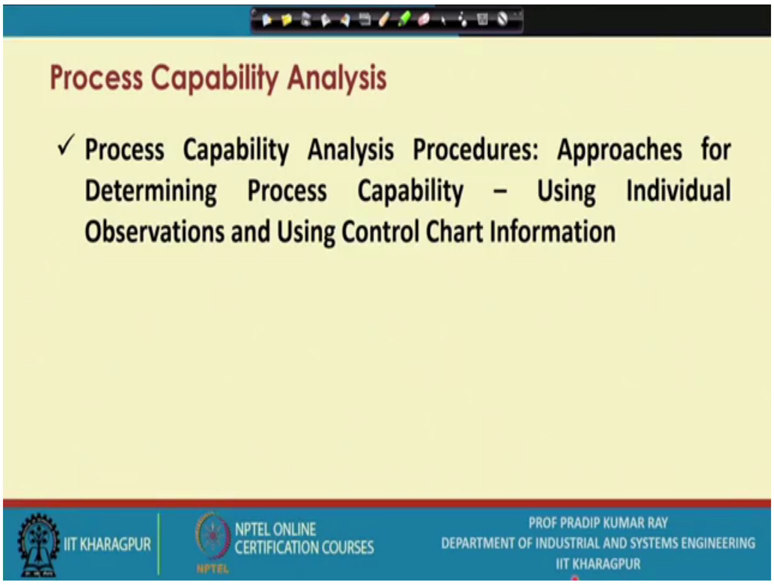
Quality Design and Control
Prof. Pradip Kumar Ray
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 28
Process Capability Analysis (Contd.)

We are now discussing a process capability analysis or the different aspects of process capability analysis.

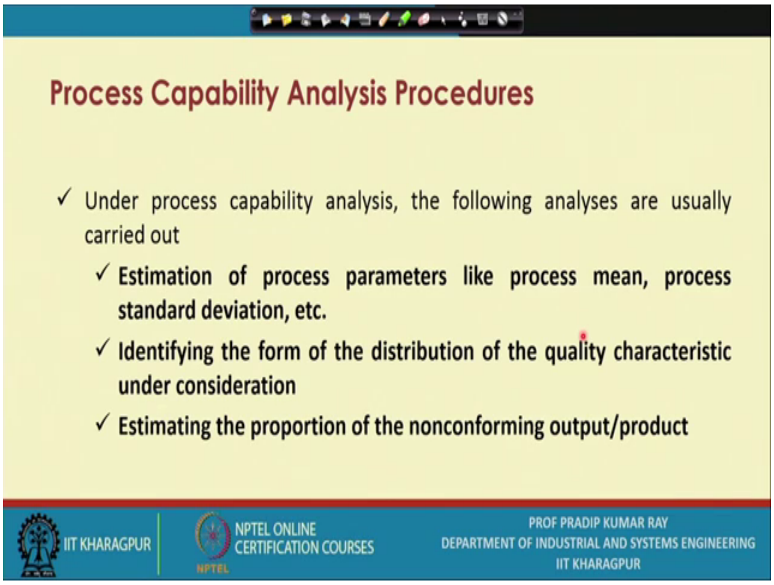
(Refer Slide Time: 00:25)



In this presentation, I will be focusing on the process capability analysis procedures and what are the approaches we need to adopt for carrying out a process capability study and; obviously, there are there could be different approaches, and these the particular approach we select based on the in which form we are able to collect data.

Now, here when you talk about approaches for determining process capability related to a particular process, either you know you are allowed to use the individual observations. So, that is the first case or in many a time you know the control chart information is also made the available. So, you can also use control chart information. So, I am going to elaborate the approaches you adopt or you use for carrying out process capability analysis.

(Refer Slide Time: 01:31)



Now certain important points you should keep in mind, you know we have identified a process for which process capability analysis is to be carried out and; obviously, you know we are running the process and while you run the process in order to produce something say parts or the components or the product itself, and what do you need to know that is what are the process parameters.

