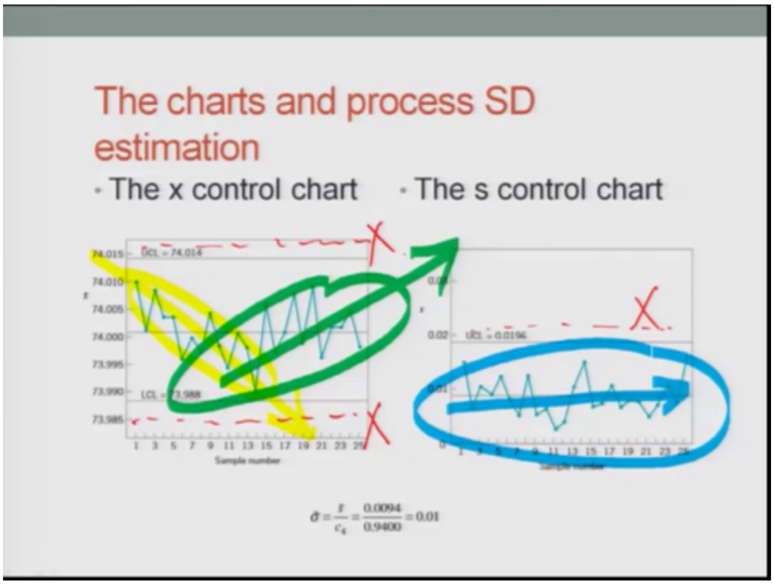
Total Quality Management - I
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 26
S square chart and MR chart

Very good morning, good afternoon, good evening, my dear friends. Welcome back to this 26th lecture for TQM under the NPTEL MOOC program; I am Raghunandan Sengupta from the IME department, IIT, Kanpur. So, if you remember we were discussing about the charts the concept even like I am going up back about 2 or 3 lectures before. So, we had x bar charts and then the R charts, considering the concept R why we are using is the range when the standard division is not known in case we want to replace R which is the actual concept of dispersion or the overall variability which is there in the readings. If you want to measure them, obviously, we need the standard deviation or the variance of the population. And if we want that, we will use the s charts.

And we also saw that how depending on the sample size the coefficient of a 1, b 1, c 1 these would change accordingly from the tables and we can use them accordingly. And if you are trying to eliminate or include more samples how would the sample size have an effect on the coefficients would also be discussed later on also and we have discussed it previously also with the change in the formulas and we will continue discussing now in the twenty sixth lecture and so on and so forth.

(Refer Slide Time: 01:46)



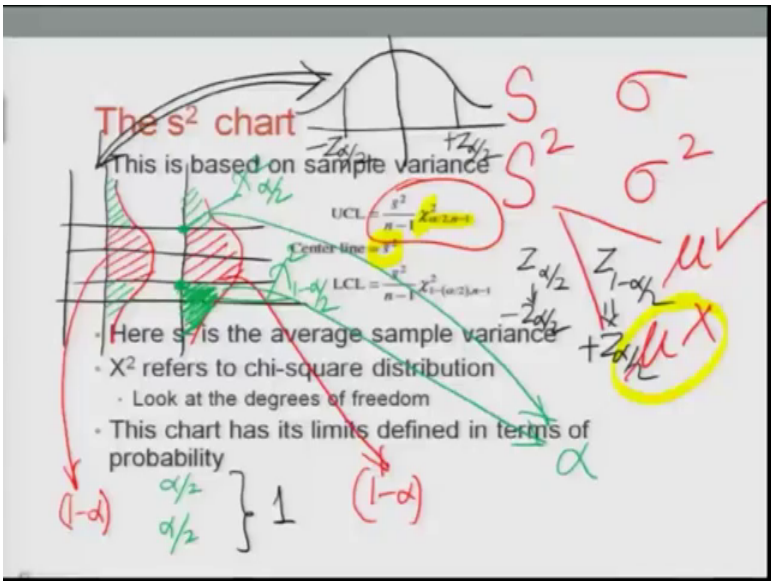
So, the chart in the process is for standard division estimation is if you remember we had change the sample size, we have recalculated the mean value. Mean value can be there x bar bar charts for the x bar charts bar bar line sorry for the error it would be the x double line as for the mean, and you have the upper control and the lower control which we will use for the x bar charts. And then you will basically have the central line which is the R bar for the R charts, and you have the upper control and the lower control depending on that. And as the sample size change you would have an effect on the x bar and the R charts also.

