Total Quality Management - I

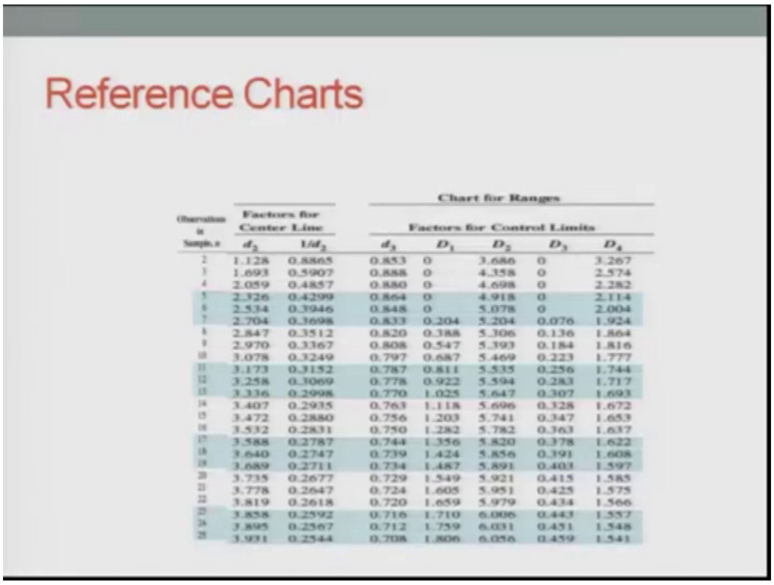
Prof. Raghunanadan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 22
Usage of X bar chart and R chart

A very good morning, good afternoon, good evening my dear friends, I always give you a introduction. So, it becomes much more lively for the discussions. So, I am Raghunanadan Sengupta from the IME department; IIT, Kanpur and I am the instructor for this NPTEL course which is total quality management one and as you know I did mention the time in again when I started the course in between also, the main bible based on which we are trying to study this course is the Montgomery book which is absolutely fantastic and all the references everything is basically take it from them. So, if you have a little bit of quantitative background and you are interested in trying to study statistics from the quality control perspective that is an absolutely fantastic fabulous book.

So, we are as I said we are and our interacting based on TQM 1 course which is total quality management one course, and this is the 20 second lecture which means we are in the fifth week and if you remember in the 20 first lecture is in the fagg end, I did discuss about the limits based on the central values which is x bar bar and some plus minus values as well as R bar and some plus minus values and the last line on the twentieth twenty first lecture I did mention that how the depending on the sample size the param the s quorum the coefficients can be decided, and we can use those coefficients for our further calculations.

(Refer Slide Time: 01:56)

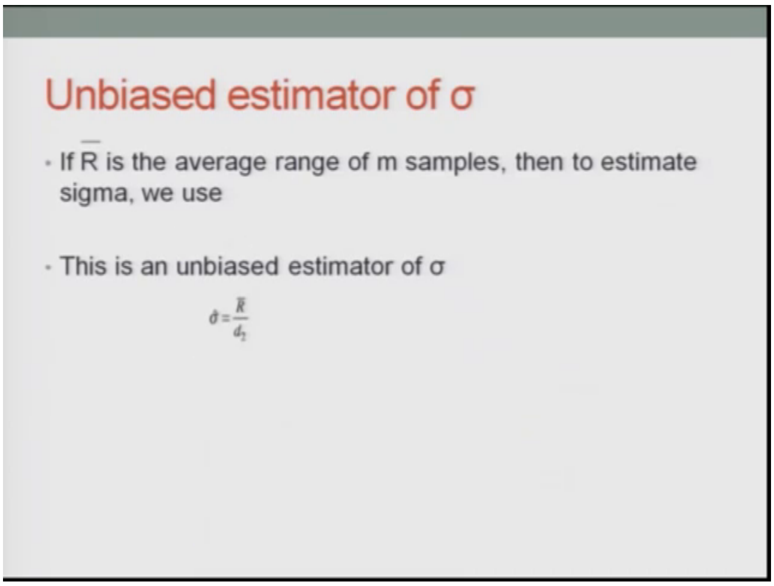


Observations in Sample, n	Factors for Center Line d_2	Factors for Center Line $1/d_2$	Chart for Ranges, Factors for Control Limits d_3	D_1	D_2	D_3	D_4
2	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.693	0.5907	0.888	0	4.358	0	2.574
4	2.059	0.4857	0.880	0	4.698	0	2.282
5	2.326	0.4299	0.864	0	4.918	0	2.114
6	2.534	0.3946	0.848	0	5.078	0	2.004
7	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

So, continuing the discussions the reference charts again you have the observations in samples which are given the factors for the central lines d 2 and in reciprocal d 2 are given and charts for the ranges depending on the suffixes for d whether small or capital R given.

Such that you can understand that depending on the sample observations in the samples which is n in science you can basically calculate and use those values for our inferior corresponding calculation.

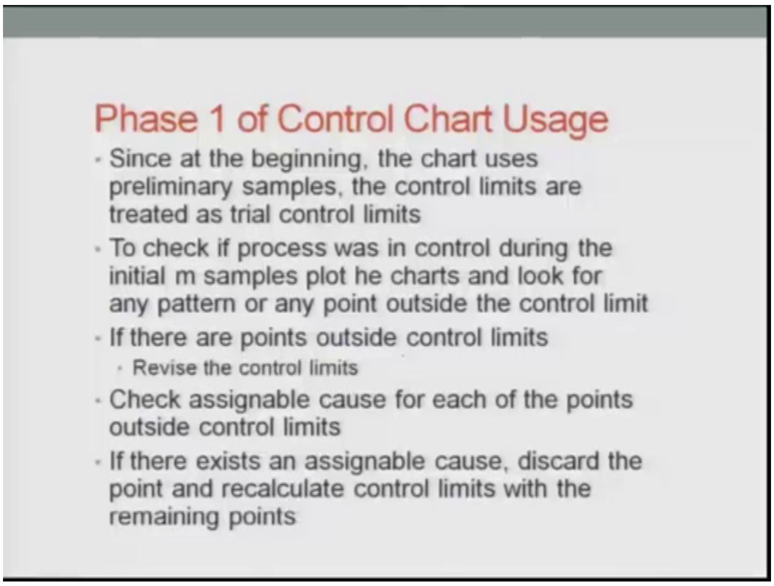
(Refer Slide Time: 02:21)



Now, unbiased estimator of sigma is very important to know that in the sense that if your calculations basically gives you some value from the sample which is not able to predict the actual population parameter base, even if the calculation seems very fine so obviously, there is some problem. So, they should be unbiased in the sense the difference on the technical difference between the population characteristics and the sample characteristics should be as close as possible to it is 0 the differences. So, if R bar is the average range of m samples. So, you have if you remember you take one sample which is of size n you take the second sample which is a size n any continue doing this. So, if R bar is the average range of m samples then to estimate the sigma, we use the formulas accordingly and which is the unbiased estimator of sigma is given by the and it is ok.