So, identification of the process parameters is a must, and then how do you relate these you know the process parameters with the kinds of operations, you carry out on a particular process. So, this knowledge is a must. Now assuming that you have these this knowledge for a particular process now, 3 kinds of analysis we need to carry out.

First one is estimation of the process parameters and as I have already mentioned that there are two important process parameters in this context; one is the process mean that you must know and the second one is the process standard deviation, there could be other process parameters also.

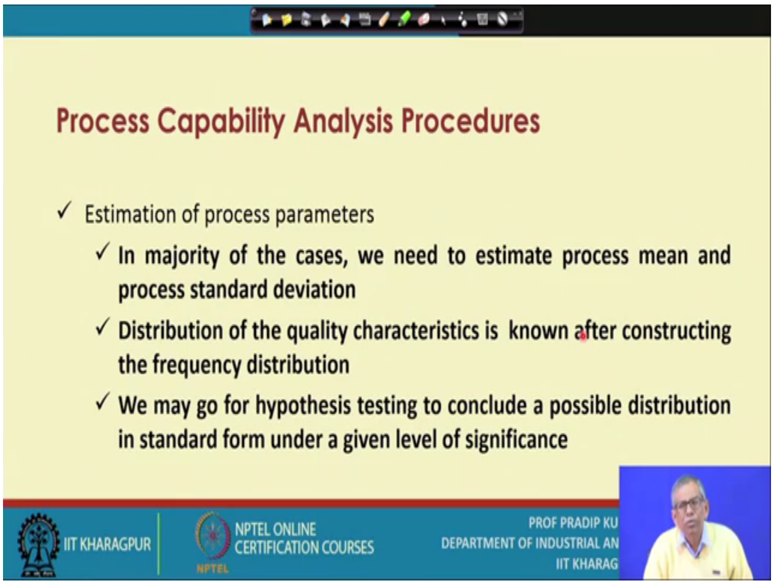
But in majority of the cases we focus on two important process parameters and we must get an estimate for them is it that is the first step in process capability analysis. So, this is these two process parameters are process mean and the process standard deviation or the process variants.

The second step what we try to do you we identify the form of the distribution of the quality characteristics under consideration you know we are already aware of that the quality characteristics or the types of other random variable you will come across. Now there must be some distributional assumption for a particular random variable, and the quality characteristic the values which you get there is all random variables.

So, the next important you know the aspect to look into, that is to identify the form of the distribution. I looking at the form of the distribution, now you can you may hypothesize that what could be you know the type of distribution for the given quality characteristics like whether it may be in the standard form like normal like exponential, like poison etcetera etcetera or it could in the empirical form also is it ok.

And the next step which we carry out we are concerned with that is estimating the proportion of the nonconforming output or the product; means ultimately you know this proportion of nonconforming output actually indicates the process performance, that is we have already pointed out.

(Refer Slide Time: 05:01)



Now while you estimate the process parameters? Now let us focus on estimation of the process parameters like the process mean, and the process standard deviation. There are few observations there few conclusions we may make like in minority of the cases. We need to estimate process mean and the process standard deviation this point we have already mentioned, and then once this is known; that means, the estimate of the process mean or estimate of the process standard deviation.

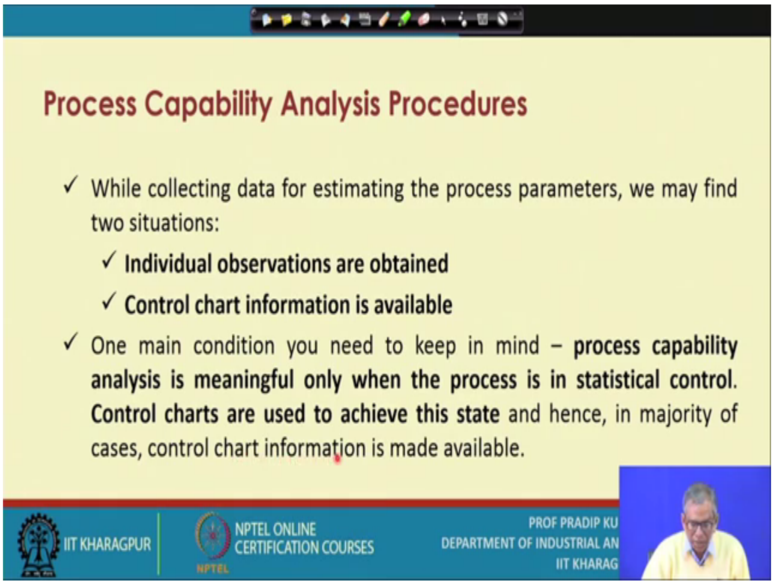
Then distribution of the quality characteristic is to be known after constructing the frequency distribution; that means, the first exercise you should carry out when you start collecting data. So, was the 100 such the data points you have collected from a running process are related to a particular quality characteristics ok.

