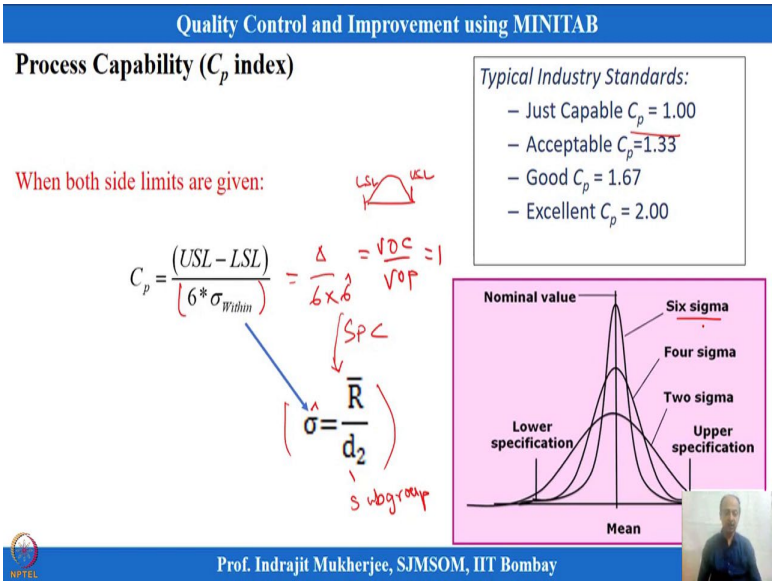
Quality Control and Improvement with MINITAB
Prof. Indrajit Mukherjee
Shailesh J. Mehta School of Management
Indian Institute of Technology, Bombay

Lecture - 13
Process Capability

Hello, everyone and welcome to session 13 of our course on Quality Control and Improvement using MINITAB. I am Professor Indrajit Mukherjee from Shailesh J. Mehta School of Management IIT, Bombay.

(Refer Slide Time: 00:40)



So, earlier session, we have just started about Process Capability. So, we are talking about C_p index. So, from there onwards we will continue in this lecture. So, let me just recap what we have told. So, capability index what we mentioned is that it is a index which shows that ratio between voice of the customer and voice of the process.

So, how do we express voice of the process? Voice of the process is 6 multiplied by standard deviation of the process and which is to be estimated and this sigma estimation is also known as within sample variability. So, in this case MINITAB; MINITAB expresses this variability as sigma equals to $\frac{\overline{R}}{d_2}$.

So, this formulation what you see over here is basically within sample variability over here. So, this is calculated based on statistical process control chart. So, this is estimation of the sigma over here, can be done by if we are drawing the X bar R chart in that case.

And, we have also told mentioned that this process capability we estimate whenever it is under natural variation; that means, common cause variability is only and the processes is stable. So, in that case we apply this one.

So, this will be the $C_p = \frac{\Delta}{6 \times \hat{\sigma}}$, and this estimation what you see is given over here in this estimation because and this is the subgroup size based on which I can define d_2 values.

So, this subgroup size is also known to us when we are collecting the samples. So, whenever the process is in statistical control immediately we can calculate the C_p index over here. So, this is $\frac{\Delta}{6 \times \hat{\sigma}}$ because 6 is the common spread that we have assumed that 99.73 of the observation will be falling in case it is a normal distribution.

So, the underlying assumption over here is it is the CTQs follows normal distribution and for that I am assuming this making a simple assumption over here and we are also assuming over here that the centering of the process is perfectly on the target. So, there is no variation from that and so we can calculate a capability index which is known as C_p index over here ok. So, and MINITAB does it automatically for you.

So, you have to only mention that whether how to calculate the sigma you mention that one and MINITAB will calculate for you. So, then also we mentioned that just capable process, if voice of the customer and voice of the process. So, this is the voice of the customer and this is the voice of the process, and the ratio is equals to 1 means I am consuming full tolerance basically. My variability of the process is consuming the full tolerance.

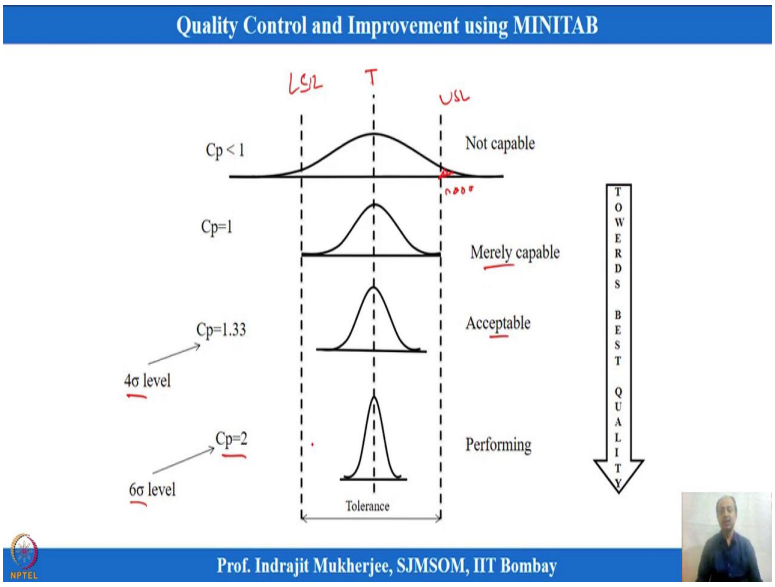
So, in other words; that means, that if this is the specification and upper specification limit and lower specification limit. So, completely I am consuming the total variability which is spread across from USL to LSL. So, this is not an industry standard that equals to one condition just acceptable like that.

So, if you have to improve that one you have to go to 1.33 and beyond for not so critical items, but if it is very critical maybe we go about to 1.67 or 2 as the process capability for specific CTQs like that. We want to reach that level of now capability like that.

So, if it is a 6σ process, it is linked with C_p value of 2 and then we have other index over here $C_p = 1.33$, about 4σ and $C_p = 1.67$ about 5σ. This way we can link with C_p values and also with sigma level of a process like that.

But, that is not the right way we will see another way of defining the and linking with the sigma levels like that ok and MINITAB gives you some options to do that ok.

(Refer Slide Time: 04:33)



So, in this case what will happen is that so, I can calculate this and also we have to remember that this is for both side specification. So, C_p can be calculated whenever USL and LSL is given; when we have one sided specification in certain scenarios, one sided tolerance, in that case C_p cannot be calculated and MINITAB also does not reveal any such calculation of C_p whenever I give one sided specification let us say LSL or USL.

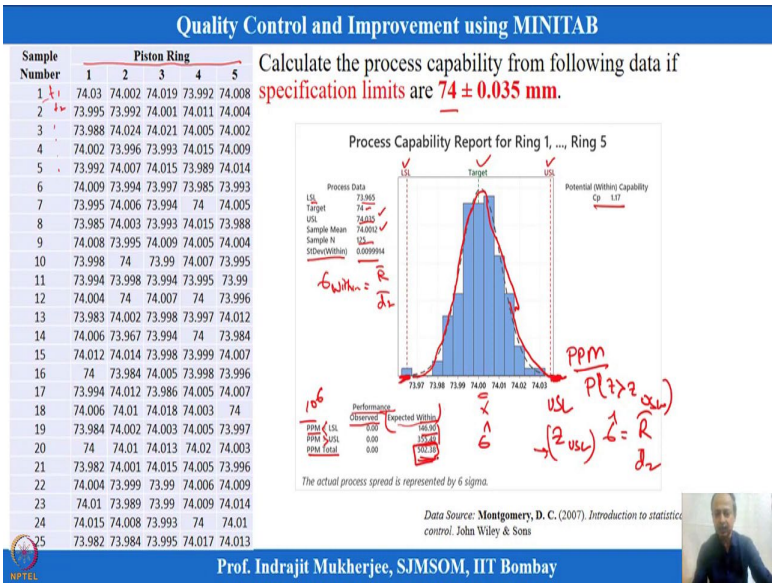
So, with some example we will try to see. So, here you can see like that what I explained last time also say if it is not capable in that case rejection some will be outside the USL condition and this is the LSL condition what you see. So, many many data points will

fall over here outside responds. So, there will be rejections like that and let us assume this is the target value so, what we are getting over here ok.

So, this is merely just capable what we told and this is $C_p = 1.33$, 4σ level what is written over here. So, some companies may accept may be considering this as the acceptance level of C_p values, but excellent performance we always say that C_p value should be equals to 2 which is equivalent to 6σ level.

So, towards excellency if we are moving for quality we should look for C_p close to 2, and at least more than 1.33. Many of the industries follow 1.33 as a basic standards of accepting a CTQ performance.

