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

Lecture - 31
u, g and h chart

A very good morning, good afternoon good evening my dear friends; welcome to this to TQM one lecture series under NPTEL, MOOC and I am Raghunandan Sengupta from IME Department; IIT, Kanpur. So, we have finished thirty lectures which is six weeks. So, we will continue another ten of them to complete the whole set of eight weeks lectures; so this is the thirty first one.

Now, if you remember in the fag end of the thirtieth lecture; we were discussing that depending on as the sample size change which means you have to find out the central line, the upper control, lower control and; obviously, it would mean that the values of the standard deviations would change because if you remember there is a factor of n which comes in the denominator.

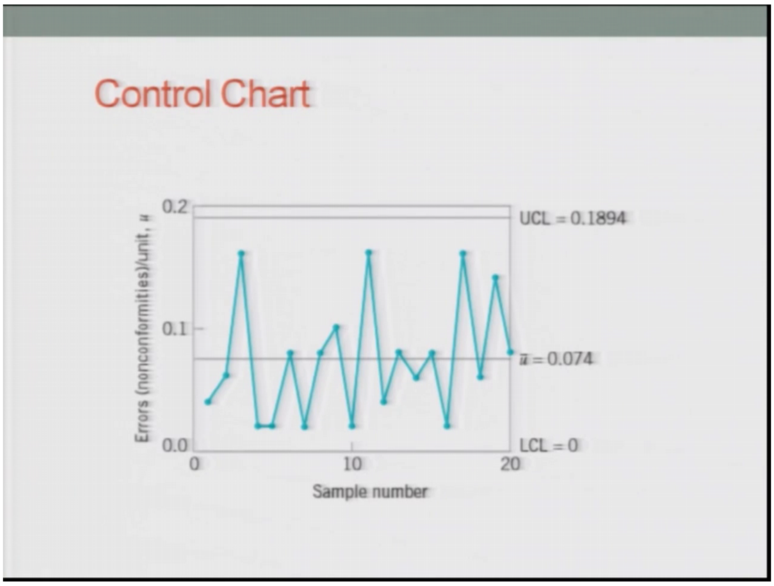
Generally, for the case of binomial distribution; so, the standard error or the sample or the standard deviation of the sample; standard error is a technical word which is used and that value is given by square root of p into 1 minus p by n. So, this n becomes n 1, n 2, n 3, n 4; depending on number of samples size which we have for each set of observations. And these p they need not be the actual value from the population; they would be basically estimates taken from the sample.

And based on that another concept which also came up was basically to find out the average sample size; like rather than changing it for the first set of observations from n 1; then it changes to n 2 for the second set of observations. You find out the average which is basically n bar and use that for our calculation.

So, in the first case when you are changing the sample size; the upper control and the lower control lines keep changing. So, if you remember; in the last class we were drawing short horizontal lines which basically denoted the changing UCL and the LCL, but for the average sample size being constant which you find out; which is n bar, you will be have a constant upper control and a constant lower control based on that you to do the calculations. So, we were just trying to complete one problem; we found out the n bar and we drew the; gave the information set based on which we can find out the calculations to find out the central line, upper control, lower control and then we draw the graph. So, this is what I am going to show now; the graph.

For the control charts when you want to find out the errors.