So, continuing the discussion, we are trying to find out the actual central line or the upper control lower control limit for the standard division estimation process. So, if it is the x control charts, here you had the concept of upper control and lower control given as 74.014 73.988. And based on that you can find out where the fluctuation is if at all and then basically make very intelligent understanding of how the process is going whether there are assignable causes, unassignable causes are there and you can take actions accordingly.

So, if you consider this chart, which I will highlight. So, in this region, if you consider, there is a downward trend. So, I am using the highlighter only. And then if I use a different highlighter with the color say for example, green, so again they would be a change and it will go in upper trends of upward trends. So, you have to basically make an analysis that downward trend, upward trend, if whether is due to some causes or whether they change in the overall workmanship. I am talking from the manufacturing prospective only. There is a change in the quality of the raw materials, cutting tool has changed or say for example, the fluid which is being used for the as the coolant has changed, whether the temperature in the room has changed whether maintenance has been one on the machine, whether the humidity has changed. So, all these things would come under the picture.

And if you consider the s chart which is basically for the standard deviation so; obviously, the trend would basically that would give you the variability. So, here I will use another highlighter with the color blue. So, the whole trend is basically coming on and average in a straight line, so that will give you that even if the means are fluctuating, the overall trend or the variability in the whole process is under control, because it is not deviating going over the upper control values. So, I will try to use the (Refer Time: 04:43) and with the color as red. So, any deviations which are happening above are not there any deviations obviously, they would not go below 0 is not there in the same case any movement upper the upper control limit below the lower control limit are not there.

(Refer Slide Time: 05:04)



Now, if I considering the s square charts, s square is basically the best estimate so called best estimate for the variance. So, what you have is basically the s, which is for the standard deviation, you will basically have the s square for the variance. Now, this s square can be of 2 types technically when mean value is known which is a tick mark and mu value is not known. So, based on that, you will do the calculation. And if I remember correctly I did mention in the formulas as we are discussing that how you can find out the variance of the sample given the mean value known and variance of the sample when the mean value is not known.

So, you have to basically use the formula in the first case as 1 by n and the summation whatever it is summation would be summation of x i minus mu whole square. And in the case when the mean value is not known you replace n by n minus 1 that means, there is one loss of degree of freedom. And in the bracket inside the summation sign, it will be x I minus x bar n whole square because the mu has been replaced by x bar n which is the best estimate.

Anyway continuing the discussions, so this s square is the chart is based on the sample variance. So, the upper control limit would be given by this. Now, this is the first time we are seeing this formula. So, I will highlight it using yellow color. So, this distribution we see here is basically known as the chi square distribution. Now, the chi square distribution, f distribution, and t distribution, which I did mention a few times are basically the sampling distribution which comes from the study related to something to with the normal.

So, technically we will use chi square distribution when we have to study something related to the variance. We will use f distribution, when we are something to study with the variance, but for 2 different populations. And when we are studying the t distribution or the z distribution, we have to study something to do with the mean. So, if it is mean it would be with depending on the assumptions, the distributions would be either z which is the standard normal distribution or the t distribution, whatever the degrees of freedom they are would be to do with the mean of the population. And chi square and f would be to do with the variance, but the chi square case we are only concentrating on the variance of one population; and in f distribution, whatever the degrees of freedoms are it has to do something with the variance of 2 different populations.