Then of the component and then what you try to do; that means, first you construct the frequency distribution. And looking at the form of the frequency distributions you may conclude at the initial stage subjectively, the what could be the possible distribution of of the quality characteristics.

We may go for hypothesis testing; hypothesis testing to conclude a possible distribution in standard form under a given level of significance; like we might say that my level of significance is say 0.01 or point 0.05. So, the distribution is assumed to be is assumed to be normal distribution is it. So, for that the standard procedure is hypothesis testing ok.

So, already we have mentioned that in statistics there are two kinds of statistics we have, one is you know descriptive statistics and the second one is the inferential statistics. So, here in this case we will go for hypothesis testing that is a part of inferential statistics.

(Refer Slide Time: 07:23)



Now, the first thing you must ensure before you carry out a process capability study for a process, that t sufficient data are made available and you are allowed to collect this data. So, while collecting data for estimating the process parameters, we may find two situations ok.

So, what are these two situations? The first one is you just go to the process you consult with the concerned person, and you ask him or her what kind of data related to a particular quality characteristics you are you are collecting as the production process you know is on.

So, the individual observation say quality characteristic is a random variable x, and you might say that last one weeks data or last one fourth nights data are made available. So, these are the individual observations and how to use these individual observations for process capability analysis. So, this could be one case.

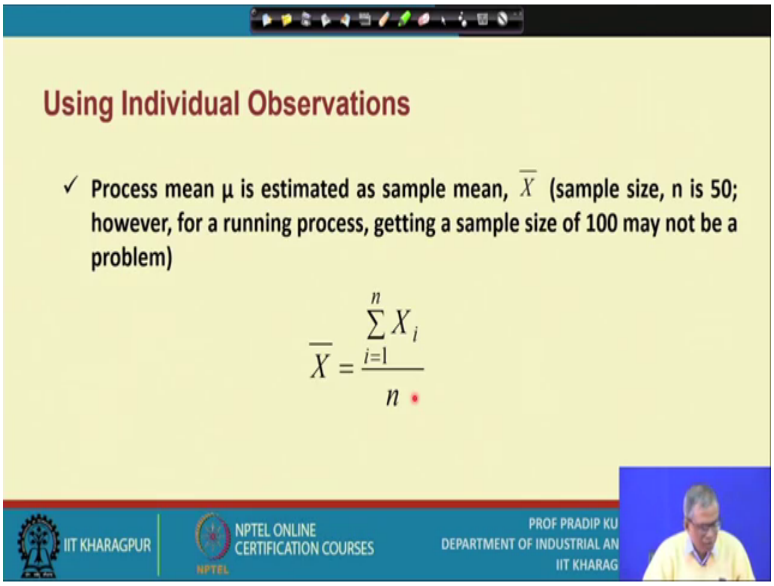
The second case is that is the control chart information is available as you may be knowing, and we have always emphasized this aspect that whenever you conclude that the process is in statistical control, and then you are supposed to measure the performance of the process. So, the process capability is indicative of the process performance this point already have mentioned, but make sure that the process is in statistical control is right.

Now, how to make that the process or how to achieve this condition this condition of the state of statistical control; obviously, you need to use certain tools and this this these tools which you must be used in for online real time control the process. So, the control chart is one such tool. So, prior to estimating the process parameter values is it in a control condition, it is assumed that you have been using the control chart or substantial period of time, and you have achieved this condition that is the state of statistical control right.