(Refer Slide Time: 05:52)



So, based on this we are just taking an example. Last time we showed that in MINITAB how we are going to do that. So, I am taking a piston ring example which is having 25 observations and each of the observations has 5 subgroup size over here and at a given time point t_1.

So, this was collected at t_1, so, this will be collected at t_2. So, t_1, t_2 observations like that and we have t_{25} observations which are at certain gaps the observations are collected at certain pre-decided intervals and can be calculated also how much should be the interval.

We are not going into that details, but that is possible. And, how many subgroup size to be taken that is also rational sub grouping is another concept which is used to define that one. So, if that is correct in that case we can analyze the data and we can analyze it using control chart techniques.

So, I am taking the same example the specification is given as 74 ± 0.035. So, tolerance is about 0.07 if you see from USL and LSL and this is the total Δ that we have or voice of the customer that is given.

And, if it is within this distribution over here what we will see is that this is MINITAB output so, this is LSL, this is USL this target I have defined let us say 74 is the target value over here and C_p index is calculated 1.17.

So, LSL is 73.965 and USL is 74.035. This we have to enter MINITAB and target is 74 we have mentioned. Sample mean is this total observation what you are seeing over here 125 and 25 observations. So, 125 observation. So, overall mean $\overline{\overline{X}}$ is 74.0012.

125 observation and you see standard deviation within is calculated over here this is based on MINITAB will ask you what basis I will calculate σ within. So, we have mentioned over here that you calculate from control chart and the formula to be used is $\overline{R}/d_2$ or from sample range chart you calculate the response.

And, also you will find that performance index will be provided over here. There is a term which is known as PPM < LSL and PPM > USL and PPM total. PPM means parts per million. So, number of items that will fall outside USL and fall below LSL that will be given as performance over here.

So, over here the dataset will be plotted over here and MINITAB we will see if something is going beyond the tolerance and that will be counted then those counts and that will be converted into PPM $\left(10^6\right)$. So, this will be converted into that number how many fall out in this much. So, in million how much it is, so, that will be calculated as observed.

An expected performance what you will find is that MINITAB will place a normal distribution curve over here what you can see dotted line and so, if this is the and based

on certain mean that we have $\overline{\overline{X}}$ and we can always calculate that corresponding to USL. So, USL will be converted into a Z value.

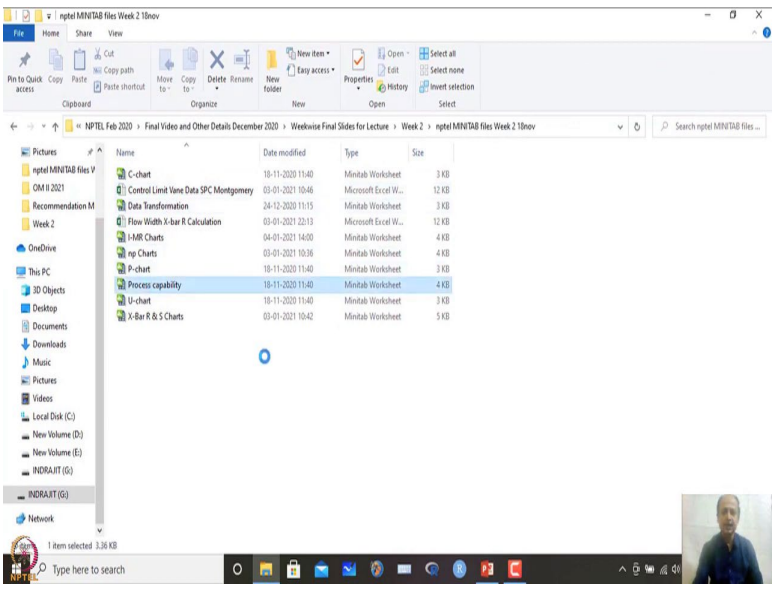
So, Z_{USL} over here and based on that it will find out that what is the probability of Z greater than Z values for USL over here and that will give me a fraction on confirming over here and that will give me the PPM that is falling outside. So, a normal distribution assumption will be used over here to see what is the expected performance.

And, also the standard deviation calculation that will be used for Z conversion will be $\overline{R}/d_2$ that. So, corresponding to value of $\overline{X}$ and standard deviation $(\hat{\sigma})$ over here what I can do is that I can calculate Z_{USL} over here and from there I can reach to this PPM level over here on this side I also I can reach on this side.

So, expected within performance is shown over here how much will be less than this. So, probability and then accordingly convert to PPM and here also we can do that. So, expected within an expected PPM more than USL also.

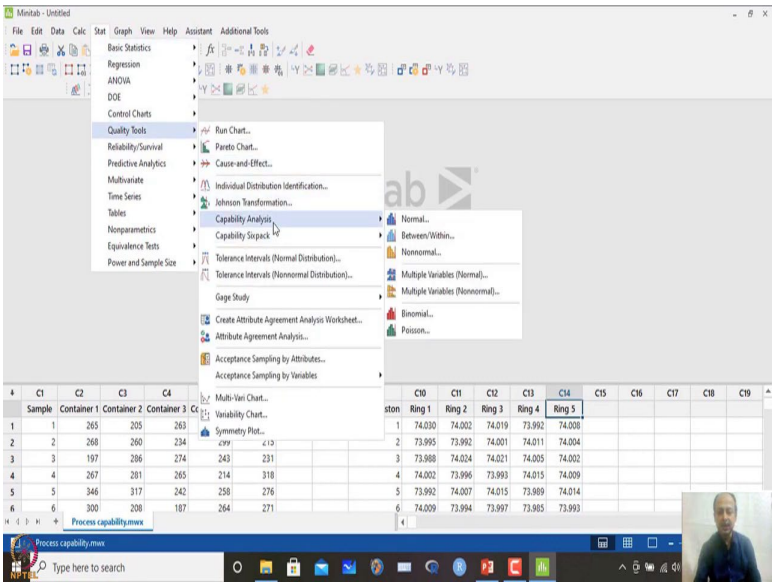
So, this is the expectation in case you are just superimposing and thinking that this is a perfect normal distribution with the mean and standard deviation that is calculated from the data set how much will be the fallout from LSL and USL like that and total summation of these two will give you the total values that you see over here, ok.

(Refer Slide Time: 10:37)



So, what are the options that we that we use over here? So, I will just show you the options. So, I will go to the dataset and in this case what we will do is that we will use process capability dataset and we will try to illustrate the same thing so that we can see how we are.

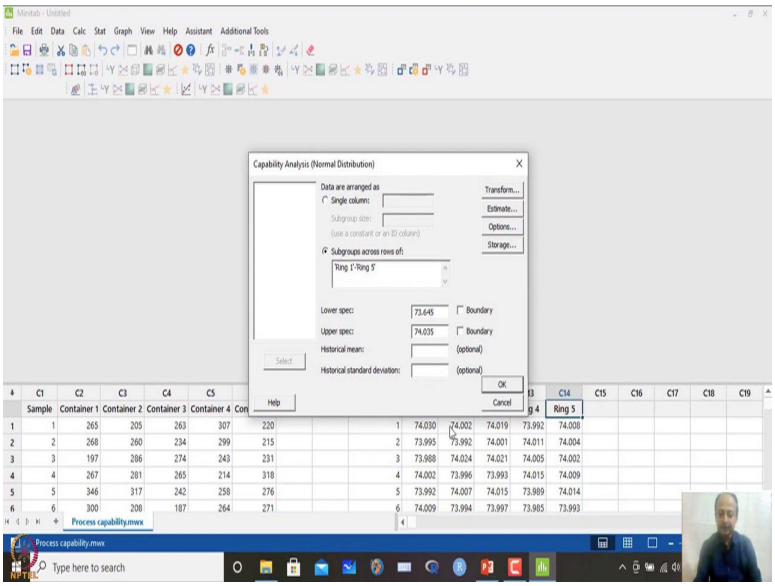
(Refer Slide Time: 10:52)



So, this data set is already I have created. So, a Minitab file I am opening. This is the observations of one dataset, this is the other observation which we will analyze now sample pistons – ring 1 to ring 5; 5 sub groups and these are the observations over here.