Now, if I remember I did mention and correct me if I wrong that in one of the classes previously I did mention that for the normal distribution, the sample mean and the sample variances are independent of each other, which is a very important characteristics for normal distribution. It is not true for the other distribution as such. So, in that case, if we consider then the sample variance is distributed chi square with a certain degrees of freedom. So, now where the chi square is coming up that will give you some information with my statement, which I just made few seconds back.

So, the chi square distribution is coming. So, here where the chi square is, so I will basically now high light the degrees of freedom and the and the efficiency level of calculations. So, this degrees of freedom is basically n minus 1. Now, here let us a take a pause, if you remember I did mention in the part when the mean value is not known, you lose one degrees of freedom. So, this n minus 1 is basically coming from that fact that you have to use the set of observations x 1 to x n, the first time to find out something to do with the mean or the populations such that you use the sample mean to basically replace than that, so that is point one.

And point number 2 is that obviously, if you are using that you are trying to utilize the sample mean as the best estimate for the population mean that is why it is n minus 1. So, in that case the central line would basically be s square, s dash or x bar is basically the best estimate. And the corresponding upper and lower control limits would basically depend on what is the overall coverage area you are trying to basically cover. So, if you remember the value of the overall efficiency or overall say for example, degrees of freedom corresponding not to end, but basically with respect to the value of alpha which you are using, alpha is basically the level of confidence which you have. So, if you are trying to basically understand that then alpha would give you the exclude and include part which is there over and above the upper control limit and this and lower control limit. So, this is I will basically try to highlight using the graph.

So, let me try to do. Say for example, I have this central line, this is the zero value, this is the lower line, this is the upper line. So, lower line and upper line I am drawing randomly, I will come to that later on. Now, consider for the first time that I am trying to draw corresponding to the normal distribution. So, let me draw the normal distribution, let me change the color. So, this is the normal distribution which I have drawn. So, now, if I consider the overall coverage in this area is 1 minus alpha. This is true for the fact that the areas, which I am trying to cover on the left and right side, these are of equal proportions because they are symmetric, hence the sum if you consider is basically 1. So, on the left hand side, right hand side, the overall coverage are alpha by 2, alpha by 2; and the middle portion between the upper control and the lower control limit is 1 minus alpha. So, 1 minus alpha plus alpha by 2 plus alpha by 2 basically means 1. Now, this was a true for the normal distribution case.

Now, if I draw the chi square, so let me try my level best to draw the chi square. Let me again change the color, so is something like this. So, considers this chi square. Now, the values which I will draw on the so this coverage area would be basically be again equal to 1 minus alpha. So, this was 1 minus alpha, this also is 1 minus alpha. Now, I will try to use the color of green, this area and this area they would add up to alpha, but in this case they are not equally distributed onto the left and right side because this is not symmetric.

So, that is why when I use the nomenclature of trying to mark technically I will be using the concept that if I am reading from the left hand side that means, I have turned it vertical and I am basically reading it from the side when I am pointing my left hand. So, that is towards minus infinity I mean coming to this side towards the mean value. So, the overall coverage would basically start increasing.

So, if this total coverage is basically alpha by 2, so at this point the value should be such whatever the chi square value is it would be such that the overall coverage area till that point is alpha by 2. Then I continue doing it. When I come to the next point the overall coverage probability would be chi square, 1 minus alpha by 2 because I have already covered that areas such that the areas still left onto the right which is this one, when I am basically making it much bolder in green color.

This area is still to be covered is alpha by 2. So, the overall area which I have already covered from minus infinity to that point is 1 minus alpha by 2, hence the nomenclature, I trying to mention what is the chi square we use the nomenclature of the value of the degrees of freedom plus the level of confidence which is there, corresponding to that point.

