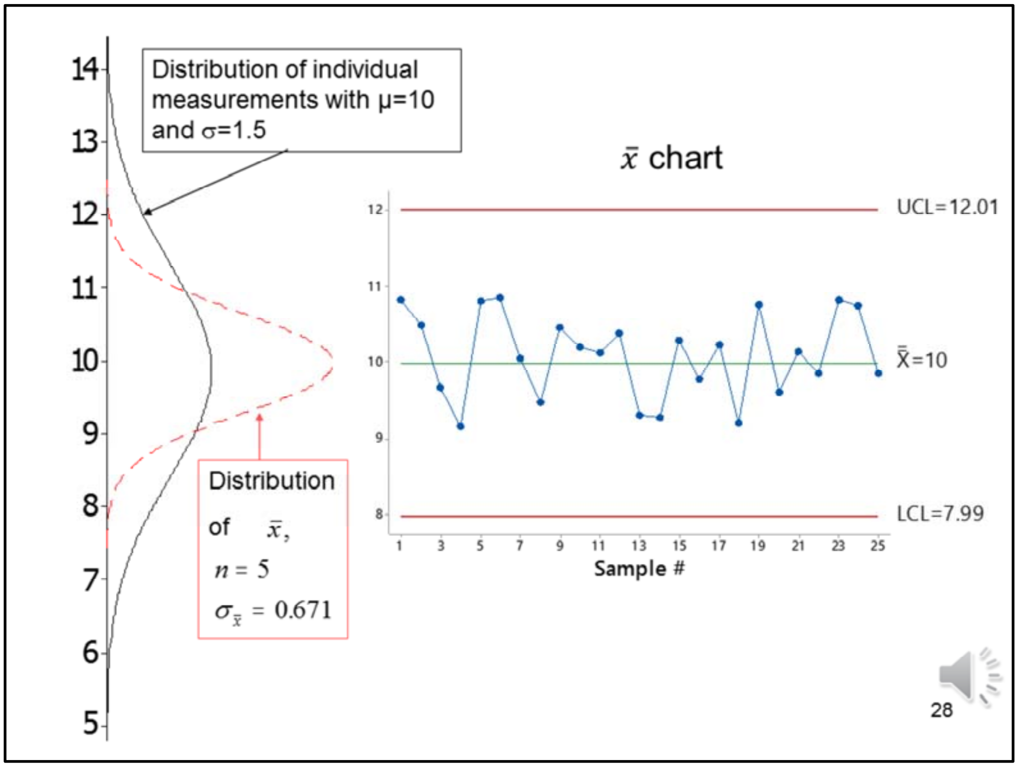
27

Control chart limits are almost always set at plus or minus three sigma from the mean. All the formulas you will be learning in the next lecture will be based on 3 sigma limits.

That means our general model for the control chart the has the upper control limit equal to the mean of the sample statistic plus three times the standard deviation of the statistic.

The center line will be the mean of the statistic.
And the lower control limit is the mean minus three times the standard deviation.