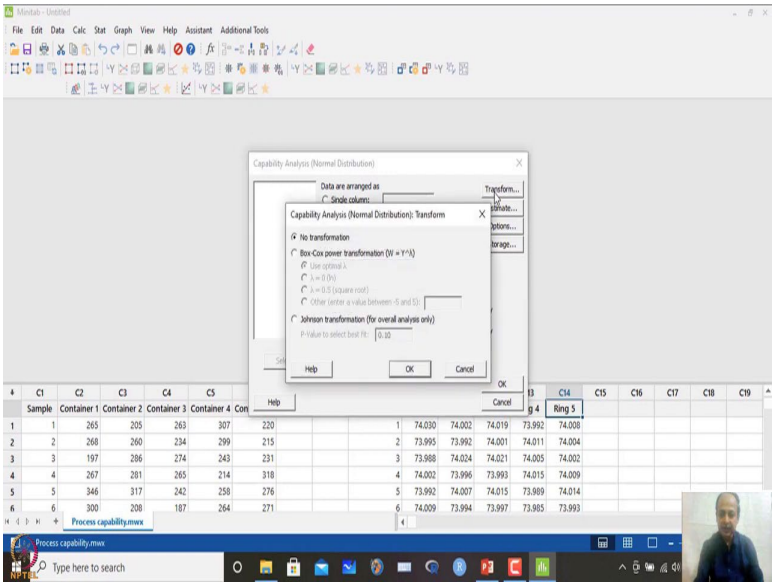
So, I go to Stat on top and what I do is that I go for Quality Tools and over here Capability Analysis. And, what I will do is that Normal capability because I am assuming normality over here.

(Refer Slide Time: 11:17)



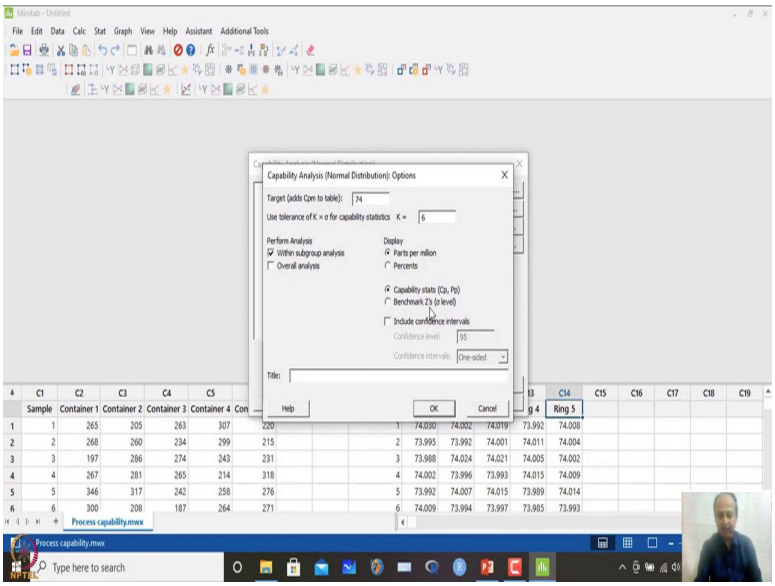
So, let us assume the data follows normal. So, I am using Normal capability. Then Subgroups across row I will use and because all observations are in different columns like that. So, I will use from here to here and then I will say select this one. Then lower specification I will write this is 73.965 upper specification is 74.035

(Refer Slide Time: 12:13)



And, then here you will find that transformation or no in case it is non-normal in that case what is to be done we will see. So, this is we are assuming normality. So, I will not click anything over here.

(Refer Slide Time: 12:22)

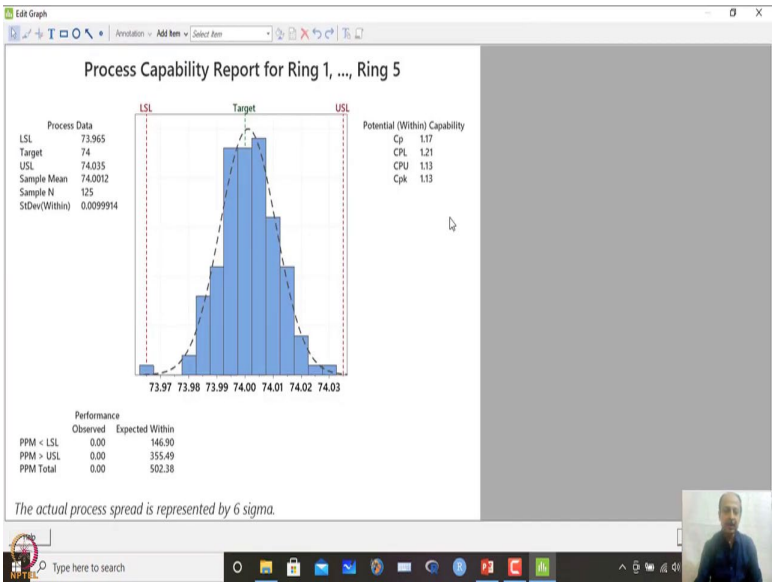


Estimation over here, so, within capability estimation which method I will follow for subgroup size greater than 1. So, here it is written that you see within subgroup standard deviation how do we calculate. So, you have to mention Pooled standard deviation, S method, R method. So, I will use the R method because R control chart we have seen X bar R.

So, used unbiased constant we will keep this one as default over here. So, and also in case subgroup is 1. So, in my case it is not true. So, this area you can ignore like that. So, I will click Ok like that. So, over here an Options; over here Target is let us say assume 74 is the target like that and within group analysis we are doing. So, Overall analysis we are not considering over here.

So, then Capability C_p values I want to see and Parts per million also I want to see other than that we do not understand now at this current time point. So, we will just click Ok and K value is taken as 6 because 6 standard deviation we are keeping over here and I click Ok .

(Refer Slide Time: 13:26)



What will happen is that, I will get this analysis what you see over here and you see that LSL I have given target I have given. So, USL is given, sample mean is calculated based on number of observation, which is 125.

Standard deviation within is calculated as from $\overline{R}/d_2$ formulation. So, this is calculated as range average and from that d_2 is taken from number of sub group size that is 5. So, based on that this is calculated and PP observed and expected performance is shown over here same thing what I explained like that Z it is converted into Z and from that we calculate what is PPM < LSL, what is PPM > USL.

So, total PPM parts per million that will fall outside the specification is basically 502 and then C_p index is 1.17. So, at this time point we are not concerned about C_{pl}, C_{pu} and C_{pk}. So, let us assume the C_p is the only measure we know. So, 1.17 and that is the value I am looking for. So, C_p 1.17 is not enough. So, 1.33 I told maybe the standard many industries follows; that means, improvement is needed over here ok.

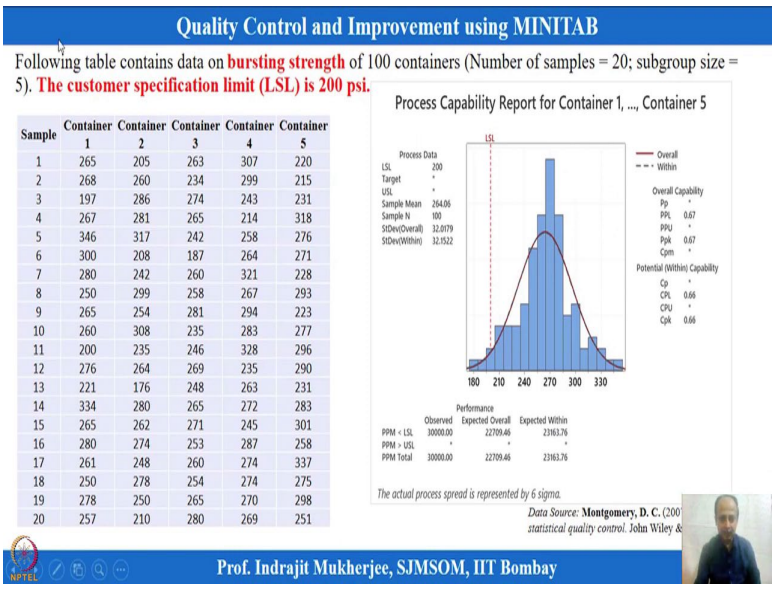
So, that gives you some basis when we have existing scenarios and whether to you take improvement initiatives or not for a CTQ that we can define from here. So, C_p values is less than 1.33 let us say and in that case improvement is needed. So, in that case I have to

reduce the variability. You see the measures of C_p index says that I cannot do anything on the tolerance I cannot do anything on the tolerance.

So, this measure what I have to do is that I have to reduce the sigma level. So, if you can; if you can reduce the sigma over here what you see over here. So, if I can reduce the sigma values that is coming over here then my C_p will go it is inversely proportional to this. So, this will go up like that ok, variation reduces C_p index goes up. So, in that case so, my variability reduction is the target over here from the formulation that we see.

So, but scenario is maybe one sided specification we are having. So, let me take this is the next example over here.

