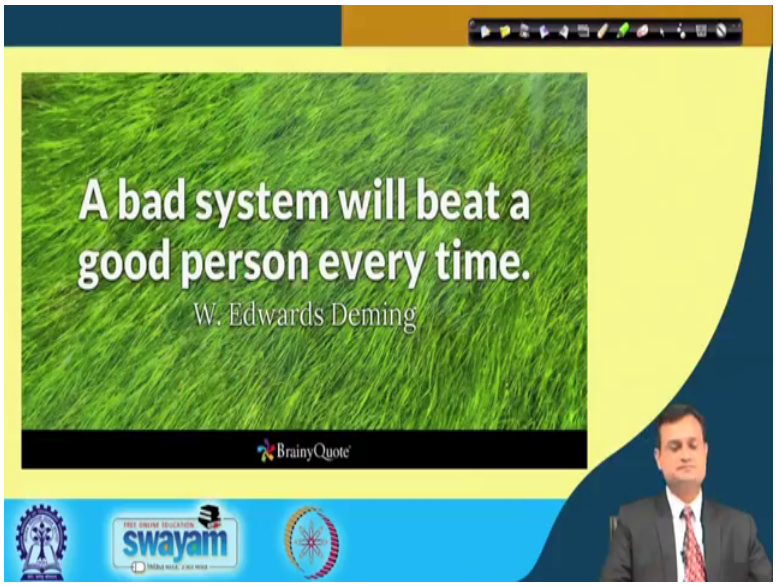
Six Sigma
Prof. Jitesh J Thakkar
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture - 27
Non - Normal Process Capability Analysis

Hello friends, once again I welcome you to our ongoing journey on Six Sigma and today we are discussing lecture 27 that is Non-Normal Process Capability Analysis. We have devoted three lectures process capability analysis for normal and we had seen the first lecture; the concepts related to process capability what is its need, what does it tell and what are the various cases three cases when my process is meeting, barely meeting, the specifications, my process is barely capable rather my process is not capable and my process is adequately capable.

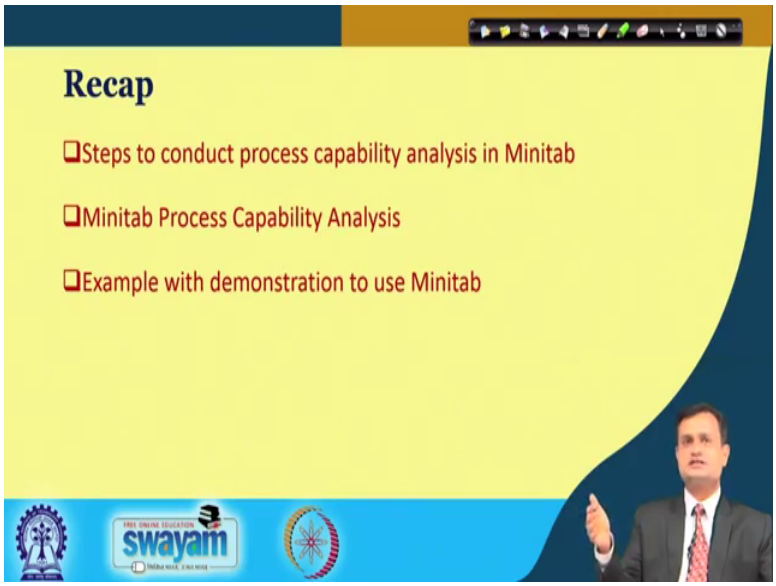
And then we have seen the various indices and followed by that we have seen the application of mini tab for conducting say process capability analysis for normal data. This lecture 27 will help you to understand: what is non normal process capability, when it should be conducted, how it should be carried out and how we can use the mini tab to easily conduct this non-normal process capability analysis.

(Refer Slide Time: 01:42)



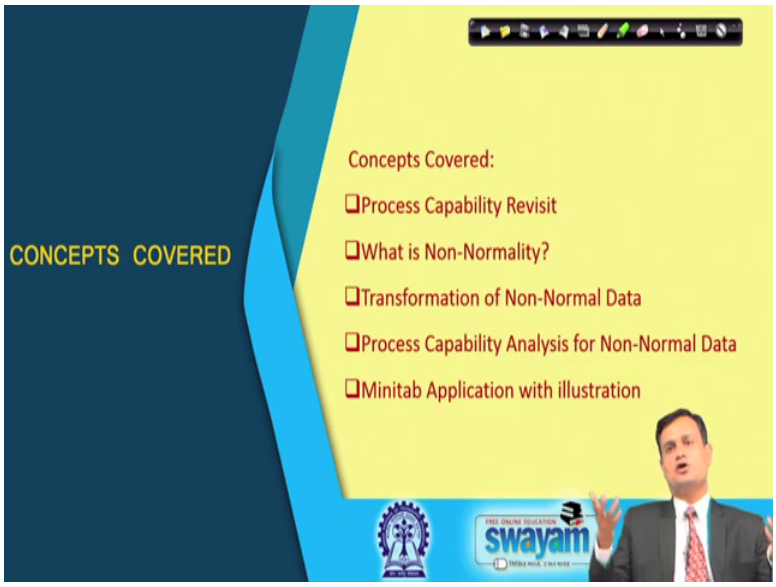
So, once again I would like to remind a bad system will beat a good person every time. So, correct the system first.

(Refer Slide Time: 01:53)



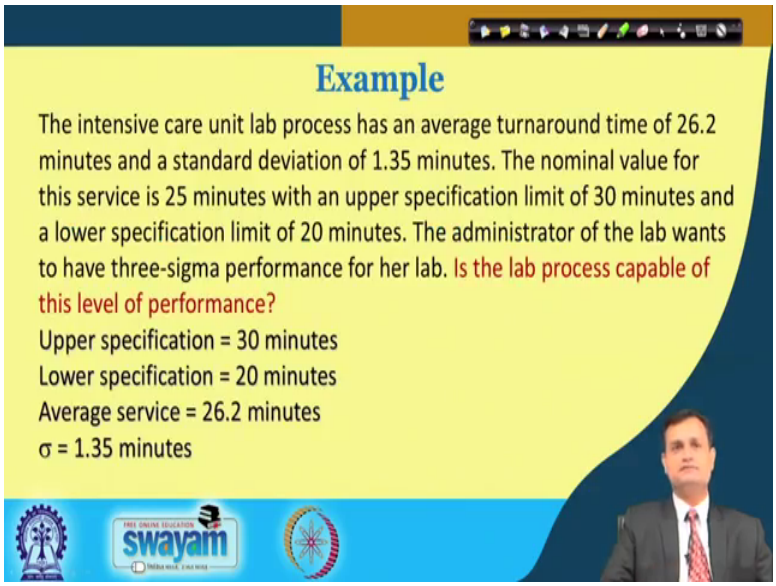
So, we have seen this that what is the use of mini tab, how it can be conducted. There is six pack process capability analysis and there is simple capability analysis and what are the outputs it can generate and how to interpret these outputs.

(Refer Slide Time: 02:10)



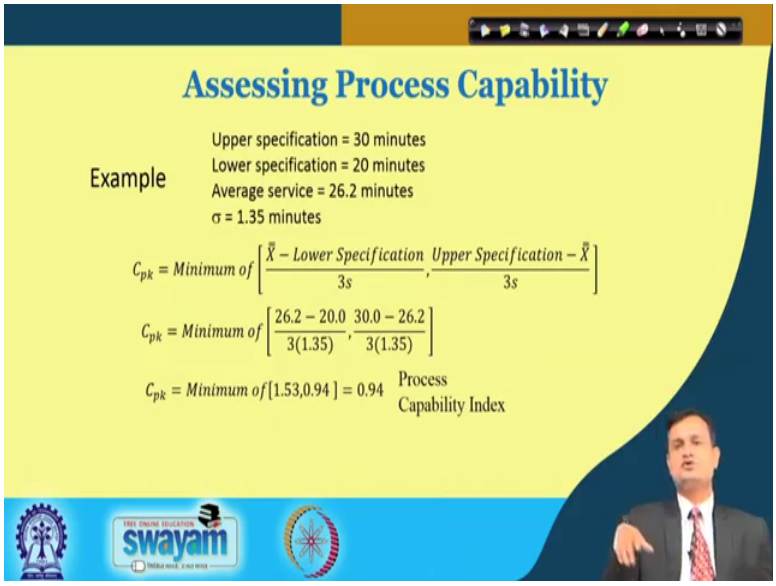
This particular lecture 27 will focus on process capability revisit just a small example. What is non-normality? Transformations of non normal data; what are the ways you can conduct the process capability analysis and mini tab application with illustration.

(Refer Slide Time: 02:36)



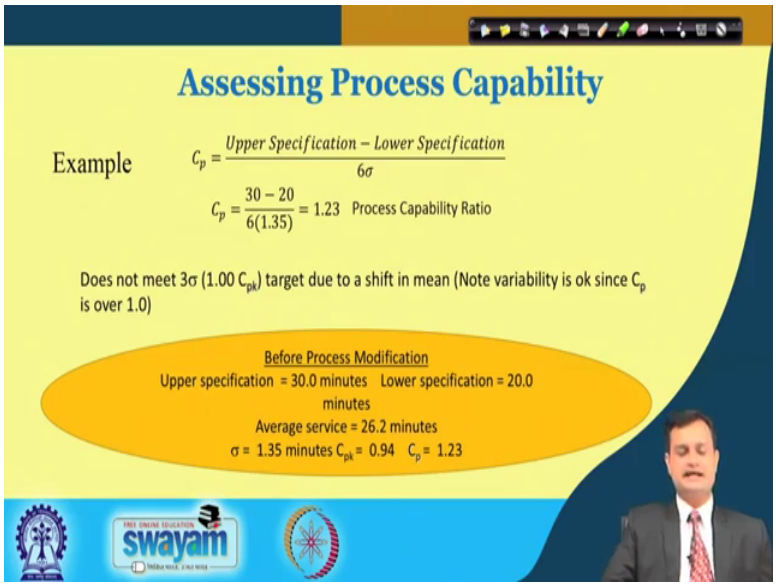
So, if we just revisit our concept with an example, let us say there is an intensive care unit lab process has an average turnaround time of 26.2 minute in standard deviation 1.35 minutes. The nominal value for this service is 25 minutes with an upper specification 30 minutes and lower specification 25 minutes and sorry 20 minutes, the administrator of the lab wants to have three sigma performance for her lab and is the lab process capable of this level of performance that is the issue that I want to investigate. Now your upper specification limit USL 30 LSL 20, average service 26.2 and sigma 1.35 minutes.

(Refer Slide Time: 03:26)



So, I will just plug in the values into my expression for C pk and it basically tags the minimum of X double bar minus lower specification divided by 3s and upper specification minus X double bar divided by 3s. So, computing this I get C pk minimum of 1.53 and 0. 94. So, I will check 0.94.

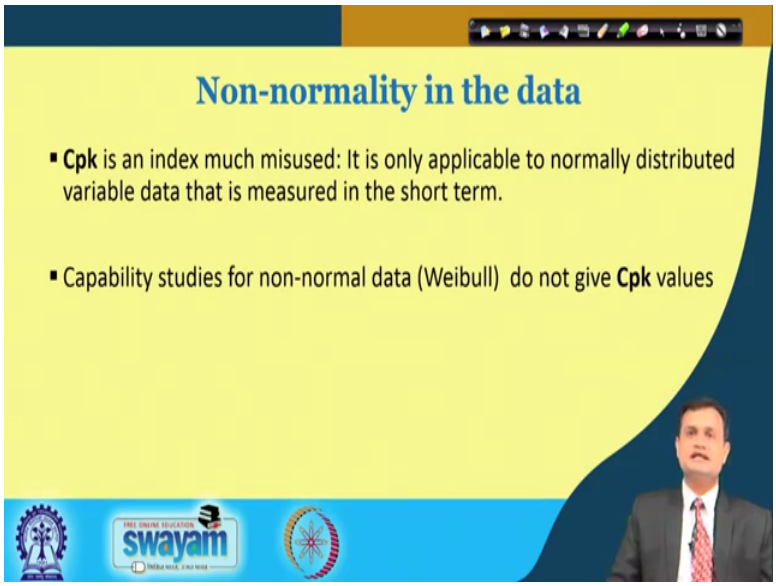
(Refer Slide Time: 03:58)



So, this is my process capability index. I would also like to find the C p which basically measures the spread and variability of the process. So, USL minus LSL divided by 6 sigma and I get 1.23. So, here I can say that does not meet 3 sigma which is typically 1 C pk target due to shift in mean. And if we note here, then variability part is C p is ok it is greater than 1, but my C pk has a problem and typically my C pk value is my C pk value here is 0.94.