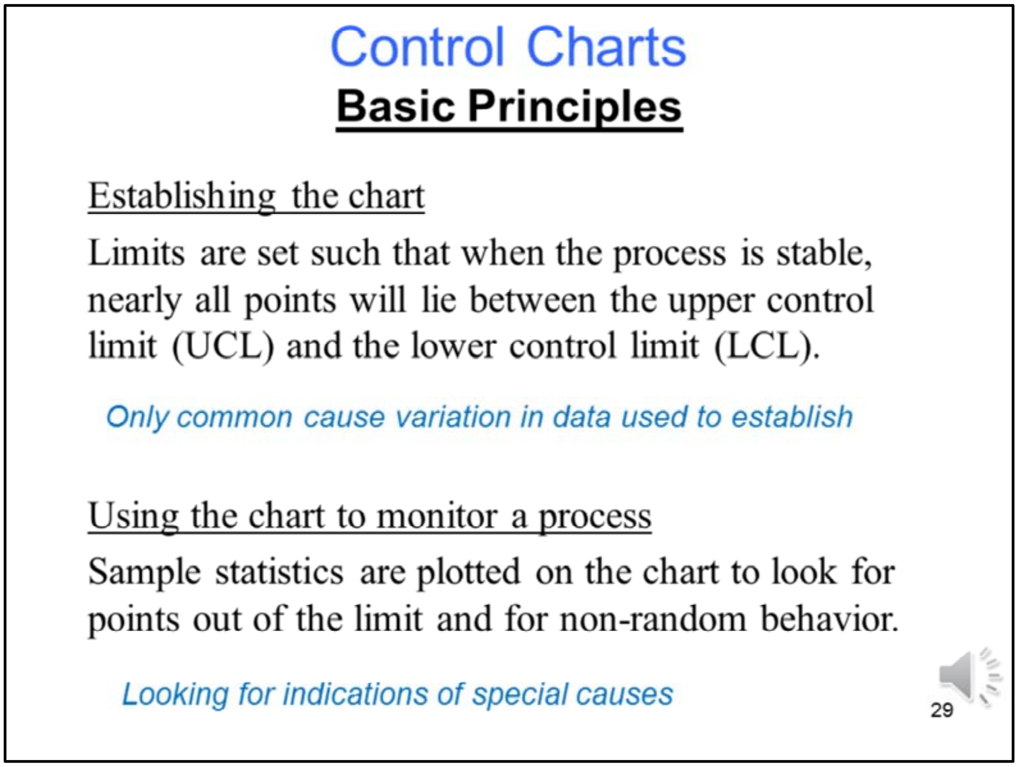
Notice that I keep referring to the sample statistic. Remember that we can have different types of control charts. For variables data will be learning about xbar, r and s charts in this course. That is why I am referring to this as the general model. All the charts have the same basic format. For this lecture I am just focusing on xbar so that you can get the concepts down.



We are able to determine bounds for the control chart, our upper and lower control limits because of the properties of the central limit theorem.

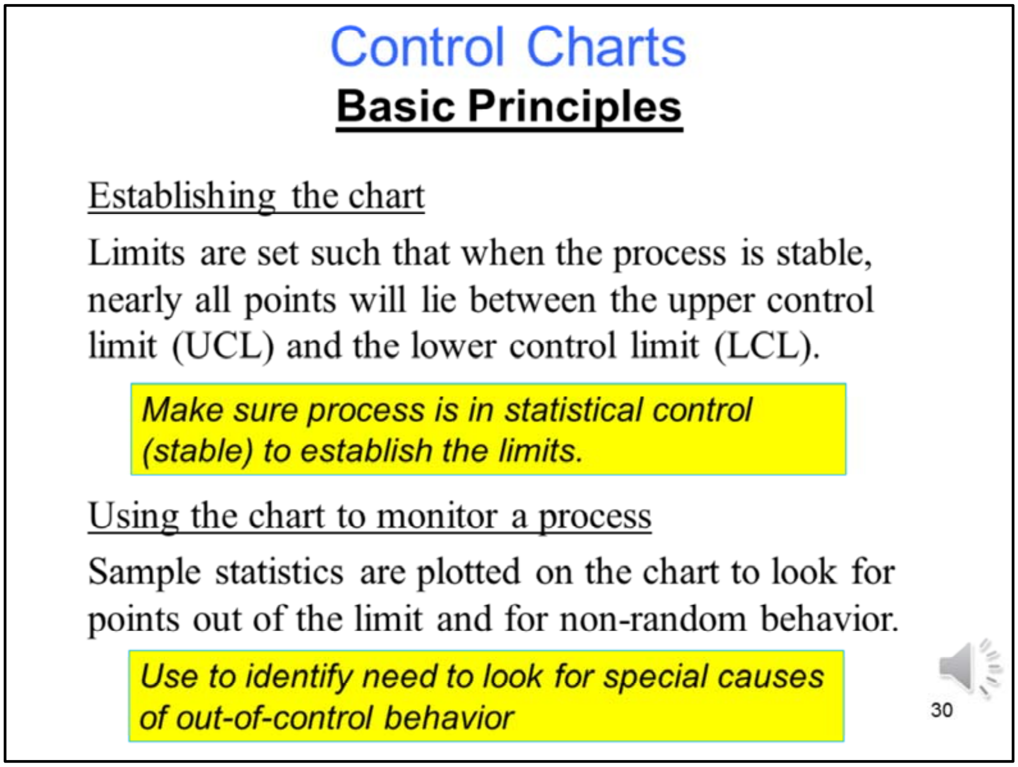
For example, with the x bar chart, we know that sample means for the same size sample are normally distributed. So we set our limits where we know the sample means should fall if the system is stable. The limits are usually set at plus or minus three sigma from the mean, so 99.73 % of the time, the sample statistic, such as the mean in this case, would fall within the control limits.