So obviously, in majority of the cases in almost all cases we will find, that the control chart information is already made available, and if you refer to this control chart you know the information at the data obviously, you will be in a position to carry out the process capability analysis. So, the second situation is very very common whenever you carry out process capability analysis one main condition you need to keep in mind.

So, this is to be kept in mind this is important. So, the condition that the process capability analysis is meaningful only when the process is in statistical control, is it this point we have been emphasizing all the time. So, control charts is used to achieve this state or this condition, and in hence in majority of the cases control chart information is made available. So, this point already I have elaborated.

(Refer Slide Time: 11:24)



Now, let us talk about estimation of the process mean. So, process mean the notation is mu and as may be knowing that we always say that the process mean remains unknown; that means, it is never known what you can estimate that is a good estimate for the process mean.

Now, how do you get a good estimate; obviously, from the population or the from the process you have to draw a sample, and how to draw the sample there are many procedures it would be a probability sampling procedures or it could be a nonprobability sampling procedures, but what we say that make sure whatever may be the procedure make sure that the sample or the subgroup is representative of the population; that means, the kinds of characteristics you have you have to identify the characteristics of the population, and you must have the same sort of characteristics in the sample is it. So, that is to be ensured.

And another condition you must ensure in many cases that is the condition of randomness is it. So, you know there are different procedures. So, in course of time will be will be telling you particularly when we will be discussing you know acceptance sampling is it. So, there is another important topic in quality design and control will be elaborating on this concept that, how to draw the sample what sort of procedures you have so that the sample have becomes a representative of the population or the process.

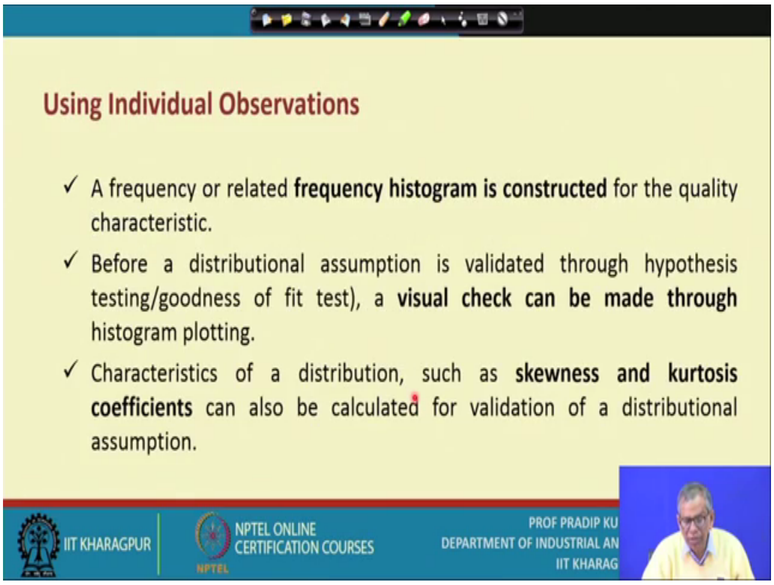
So; obviously, you know the here from the population or from the process; that means, what is the population that; that means, suppose the total the number of produced units in a shift say 8 hour duration, you said 2000 units is it 2000 units of a component of a part. So, that is basically the population size; obviously, it may not be necessary or it may not be practical to get 2000 say mess the data points.

So, what do you try to do, you try to get a sample representative of the population and the now you know you consider the data points in the sample. So, here what do you try to do; that means, the process mean mu is estimated as sample mean and the notation is x bar. He just always we have these notations, you refer to X bar control chart we will find that this notation you use.

So, sample size n is restricted to 50, though sometimes we say that you know n is greater than equals to 30 may be considered to a large sample. So, the sample size should be as large as possible and in a running production system in a running process, getting a sample size of 100 may not be a problem is it. So, this we have observed like the practitioners they will say give us just the two days' time. So, I will have a sample of 100 units or even two 100 units is it ok.