By the way let me also mention that and in to clear terms. So, this value of sigma hat which you see. So, this is like which basically is an estimate. So, that is basically coming from the sample and the actual population one is basically given by standard deviation which is sigma. So, what you do is then use this sigma hat in order to predict sigma. So, there is an unbiased estimate of sigma is given by the ratios of the range divided by the average range divided by d 2. So, d 2 would basically depend on the sample size. So, between depending on the sample size you can find out d 2, use that to find out sigma hat that will give you a best estimate for the population parameter sigma.

(Refer Slide Time: 04:18)

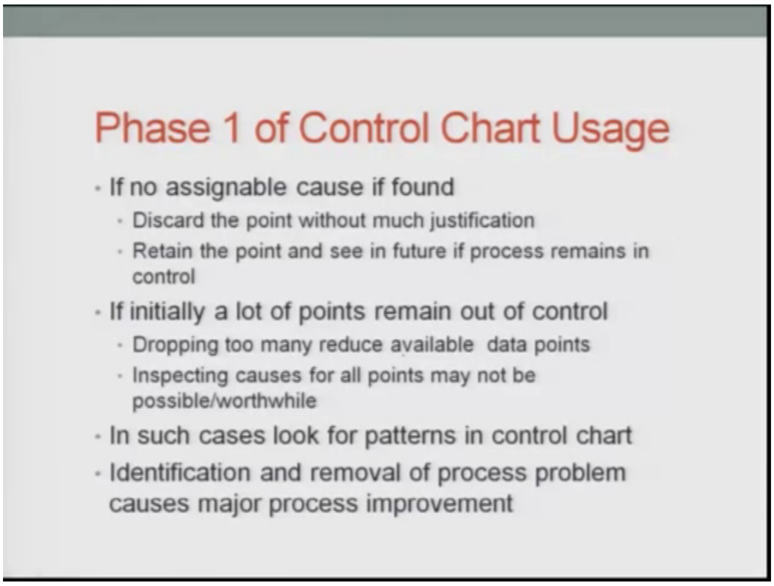


Now we will consider the phase one of the overall processes of the control charts.

Since at the beginning the charts uses preliminary samples the control limits are treated as trial control limits which will be finalized later on. To check if process was in control during the initial m samples plot the charts and look for any pattern or any points which are outside the control limits. So, technically that if there is no variability, but there has to be some stabilization on the process you will continue taking the observations product. So, once the variations technically become stable, you will be able to basically proceed with the calculations. To check if process was in control during the initial ms samples plot the charts and look for any pattern or any point outside the control chart. So, if there are; obviously, they should be such warnings which you should take care in order to basically find out the estimation process is best without any error.

If there are points outside the control limits, revise the control limits and do your calculations accordingly. Check assignable causes for each of the points outside the central limit on the control limits, if there are if there exists an assignable cause discard the point and recalculate the control limits with the remaining points and continue doing it likewise. So, if there are some assignable causes which you know should not be taken you basically take that set of observations out.

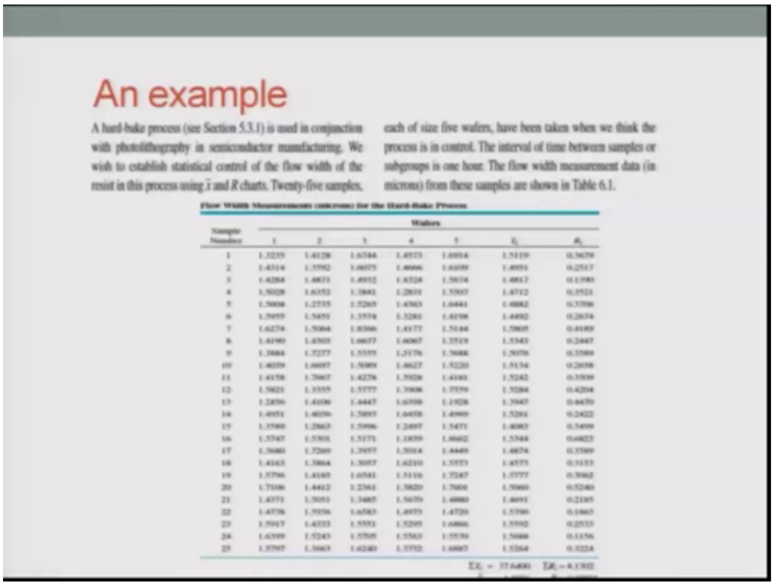
(Refer Slide Time: 05:43)



Phase 1 of the control charts containing if no assignable cause is if they are found discard the point without much justification, retain the point and see if in future if process remains in control and then you take decisions accordingly. If initially a lot of points remain out of control then it means that you should dropped, you dropping too many would reduce the availability of the data points.

So, if you basically drop many of them that sample size we rarely decreases which gives you a problem because the for a small sample size trying to predict for the population parameter becomes or population characteristics becomes difficult. So, inspecting causes for all the points may not be possible or worthwhile because that will entail money. So, in such cases look for patterns in the control just, take the decisions accordingly identification and removal of process problem causes major process improvement. So, if their problems are eliminated at one we are able to improve the process capability and the improvement of whole process problem takes place.