All the numbers in this example should look familiar to you. You calculated them.



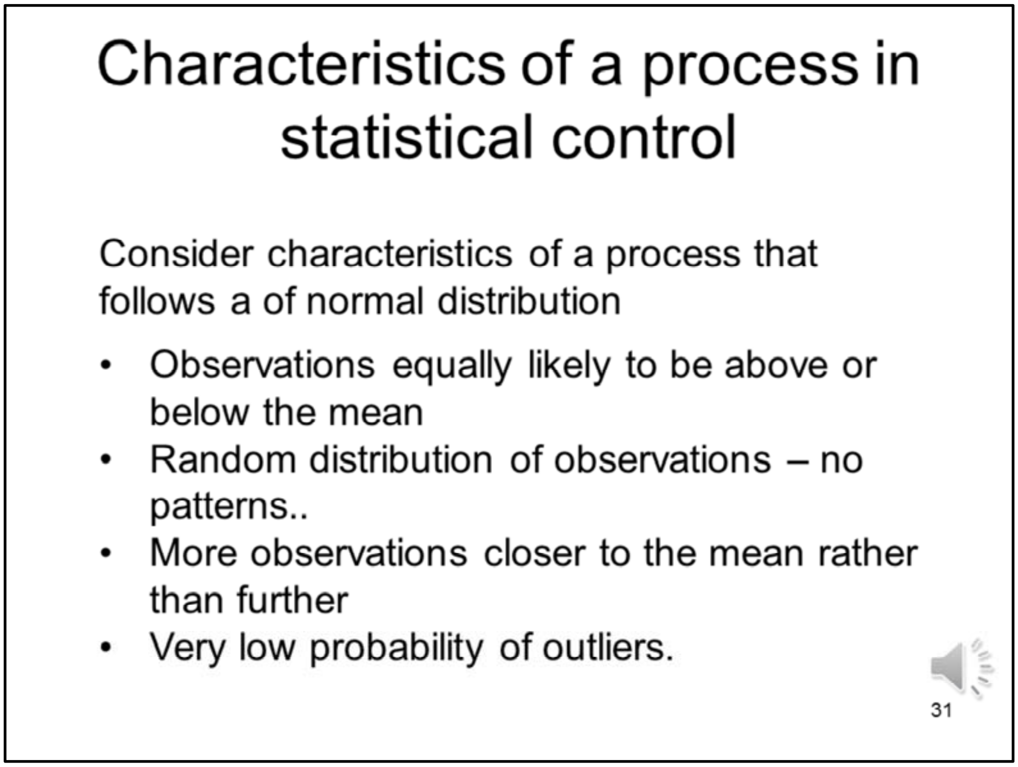
So when we establish our control limits, nearly all the points will lie between the upper and lower control limit. You will see the abbreviations UCL and LCL used. We do have to make sure the process is stable when we set the limits. So we only want common cause variation in the data used to establish the limits.

Once the control limits are established, we can use the chart to monitor the process, by plotting sample statistics on the chart to look for points out of the limit and for non-random behavior. These would be indications of special causes acting upon the process – these causes would need to be investigated.



When we establish the chart it is imperative that the process is in statistical control to establish the limits. We want only common cause variation in the data. So we need to understand what being non-stable or out of statistical control looks like.