So, if I know basically make a sense with the first set of diagram for the normal distribution then it will be easier for us to understand. So, let me try to make the normal distribution in this slide only. So, the normal distribution which I am drawing is what I have already discussed. So, this is the technically I am just turned it anticlockwise 90 degrees in order to make sense. This is the mean value. This is the z values which are there. Now, what is interesting to note is that this would be z alpha by 2 minus this would be plus z alpha by 2. Now, the interesting fact is these are symmetric. So, as they are symmetric, if you are covering z alpha by 2 technically the overall coverage from the left to that point and for z 1 minus alpha by 2 onto the right till infinity, the overall areas are same which is alpha by 2. But the interesting fact is that their symmetric, because this is the this is the normal distribution hence we use in place of z 1 minus alpha by 2, you use plus; and in case of z alpha by 2, you use minus.

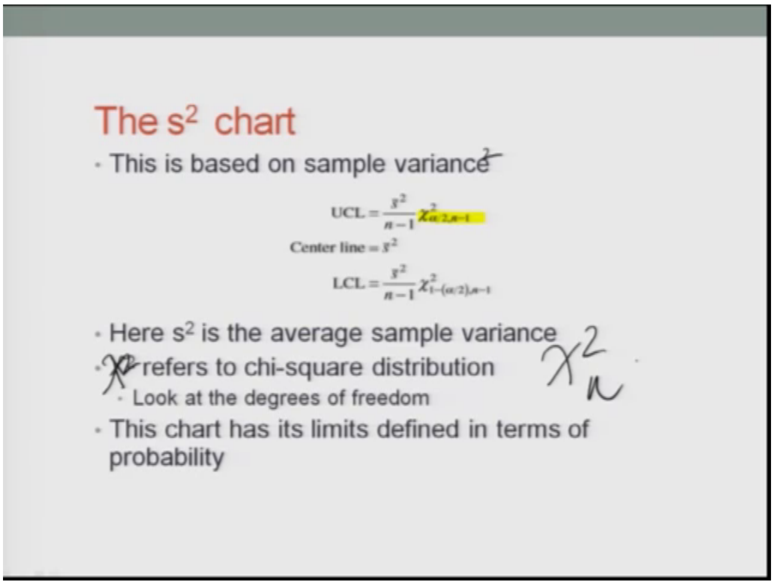
So, this is the actual value only specific to normal distribution, but the nomenclatures are as you are covering from the left the alpha by 2 value or alpha 1 value, alpha 2, alpha 3, alpha 4 value or whatever it is the total area you have covered till that point. So, this has become a little bit too messy. So, what I will try to do is that I will try to remove this. You will just give me 2 minutes let me remove this or else it gives very untidy and it becomes very difficult for me to explain and for you to understand also, yes. So, it is cleaned looks quite neat and clean, but even though I had covered a lot of concepts here, but we will come to the later one.

So, technically the values of chi square alphabet to n minus 1 would basically signified to from which direction you are coming and at that point what is the overall coverage. Degrees of freedom again I am repeating is n minus 1, because I have to ask what degrees of freedom based on the fact that I am trying to find out the sample mean as the best estimate for the population mean, hence it is 1 by n minus 1 when you are trying to find out the sample variance. The central line remains as s bar square which is basically the average values of the samples.

Now, if I count go to the lower control limits, obviously, this would depend the degrees of freedom remains same. The overall coverage would now be based on the fact that whatever total coverage we have covered, so that would be one this is the total area of the property is 1 minus alpha by 2 because those that has been divided in that proportion that the central area covered is 1 minus alpha. And the left and the right portion total sum of that is alpha obviously because that would be sum of them. In the normal case, I may again repeating they are equals that is why I mentioned alpha by 2 alpha by 2, but in the chi square case it may not be equal because this is not so called symmetric distribution that is why you use the actual general nomenclature as mentioned.

So, let me read the points again. I went a little bit slow trying to draw the diagrams, but please bear with me the overall concepts which are covered if you actually go through its slowly you will understand or refer to any good statistic books, this thing could be made very clear in the book that how the general concept of chi square t and f distribution are covered in a very simple manner. So, that once you fix in which direction you are trying to basically cover that probability and how you are going to mention the nomenclature of the names things would become very, very easy.