(Refer Slide Time: 06:44)



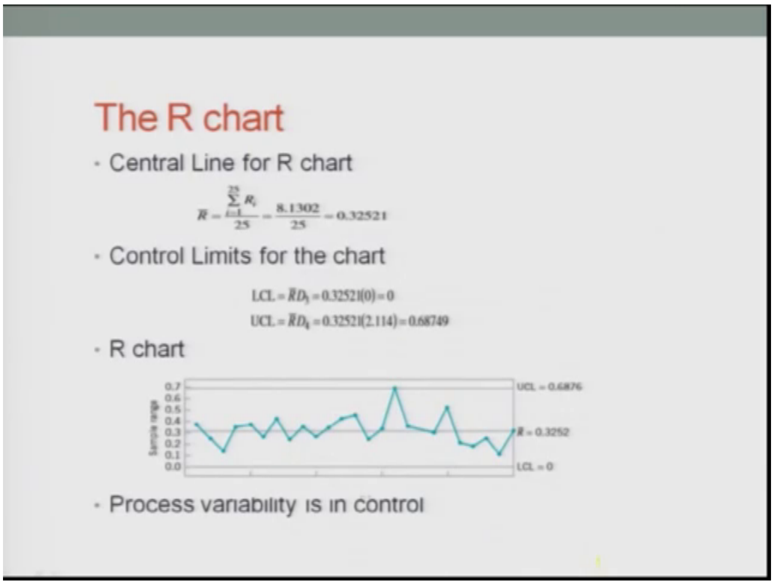
So, let us consider an example a heartbreak process. So, this is basically taking from a section is 5.3.1 from montgomery book a hard brake process is used in conjunction with photolithographic in semiconductor manufacturing, we wish to establish statistical control of flow width of the res and of that process such that it resists in the processing using x bar and r charts 25 samples are taken. So, this information is given the table for in the first column you have the sample number. So, 1, 2 till 25 each now you will be asking; what is the number of observations which is there in each sample. So, that is given as 5. So, if you read it where I am hovering my pen, it sees that 25 samples each of size 5 is taken. So, n is 5, m is 25 has been taken when we think of the processes is in control the interval of time between samples or subgroups is one hour. So, if you remember I mentioned 9 ‘o’ clock, 10 ‘o’ clock, 11o clock so on and so forth.

So, the time difference between the set ofs and observations you are doing is basically given as in one hour, the flow width measurement data in microns for these samples are shown in figure 16.1. So, along I am again mentioning along the first column you have the sample number each sample is of size 5 and the wafers basically basic informations are given. So, the cell 11 which I am marking now where I am putting my pointer is basically the reading for the first sample the first observation. Similarly if I take up the first sample the fifth observation it is given by 1.6914. So, based on that what I do is that I find out the sample averages.

So, this one point where I am again my highlighter is is pointing I am not trying to draw it, this 1.5119 is basically the sum of all these things divided by 5 which is the average of that particular sample and the ranges which is given as 0.3679 is basically difference between the maximum and the minimum. So, we basically find out those values for the second sample, third sample till the twenty fifth one similarly we do and try to find out the ranges for the second set of observation third sets the observations till the last one. And if you see at the fag end of this table, what I have what the problem does is basically gives us the x double bar which is the average of the average. So, what we need to do is that the first observation which is given here is 1.5119 is x 1 bar which is one means the first set of observations of 5 and the bar means the average of that.

Similarly, if I consider 1.5805 that would give me the average of the seventh set of observations where the set of observations are 5 similarly I have the have the have this ranges which I have already mentioned. So, x bar bar basically means sum of all these things divided by 25, which comes out to be one let me highlight it would be easier. So, this is basically the average of the average and this is R bar for the average of the range.

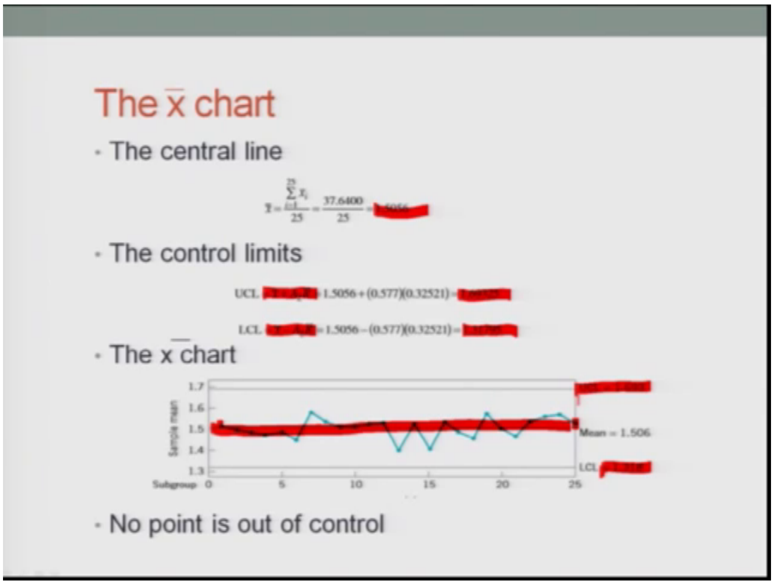
(Refer Slide Time: 10:16)



So, now, of the central line would be given by the for the average would be given for this range would be given by the sum of them divided by 25, which comes out to be 0.32521 and the control limits for the chart depending on the values of the coefficients which are already discussed in the table which we did few minutes back, would give you the lower control limit at 0 the or upper control limit as 0.68749 which I now draw it on the table. So, what we are doing is that we are trying to basically plot the observations set number along the x axis. So, the first one where I am highlighting is the first sample of 5 observations, if I mark this where I am now hovering my pen it is basically the last observation for the 5 set of observation which I have. So, all these dots which are there are basically the sample average, the central line which is there sample average means for with respect to the range the central line which is there which I will try to highlight using the red colour. So, this is basically R bar which is as given as 0.3252. So, from where does it come we have already calculated it accordingly.

So, LCL which is the lower control limit is 0 which is marked here, the upper control limit is 0.68749 which is given here. So, you can find out that how the overall fluctuation of the range values occurs within these limits of upper control and lower control with the central line. Now if I go to the x bar charts I do the same calculations, I find out x double bar which is the sum if you remember the second last column were one or all the averages x 1 bar x 2 bar x 3 bar till the last one.