So, in a running production system a repetitive manufacturing, this it will not pose any problem. So, then the sufficient number of data points you will have in the sample next what do you do? You just calculate the sample mean is it or the sample mean that is sigma i equals to 1 to n, X i divided by n is it. So, it is an arithmetic mean you calculate.

(Refer Slide Time: 15:31)



Next what do you do; that means, a frequency or related frequency histogram is constructed for the quality characteristic is it. So, this is. So, the next exercise you carry out now the before a distributional assumption is validated through hypothesis testing. So, you go for the goodness of fit test.

Now, we say that the random variable may be assumed to be a normal random variable. So, how do you how do you conclude this; obviously, go for hypothesis testing is it and as I said there are many kinds of the tests you have like say the chi squared tests you have. So, later on n we will be referring to that and but prior to the carrying out this goodness of fit test a visual check can be made through histogram plotting; that means, first you look at the histogram ok.

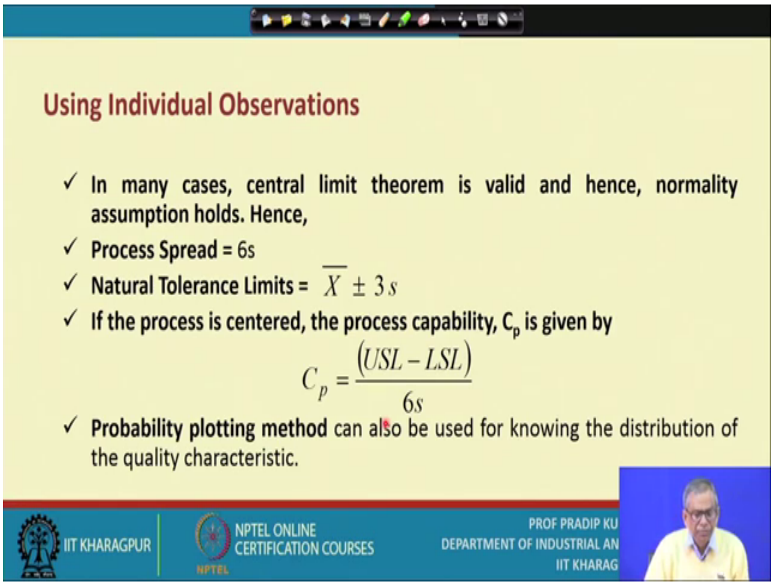
And then you conclude that what could be the possible at the distribution; this is subjective, but the subjective assessment is a must before you go for objective analysis is it. So, in this case if you plot the histogram, now you will have some informations like by looking at the shape you may conclude whether the you know weight whether the distribution is unimodal or multimodal, whether the distribution is symmetric or not is it what is the peakedness of the distributions is it ok.

And similarly you know the skewness, whether it is skewed to the right or the skew to the left. So, these sort of observations or say subjective judgment you may make regarding the distribution regarding the random variable under consideration. So, this is the must.

So, characteristics of a distribution such a skewness is it, and kurtosis coefficients can also be calculated is it. Like say in terms of the moments of the distributions; that means, as the first moment, second moment, third moment, fourth moment. As you collect more sort of information, you may be in a position to compute the third moment of the distribution as well as the fourth moment of the distributions.

So, when you calculate you have the data with which you can calculate the third moment; obviously, know the skewness coefficient you can calculate and similarly if you have sufficient information. So, that you can have the fourth moment then; obviously, the kurtosis also coefficient also you can calculate. So, that you may have you know a fairly good idea about the variability about the characteristics of the distribution is it. So, this exercise we do routinely.

(Refer Slide Time: 18:39)



Next what do you try to do; that means, you know we calculate along with the mean we also calculate the standard deviation is it. So, the standard deviation part as you know that the sample standard deviation, that is you know you have n number of data points and make sure that, that is an unbiased estimate of the standard distribution you must have. So, I have mentioned about the mean, but subsequently with the same set of data in the sample you calculate the sample standard deviation is it ok.