(Refer Slide Time: 15:47)



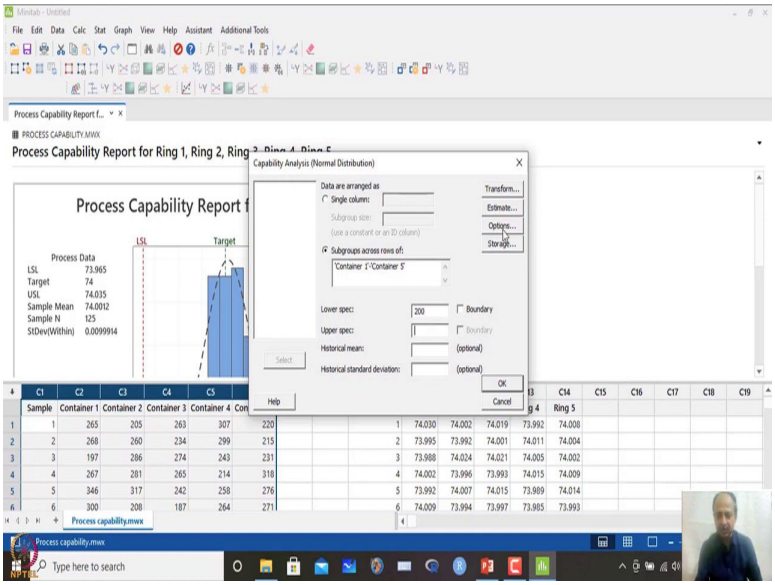
Sample	Container 1	Container 2	Container 3	Container 4	Container 5
1	265	205	263	307	220
2	268	260	234	299	215
3	197	286	274	243	231
4	267	281	265	214	318
5	346	317	242	258	276
6	300	208	187	264	271
7	280	242	260	321	228
8	250	299	258	267	293
9	265	254	281	294	223
10	260	308	235	283	277
11	200	235	246	328	296
12	276	264	269	235	290
13	221	176	248	263	231
14	334	280	265	272	283
15	265	262	271	245	301
16	280	274	253	287	258
17	261	248	260	274	337
18	250	278	254	274	275
19	278	250	265	270	298
20	257	210	280	269	251

And, we have an example over here that is container over here. So, here in this example what you see is that 5 container subgroup size is 5, but LSL is only given 200 over here. So we have 20 samples with 5 observation. So, total 100 observations we are having 5 subgroup size. So, total 100 observation and lower specification limit is given as 200 psi.

If you see on the right hand side, it has not calculated potential within capability you see C_p is not calculated and star is given over here. So, when I do this one so, let us let us go to the MINITAB file and let us close this one and I will this is the sample observation C1

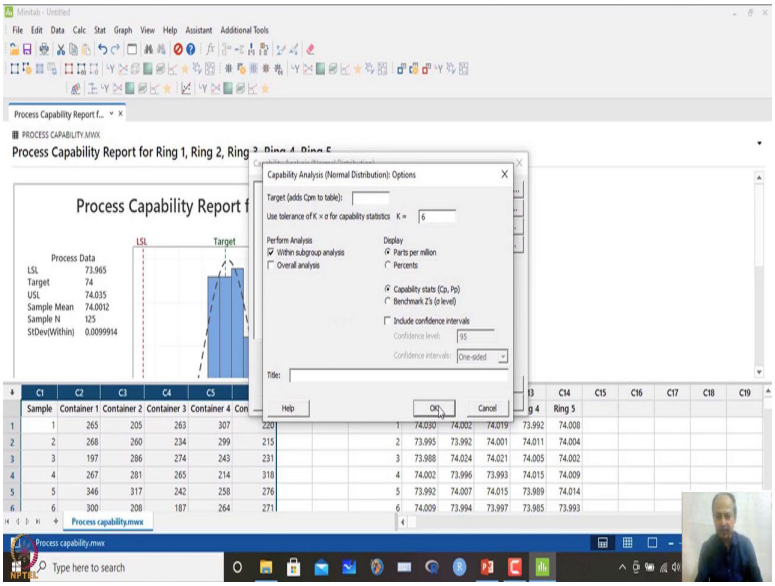
to C6. I go to Stat and what I do is that Quality Tools and in this case Capability Analysis, Normal capability analysis over here.

(Refer Slide Time: 16:34)



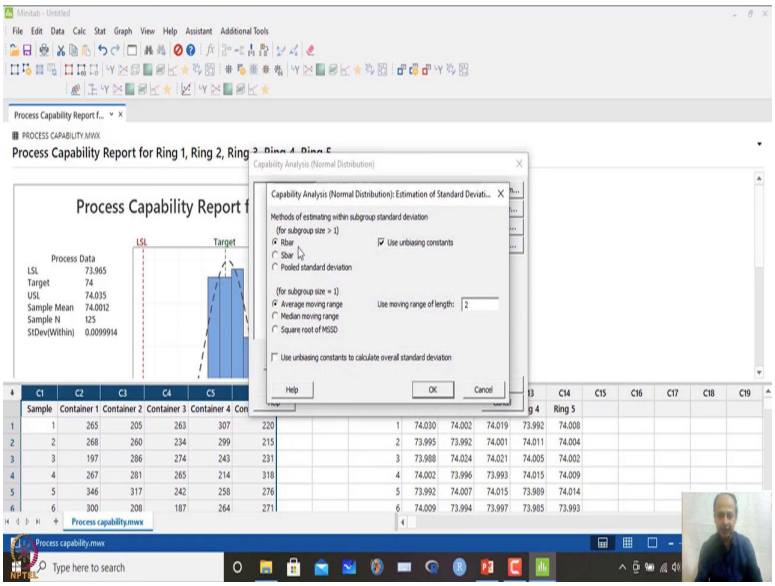
And, I I want Container data to be placed over here and LSL is given as 200 over here and upper specification limit is not there.

(Refer Slide Time: 16:47)



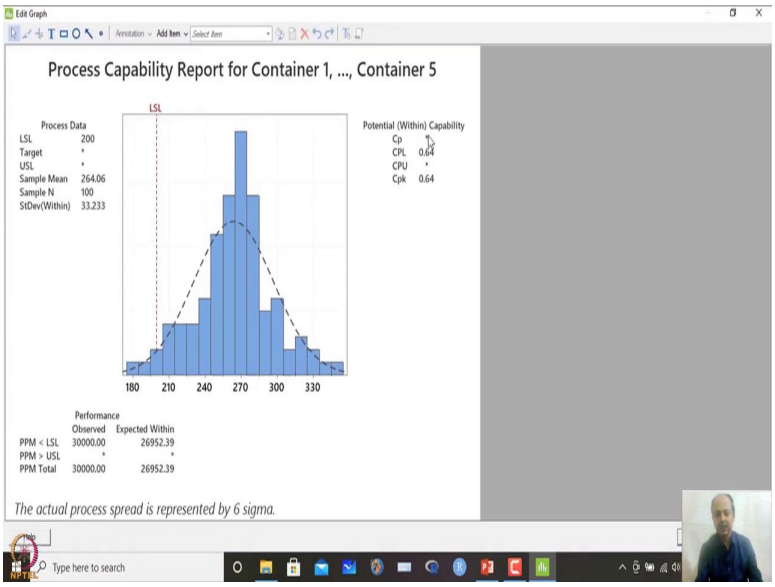
And, Options over here Within that is fine, target we have to change options, Target we will remove because this is a another example.

(Refer Slide Time: 16:58)



So, estimation we mention $\overline{R}$. So, it will do like that.