So, little bit less we have the standard that, it should be around 1.2, 1.3 or at least 1 and in this case my variability is Cp is ok, but there is a shift in mean and hence my C pk has some problem.

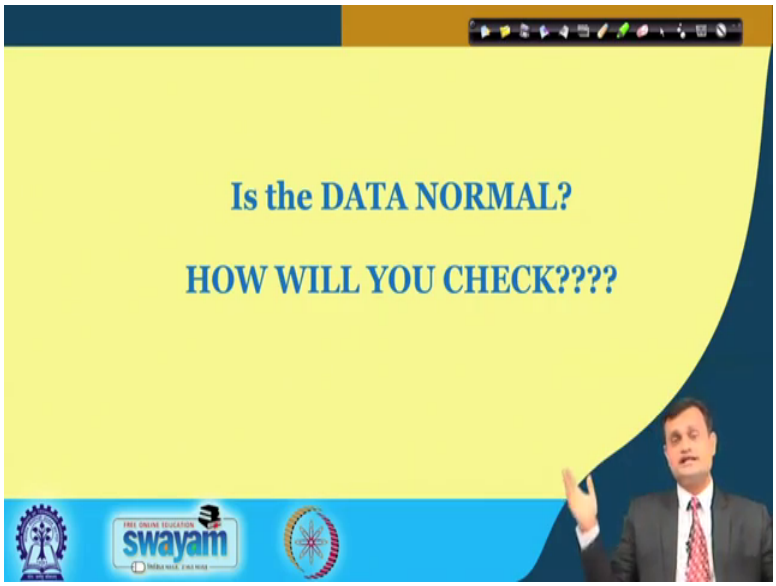
(Refer Slide Time: 05:19)



So, you have the upper specification 30 lower specification 20, average service 26.2 sigma is 1.35, Cpk 0.94, C p is 1.23 before any modification now, you need to take the corrective action short term or long term. As I mentioned short term you can just shift the mean and try to avoid producing the scrap that will give you the cost saving, but this is just the short term temporary measure. For long term you try to replace the part, set the process, develop the skill of the operator and see that your process can remain at a particular targeted value for a long operation period of time. Now let us talk about the non-normality in the data which is the focus of this particular lecture.

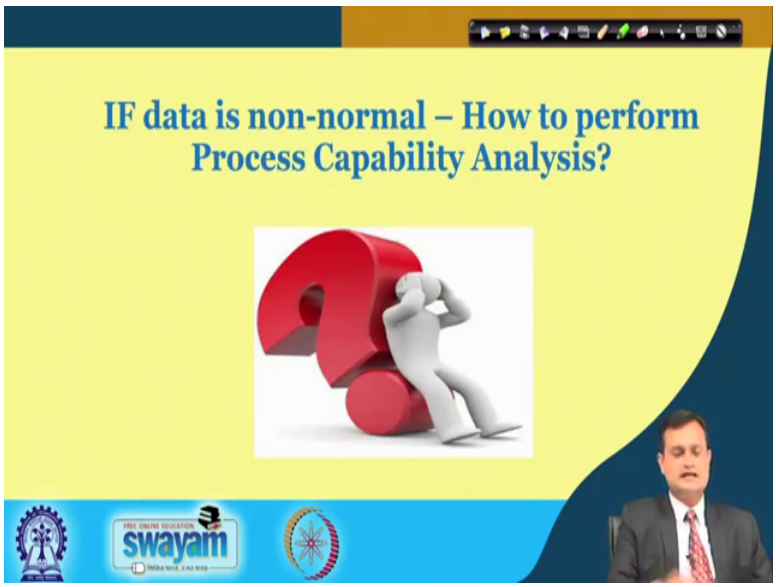
So, Cpk is an index which is many a times misused. Remember that this index is only applicable to normal distributed, where normally distributed variable data and this is typically measured for short term. Capabilities studies for non-normal data, let us say Weibull distribution do not give Cpk values and if you will just try to fit let us say this nonnormal data and try to figure out or find out the values of C p and Cpk then it would be misguiding misleading. So, my assumption about the normality of the data must be checked first and then only the question of process capability analysis comes. If the data is non-normal you must conduct the non-normal process capability analysis, if it is normal then you go for normal process capability analysis.

(Refer Slide Time: 07: 06)



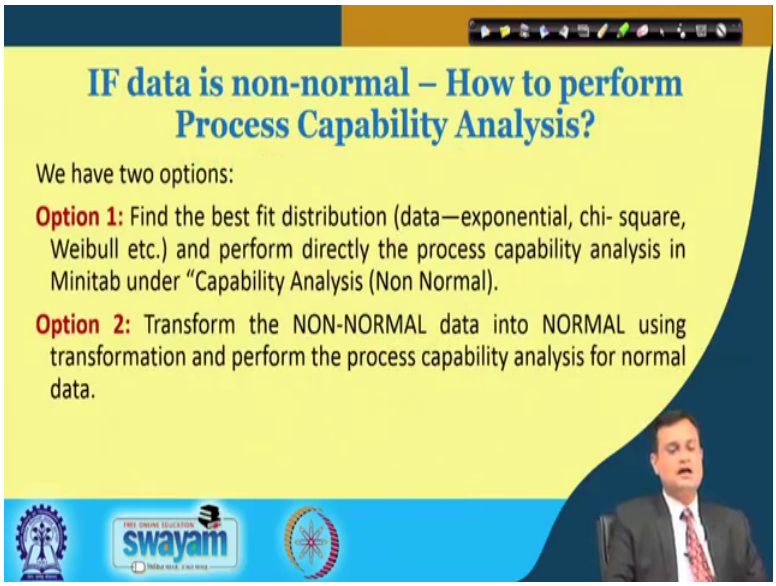
So now, the question is that, if I have the data non-normal or if I have to check rather my data is normal or not normal then how will I check?

(Refer Slide Time: 07:20)



So, there is a procedure and that can help us to check whether my data is normal or not normal, non-normal and what to do. So, if data is non-normal how to perform process capability analysis that is the second question. So, you will just struck your head and think that now my data is not-normal what to do.

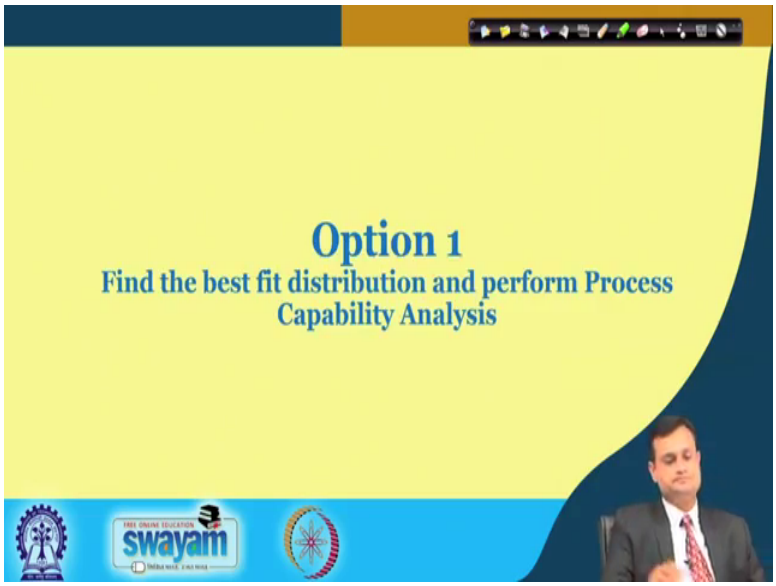
(Refer Slide Time: 07:44)



So, if data is not normal, it does not follow typically the normal pattern when you look at the histogram or normality plot, then how to perform is a question. So, here I would recommend that there are two options possible; option number 1 find the best fit distribution let us say your data fits to exponential, chi square, Weibull and perform directly the process capability analysis in mini tab under capability analysis non normal. So, this provision is that, but for that you need to figure out that what is the best fit distribution for your data.

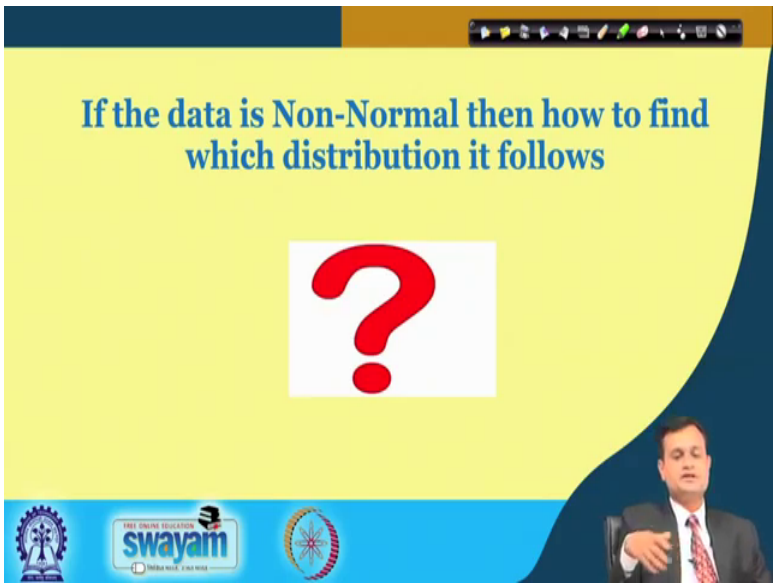
If you do not want to do this, the option 2 is to transform your data which is not normal into normal using some kind of transformation, mathematical transformation and then perform the process capability analysis with the simple normal data assumption. And, you can get the same result both way either option 1 option 2.

(Refer Slide Time: 08:58)



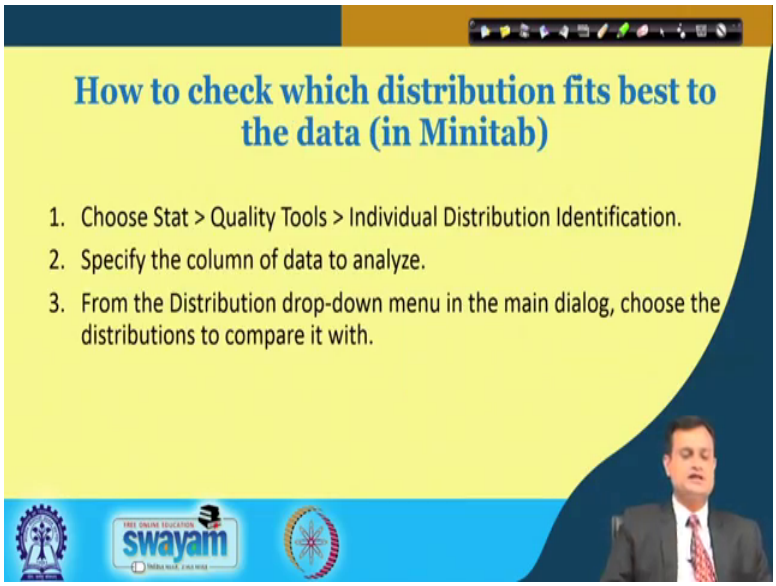
So, let us try to see that how to execute different options. So, option 1 find the best fit distribution and perform the process capability analysis. So, I am interested in finding the best fit distribution; it may be exponential, Weibull, triangle whatever I want to see that my data fits best fit to which particular distribution.