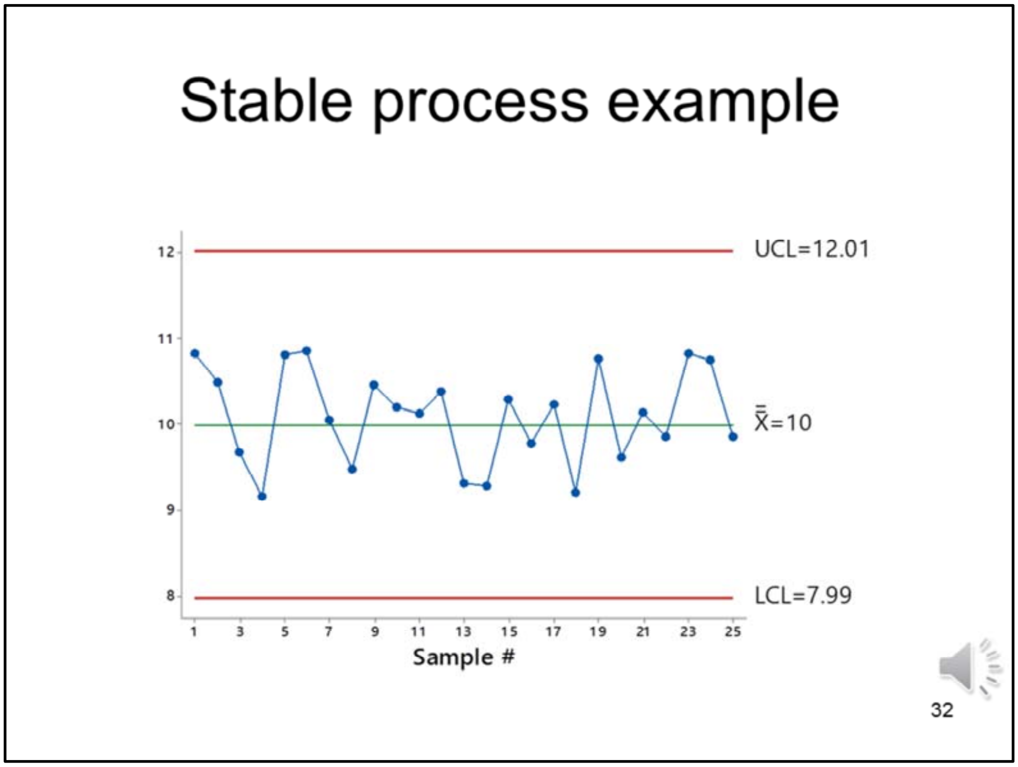
Also when we are monitoring the process, we need to look for that non random behavior that indicates that we may have special cause variation in the data.



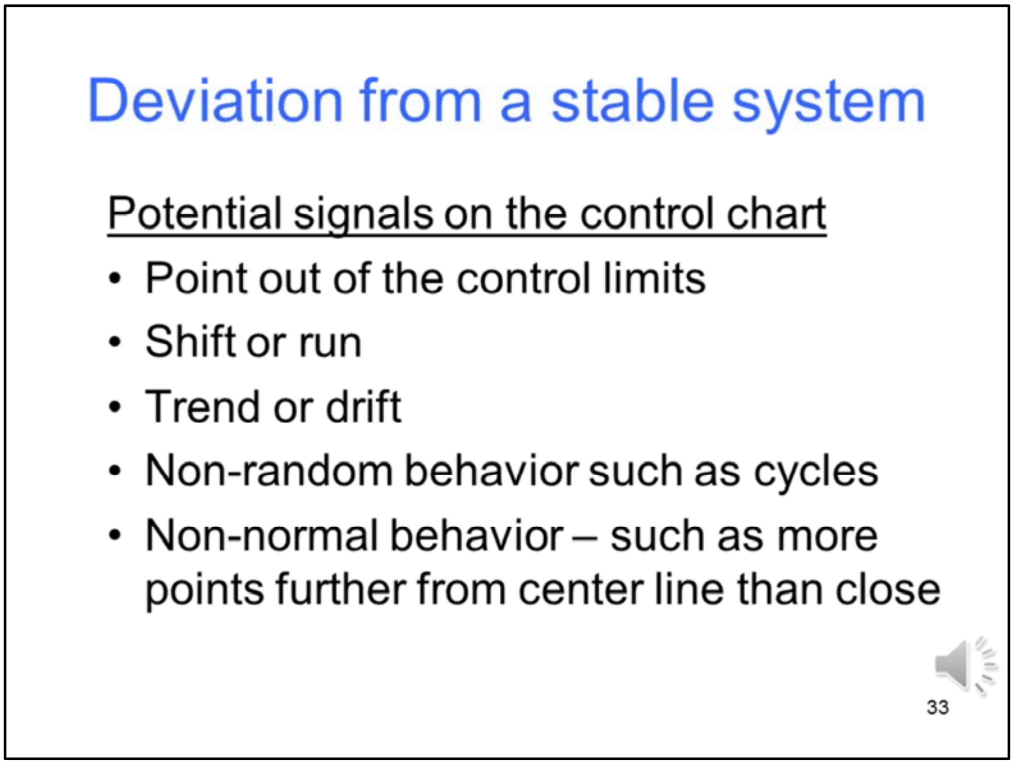
A process in statistical control will show signs that the data comes from a normally distributed process.

We talked about some of these at the beginning of the lecture.