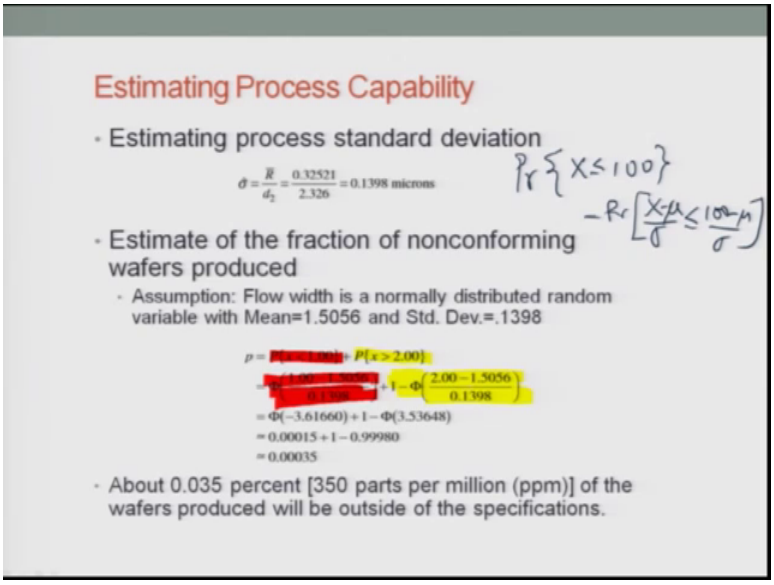
(Refer Slide Time: 12:14)



So, I find out the averages which comes out to be again I am using the red colour highlighted is 1.5056 which is this line, and that upper control limit and the lower control limit are given by the formulas as given here and here.

So, this x bar double bar is coming from 1.5056; this are bars you have already calculated and this A2 which is plus minus would depend on the sample size, which we find out from the table. So, these values upper control comes out to 1.6932 which is plotted here and another value is 1.31795 which is plotted here and then you see the plots have been made. So, the plot where I am hovering now my pen is basically the average of the first sample, if I consider this which is the last one would be the basically be the as average sample average for the last set of observations of 5 taken together which is the 25th one.

(Refer Slide Time: 13:13)



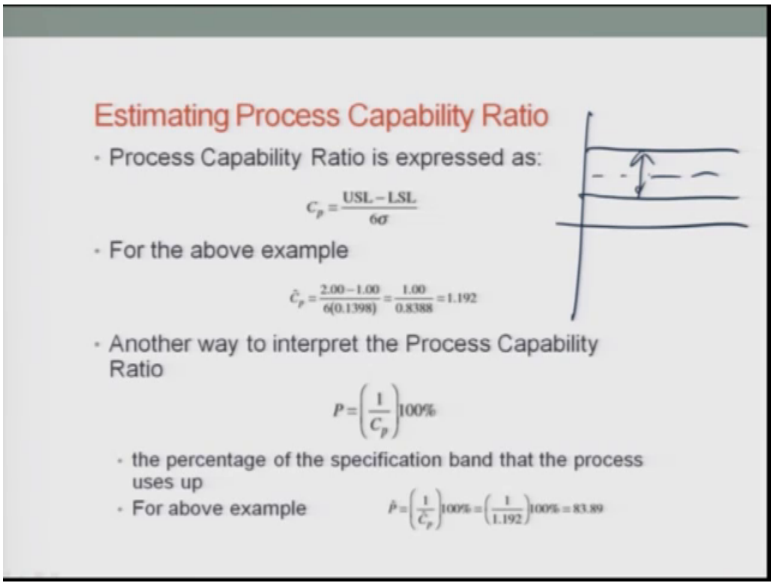
Now you want to basically find out estimating the process capability; estimating process k standard deviation basically can be found out from the formulas which is the best estimate.

Which is sigma hat is equal to R bar divided by d 2 and that comes out to be 0.1398 which is about 0.14 micron, estimate of the fraction on the none coughing wafers produce is found out. So, you want to find out what is the control limit or the dispersion or the probabilities. So, assumptions are flow width is normally distributed which is a major observation observations and comment based on central limit theorem, you find out the means which is 1.5056 and the standard deviation here for already found out is 1.39 you use that is concept of mu minus z alpha by 2 into whatever the coefficient should be depending on 1 sigma 2, sigma 3 sigma and the other values would be plus one of that because you are going on the right hand side.

So, you want to find out the probabilities, these probabilities are very simple in this the first formula gives you the probability that the x values are less than 100, and the second value gives you the probability of the x are greater than. So, you use the standard normal deviate and find out this those values. So, standard normal deviate formulas I will just write it down. So, this will be probability of X less than 100. So, this will be probability of x minus mu by sigma less than equal to 100 minus mu by sigma. So, these are what I am doing is I am converting from the standard normal due to the standard normal deviate using that.

If I have, in this case let me use an easier format to note I use a highlighter occur that two o. So, this from now is this and which means once you use the standard normal deviate and the colour scheme for this would give mu. So, once you have this you find out the values for the standard normal deviate table, the values comes out to be about 0.35 into 10 to the power minus 5 which means that about 0.035 percentage or 350 parts per millions ppm of the wafers produce would be outside the specifications.

(Refer Slide Time: 16:04)

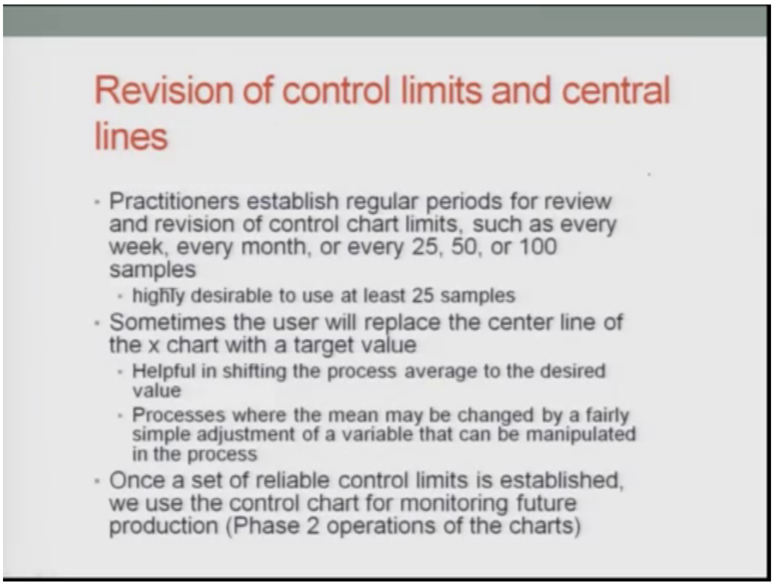


Now, you want to find out the estimating the process capability ratios.