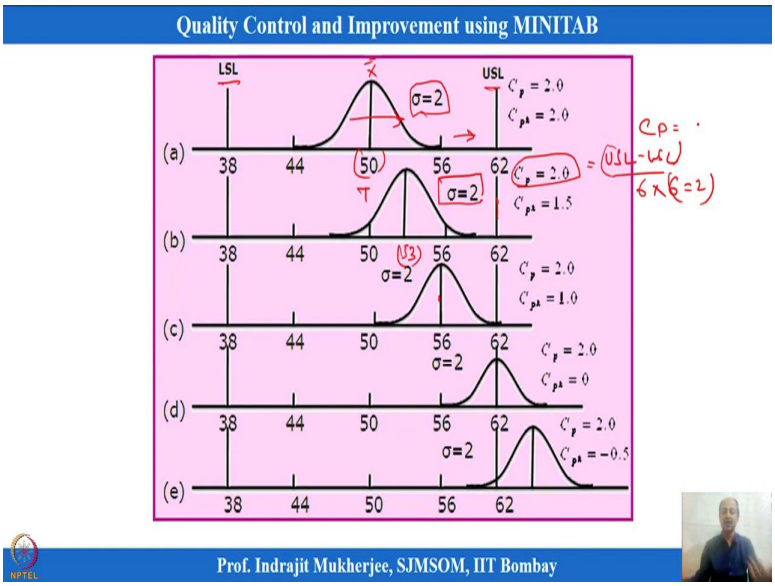
(Refer Slide Time: 17:02)



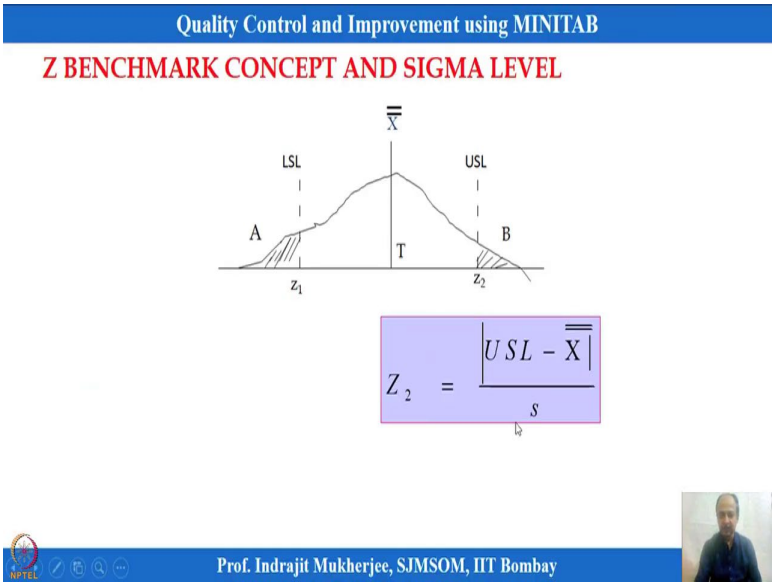
But, what you see is that when I click this one results what you see C_p star information is given over here. So, star means it cannot calculate that one because this is one sided specification. So, if it is one sided we cannot calculate C_p values for that something else has to be done ok.

So, this cannot be done. So, because I cannot see Δ over here so, in that case in case you have Δ then target values and in that case can be calculated, but we are not looking into that. So, one sided specification we will assume that C_p index cannot be calculated for that. So, some other index has to be done for those scenarios like that ok.

(Refer Slide Time: 17:43)



(Refer Slide Time: 17:44)



So, that I wanted to mention. So, then let us go back and see these diagram over here. So, this is the concept that is will be used by MINITAB. So, Z conversion of this. So, Z

conversion I mentioned that $\overline{\overline{X}}$ will come from control chart and then σ will be known which is S over here.

So, when I get these values what I can do is that what is the B probability of this area over here, B can be calculated, A can be calculated and based on that we can also calculate a Z benchmark concept which is used in calculating the sigma level of the process like that ok. And, MINITAB gives you this option we will discuss more about this ok.

But, what I am trying to say is that we can calculate capabilities and we can also calculate that what is the performance expected performance for that some Z conversion is required. So, Z and corresponding to this Z what is the probability that can be converted into PPM on both the sides like that.

So, this area and this area can be calculated and then that can be converted into PPM and that is what you see in MINITABs expression what you see over here. So, expected performance over here this is the portion what you what you can see.

So, in case this is normal, what is the expected performance in median. So, in long run what do you expect. So, if this is normal so, in population what do you expect? This is sample information. Sample information is giving you a performance observation is around 0, no fall out over here, but if you see if we consider the normality assumptions over here then in that case some fall outs are expected that is around 502 ok.

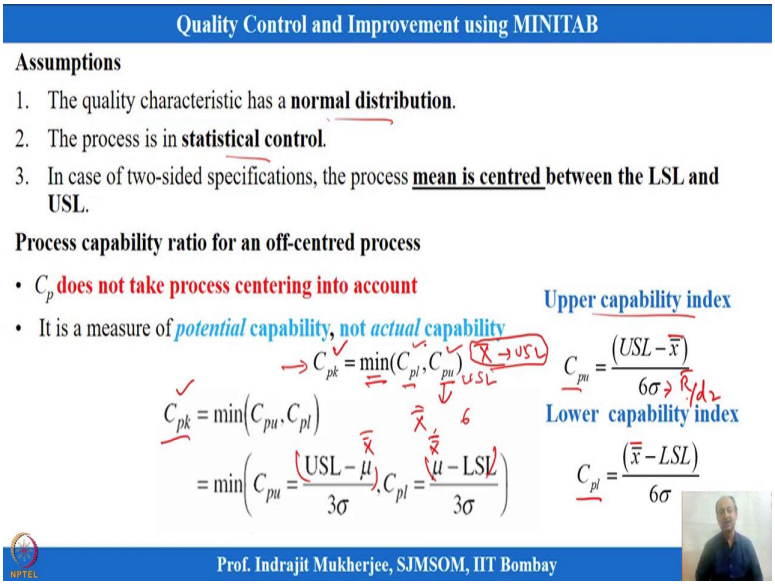
So, that is the; that is the idea of placing this expected within performance like that in PPM. So, some idea you will get how much PPM fall out will happen ok parts per million. So, this is the way they calculates and but what you see is that σ is 2 over here which is freezed and this is the specification what you see upper controlling with line and lower controlling with line over here.

And, in that case what you see is that σ is kept same over here, but the distribution is shifting. So, location was here which is the target value let us assume T, but it has moved from here to somewhere over here. So, this may be 53 let us say. So, mean is shifting basically. So, $\overline{X}$ was somewhere over here which is on the target so, specifications was like this.

So, σ remains same and in this case, but the mean is shifting over here towards upper specification line. So, it will move to over here, but C_p index you see remain same because $C_p = \frac{(USL - LSL)}{6 \times \sigma}$ .So, standard deviation remains 2. So, it is not changing and USL this is also freezed, but the whole distribution is moving from this end towards this on the higher upper specification side.

So, accuracy part is not considered in this formulation what you see of C_p. So, what we have to do is that we have to revise this formulation, so that we will penalize if it moves to the upper specification or lower specification limits like that. So, even if C_p remains 2, we can expect that this is deteriorate our performance is deteriorating, but that is not captured in C_p.

(Refer Slide Time: 20:50)



So, what they do is that they uses another index which is known as C_{pk} index what you see over here which is the minimum value of C_{pl} and C_{pu}. This MINITAB will calculate automatically. So, this is for USL how much is $\overline{X}$ from the USL and how much is $\overline{X}$ from LSL.

So, if the difference between these two comes down, and then this C_p index also comes down like that. So, whichever way you move it will basically penalize the C_p values like

that ok and minimum of these value is taken so as to ensure that we want to improve the minimum one.

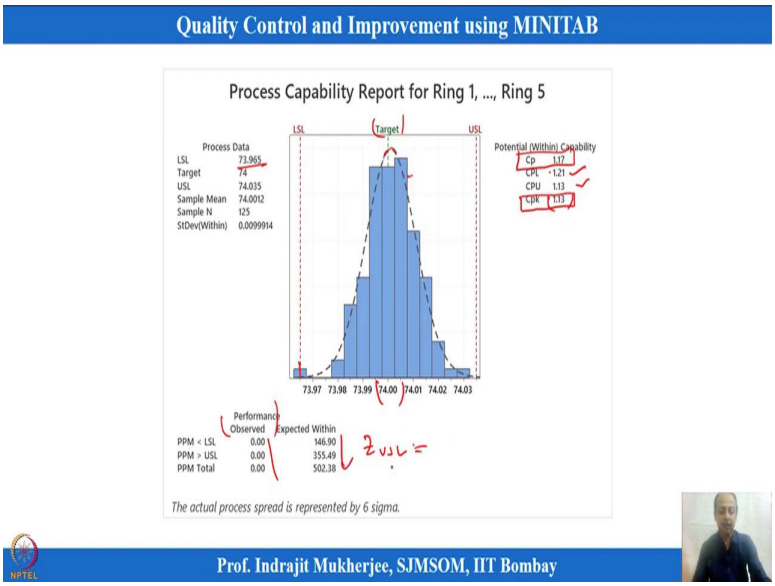
So, in this case so, C_p index takes care of this centering concept of where is $\overline{\overline{X}}$. So, not only sigma is considered, $\overline{\overline{X}}$ is also considered over here. The formulation is USL minus $\overline{\overline{X}}$ we can think of over here and this is $\overline{\overline{X}}$ assuming that is the population average what we are getting.