(Refer Slide Time: 09:24)



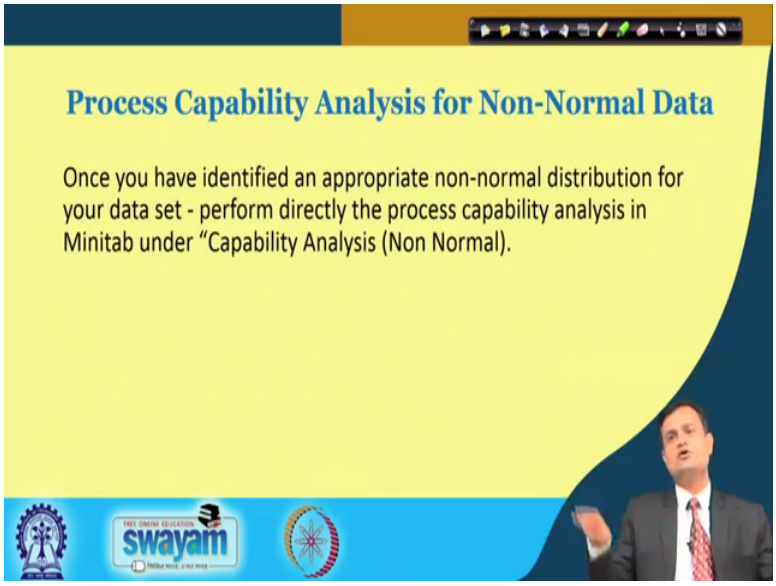
So, if data is non-normal then how to find the distribution.

(Refer Slide Time: 09:30)



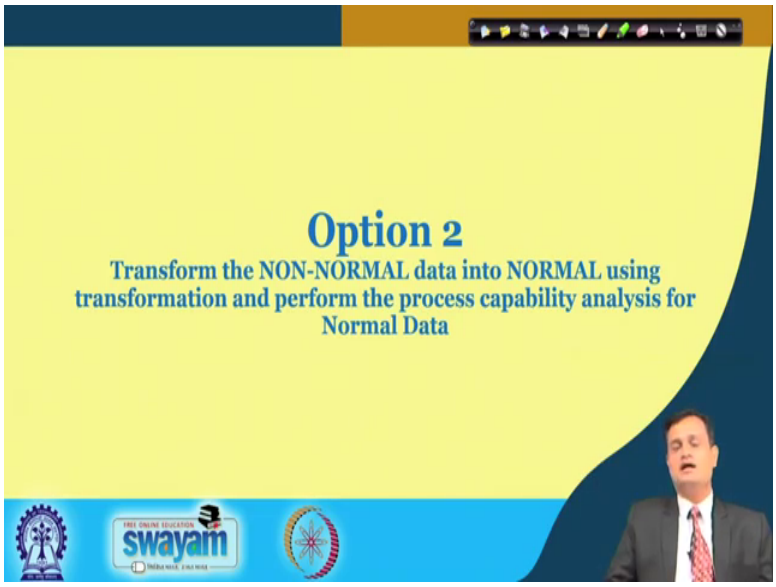
So, you can use mini tab very easily to do this exercise and choose the stat quality tools, individual distribution identification, specify the column of data to analyze and from the distribution drop down menu in the main dialog you can choose the various distributions to compare with. And, then your mini tab will give you the best fit for various distribution comparing and you can find out that which distribution your data basically is resembling.

(Refer Slide Time: 10:06)



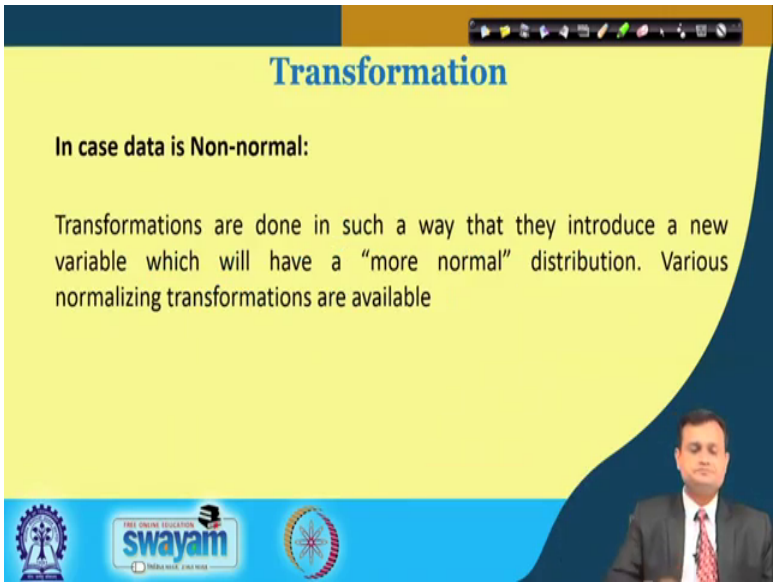
So, process capability for non-normal data once you have identified figured out your non-normal distribution, best fit distribution then directly go to the option capability analysis non-normal perform it and you will get the results.

(Refer Slide Time: 10:26)



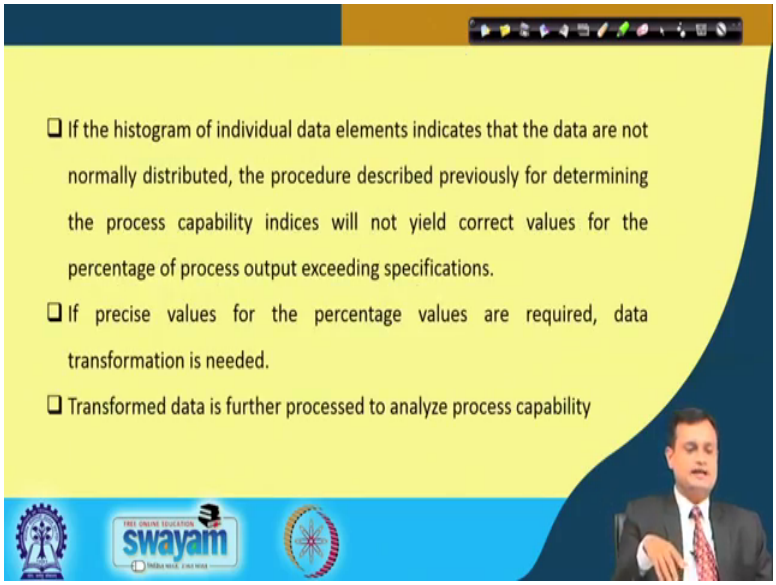
So, option 2 is basically to transform. Now, when I cannot do the regular capability analysis which is made for normal data and if my data is non-normal I have to transform the data this is option 2.

(Refer Slide Time: 10:45)



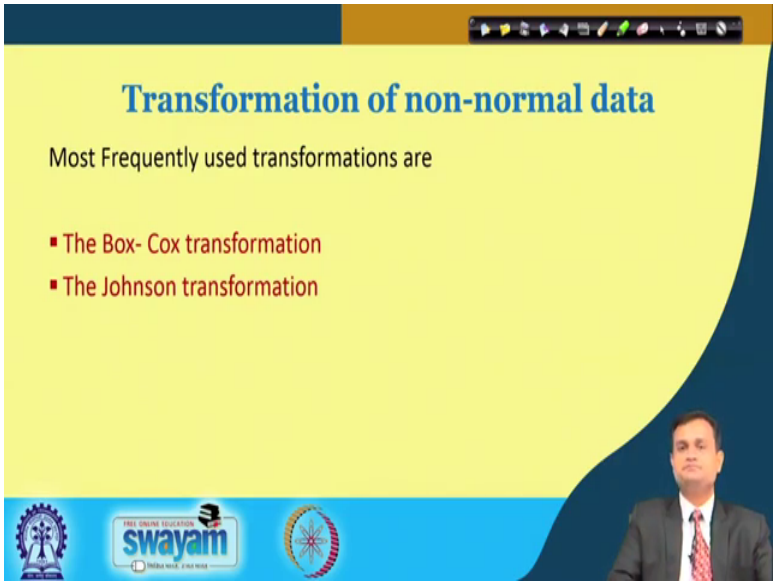
So, you have the non-normal data and try to make it more normal by transforming it through some mathematical function.

(Refer Slide Time: 10:56)



So, you can a first check through histogram or your normal probability plot that whether your data is normal or not, and if you feel confident that yes my histogram as well as normal probability plot indicates that my data is not normal, then you transform the data.

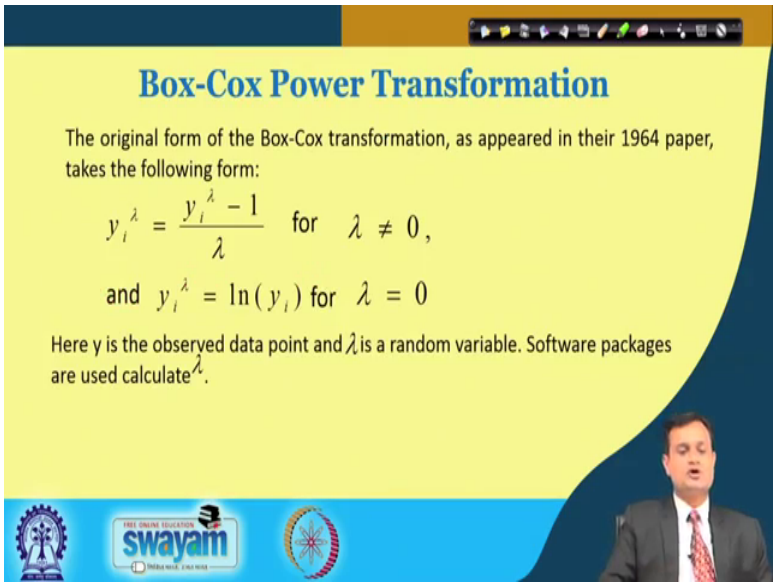
(Refer Slide Time: 11:18)



So, you have two options to transform the data two transformations rather one is Box Cox transformation other is the Johnson's transformation, it is extremely difficult to do it

manually I will give you the mathematical expression for your understanding. But, we take the help of software which has inbuilt function mathematical function and that helps us to transform the data very easily.

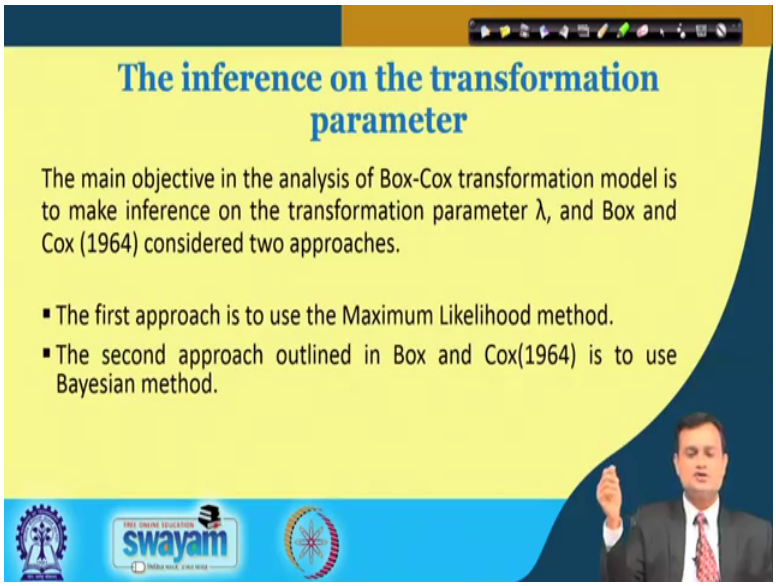
(Refer Slide Time: 11:48)



So, Box Cox transformation, the critical element here is the lambda which raises the power of the data to that exchange where your data becomes more normal. So, just see the expression that y i rise to lambda is equal to y i rise to lambda minus 1 divided by lambda for lambda not equal to 0.