So, what you try to do; that means, individual values you must subtract from the individual values the mean value; that means, X bar and the you know the square of this deviation and you add them for all the individual values and then you divide by n minus 1 is it, and that is that is essentially the variance and then for the given sample of size n. So, you calculate this positive square root of the variance and that is the standard deviation is it. So, you can get that formulation there is absolutely no problem is it.

So, with respect to the X bar you calculate the standard deviation, and first you calculate the variance and then you calculate the standard deviation. In many cases central limit theorem is valid and hence normality assumption holds particularly for running production systems, and the system is such that when you get the value of x; x is basically you know this value or say you know the notation could be y or that is say the output say the quality characteristic.

Now, in many cases we will find in the repetitive manufacturing or for that matter any kinds of manufacturing systems when this value you get in the you know the downstream stage is it. Now obviously, there are many factors affecting the value of y, factors in the upstream process there are many factors affecting. If this is the you know this is the situations; that means, you may always assume that the y or the output quality characteristics may be normal is it. So, this is the one the physical basis you must aware of before you go for normality assumption.

And the second one is you know the central limit theorem you should be aware of is it because many sorts of tools and techniques the expressions you come across n quality control and n quality control exercises, and the basic assumption is the normality assumptions. There are cases where you may have non normality assumptions, but in majority of the cases you assume normality.

Central limit theorem states that whatever may be the distribution of x, x is a random variable. The distributions of any you know the statistic of x whether it is a mean or whether it is a standard deviation may be assumed to be normal when the sample size is very large. So, we are dealing with a large sample sizes running production systems.

So, how do you define the process spread, I have already have mentioned. The process spread when you assume that you know upper natural tolerance limit, lives say at say 3 sigma away from the mean with this assumption, and similarly the lower natural tolerance limit related to a particular process lies say minus 3 sigma away from the mean; that means, on the other side of the mean.

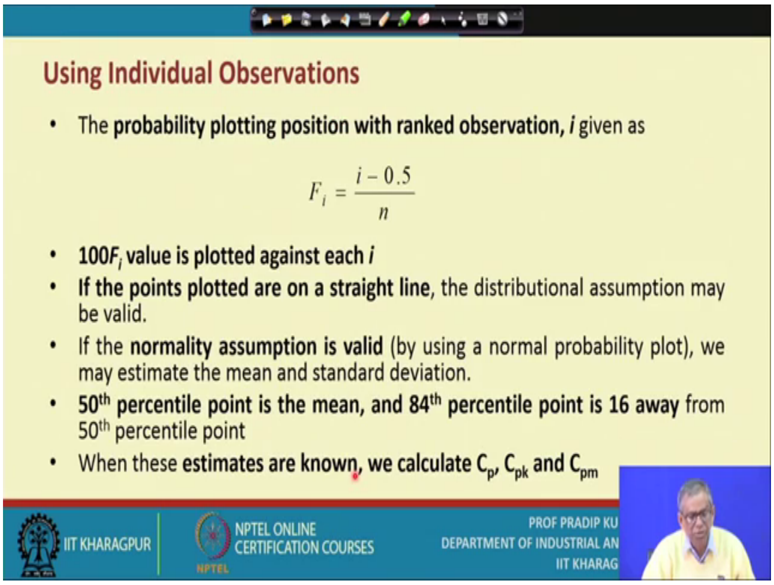
So, if these assumption is valid, then the process spread is; obviously, 3 sigma plus 3 sigma that is 6 into sigma is it. So, this is just one case otherwise it is essentially you know 2 k sigma. Now the k could be in in this case we have assumed the value of k is 3, but it may be you know in a special case you may assume it to be say 2.5 or say 2 or even 3.09 and many. So, depending on a.

Suppose you will you will face a you know a critical situation or say suppose you know the specific situation. So, the value of k could be 2.5 or choice is yours, but otherwise unless otherwise specified you assume it to be 3. So, that is why it is 6 into s. So, what is s? S is the sample standard deviation why not sigma? Because sigma value actual value of sigma is never known what is known is, actually the sample standard deviation and the sample is assumed to be represented a one.