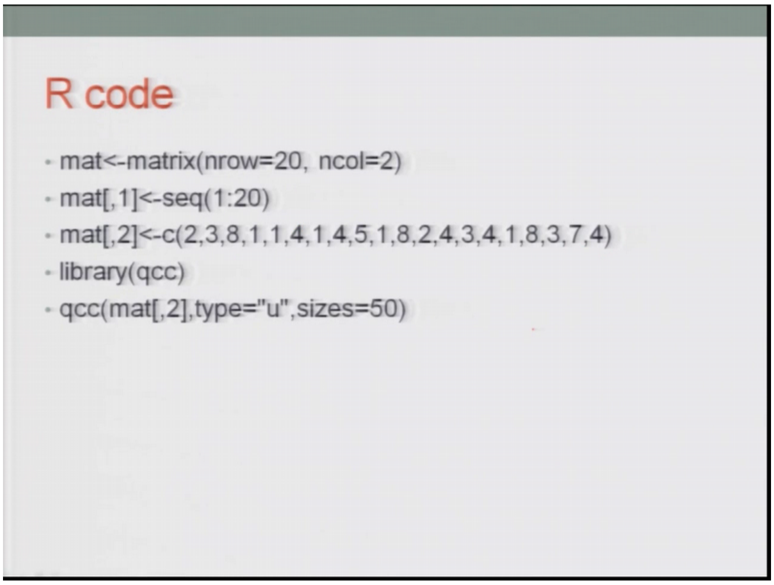
(Refer Slide Time: 02:54)



Now, also I am again mentioning it and the second time; generally in the nonconformity problems were also idea based on which we can draw the or basically have a look at the control charge. So, this was basically nonconformity which were basically studying. So, if you measure the nonconformity along the Y axis; along the X axis as you see; you have the sample number.

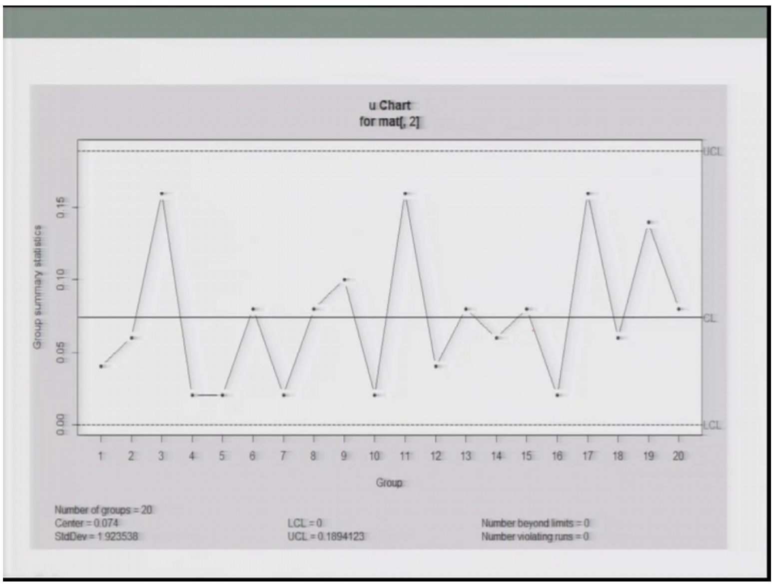
So, if there are averages which is the average value of the sample size; you see the changing errors which are given in the line diagram. So, this is the line diagram which we have; where I am just hovering my pen, I am not drawing over it; so, it goes like this. So, they are always between the upper control limit; which is 0.1894 and the lower control limit; which is 0 and the average number of errors who comes out to be 0.074; which I am just circling.

(Refer Slide Time: 03:59)



So, if you write the R codes. So, this R codes are giving just as a boost up concept which many body would be interested to take up. And if you learn R; it really helps in trying to solved different type of problems either in statistical quality control, total quality management or in the area of statistics and all this things. So, you have a matrix of size 20 cross 2 and you have the sequences of numbers given from 2, 3, 8; till the last three being 3,7,4. And then basically you find out the average values draw the upper control, lower control and basically name what is the nomenclature of the graph.