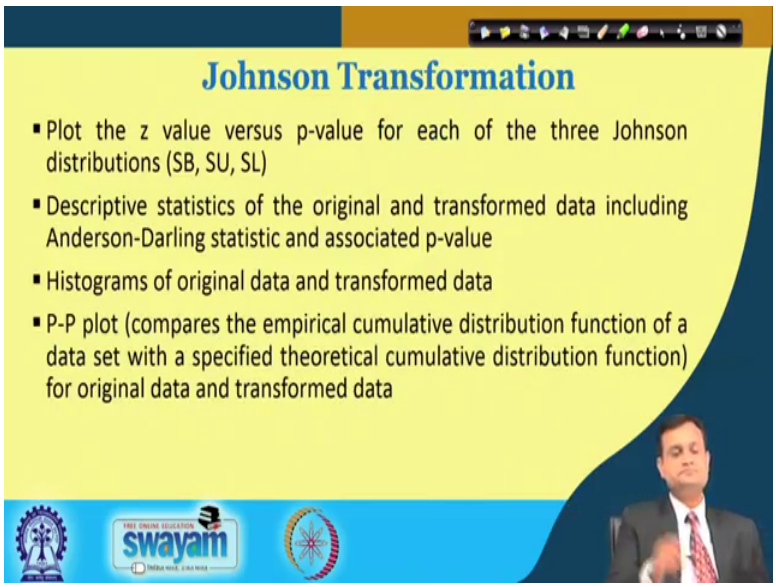
And if you have lambda equal to 0 then y i rise to lambda is equal to log y i. So, lambda typically is a random variable and software packages are typically used to calculate the lambda and y is basically your absorbed data point, typically your data which is non normal and this lambda will help you to transform your non-normal y i data into more normal.

(Refer Slide Time: 12:41)



So, you have certain influence on the transformers and parameter. So, the main object you have Box Cox transformation model is to make inference on the transformation parameter the critical parameter lambda, and Box Cox in 1964 this particular method was proposed he advocated two approaches. Number 1 is to use the maximum likelihood method you may have to read a little bit on this. And, the second approach is to use the Bayesian approach and considering the previous data posterior distribution and you can little bit read on these to just get familiarized with this terminology.

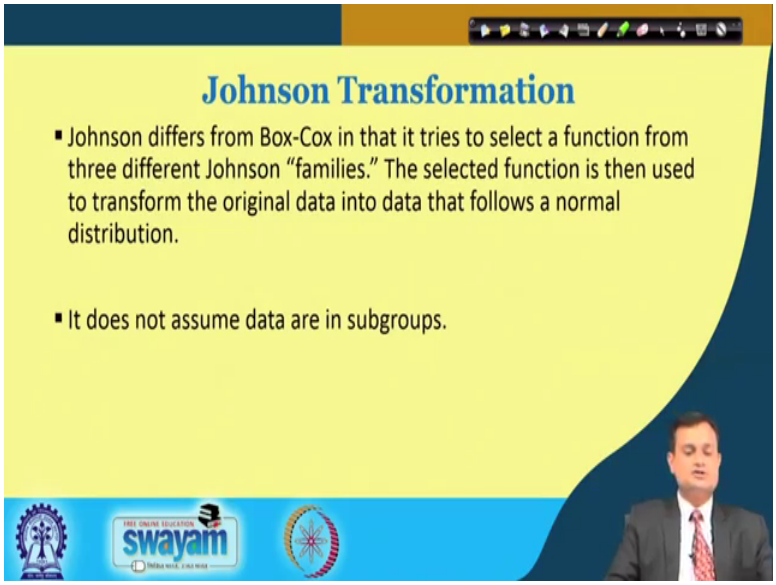
(Refer Slide Time: 13:27)



So, you have another option that is Johnson's transformation and the Johnson's transformation typically plot the z value to your normal critical value versus p value that is the probability value for each of the three Johnson distributions. So, he has given a typical family or three distributions SB SU and SL. So, descriptive statistics of original and transform data including Anderson darling statistics which basically checks the normality of the data and the associated p value.

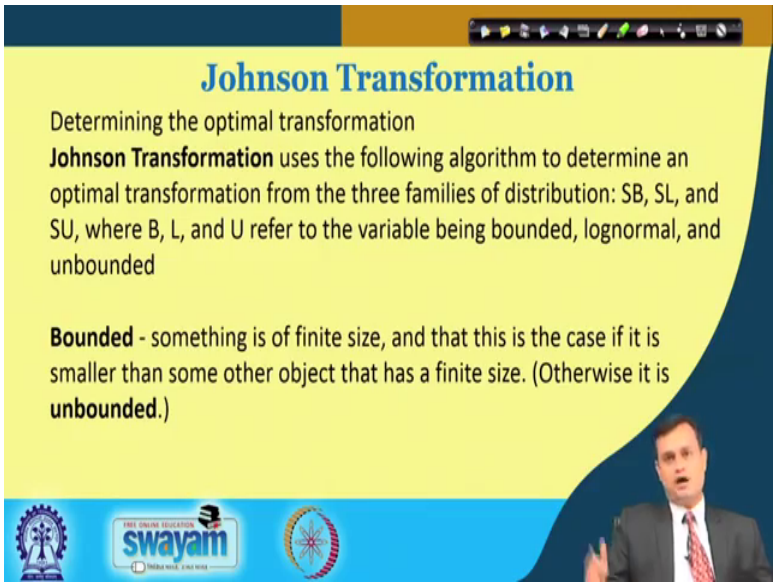
So, histogram of the original data and the transformed data can also be compared for visual purposes and you can have the PP plot which typically say compares your empirical distribution function and theoretical cumulative distribution function. So, empirical distribution is based on the sample, how your events are occurring and the theoretical distribution is based on the large data set available. So, we try to compare these two and see that to what extent my transform his son is acceptable.

(Refer Slide Time: 14:47)



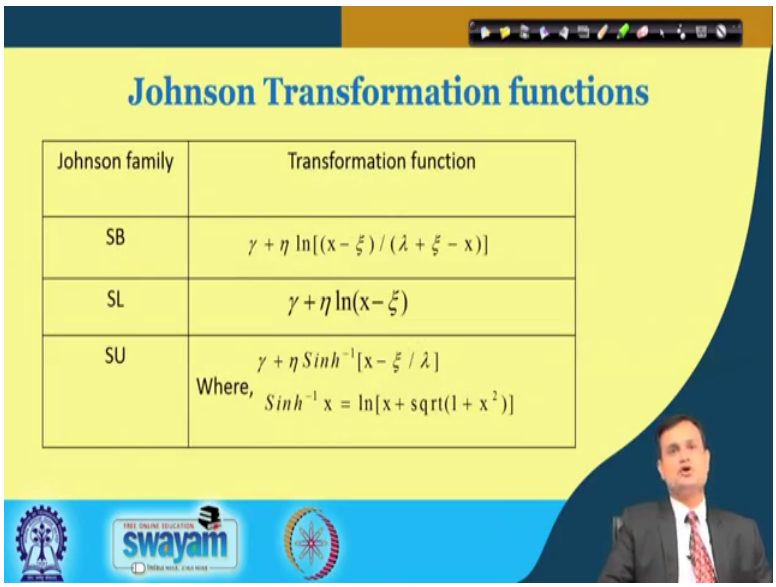
So, you have Johnson's transformation typically differs from Box Cox and the difference is that it tries to select the function from three different Johnson families. SB SU SL and the selected function is then used to transform the original data in case of Box Cox that transformation component was lambda, here you have three distribution families and typically it does not assume the data in subgroups.

(Refer Slide Time: 15:23)



So, you have typically this Johnson transformers families, which is SB, B stands for bounded L stands for log normal and U stands for unbounded. So, bounded is something of finite size and this is the case if it is smaller than other object, that has a finite size otherwise it is unbounded.

(Refer Slide Time: 15:49)



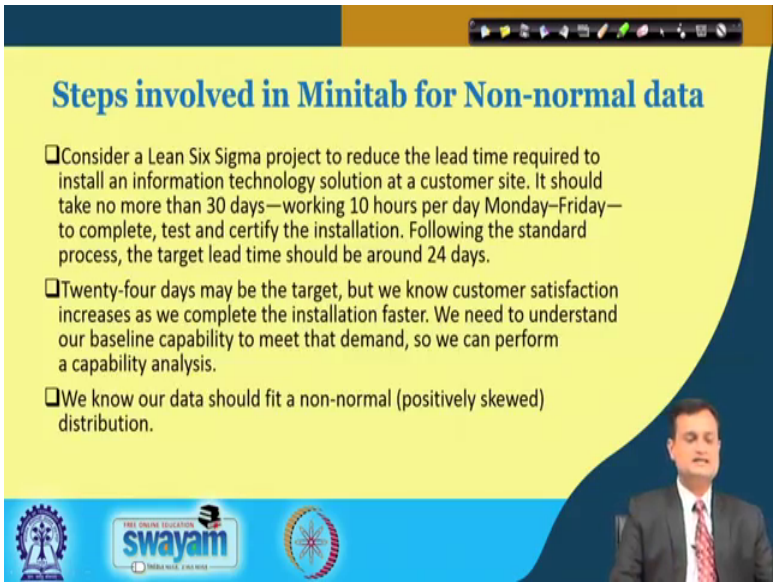
So, you can see the expressive mathematical expression and as I mentioned that it is difficult to do it manually, it is always better to use the software. But obviously, we should appreciate the logic behind the software has basically created a platform and

using it. So, you have SB, which is typically express in terms of a transformers and function lambda plus eta log x minus eta divided by lambda plus eta minus x and same way your SL and SU. So, I am not going into the details of these mathematical expressions, but I would like to make a mention that depending upon your nature of data it will select one of these mathematical function and transform the data to the more normal data.

(Refer Slide Time: 16:45)

So, let us try to see how we can conduct this analysis that is non normal process capability in mini tab, I have highlighted here two sources from which you can learn this application into greater detail I also use them for explaining this application to you; so, illustrative example 1.

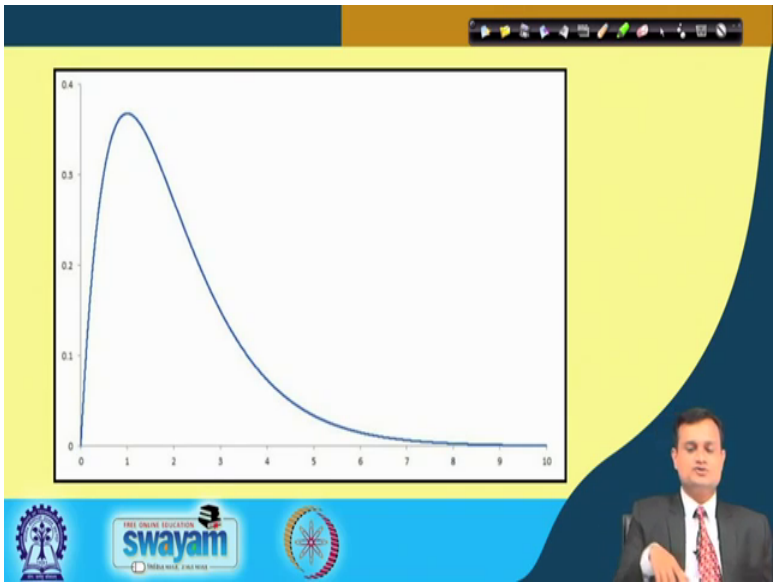
(Refer Slide Time: 17:09)



Let us say I have the minute app software and typically I need to follow the steps like this. Consider typically a lean six sigma project to reduce the lead time required to install typically an information technology solution at a customer site and it should take no more than 30 days. So, working is for 10 hours per day Monday Friday to complete Monday to Friday and test and certify the installation.