So, process capability ratio is expressed as the difference between the upper control and the lower control divided by 6 sigma. 6 sigma is coming from the; oh total width. So, if it is plus 3 sigma on to the on from your side plus three sigma to the right and plus sigma on to the left. So, ma for my case it will be left and right. So, it is just the reverse. So, if I am basically putting about 99.97 of the total coverage between this plus minus 3 sigma so; obviously, I would divide by 6 sigma. If I am trying to basically find out to a our coverage probability above for about 95 then obviously, it will be plus minus 2 sigma and if it is basically a coverage of about 67 it will be plus minus 1 sigma. So, that is why the 6 value which is there in the denominators dignity signifies what is the level of officer's coverage which are covering.

So, basically it means the upper control them and the lower control which is the differences which you have. So, if I check the. So, this is the upper control this is the central line this is the lower control. So, these is the differences which we have for the other example if you want to find out the process capability, you find out the differences of the upper line and the lower line and divided by 6 of the sigma which is the standard deviation which you have already found out as 0.13 in 698 and it comes out to be 1.192. Another way to interpret the process capability would be, that process capability will be given by one by Cp and the percentage of the specification banned that the process uses can be utilized accordingly. So, for the above example the concept of process capability gives us the value of about 83.89 value.

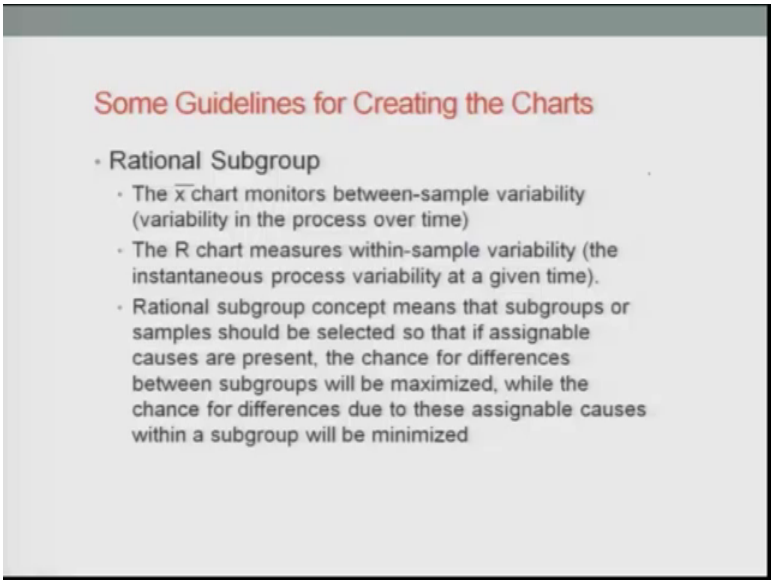
(Refer Slide Time: 17:56)



So, just trying to recap practitioners established regular periods for review and revision of control central charts such as every week, every month, every day or whatever the frequency is. Highly desirable it is basically highly desirable to use at least 25 samples because each sample set may have 5, 10, 15 whatever, but the larger the total set of observations are throughout the samples which basically would mean that you have been able to find out the characteristics of the sample based on a larger sample, such that it will mimic the population to a much greater extent. So, if you remember the biasness and all this thing should technically decrease. Sometime the user will replace the central line of a x bar charts with the target value.

So, it is helpful in shifting the process average to the desired values and make calculations accordingly. Processes where mean may be changed by a fairly simple adjustment of a variable that can be manipulated in the process can be utilized, but it will obviously, have some justification why you are trying to do that. Once a set of reliable control limits is established we use the control chart for monitoring the future production which is basically the phase two of the utilization of the chart. So, was you use phase one the concept of chart or drawing in the charts and bring a control, and then use this control to go to the second stage to find out whether the process capable what is changing or it is fixed.

(Refer Slide Time: 19:27)

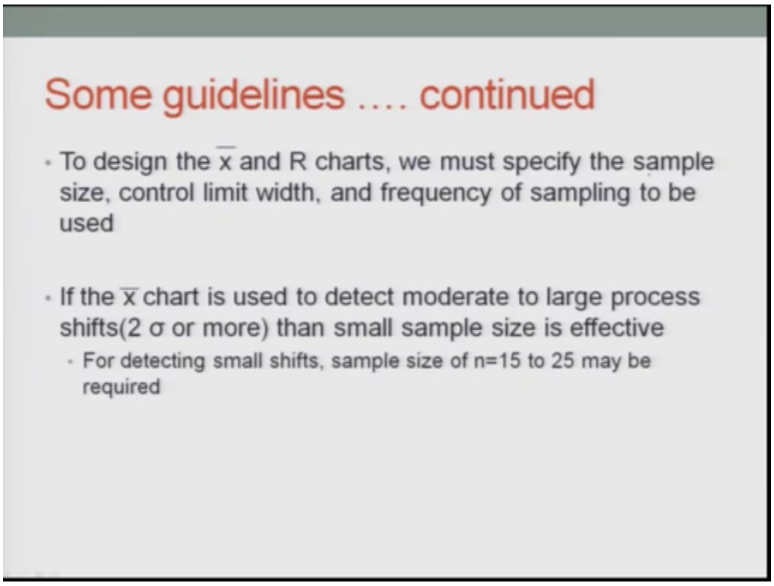


Some guidelines for creating the charts or x bar and R charts size national for the subgroup should be done.

Such that x bar just monitors between sample variability, variability in the process over time, the R charts or the as done here would basically measures within sample variability. So, what you are doing is that within the sample variability I am trying to my measure by finding out the difference between the maximum and the minimum the inst and the. So, it we basically means that the averages which you are trying to use from the x bar basically means for the sample averages throughout the sample, and the R obviously, R bar would come later, but the R which you are trying to use would be the variability for each sample then you average it out for the whole set of often m number of samples which would technically be equal to in the respect then the information which you I am getting from such samples which are being combined together, should give you us the maximum amount of information about the population.