- Observations equally likely to be above or below the mean
- Random distribution of observations – no patterns..
- More observations closer to the mean rather than further
- Very low probability of outliers.



So there is nothing in this data that indicates to me that the process is not in statistical control. No you do not have to go counting points above and below to make sure you have exactly the same number above and below the center line. Remember, you are taking samples, when taking samples, it is not likely to exactly mirror the population. Usually a visual review will suffice.



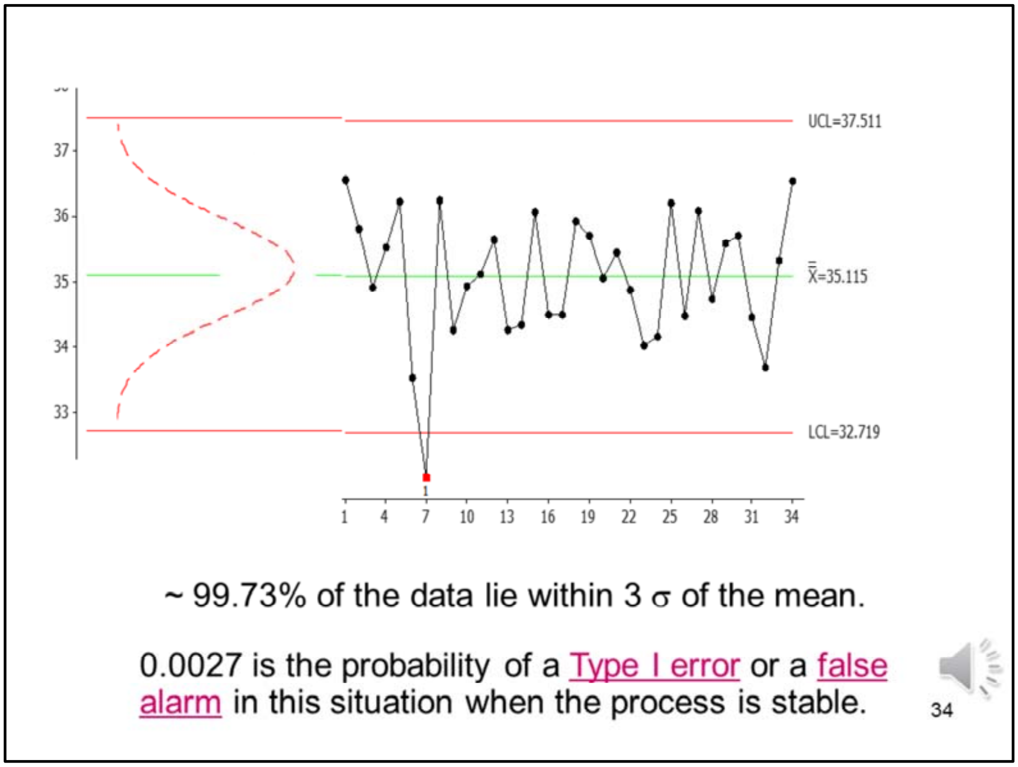
Some of the potential signals that a process is not in statistical control include

A Point out of the control limits – this one we have already discussed some.

A Shift or run above or below the center line. Trend or drift of the points. These behaviors are not random within definable bounds, so the process would not be considered in statistical control

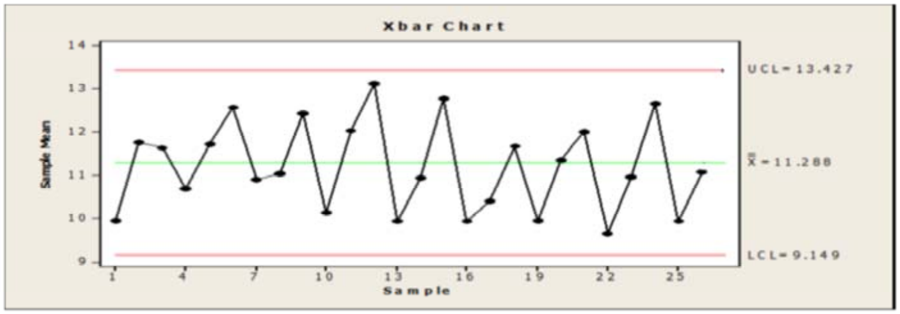
You can also have other non normal behavior such as cycles, or points not being clustered around the center line, or even too many points close to the center line.

If any of these conditions occur we would need to investigate before we could conclude the process is stable.