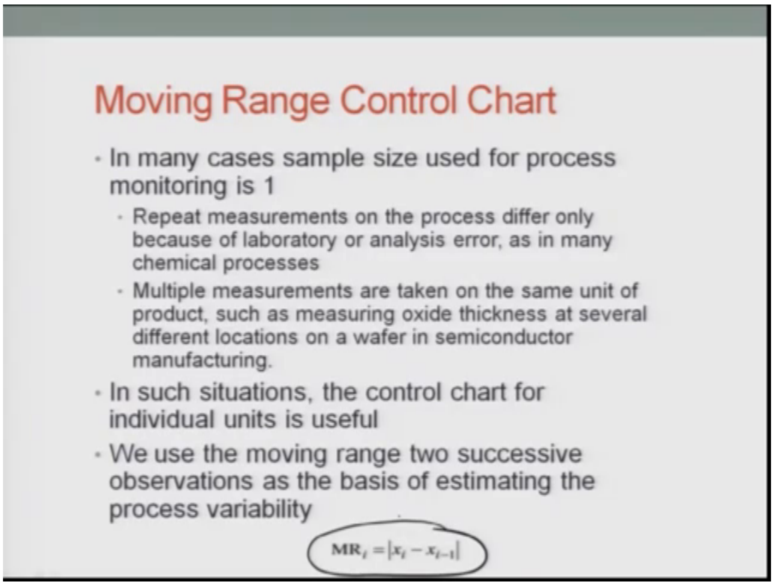
(Refer Slide Time: 19:33)



So, highlighting the bullet point, here s square is the average sample size it should be s bar square as the average sample size. And this chi square which we denote so this is basically chi square which we are denoting differs to the chi square distribution with some degrees of freedom which is n. So, we basically have to look at the degrees of freedom. This charts has its limits defined in terms of the probability, so that probability level of coverage well is basically the alpha value. So, if I am very much confidence, less confident the values of alpha would be changing accordingly.

So, moving rage control that means, now the ranges the average values change depending on the actual input data which is coming. Like say for example, in the morning 9 o clock I took 10 readings based on that I have to give the charts. Say for example, at 11 clock in the morning, I have taken another 10 readings, now I want to basically include the first 5 and from the second reading and include the last five from the first reading excluding, the five first five of the first reading. So, this (Refer Time: 20:34) of the new sample would be five from the first that is the last five from the first and the first five from the second one and when we combine obviously, the overall mean value upper control and lower control limits would change so obviously, you have to make discussions accordingly.

(Refer Slide Time: 20:52)



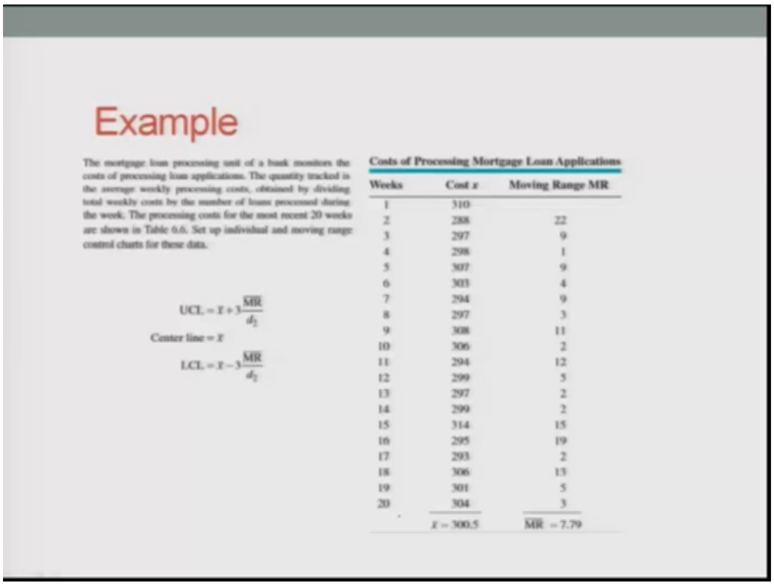
So, in many cases samples are used for process monitoring would be repeated. So, repeat measurements on the process differ only because of the laboratory on analysis on error as there may be in say for example, chemical process, manufacturing process and all these different type of a manufacturing process. The multiple measurements are taken on the same unit of products such as measuring the oxide thickness of some process which is being done or some paint thickness.