So, minimum of these two values what you see will give you the C_{pk} index over here. So, the I want to penalize if it is moving away from the target. So, that is the objective over here and upper capability index is calculated C_{pu}. This MINITAB automatically calculates for you because average is known to us and in that case we can calculate $\sigma = \frac{\overline{R}}{d_2}$. So, in this case this is also not a problem to calculate.

And, then the MINITAB will calculate C_{pu} and C_{pl} and based on that minimum value it will report C_{pk} index over here ok. But, the assumptions over here what is considered is one of the assumptions is normal distribution assumptions that is taken; process is under statistical control that we told that only assignable. Assignable cause is not there so, under control. So, everything is stable and the mean is centered over here, then only we can calculate C_p index like that ok.

But, if mean shifts in that case we need to consider this formulation which is revised one and it is measure of potential capability what is mentioned not actual capability. So, in this case just an short term capability you can think of. So, in this case what we can do is that.

(Refer Slide Time: 22:48)



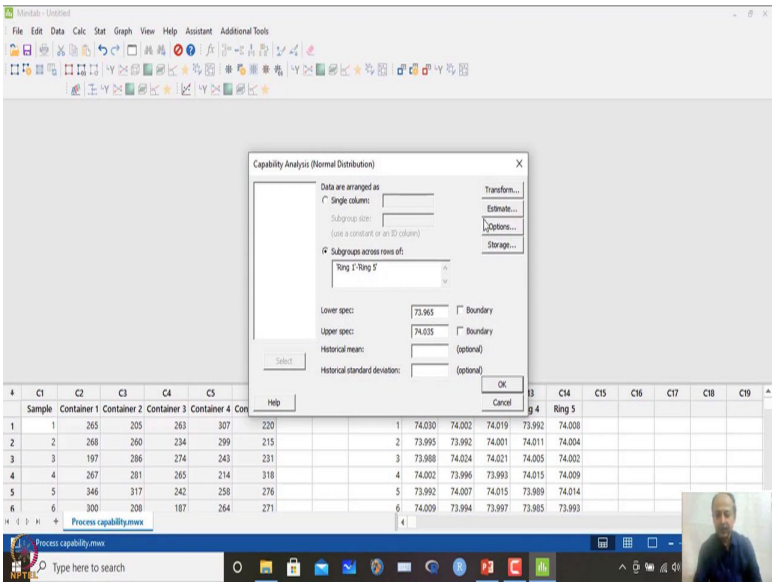
And, this can be calculated. So, here when you give the command over here you will find a C_{pk} index that is shown over here 1.33. C_p index what we have seen is 1.17 and it will calculate C_{pl} which is based on the previous formulation what we have shown.

So, that can be calculated and this is shown over here what you see C_{pu} and minimum of these two is taken as 1.13. So, histogram is drawn over here with the overlapping normal distribution this is the target which was placed at 74 and this is the upper specification limit with 74.035 and this is 73.965 lower specification limit.

And, you will see performance, observe not a single observation falling outside this one as per the real data. And, but if you superimpose this and convert into Z and Z_{LSL} and Z_{USL} like that and that some probability we are getting and that can be converted into PPM and this is the number that we get ok.

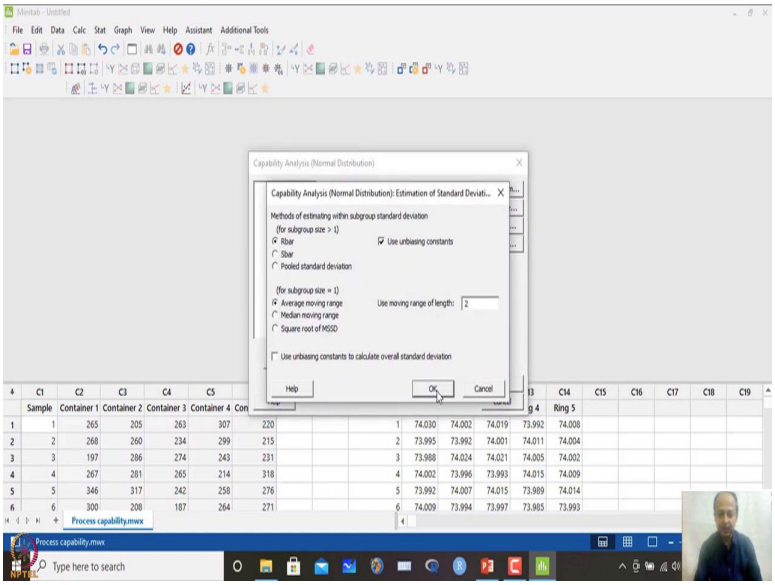
So, how do you calculate C_{pk} index over here? So, I am just taking you to the same example that is given over here and we will delete other things. Then go to I am showing you again Stat, Quality Tools and Capability Analysis, Normal capability analysis.

(Refer Slide Time: 24:27)



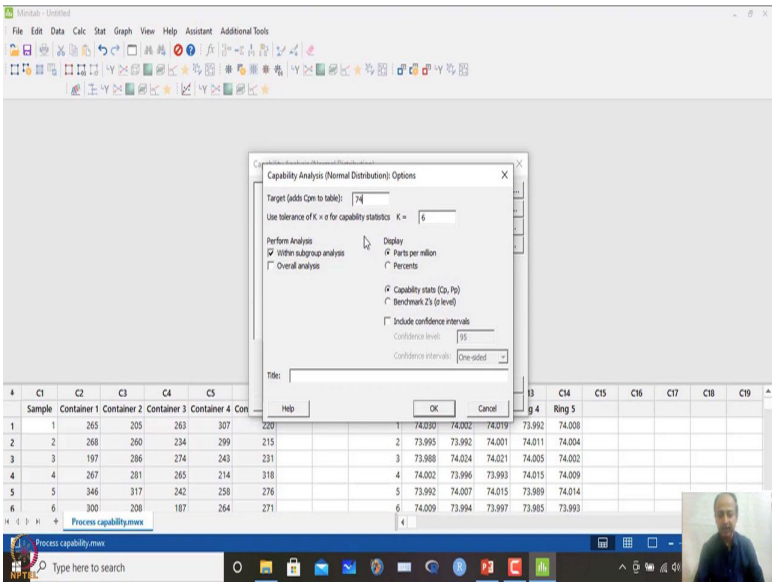
And, what do you do is that you Ring 1 to Ring 5 and select this one, then I change the specification as 73.965 and this is 74.035.

(Refer Slide Time: 24:45)



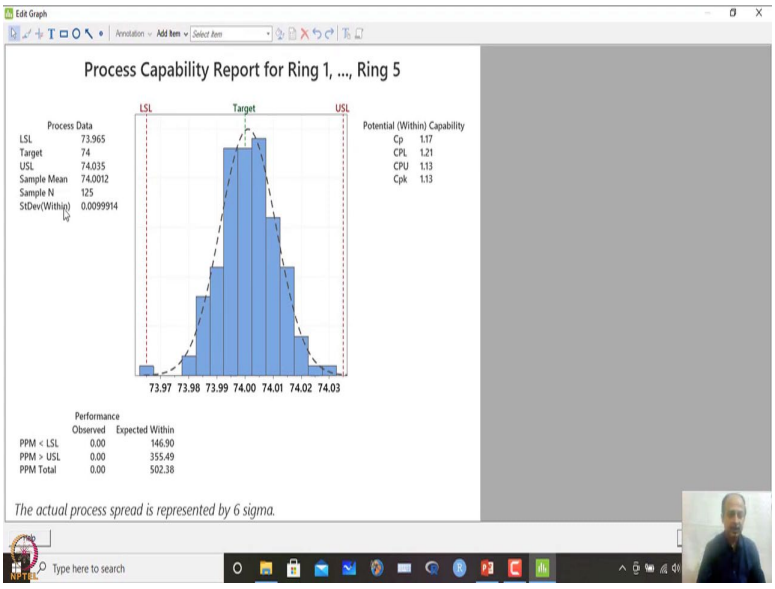
And, then Estimation over here $\overline{R}$ is the estimation that we will use, other things remains same.

(Refer Slide Time: 24:50)



And, Options over here target value let us say is 74, we can just mention about observation like that. And, Parts per million will be reported and Capability index will be reported over here, we will concentrate on that.