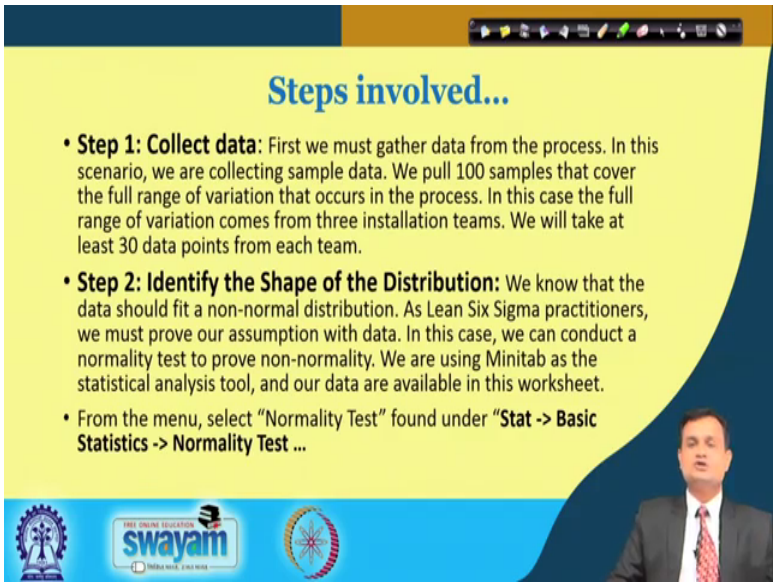
So, following the standard process the target lead times should be 24 days and I need to conduct the non normal process capability to see that to what extent I am capable enough to meet this target. So, 24 days maybe the target, but we know that customer satisfaction increases, if we complete the installation faster and we basically need to understand or figure out that our baseline capability to meet that demand what is it. So, we can perform a capability analysis.

(Refer Slide Time: 18:22)



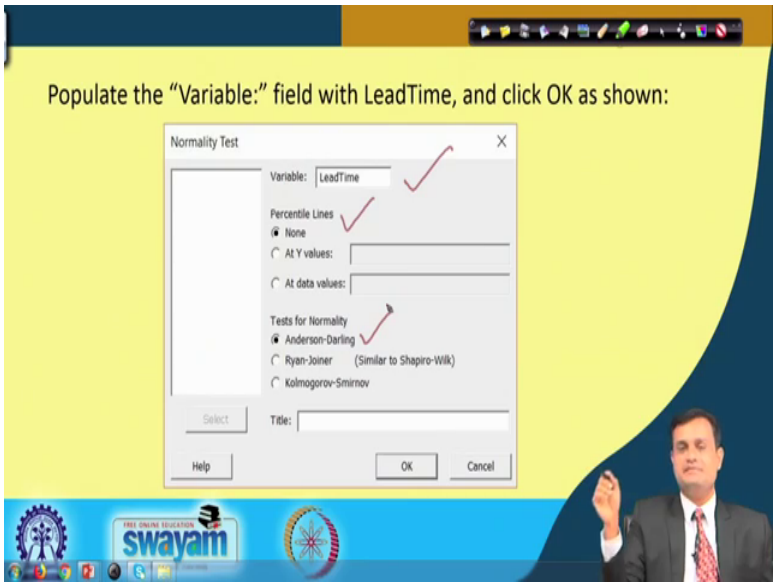
Now, just see this and you would find that your distribution is positively skewed and it does not out the normal distribution pattern of the data and I have to conduct the non-normal process capability analysis.

(Refer Slide Time: 18:41)



So, steps are very simple to follow, you collect the data identify the shape of the distribution. So, for that step 2 you go to stat basic statistics normality test.

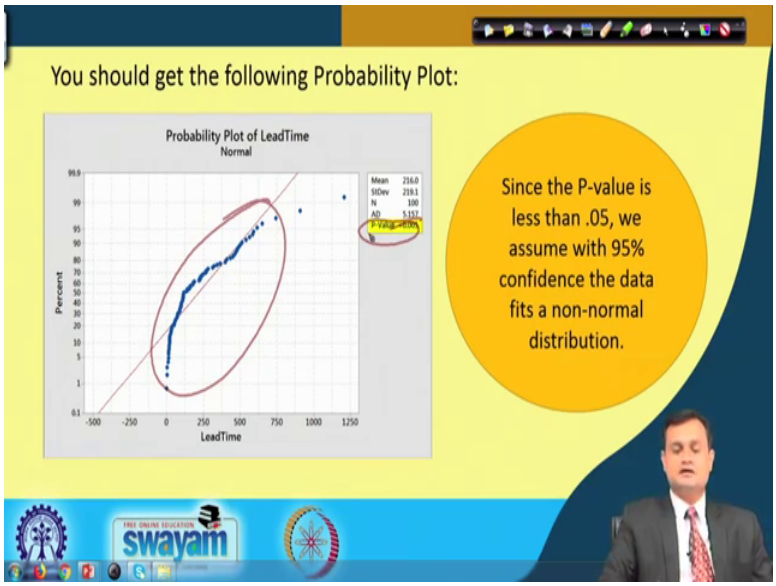
(Refer Slide Time: 18:54)



So, I am showing you the window also so, that you get clear on idea about executing the non normal process capability in mini tab. So, you have this particular window where you have percentile lines. So, I am saying none you are the variable of interest, I am putting lead time test for normality I am selecting Anderson Darling. So, I would like to just mention here when we would be discussing the various test in the subsequent lectures statistics is the tool by which inferential statistics by which I can check my hypothesis, I can check my claim at certain level of significance and say that my assumption or my claim is true or it is false.

So, when I want to do this, with certain kind of assumption then there are various trades available for this purpose I would use the Anderson darling test which is widely used to check the normality. So, in lecture one I clearly mentioned that to learn this course you need to have as a prerequisite some basic knowledge of the statistics anyhow we are discussing the important concepts, but it is always desirable that you have some knowledge about the basics of the statistics including distribution and probability theory and other. So, that part you should keep in mind.

(Refer Slide Time: 20:45)



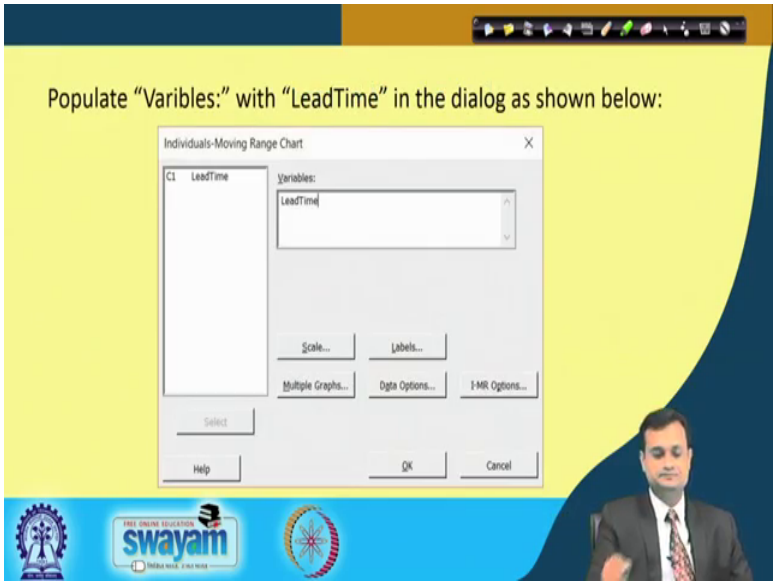
Now, let us see that what is the next step in my analysis using mini tab. So, I can see here very clearly that my normal probability plot shows that my data is not fitting the straight line, then my assumption about normality of the data is not valid, it means my data is non normal. You can also check the P value we have not discussed the concept of P value so, far, but you can just see that your P value is point 0.05, which is quite less than the level of significance you might have assumed 0.05.

And, hence it is in the rejection region we will see this in the subsequent lecture how to test the hypothesis and again this inferential measure indicates that my assumption about normality is not true and my data is non normal.

(Refer Slide Time: 21:56)

Now, step 3 verify the stability. So, typically in lean six sigma project which we are trying to study for installation of the software, I have a data set, but I would like to check the stability through individual moving average chart and this will tell me that over a period of time my processes in the statistical control or not. If not then there is no point in conducting the process capability analysis whether normal or not normal for the collected data.

(Refer Slide Time: 22:32)

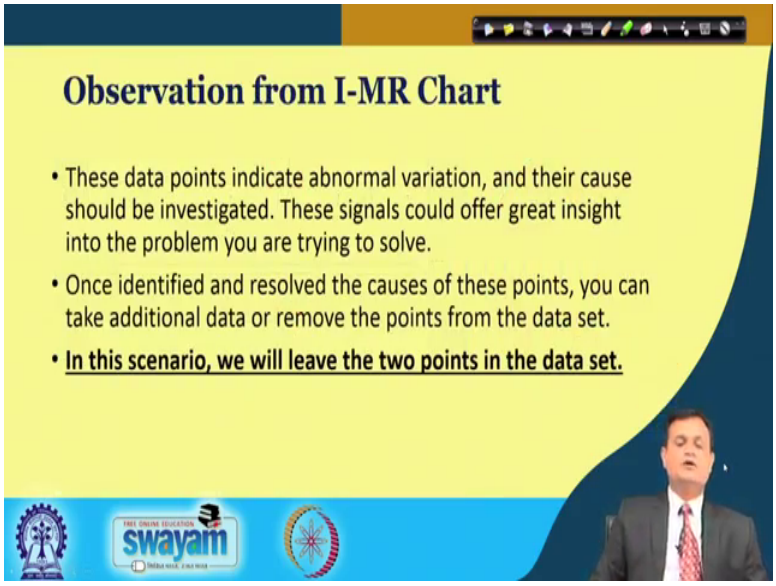


So, I can do this very easily in the mini tab.

(Refer Slide Time: 22:37)

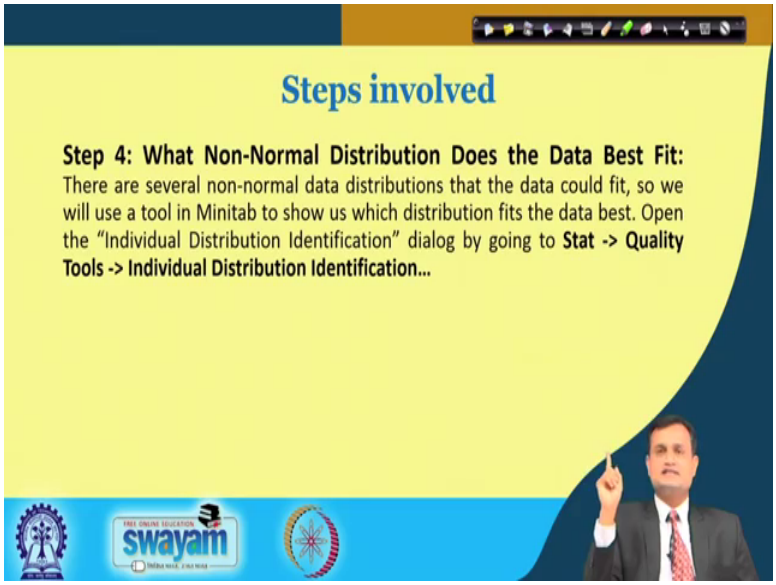
I am putting the variable each time, I can generate this chart already I was shown in the previous slide and I can see that yes there are some of the points which are falling outside and that is a stability issue. So, either I can discard this point and do it again or I can take certain corrective actions to see that my process becomes statistically controlled and stable so, that whatever data I collect that are adequate and good enough to continue conduct the process capability analysis.

(Refer Slide Time: 23:18)



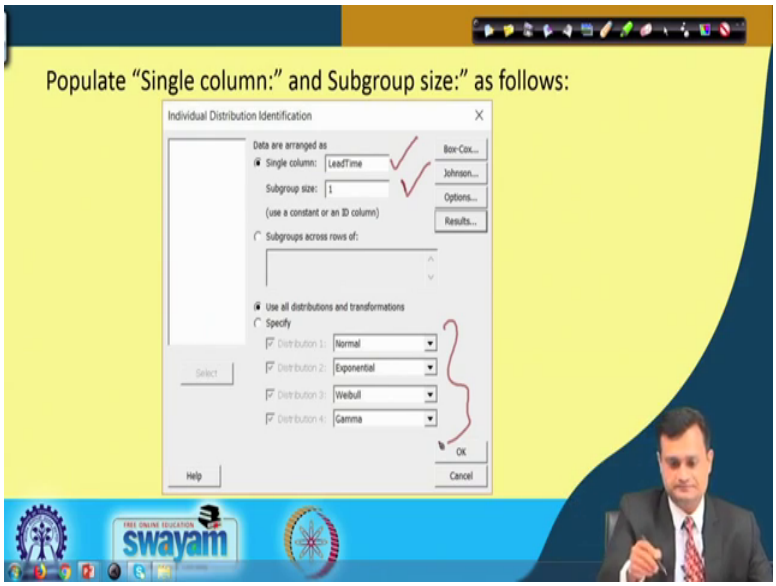
Now I have the observation that these data points which are indicated outside the upper control limit, they indicate the abnormal variation and once identified resolve the cause of this points as I said take the corrective action. And, you can take the additional data or remove the data points from the data set and then you can conduct the process capability analysis for the remaining data.