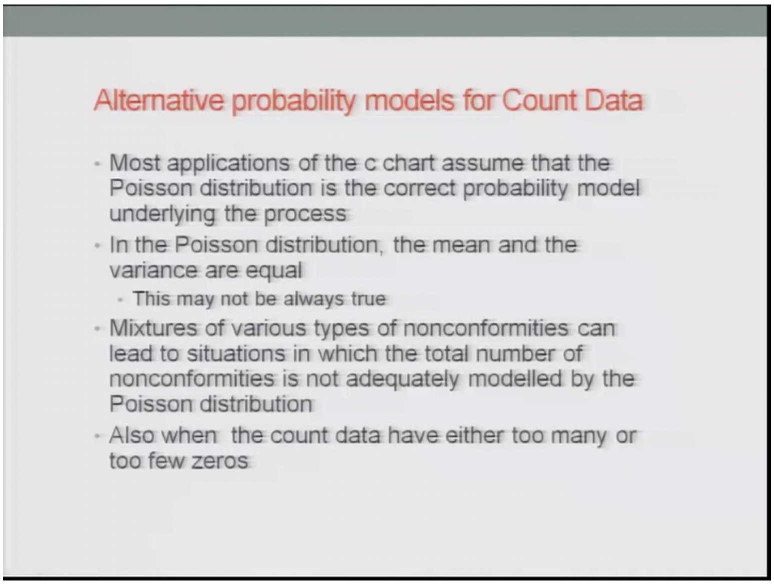
(Refer Slide Time: 04:42)



So, this is the group summary statistics are given along the Y axis; the groups or the sample number is given along the X axis. You have the upper control limit, the lower control limit; upper control limit is basically 0.1894123.

So, if he keep it to three decimal places is more than enough to solve the problem. The lower control limit value is 0 and the central line is given as 0.074 based on that you will find out; so, there are any aberrations and the set of observations of there out of control. So, for this diagram is they are not out of control.

(Refer Slide Time: 05:18)



So, we basically will now study alternative probability models for count data. So, what are these I will try to explain? So, most applications of the c charts; the conformity charts which we does considered assume that the Poisson distribution is 2.

Now, again to step back I am not changing that course of discussion, but remember that whatever the distribution be maybe; it is binomial maybe is Poisson, maybe is exponential. If you keep taking observations, then in the long run if you draw the histograms; the histograms will slowly mimic the normal distribution. Now, which means that whatever the underling distribution was; the conglomeration of the samples which we were going to take; the sample characteristics would slowly basically mimic the normal distribution.

And the corresponding mean and the standard deviation on the normal distribution would be something to do with what is the original distribution based on which you are trying to draw the standard normal distribution. So, this was basically the essence of the central limit theorem.

So, continuing that let me read the first bullet point again; most application of the c chart assume that the Poisson distribution is the correct probability model underlying the process. So, they would be other distribution for other cases, in the Poisson distribution the mean and the variance are equal; which is true. So, this may not be actually true in all the cases, that is absolutely true. So, what we will try to do is that try to find out some other distribution which basically mimics the so, called population characteristic based on which we are trying to do the study. So, mixtures of various types of nonconformities can be lead to situations in which the total number of nonconformities is not adequately modelled by the Poisson distribution.

So; obviously, it would be mean that the Poisson distribution based distribution or based on which we are trying to draw the; find out the limits and then try to find out that how the overall process graph changes; me have a never error because the Poisson distribution by itself is not true, also when the count data has either too many or too few zeros; maybe a problem.

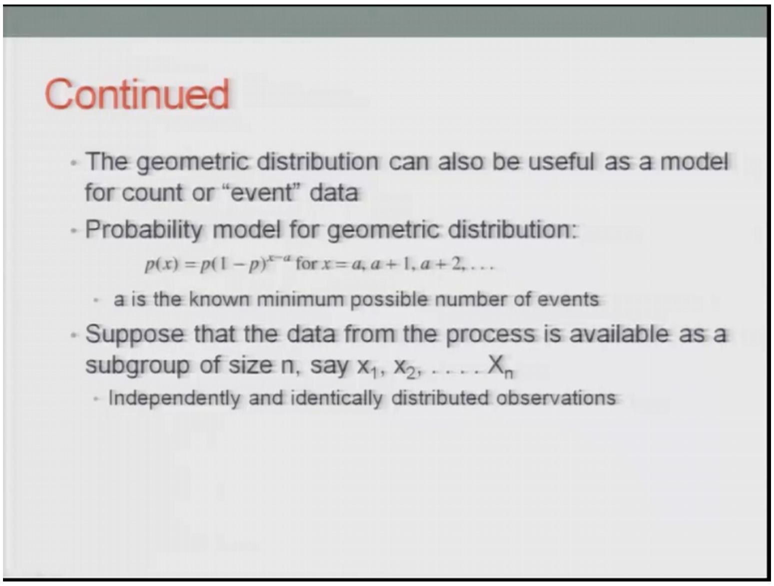
Because in the Poisson distribution is a discrete one and in the discrete case; we want to find out: what are the numbers. So, if there are numbers so; obviously, they would basically denote; consider like this, the number of defects or the number of good items you want to find out. So, if they are too many zeros or too less few are zeros then; obviously, the Poisson distribution may not be the actual case, which will give you that information.