So, rational subgroups concept means that subgroups or sample should be selected. So, that if assignable causes are present the chance for differences between subgroups would be maximized while the chances for difference between those to be assignable causes within us a subgroup would be minimized; obviously, you do your calculations accordingly.

(Refer Slide Time: 20:46)

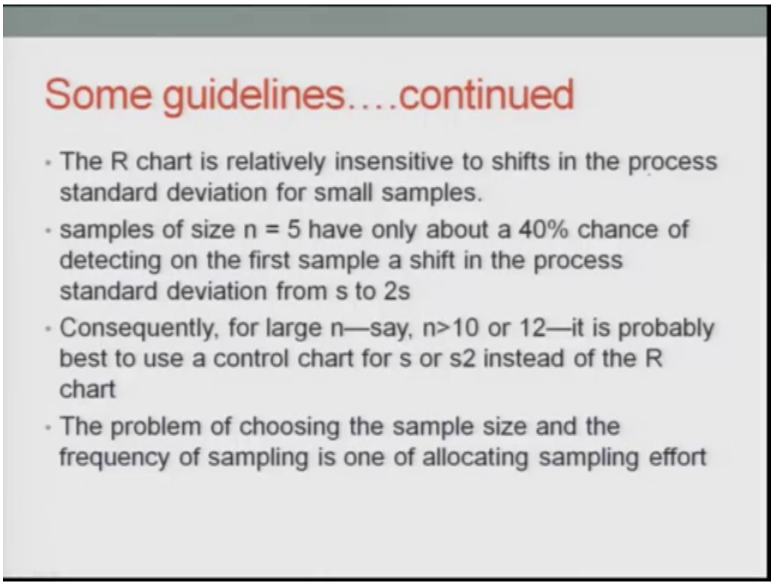


Some guidelines as were saying to design, the x bar and R charts we must specify the sample size control limits widths and the frequency of the sample to be it cannot done. So, whether its weekly monthly hourly on based on seconds that will basically the decided beforehand we will also decide on this number of observations which are there in each sample and you will also consider the number of samples, it has been mentioned it should be 25 or more will try to basically understand that also and keep that in mind.

if the x bar chart is used to detect them and detect moderate to large processes shifts of two sigma or more then the smaller sample size is effective, because in that case what you are trying to do is that you are trying to basically find out small sample size and take technically take more of them in order to basically elevate your or basically overcome the problem of trying to take all the set of observations at one go. For detecting small shifts sample size can be about in the range of 15 to 25 so obviously, you will have the m size also depending on that you which can be made such that the overall number of observations which you have which is n into m would give you a good idea.

That about the characteristics which you find for this whole set of observation which is n into m would in the long run mimic the population parameter or the characteristics to the maximum possible extent.

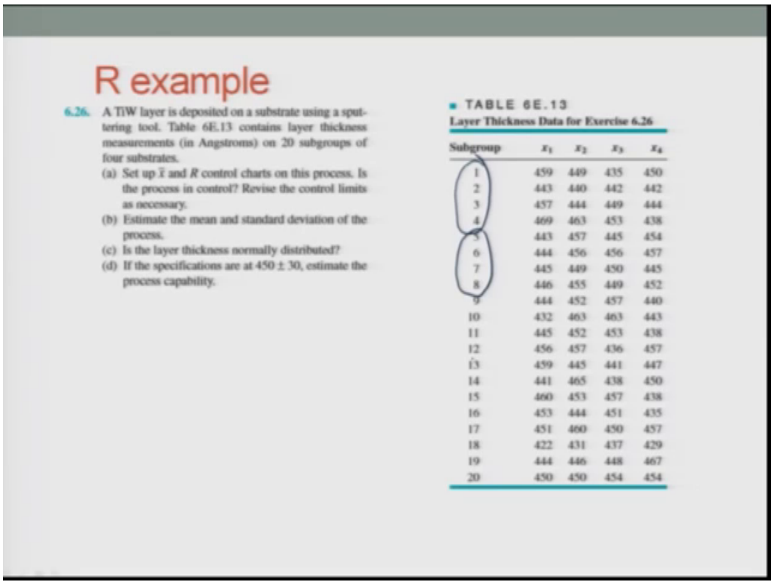
(Refer Slide Time: 22:19)



The archers is relatively insensitive to shifts in the process standard deviation for small samples, sample of size n which is 5 have only about 40 percents chance of detecting on the first sample at shift in the process standard deviation from s to 2 s. So, there are 60 of them would give 40 would not give you that information. Consequently for large n say n greater than ten or 12 it is probably best to use of control charts for s or 2 s is instead of are accordingly.

So, the problem of choosing the sample size and the frequency of the sampling is done one of the allocating the sample effort based on which you can increase your efficiency for the sampling.

(Refer Slide Time: 22:59)

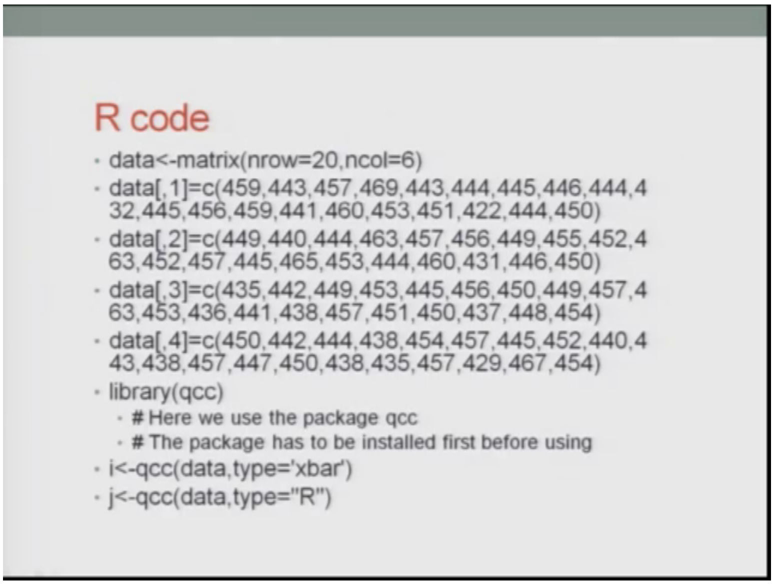


Subgroup	x_1	x_2	x_3	x_4
1	459	449	435	450
2	443	440	442	442
3	457	444	449	444
4	469	463	453	438
5	443	457	445	454
6	444	456	456	457
7	445	449	450	445
8	446	455	449	452
9	444	452	457	440
10	432	463	463	443
11	445	452	453	438
12	456	457	436	457
13	459	445	441	447
14	441	465	438	450
15	460	453	457	438
16	453	444	451	435
17	451	460	450	457
18	422	431	437	429
19	444	446	448	467
20	450	450	454	454