We have already talked about points being out of the three sigma limit. While we are looking at 3 times out a thousand, it will happen. If we are taking a sample every hour on a 3 shift seven day a week operation, it will happen on average, about twice a month. But the probability is low enough that any point out of the 3 sigma limits should be investigated.

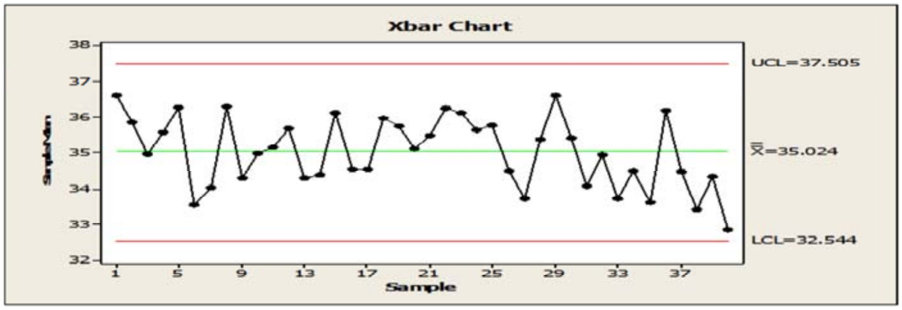
When you do get a point out when the process truly is in statistical control, that is called a type 1 error. We will discuss this more in a later lecture.



This process would not be considered stable because

A. The data is trending upwards.
B. It shows a point beyond the upper control limit.
C. The pattern of the points appears to be cyclical.
D. There is a run of points above or below the mean 35

C – Correct. The points seem to be going up and down in a pattern. One place this is often seen is during three shift operations, either due to differences between operators or to environmental conditions.



This process would not be considered stable because

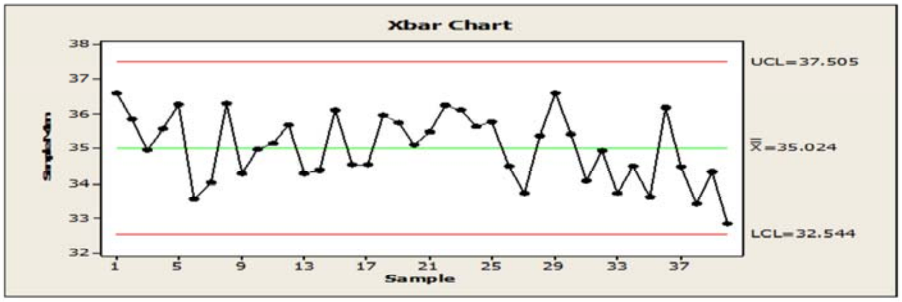
A. The data is trending upwards.

B. It shows a point beyond the upper control limit.

C. The pattern of the points appears to be cyclical.

D. There is a run of points above or below the mean 36

D correct. Those 8 points in a row above the mean would typically call for investigation.



The probability of 8 points in a row above the mean is

A. 0.5^8

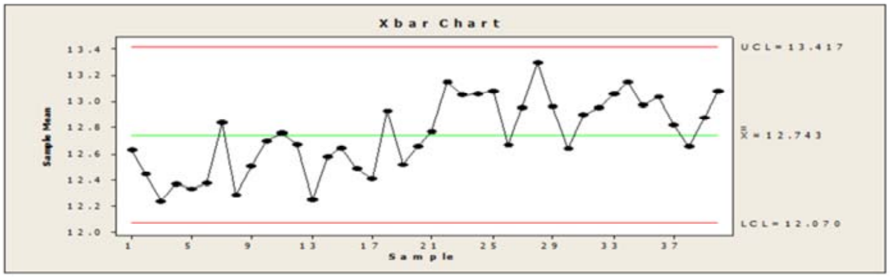
B. 0.39%

C. The same as the probability of 8 points in a row below the center line.

D. All of the above.

37

D correct. The probability of 1 point being above is .5 the second is .5 etcetera. So for independent events we multiply so it is .5 to the 8th. Which is .0039 or 39%. The same would be true for a run of points below the center line.