Now, remember that one of the basic theories in probability was that in case if you have a distribution, consider for the binomial case; if the sample size n. So, this n and the sample size which we take for the drawing this charts are different; technically different. So, if the sample size and the parameter p which is the proportions of good items or the proportions of bad items; whichever you look at and whichever you analyze the problem, if n is very large and p is very small then; obviously, it can be mimic by the standard normal deviate.

So, there is a theoretical proof of that, but I am just stating it where things would be much clearer that depending on some assumptions, whatever the underlying distribution is whether it is grid or continuous; it can be mimic by the standard normal distribution or the distribution which is normal with a certain mean and a certain variance.

So, what are the assumptions that would become very relevant depending on the problem?

(Refer Slide Time: 09:08)



So, to continue; so, geometric distribution can also be used and as a model for count or event data. So, how many number of times that thing is occurring; so, it can be say for example, I want to find out that and considered again going back to the same example may not be the exact very best example for this discussion, but let us consider this. Say for example, you are checking number of very high advanced technology equipment.

Again, I am saying that you are using it for a fighter jet or using it for in a phase maker or using at a very high ended things considered a bridge over which lots of traffic goes or a train pass. So, any error; obviously, would be disaster. So, what you want to check is that the moment we find out one error; obviously, we will reject that; so based on that we are trying to use the model which is the geometric distribution.

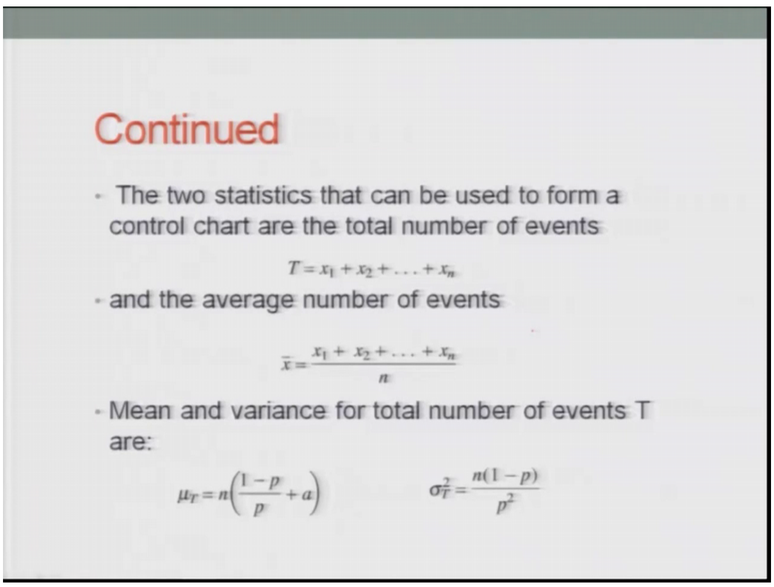
So, you want to find out that if an error has a probability of p. So, obviously, the probability of not and not having in error would be 1 minus p. So, the moment the first error occurs you will basically take corrective actions and basically make a note on that. So, the probability model for the geometric distribution is given as p and 1 minus p which is basically x minus a. So, this x would be basically depend on the number of observations you are taking and a can be 1, 2, 3; x can be a, a plus 1, a plus 2, a plus 3; depending on how you have framed the problem.