So, let us considered an example a tum tungsten wafer layer is deposited under some substrata, using a sputtering machine which is give you and the data related to that is given in in table 6 E 13 which consists a layer thickness measurements in angstrom, and 20 sub groups of 4 substrates. So, so 20 what you have is basically the first substrate is 1 to 4 second one is 6 to 8 and so on and so forth. So, what you do is that you need to find out and set the x bar charts and R control charts on this process now you want to answer is this process in control if not what are the problems.

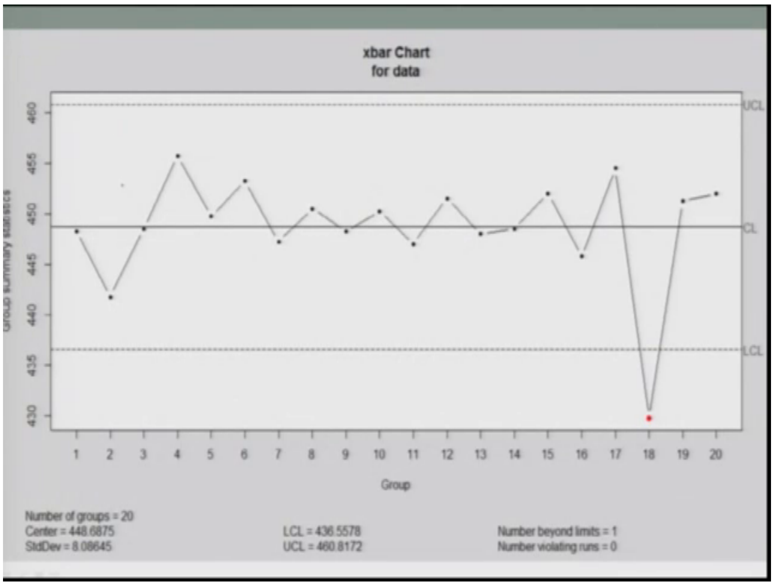
So; obviously, you have to take some decisions accordingly, it revised the control limits and do your calculations again, estimate the mean and the standard deviation to the processors that you will try to find out the sample mean which you find out the sample standard division is you will find out, whether it mimics the population parameter to the maximum possible extent. Is the layer thickness normally distributed that is the question we need to answer and third the last question we need to answer is that specifications if they are at 450 plus minus 30 values, then estimate the process capability and comment intelligently.

(Refer Slide Time: 24:19)



So, what we do is that we write a R code. So, the data is basically stored in matrix with n rows which is 20 and n columns is 5, 6 as already mentioned in the part of R, as well as if you have seen that dataset which is given here. So, the values are all given. So, if you want to basically find out the first column and all the values of the rows you will give gives this function as data comma 1 and it will give you all those values accordingly. So, you basically call this library functions and try to basically for m fit the functions as x bar and r and you do your calculations accordingly.

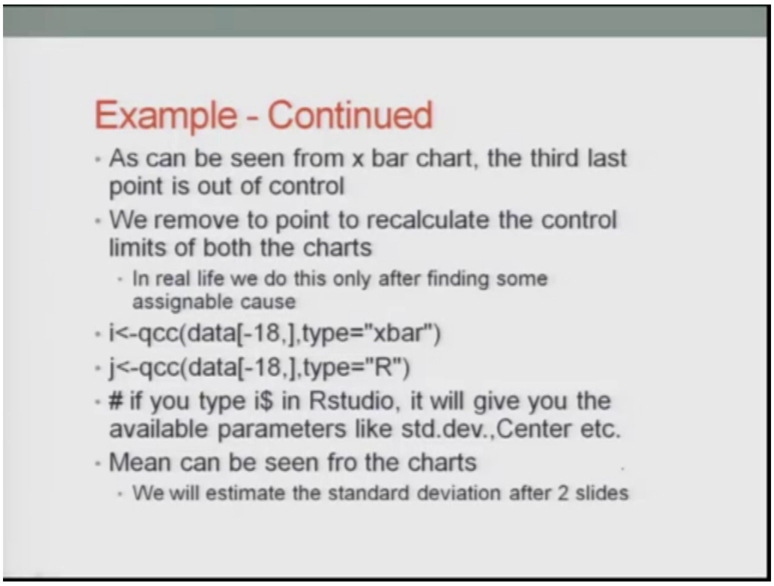
(Refer Slide Time: 25:02)



So, once you run the code the distribution or sorry not the decision the process capability chart would look like as given here. So, here the groups are there. So, e and you find out the number of groups is 25 and you find out that there are 4 of them or each, the central values the standard deviations are given here, I am not going to the calculations you can find it out the upper control limit and the lower control limits are given. So, based on that if I study the graph is always between the upper control limit where I am basically marking, and the lower control limit which I am marking.

But; obviously, this is one outlier. So, if there is an out lie you have to basically think that how actions can be taken accordingly. So, what you do is that eliminate some of the observations redo the calculations once you do the calculations the graph which gives you the information, which is in front of me shows that the outliers have been removed . So, this can be done very easily in excel sheet in R programming whatever it is.