(Refer Slide Time: 25:03)



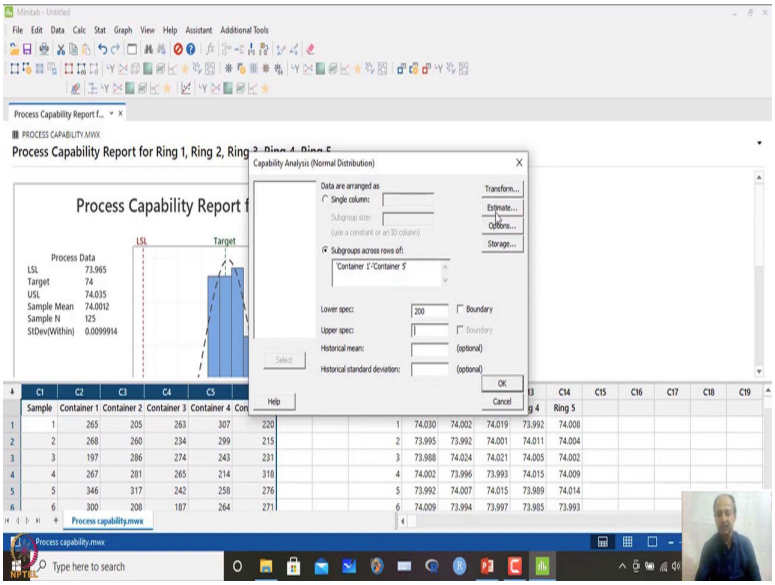
And, click Ok. What will happen is that you will get this information and this is the graph that will be reported over here and you see standard deviation is calculated. This is

the standard deviation that you observed over here and this is 0.0099, this is the within standard deviation; that means, $\frac{\overline{R}}{d_2}$.

And, this is the C_p index 1.17 and this is the C_{pk} index. C_{pk} is minimum of C_{pl} and C_{pu} that is the formulation we have and minimum is 1.13. So, a MINITAB is reporting C_{pk} value 1.13 what is as per the theories that is expected that is the theory that is expected ok. So, within performance is also given over here.

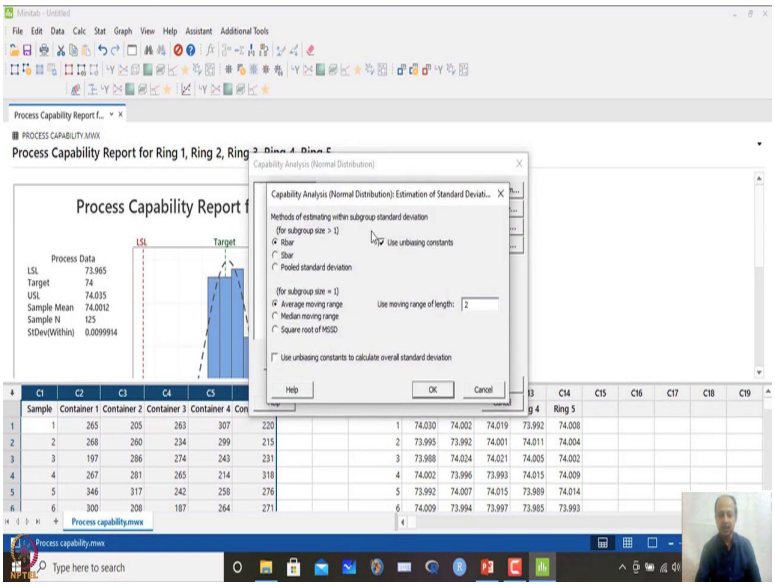
The actual process spread is represented by 6 standard deviation that MINITAB is expressing because the MINITAB has used 6 because I have given 6 as the option to calculate this one. So, that is reported over here. And, we can take the second example. So, if one sided specification, what will happen?

(Refer Slide Time: 26:13)



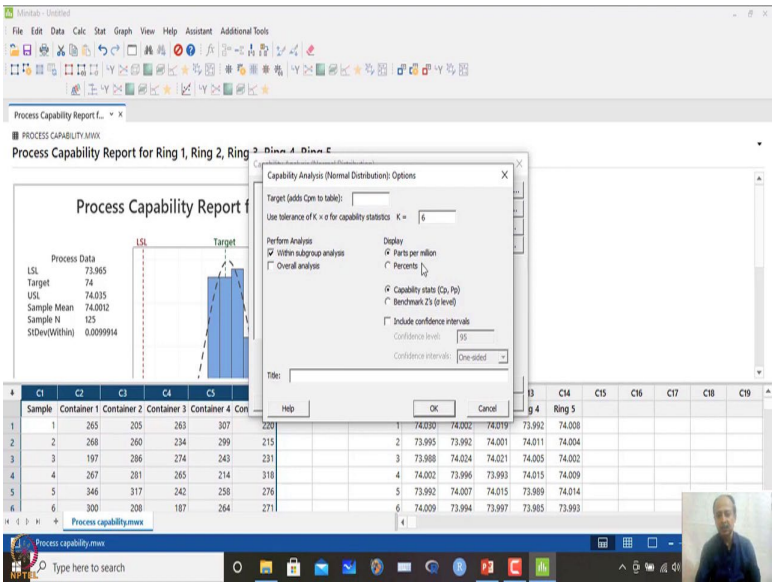
So, this example is has a one sided specification. So, I will go to Quality Tools and then I will go to Capability Analysis, Normal over here and then I will give Container this example up to this point and I will select those and in this case I will only mention 200. I will not mention the Upper capability Upper specification.

(Refer Slide Time: 26:20)



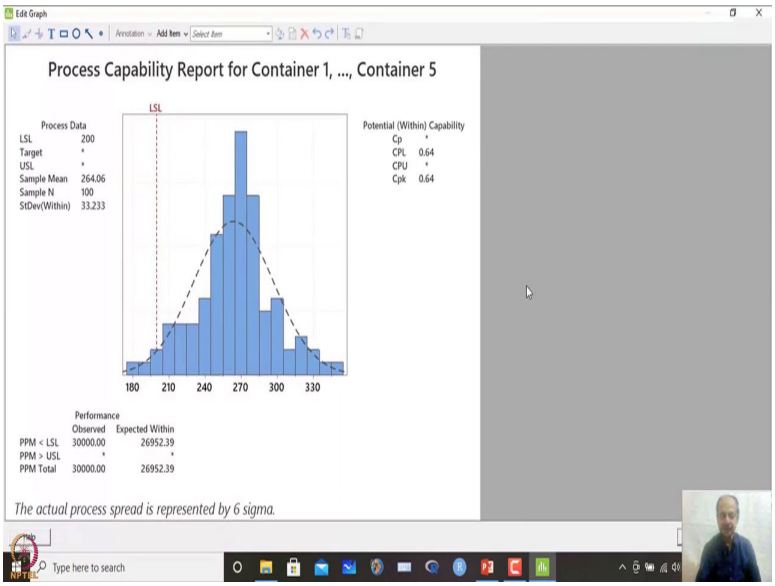
And, then in Estimation we are mentioning $\overline{R}$ formulation.

(Refer Slide Time: 26:24)



And, then Options over here, the Target value I am not giving; Within subgroup analysis we want to do and Capability I want to report.

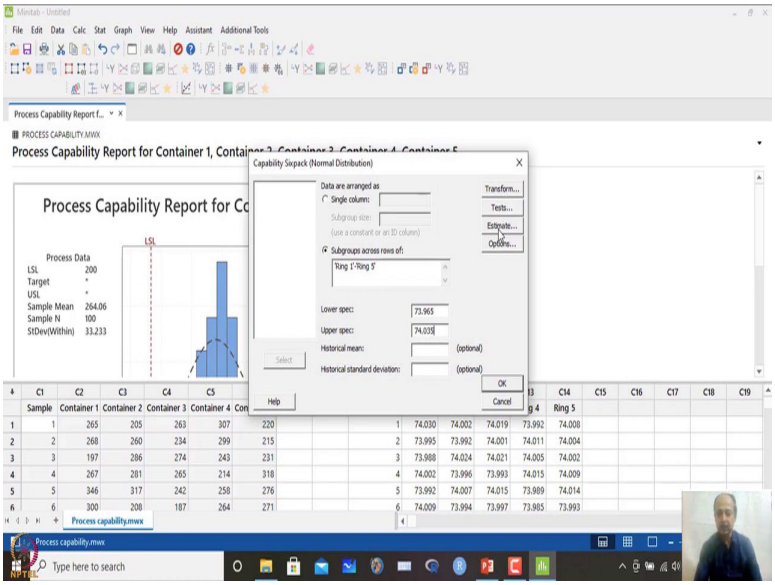
(Refer Slide Time: 26:32)



So, if you keep one sided specification what happens is that you do not get values of C_p over here and what you get is C_p lower C_{pl}, C_{pu} is not there. So, C_{pk} will be just C_{pl} over here ok. Similarly, if upper specification is given in that case here LSL was given as 200. So, that is why you are seeing this one.