So; obviously, the natural tolerance limit will be X bar plus minus 3 into s; that means, here the mu is replaced with X bar that is the sample mean and a and sigma is replaced with the sample standard deviation. So, if the process is centered we have already defined what is a centering of the process, the process capability C p or process capability ratio it is also known as pcr C p is given by USL minus LSL; that means, this is a two sided specification limit double specification limit for v is the upper specification limit and l and the lower specification limit these two values are known divided by 6 into s.

So, the probability plotting method can also be used for knowing the distribution of the quality characteristics. So, this is another method that means like say many a time if you want to the test for normality.

(Refer Slide Time: 24:59)



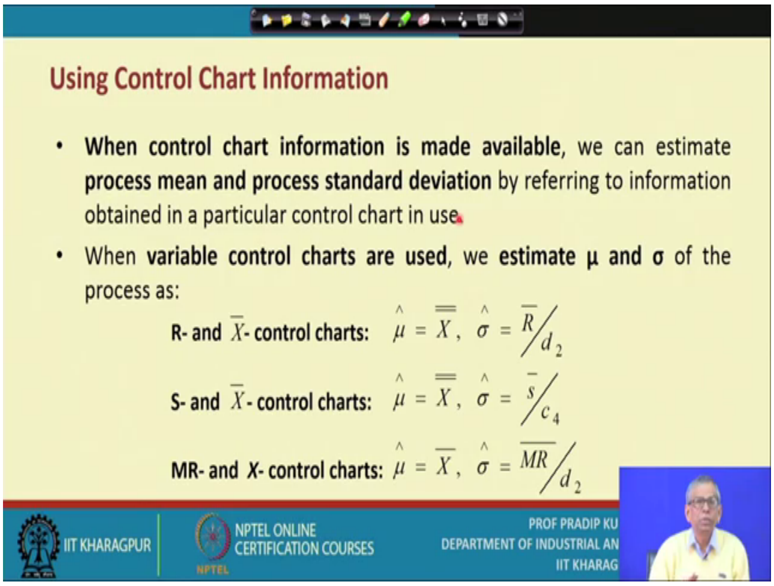
So, I do not even use normal probability plot. So, the probability plotting method is also widely used is a it is like. So, few simple steps I will explain; that means, suppose you deal with n number of observations x i, i varies from 1 to n that is the situation, what I actually you have to do; that means, you have to rank them; that means, from say the highest to the to the lowest one you have to ranked.

And then the ranked observations that is i, ith observations is ranked observations. And then against a ranked observation, you calculate fi is it that is i minus point 5 divided by n why it is point 5? There is a there is a rationale behind it. So, as a student as a learner you must know how did you get this 0.5 is it ok.

So, please find out the reasons of that why we subtract 0.5 from high, why not 5 why not 50 and all those how do you get these values 0.5. So, there is reason behind it. So, I expect the student will be knowing the exact reasons divided by n; that means, n is the total number of observations. So, 100 if I value is plotted against each i and if the points plotted are on a straight line the distributional assumptions may be valid is it if the normality assumption is valid by using a normal probability plot is it we may; that means, you have to use I know the log scale basically we may estimate the mean and the standard deviation.

So, what is that estimate? Fiftieth percentile point is the mean and 84th percentile point is one sigma away one sigma away from fiftieth percentile point this is one stigma when these estimates are known we calculate C p C pk and C pm.

(Refer Slide Time: 27:32)



Now, this is one aspect; that means, you are dealing with individual observations. When you use the control chart information, is it already you have been using a control chart and by using the control chart, you conclude at this point in time when the process capability analysis has started you assume that the process is in statistical control; that means, the process behavior is not being influenced by the assignable causes or the external factors of the assignable causes.

When control chart information is made available in many cases we do; that means, the first exercise we carry out are you in a position to apply the control chart. If you are successful in using or in in developing as well as the using control chart for your process, then the next question comes you have been using the control chart very nice, but are you in a position to measure the performance of the process is the performance of the process improving so; obviously, you have to carry out the process capability analysis. So, this is the context.