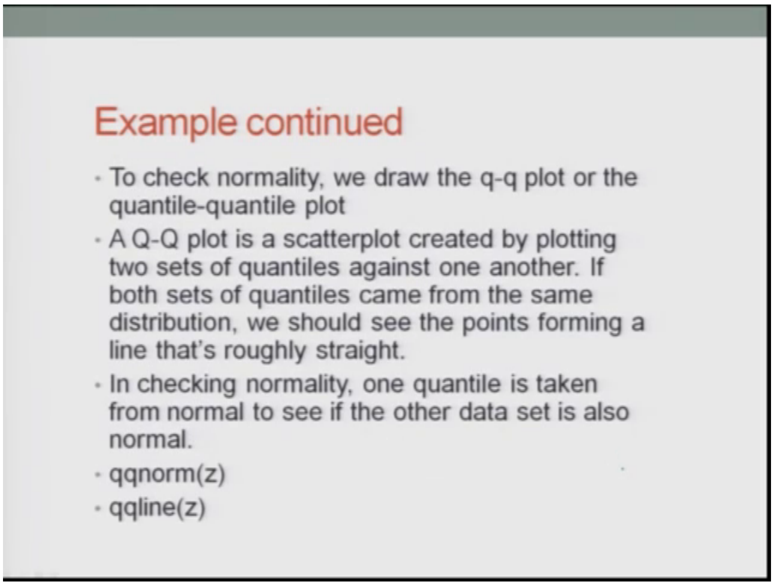
(Refer Slide Time: 26:09)



So, continuing with the example as can be seen from x bar charts, the last point is a out of control we choose the red one. We remove the point to recalculate the control limits on both the cases and in real life we can do this only after finding some assignable causes and arbitral ally you cannot remove it; obviously, there is some reason why it has happened and we want to understand that.

So, once you do this the all using the Rs from packages and the values which are given for the readings you can easily very easily draw this graph do some statistical analysis in order to basically find out whether they are assignable causes are known as assignable causes based on which the outliers has happened. So, mean can be seen from the charts we will estimate the standard deviation of the after 2 slides.

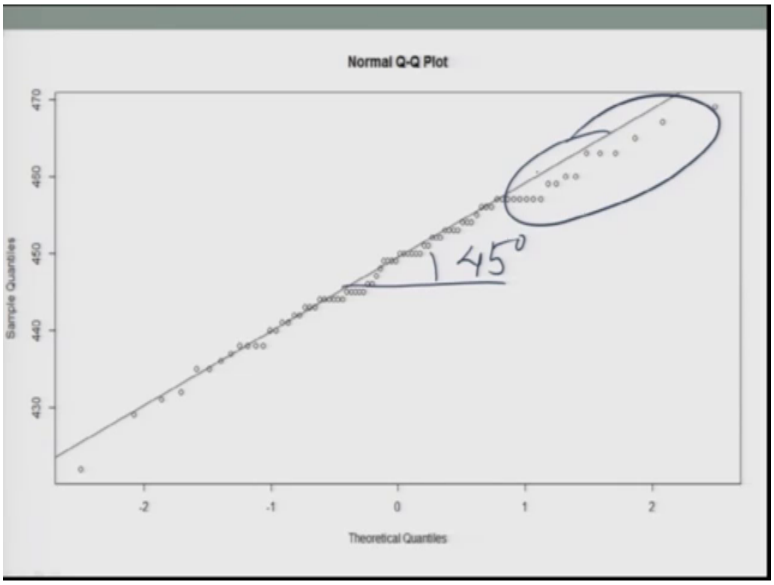
(Refer Slide Time: 26:59)



So, continuing with the example to check normality we drew the draw the q-q plots which is quantile-quantile plots, which anybody can check in from very good book a q-q plot is a scatter plot created by plotting two sets of contents against one another, if both sets of contents came from the same distribution we should see the points forming a line which is the roughly a in a straight line and 45 degrees both with the x axis and the y axis.

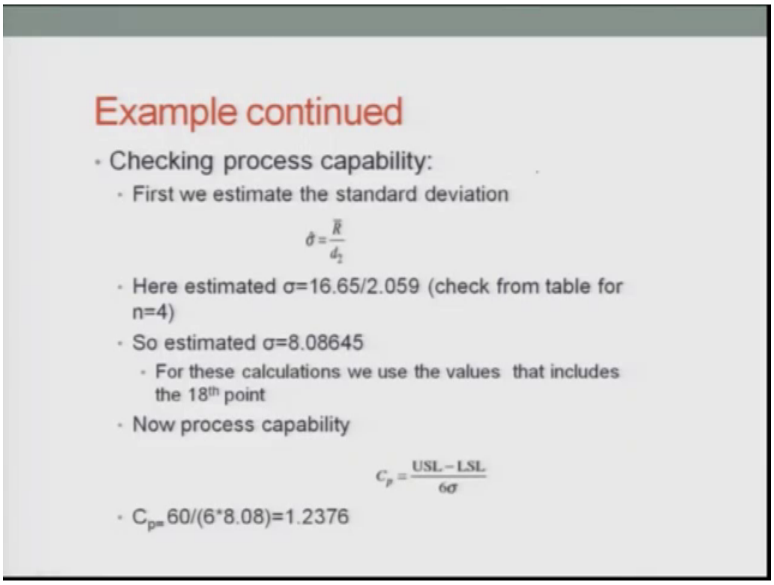
If checking normality one quantile is taken from the normal to see if the data set is also normal we can do that and you can basically check the normality test also and basically comment accordingly.

(Refer Slide Time: 27:43)



So, now, I have plotted the q-q plots. So, this is the 45 degrees line which he was talking about. So, along the x axis you have the theoretical officer's quantiles and along the y axis you have the sample quantiles. So, you will see that they are almost the same only the outliers is happening here technically; obviously, if you go into the depth; you can find out more outliers are come at intelligently, but the outliers are happening on the on their values which are very high.

(Refer Slide Time: 28:15)



So, continuing the example checking process capability gives you the first we estimate the standard deviation given by R bar by d 2 which comes out to be about 16.65 divided by 2.04.

So, n is 4 you remember that it comes out to be about 8.1, for this calculation we use the values that include the eighteenth point also because you remember if the outliers are removed depending on the assignable and non-assignable causes you should basically be satisfied yourself that the step you are taking is right or wrong. Now once you find out the process capability it comes out to about 1.2376, I will discuss all these things in more detail with more different type of charts and with this we will conclude this twenty second lecture and continue with the twenty third and more for the TQM class have a nice day.

Thank you.