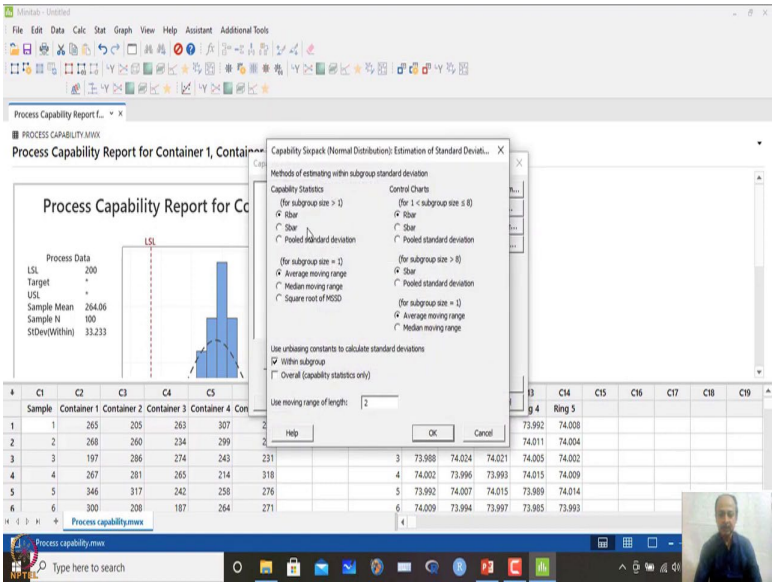
But, if only upper specification is given it will calculate C_{pu} and that will be the value of C_{pk} basically. So, one sided specification; C_{pk} will be reported. But, C_p will not be reported like that and C_{pk} is the index that we will consider over here. So, that is the way we are doing capability analysis.

(Refer Slide Time: 27:28)



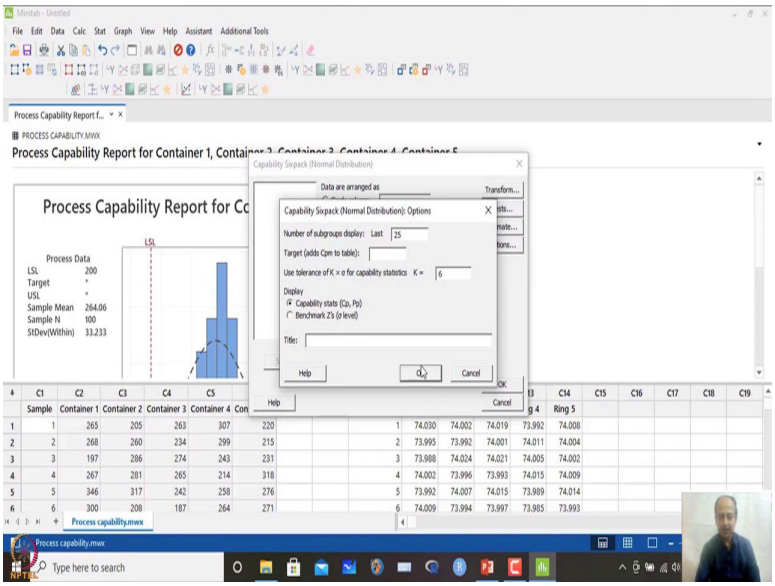
Another option you have in quality tools is that capability Sixpack. Here also there is a Normal. Here we you will get one more information that is Control Chart information also you can see. Control Chart let us say Ring 1 Ring this is the one and I give specification over here 73.965 and this is 74.035.

(Refer Slide Time: 27:39)



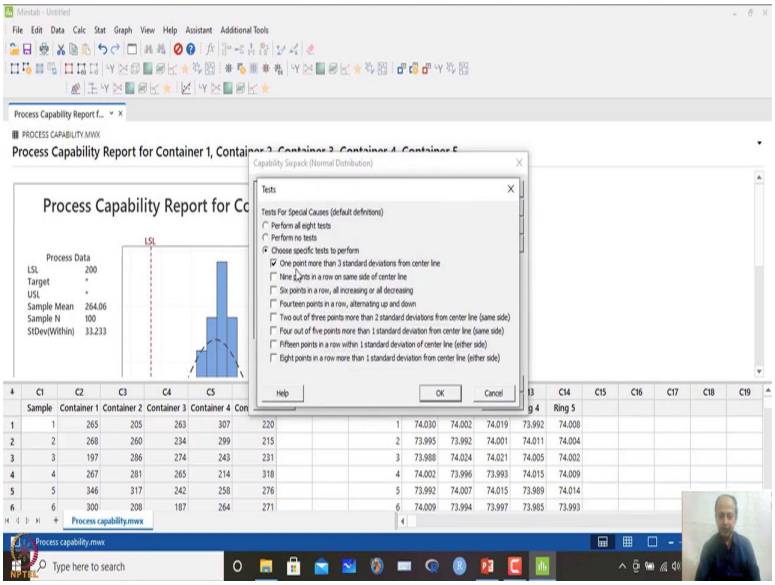
And, then I do Estimation same method over here. I keep it as it is. So, subgroups size it is not one. So, Within capability analysis. So, I am doing that one.

(Refer Slide Time: 27:51)



And, then what I do in options what I do is that I do not add Targets over here because we do not need to calculate C pm. So, that we are not interested into.

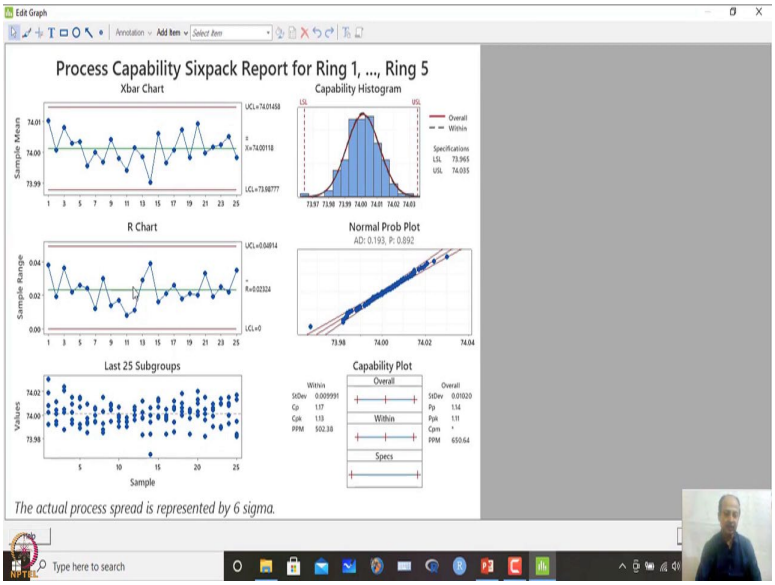
(Refer Slide Time: 28:10)



So, in this case K equals to 6. Test over here what I see one point going outside see sigma that is the condition in control chart that it is asking. So, we are doing that one. Either all 8 tests that is defined western electric we are not doing that. So, we are taking only one condition over here or other patterns we are not considering.

Any point going beyond plus or minus 3 standard deviation, that we are considering as abnormal scenario. So, I click Ok.

(Refer Slide Time: 28:31)



And, what will happen is that you will get this type of Sixpack many things you will get over here. So, in this case what you will get control charts X bar, R chart is plotted over here everything is in control and based on which upper specification lower specification is given. So, within specification C_p value 1.17 and C_{pk} is 1.13 that was the value we have calculated earlier and this shows any abnormality in the 25 subgroup size any observation that is very peculiar ok from the central line.

So, nothing peculiar is observed over here. So, and in this case random samples we can say. So, these are there is no problem over here. You will also find a normal probability plot that I told that we use for seeing whether the data set is normal or not. So, in these case what is observed from this dataset we have not gone into that details over here. But, we will discuss that in our next session in that.

And, we need to see how to check normality assumptions and over here you will find a value which is mentioned as p and that p value we see to check whether the normality assumption is violated or not because C_p and C_{pk} index depends on whether the data set is normal or not and what we are seeing is that data follows normal over here and all the

information in one go Sixpack means all the information in one go you are getting over here.

So, you can use that one as the option to see whether X bar, R is satisfactory then only we will go for C_p, C_{pk} index and then we will see whether to improve the process or not to improve the process like that. So, what we will do is that we will stop here we will start our next session from here and we will try to see some more information on sigma level and out calculate that one.

Thank you for listening to this lecture. We will return back in the next lecture with this topics to continue here.