(Refer Slide Time: 23:47)



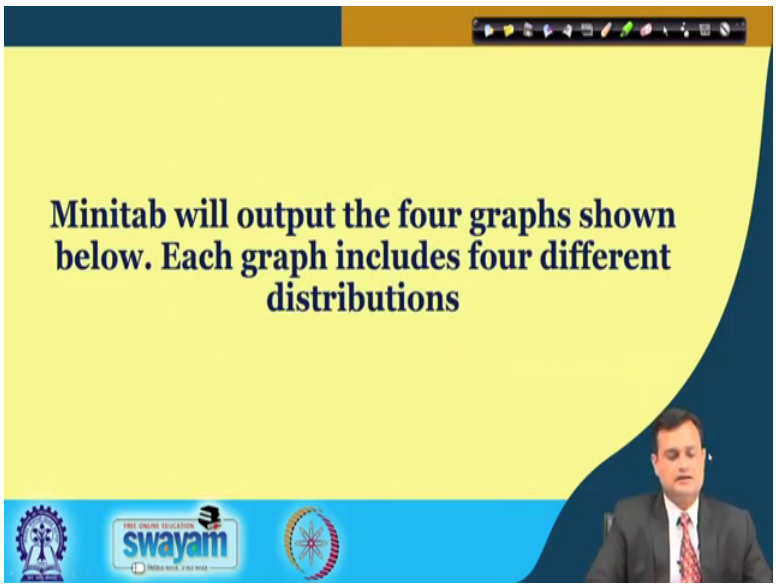
Step 4: what non-normal distribution does the data based fit? So, as I mentioned you have two options available option number 1 you can find out what is the best fit distribution and directly use the option for non-normal process capability in mini tab and conduct the analysis or else you can transform the data to normal and then you use the normal say process capability analysis in mini tab. So, both in a way it is ok.

(Refer Slide Time: 24:19)



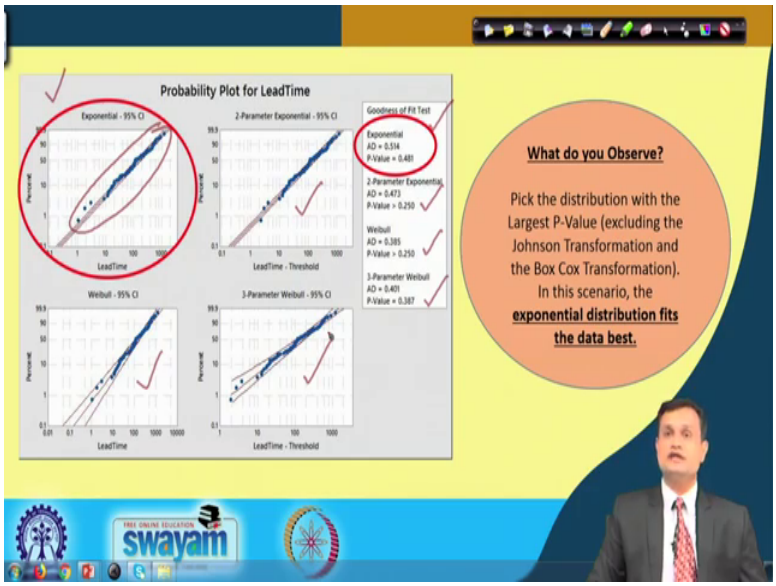
So, here I am just trying to see that what could be the best fit distribution and I am trying to do it with the option 1. So, just see this that my variable of interest is lead time subgroup size I have considered one and I am trying to select the various distributions normal exponential Wiebull gamma and my say output will help me to understand what is the best fit distribution.

(Refer Slide Time: 24:57)



. So, here mini tab will give the output for the four graphs typically shown in this slide.

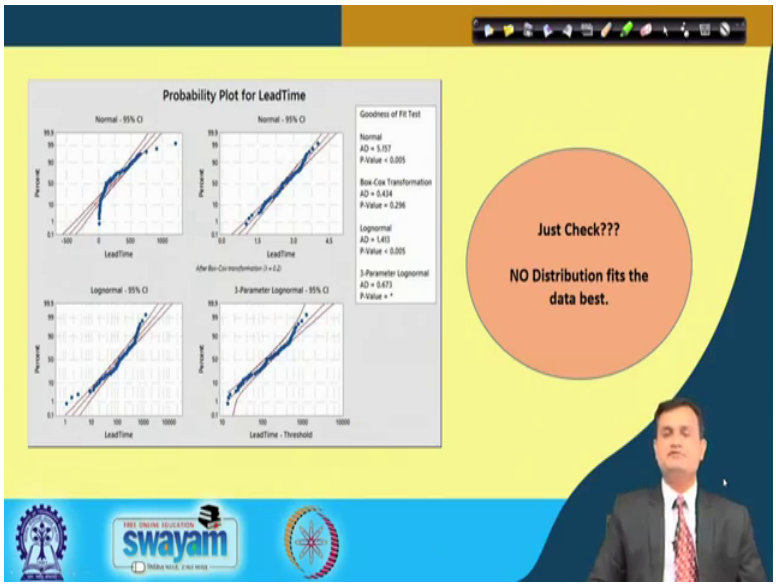
(Refer Slide Time: 25:03)



And what you can see here that, exponential non my probability plot exponential the data best fits. Also you can verify this here that Anderson Darling value is 0. 5 and 4 which is quite acceptable we have not discussed this test in detail. So, leave it, but you can at least see for the P value 0. 481. So, this is greater than the alpha which is 0. 05 in this case and it is acceptable.

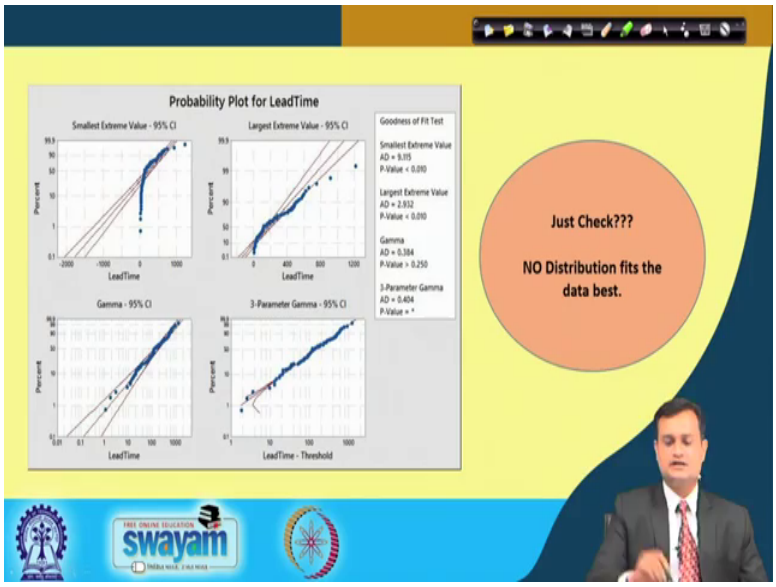
So, visually also it is acceptable as well as statistically also it is acceptable and in all other cases if you see then P value as well as Anderson-Darling value they do not conform, and if I look at the nor probability plot again it is not conforming that my data is fitting to that particular distribution. So, here are my data best fits to the exponential distribution and fine I can conduct the non-normal process capability choosing the exponential distribution.

(Refer Slide Time: 26:23)



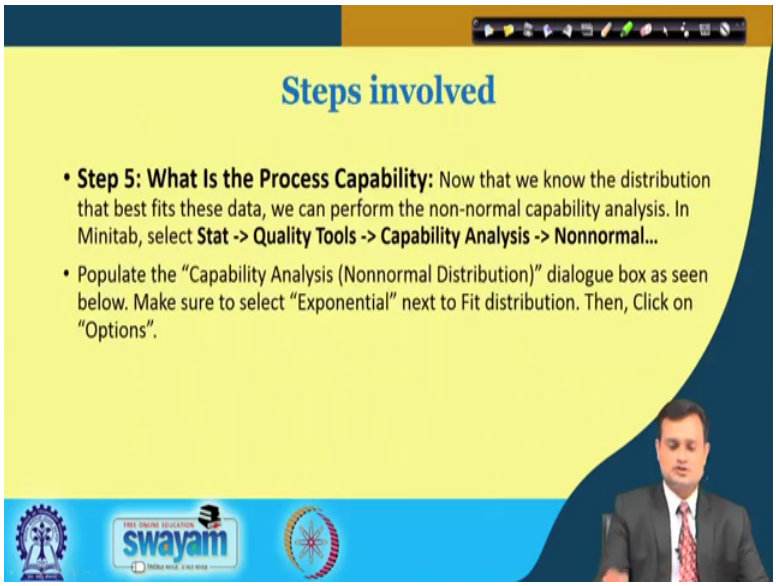
So, you can also compare it with the other distribution and just check it the best fit.

(Refer Slide Time: 26:29)



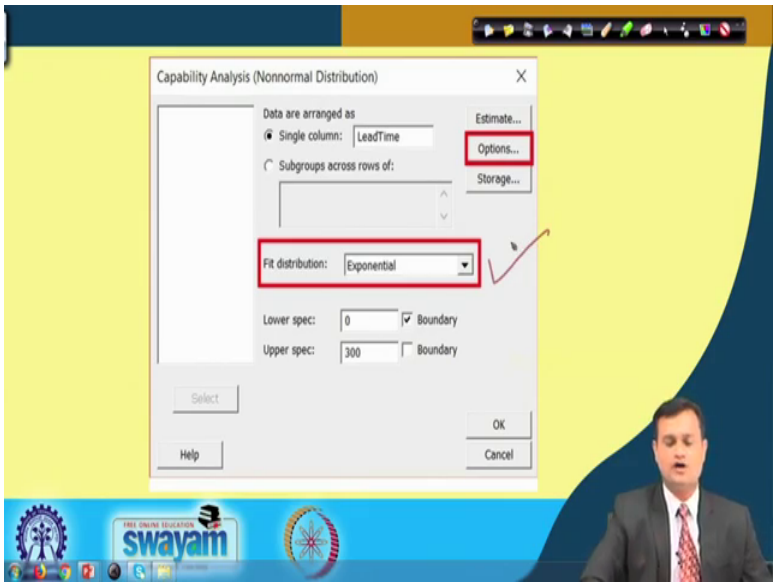
Just check for other distribution. So, no way it is meeting just visually you can see the probability plot and you can figure out that it is not fitting that particular distribution.

(Refer Slide Time: 26:40)



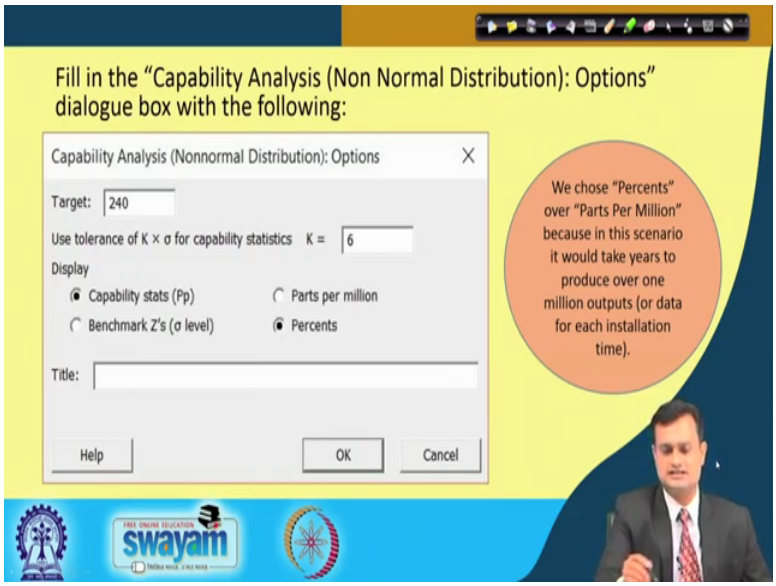
So, now step 5 what is the process capability? So, you use the options stat quality stat tools, capability analysis, non-normal I am exercising option one because it is my non-normal process capability, I have figured out what is the best fit distribution.