Suppose that the data for the process is available as a subgroup of size n and the observations are x 1, x 2, x 3 till x n. So, they would be independently and identically distributed; based on that you can find out what is the expected value, what is the variance and then considering the central limit theorem is true, you basically use the standard normal deviate, but use the normal and the variance of the standard normal deviate or the normal distribution as those coming from the actual distribution based on which you are trying to draw the normal distribution; which in this case is basically the geometric distribution.

But, obviously, would mean that technically if you would take more and more observations; later on then the normal distribution would basically be the best distribution would which would mimic the actual population distribution.

But with changed mean and changed variance which I am again mentioning are the actual expected value and the variance for the population distribution from where you are trying to pick up the observations.

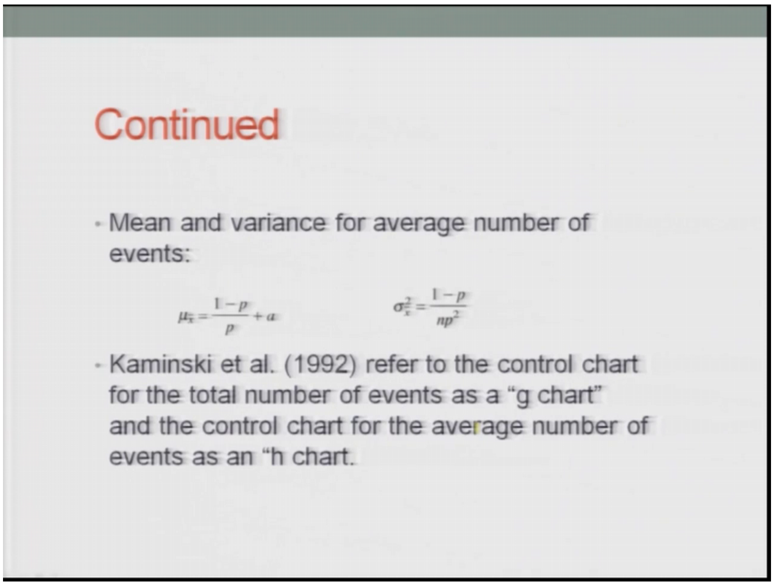
(Refer Slide Time: 11:51)



So, there can be basically two statistics which you can use, they would basically be the total number of events. So, the total number of events would be counted at x 1, x 2, x 3, x 4 till x n and generally we would be interested to find out what is the average number of total number events. Events may be either say for example, good items are passing, events can be either bad items are passing; events can be say for example, white colour boots being produced. So, depending on whatever your outlook is.

So, you find out the average; average would be basically sum of x 1 to x n divided by n and then if you find out the means and the variances, they would be given as mu suffix T and sigma square suffix T as given here; which I will try to highlight. And you will use this mean and the least standard deviation as the mean and the standard deviation for the actual normal distribution based on which we will do the calculation.

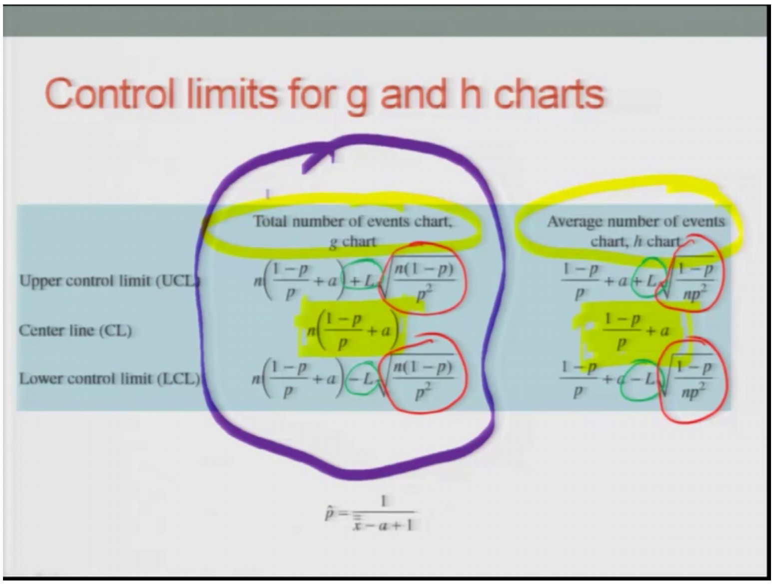
(Refer Slide Time: 12:56)