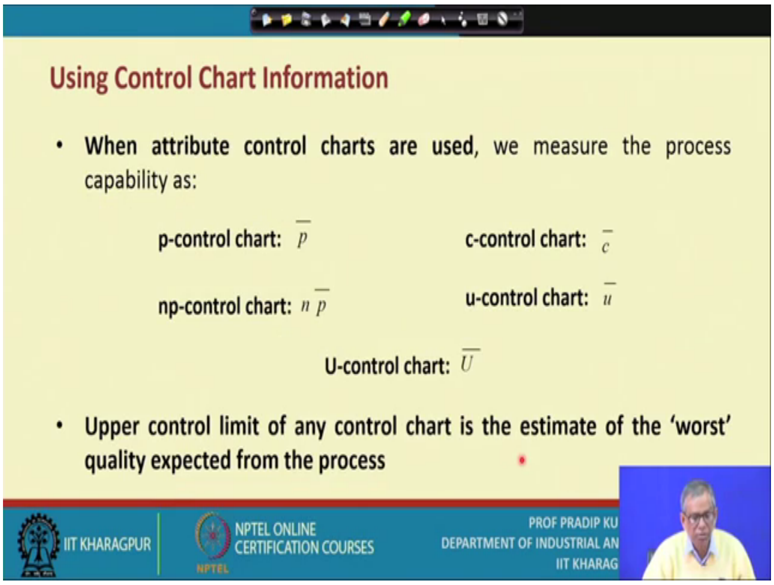
So, in almost 100 percent of the cases what we have found that the control chart information is available. We can estimate the process mean and the process standard deviation by referring to information obtained in a particular control chart in use. So, it is very simple when variable control charts are used as you know there are two kinds of control charts we have been using like some variables control charts.

So, the variables control charts you know the using the variables data which are measurable, for which gauges are used or the measuring instruments are used. So, you estimate mu and sigma of the processes as if you use R and X bar control chart as I have already pointed out that these two control charts must be used simultaneously. If someone say uses only R chart around the X bar control chart. So, absolutely you know the purpose is not served is it ok.

So, the first you use the r chart and then with the same set of data you used for constructing the X bar chart and you start using X bar chart. So, if you have been using is; obviously, what is the estimate of me process mean that is x double bar; that means, the mean of the means is it. So, this already have already computed similarly what is you know the sigma estimate of sigma that is, you know an estimate of the process standard deviation, that is R bar already you have computed while you refer to say R control chart divided by d 2 and d 2 is basically a control chart factor control chart related factor d 2 and use the function of n, and you can read this value from the control chart factor table is it.

Similarly, instead of using r control chart suppose you we use s control chart. So, that is an estimate of process mean is X double bar and estimate of the process standard deviation is s bar by c 4 again c 4 becomes a function of n. If you use the you know the when you deal with. So, the individual observations or the individual units; that means, sample size is one; obviously, MR that means, moving range control chart and x control chart you will be using. So, what will be the estimate of mu that is X bar is it the individual observation sample size is one for each sample, and the estimate of sigma is MR bar by d 2 is it ok.

(Refer Slide Time: 31:34)



So, when you use attributes control charts, suppose you are using p control chart. So, or the process capability is p bar. If you use c control chart process capability is c bar, if you use np control chart; that means, the sample size is constant the process capability is n into p bar we you refer to all these control charts and u control chart is u bar; that means, these are basically the center line. And if you use capital u control chart demerits per unit control chart or controlling. So, this is capital u bar upper control limit of any control chart is the estimate of the worst quality expected from the process its clear.

So, here directly you know the measurements are missing is it. So, you are dealing with attributes data. So, what you say that, here is an attributes control chart I am I know that what is the center line. So, the central line is indicative of the process performance is it and similarly you have the upper control limit. So, upper control limit is basically indicating the worst possible quality level is it ok.

(Refer Slide Time: 32:57)

So, these are the ways you can estimate the process capability under various situations. So, we will deal with the other aspects of a process capability in the subsequent lectures.