(Refer Slide Time: 27:00)



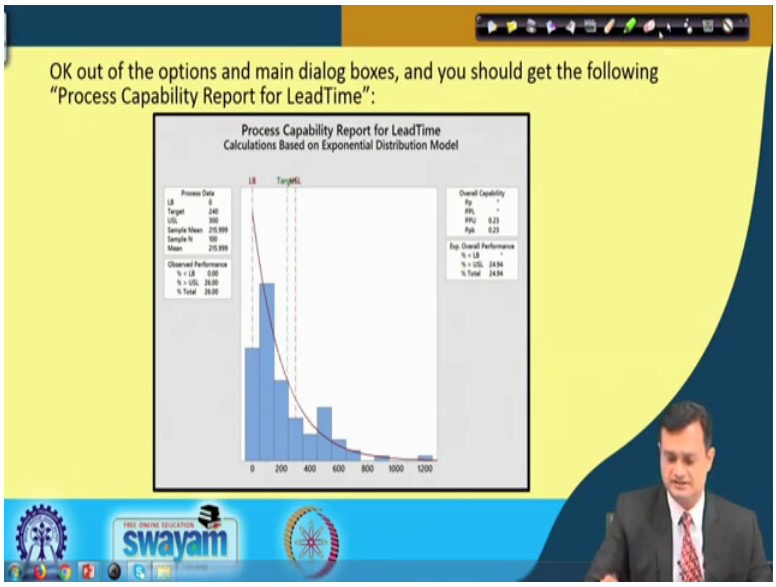
So, you have chosen here this exponential distribution which you found in the previous analysis.

(Refer Slide Time: 27:18)



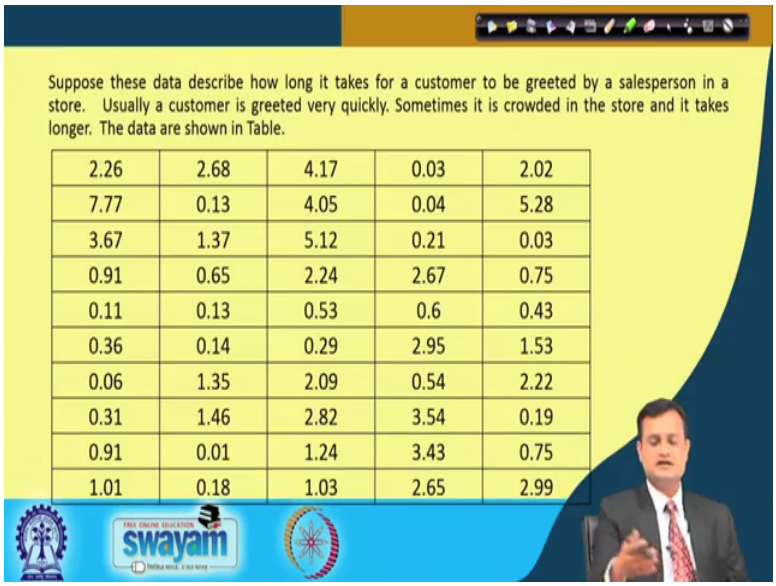
And now with this you conduct basically the process capability analysis. So, you put the target value, you put the k value, that is your tolerance came to sigma. So, six sigma let us say capability stats Pp percent and you try to find out the various values.

(Refer Slide Time: 27:39)



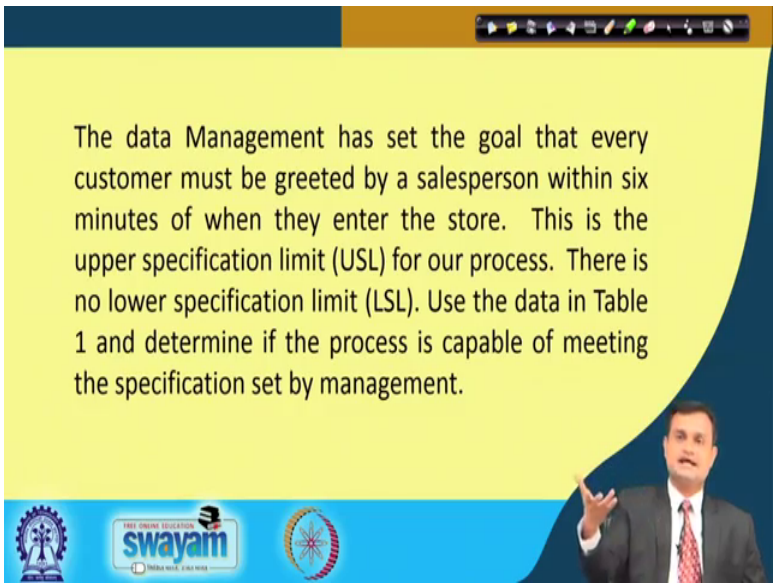
So, you will see that PPU is point 2 3 pp k is point put 23. So, my process is not capable to meet the required specifications or my target of completing a project and there is a need to improve upon my implementation plan I included another example you can quickly go through this, but the same procedure is applied.

(Refer Slide Time: 28:09)



I have the data set describing the customer to be agree they are greeted by a sales person and how long this process takes. So, usually customer is greeted very quickly, and if you keep customer waiting and then you greet then it will lose the value of greeting and welcoming.

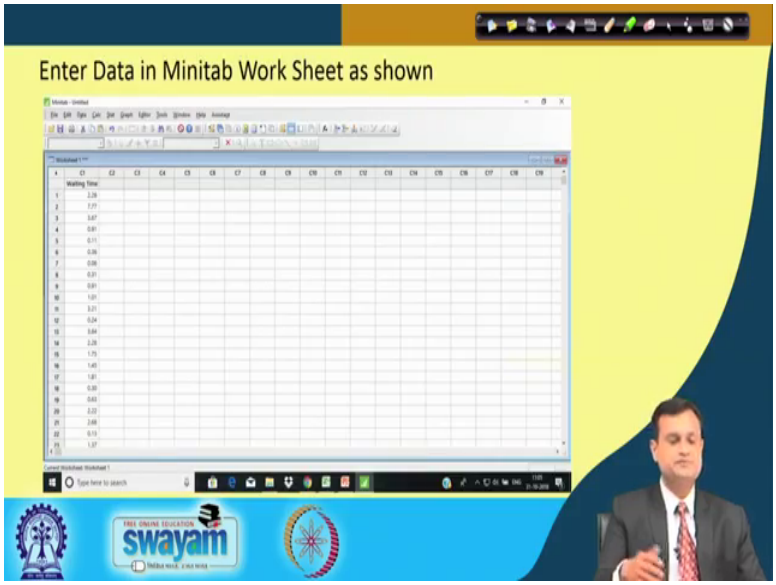
(Refer Slide Time: 28:33)



So, this is the data I have and this data said you can use in fact, to conduct the analysis as a hands on exercise I have given here the USL this is the upper specification limit for our process and within six minute it should be done and there is no lower specifications as

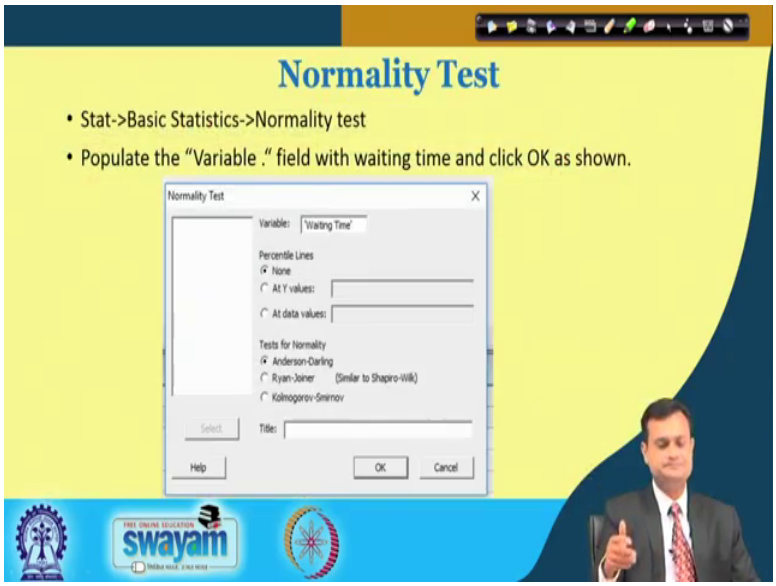
early as possible, the moment customer comes you greet welcome him and let us say you have the data given in your table one.

(Refer Slide Time: 29:03)



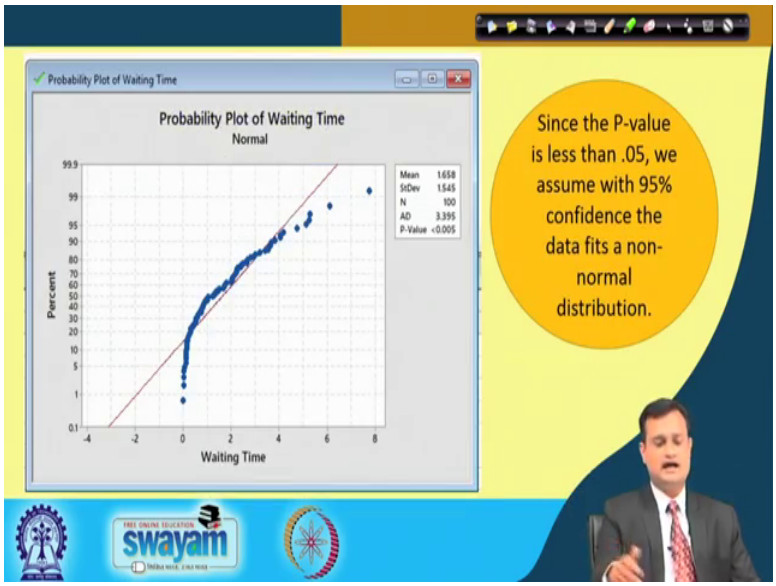
So, open these window enter the data.

(Refer Slide Time: 29:06)



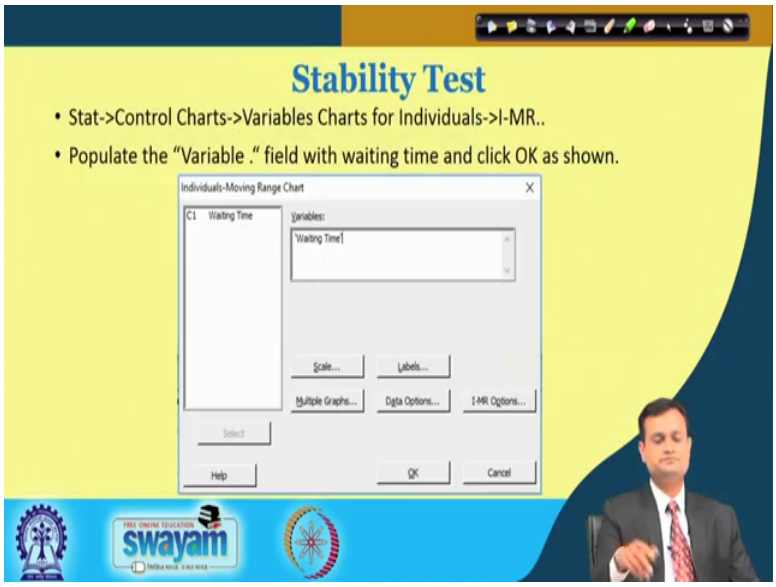
Choose this normality test first let me check whether my data is normal or not normal.