Or say trying to basically find out the Young's modules of a particular products because you will be utilizing that for some designing the tie rod of a car or trying to basically find out that how the trusses can be build for the bridge over a river. So; obviously, you have to do multiple readings. So, in that case you need to basically find out the changing limits or upper control lower control accordingly.

So, continuing with what I am reading such that as measuring the oxides thickness at several different locations on a wafer in a semiconductor manufacturing unit, because based on that you will be able to know that what type of connections would be needed, what is the voltage. Those voltages would be very small, but any small fluctuating in the voltage in the overall system of the computer or say for example, the laptop or even the android phone would basically spoil the whole system. So, you have to be very careful, what is the thickness and what is the voltage fluctuation? What is the ampere of current which you can flow and all these things?

In such situation, the control chart for individual units is useful. So, basically we use the moving range and for 2 successive observations at the basis of estimating the process variability and based on that we find out that how the range charts can be drawn. So, in this case, so for the moving range you will basically find out the differences and the mod of the differences of 1, 2 readings, 3 readings, four readings depending on what you think at what stage the changes would apply. And use those mod of the readings which is the mod of the range in order to basically calculate how the R charts and the subsequent calculations can be done.

(Refer Slide Time: 23:02)



So, in an example, let us continue the mortgage loan-processing unit of a bank monitor the cost of processing loan applications the quantity tracked is the average weekly processing cost. So, you have unit of set of officers who are doing that work now; obviously, you have to pay then salary there as processing cost and so and so forth. So, that is the weekly processing cost or the yearly processing cost or the multiple processing whatever it is. This is obtained by dividing the total weekly cost by the number of loans processed during the week. So, you process in one week 20, in another 30, in the next week 5, so you have to find out the average cost on the long run.

The processing cost for most recent 20 weeks are shown in the table. So, the first column basically consist the weeks, the second column consist the cost, and the third column basically finds out the moving average range. So, in this case what you are trying to find out is the difference mod of the preference. So, 310 and 288 is basically 22 then 289 and 297 is basically 9. So, you keep basically finding of the differences of those 2 consecutive values and basically find of their modes. So, it could have been say for example, for 2 or 3 depending on type of readings you want to do.

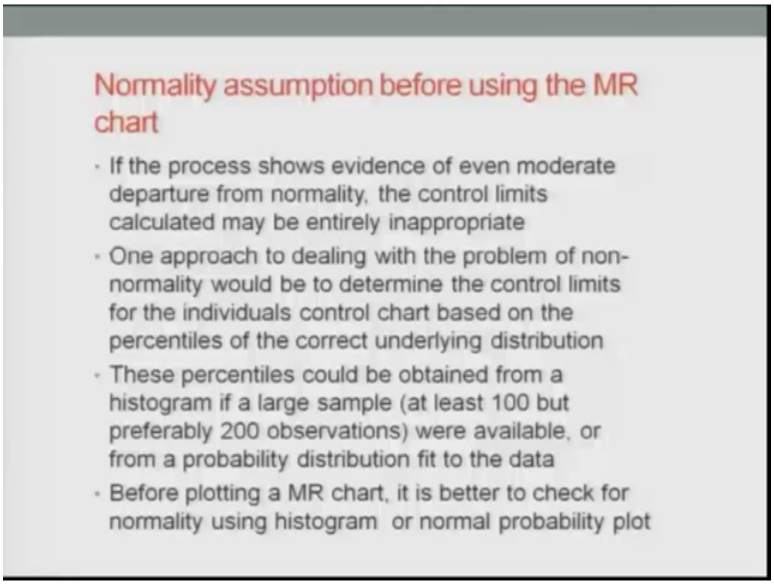
So, once you find out the moving average range, moving range not the average sorry, the average values corresponding to the second column which is the cost comes out to be about 300 point something I am just hashing the decimal place. And for the moving range the value comes out to be about 7.79. So, once you use that depending on the calculations of the values of D 2 you which will have you will be a basically be able to find out the central line which is x bar, the upper control value and the lower control value.