This process would not be considered stable because

A. The data is trending upwards.

B. It shows a point beyond the upper control limit.

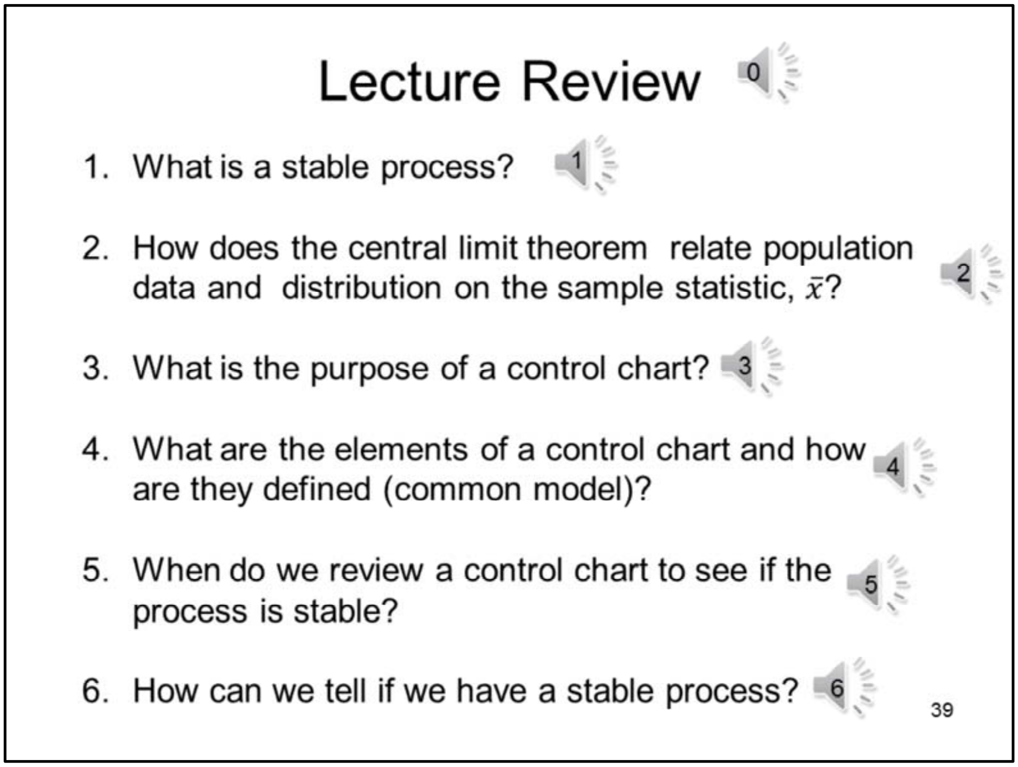
C. The pattern of the points appears to be cyclical.

D. The mean appears to have shifted.

38

A. Incorrect. It is more of a shift rather than a sustained trend. The first half of the points are distinctly lower than the second half of the points. It is definitely not a stable system.

D. This is the best answer. The first half of the points are distinctly lower than the second half of the points. It is more of a shift than a trend. But it is definitely not stable.



0) In the next lecture, we will learn how to take samples to determine control limits and use the charts to monitor production. Before then, you need to make sure you understand the concepts from this lecture. From this lecture, you should be able to answer the questions below.

1) With control charts, you will be answering the question “ is this process in statistical control? Or “Is the system stable?” What that means is that there is only common cause variation present. When there is only common cause variation the data will be random within identifiable bounds, and the process will be stable over time.

2) A population has a mean of $\boldsymbol{\mu}$ and standard deviation, $\boldsymbol{\sigma}$. If multiple samples of size n are taken from that population, the sample means, $\bar{x}$, are normally distributed with the mean of $\boldsymbol{\mu}$ and standard deviation of $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. This

value is called the **Standard Error of the Mean.** You should be able to identify in a problem if you are working with the distribution of individual population values, of the distribution of sample averages.

3) A control chart is a line graph which displays characteristics of a processes behavior over time. The main question the charts are answering is " Is the process in statistical control?" or "Is the process stable?". If the chart indicates it is not in statistical control, action should be taken.

4) A control chart will have a center line which is the mean of the sample statistic. It will have upper and lower control limits set at 3 sigma where sigma is the standard deviation of the sample statistic.

5) We have to make sure the process is stable to establish the control chart in the first place. Then when we use the established limits to monitor the process we are constantly reviewing new points added to see if the process is still in control.

6) A stable process will follow the characteristics of the normal distribution on the control chart. Any non-normal behavior, such as points out of the limits, shifts, trends or cycles could indicate that there are special causes of variation and that the process is not stable.

Related Assignments

Please see Blackboard for related assignments