(Refer Slide Time: 29:14)



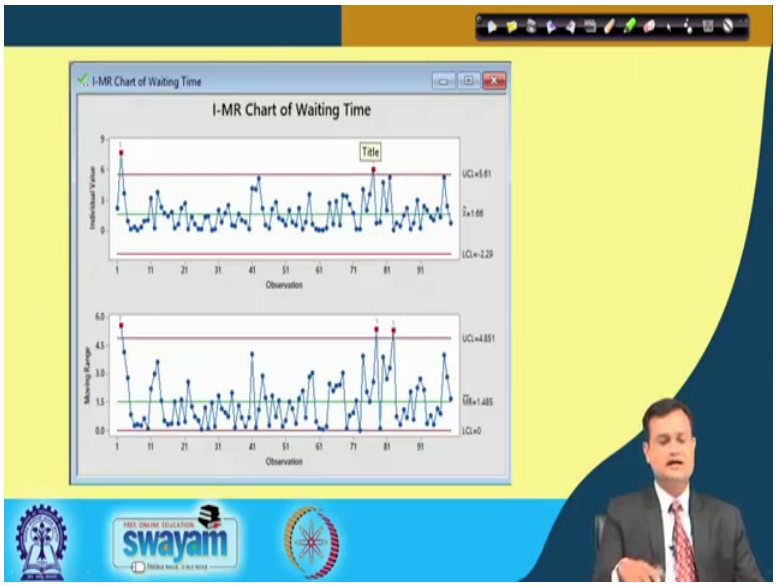
I will get this probability plot clearly I can see it is not normal.

(Refer Slide Time: 29:20)



Then I can check the stability.

(Refer Slide Time: 29:23)

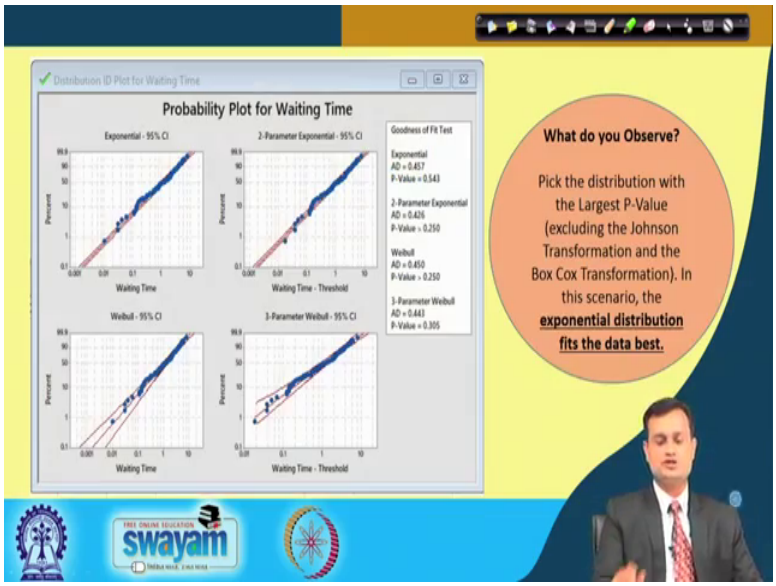


I will see that couple of points are falling outside indicated in red. So, take the corrective action remove this point consider the other points.

(Refer Slide Time: 29:33)

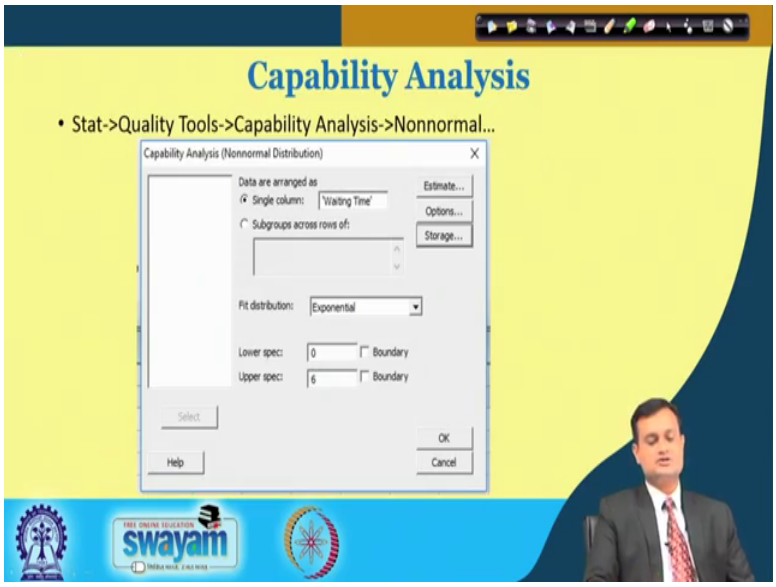
Go to stat quality tool individual distribution.

(Refer Slide Time: 29:37)



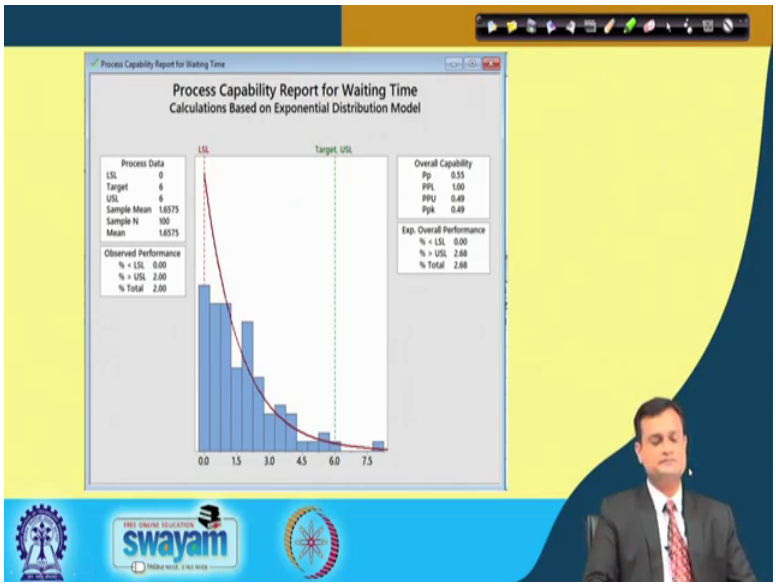
Conduct the analysis again I can see that my best fit distribution here is exponential. So, I would consider this as the best for distribution for conducting my process capability analysis.

(Refer Slide Time: 29:56)



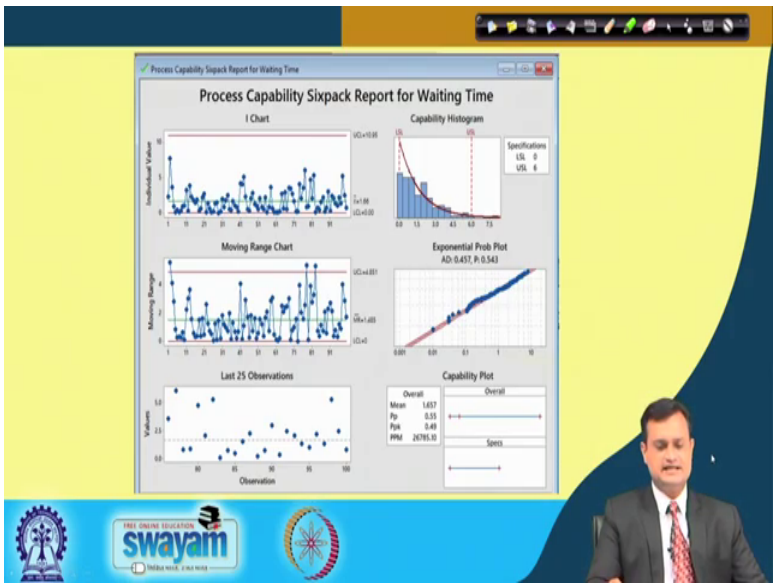
So, then you put the consider the exponential define the variable in column upper specification, that is lower specifications zero immediately you can read welcome the customer.

(Refer Slide Time: 30:08)



You will get the analysis with says Pp is lower than 1 not acceptable PPL is just one both PPU when people are less than 1. So, the sales people or the people at the desk they are not very well trained or the rush of customer is very high I need to really correct my process of greeting welcoming the customer so, that, that can give customer a better satisfaction.

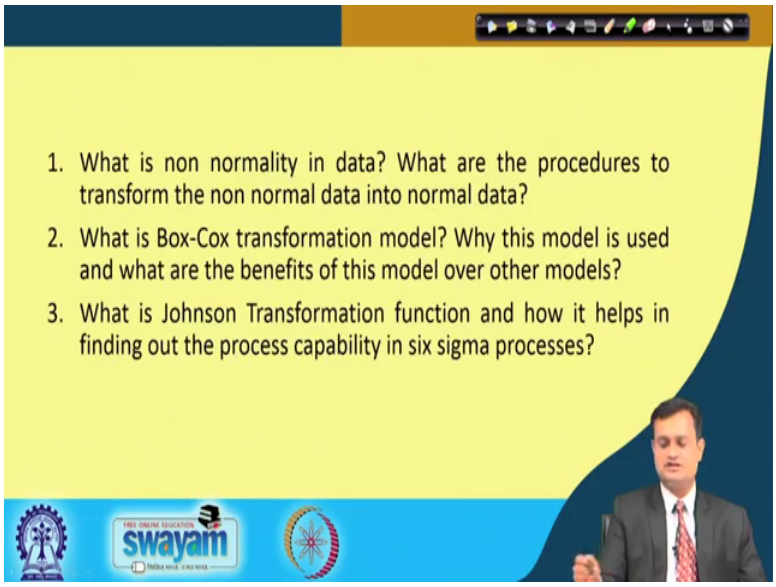
(Refer Slide Time: 30:51)



So, this is the capability analysis six pack that will also help you to understand the capability of your process, but how could we give use six different say display in one

window and r chart I chart stability of the process then observation plotted over a period of time there is there a trained exist and then normality through histogram as well as best fit distribution and then process capability indices.

(Refer Slide Time: 31:25)

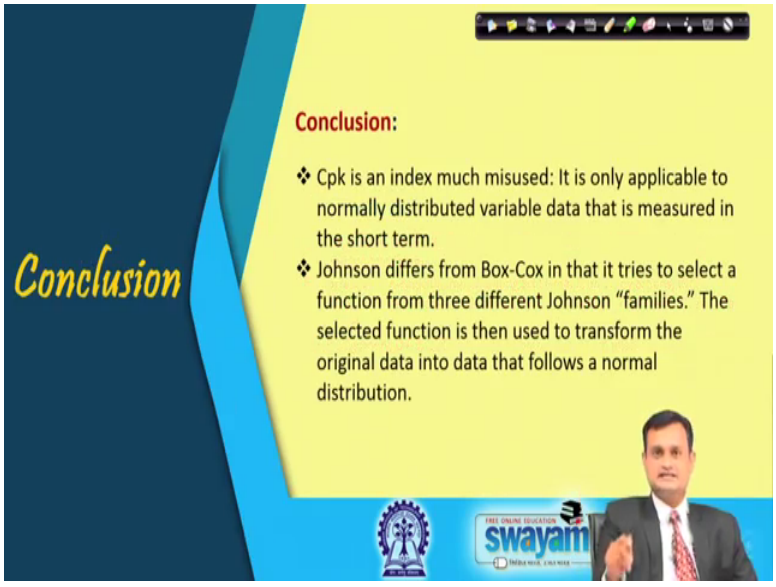


Let me just end the session with a small think it. What is non-normal process capability and what is the non-normality in the data and what are the procedures you would like to follow to convert your normal data into the normal data. What is Box Cox transformation, why this model is used and what is Johnson transformation function, how it helps in finding out the process capability in six sigma processes.

(Refer Slide Time: 31:53)

You can use this references which has given the details of non-normal process capability, but I would insist that you use the same data or some other data and have hands on exercise in mini tab. So, that you feel confident in conducting as well as interpreting the results of non-normal process capability analysis.

(Refer Slide Time: 32:18)



So, Cpk is an index, misused only use it when the data is normal otherwise convert your data non-normal into normal or choose the best fit distribution automatically your

software will take care. And, it will convert it into normal and give you the best result appropriate result of the process capability.

Thank you very much, keep revising be with me enjoy.