Now, again remember the value of 3, which is there here in both the upper control and the lower control would basically define the level of confidence you want for the upper value and the lower value. And if you go back to the fundamentally those who are z tables, you will find out this values of plus 3, minus 3 or plus 2, minus 2 or plus 1, minus 1 would basically signify the six sigma confidence the four sigma confidence and the 2 sigma confidence level. So, obviously, you can take the values of that parameter that coefficient which is 3 in these 2 calculations for the upper control level and the lower control level would be dictated accordingly.

(Refer Slide Time: 25:44)

So, this is the charts which you have for the moving range. So, if you find out the moving ranges charts, so they give you the values accordingly. So, x bar is 300 and then upper control is 321 and the lower control limit is 279 based on that due to the calculations, and find out whether the range is are such that some of the readings are over the range or below the range or where there is a particular trend, based on which you can pass some actual conceptual judgment whether the process is under control or outer control or whether their assignable causes unassignable causes or whether the process has some variations which we are not able to notice. But which would definitely should take into consideration when trying to analyze the overall capability of the process.

(Refer Slide Time: 26:59)



Now, as I did mention that for many of the cases the central limit theorem basically takes precedence. So, the normality of the readings if they are not true, but we will consider that as we keep taking more and more observations the in the long run normality would basically be the deciding factor. So, based on that we are passing the next sets set of judgments, which are normality assumptions before using the moving range chart is applicable and they are as follows. If the process shows evidence of several even moderate departure from normality the control limits calculate may be entirely in appropriate and we may not be able to use that accordingly.

One approach is dealing with the problem of non-conformity would be to determine the control limits for the individual control charge based on the percentiles of the correct underline distribution. So, you will basically you will try to find out what is the correct distribution, and based on that you will try to find out that what the distribution you should use. Now, this as I mentioned that in the previous bullet point is that if the variability for small change is quite start then be rest issued the central limit theorem would not be holding true.

And you have to basically use the actual central value, upper control, lower control value depending on what is the actual distribution based on which you are trying to work, so that is very important. So, if the variability is not much high due to not much of a change of like moving range values are not fluctuating much variability is not that high then be rest issued that central limit theorem would come into the picture and normality would hold true.

I would say that rather than using the word central limit theorem, let me stick to the point then normality would hold true. And we are very sure that normal distribution characteristics can be utilized and we can do the calculation, but if the variability your ranges are fluctuating very high then; obviously, it means that normality distribution per say is does not ruled and you have to use the distributions as applicable for the cases for our study.

So, these percentiles could be obtained from histogram if a large sample is there were available or from a probability distribution fit of the data. So, that we if we know what is the distribution we are aware of its so called central value which is the mean and the variability and use those variance and the central values of the means to do the calculations.

So, before plotting a MR chart is better to check for the normality assumptions for the distribution. And if your assured the normality assumption holds true then we can basically be assured sure that using those concept of normality would definitely give you much better results or else you have to basically take the distribution as it is by drawing the histogram charts and then fitting of the (Refer Time: 29:18) curve charts and you will try to utilize those mean and those variances of that particular distribution for our calculation.

With this, I will close the 26th lecture; and continue with discussion in the 27th and onwards for the corresponding charge, which we will consider later on in our course.

Have a nice day, thank you.