Now, if I want to find out the mean and standard deviation for the average number of events; they would basically be divided by n and we get the mean value and the variance of the standard deviation. Standard deviation would be would be the square root of the variance as given here. So, Kaminski refers to the control charts for the total number events as a g chart and the control chart for the average number events is basically h chart; based on that you can basically have the g chart and the h chart, which will give you lot of information that how the average value and the standard deviation can be made or you can utilize that who draw the process control charts; which will give you a lot of information how the process is going.

Whether there is a assignable causes and non assignable causes and he can take corrective actions as needed.

(Refer Slide Time: 13:44)



So, control limits for the g and the h charts as just discussed. So, you have the central line which for the two cases would be like; in the total number of events charts for the g charts would be given by this; which I am getting using yellow highlighter. And then you will find out the corresponding, the standard deviation of the variance and use them as the control limits like; mu plus minus k sigma. Now, this k value which is I am saying is again is basically the overall confidence level or the level of confidence you want to use.

So, then you will see that you find out the value of the standard deviation; again the standard deviation. So, this plus and minus; I will again use a different colour to make our understanding much clearer. So, this minus and this plus would basically denote to what positive value on to the from your side if you are looking, on to the right hand side if you will go and what is the negative deviations of the dispersion we will have the negative sign.

So, these if you remember they would technically be the standard deviations would be square root of n into 1 minus p by p square. Now, if you come to the values of the central line and the upper control and the lower control depending on. So, let me again change the colour; so, this is the average number of events chart, which is h then let me again highlight the average value.

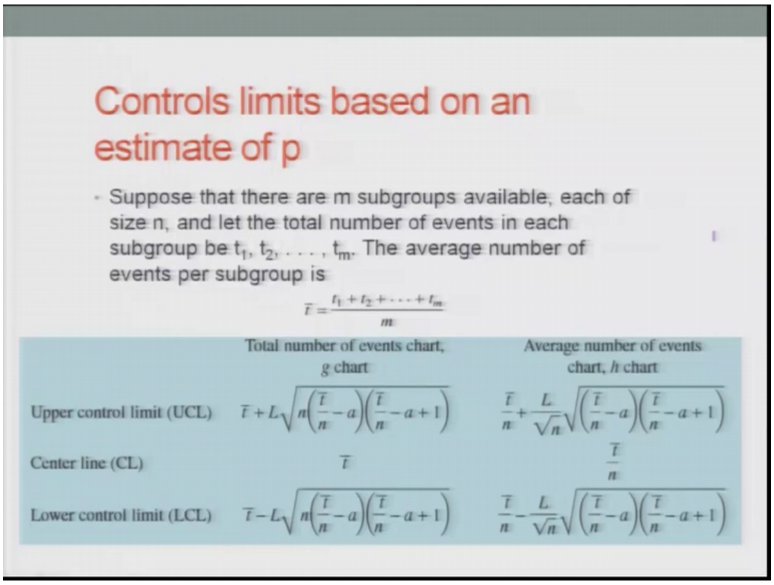
So, average value is basically see as you divide by n so; obviously, those changes would be there. So, it is 1 minus p by p plus a; so, you use that as a central line and again let me highlight the values, which I have for the standard deviations are this values; which is square root of 1 minus p divided by n in by p square. Now, why it is coming n square; n is being divided because n would basically for the variance would become a square.

So, once it goes into the denominator; it would be calculated accordingly and continuing our discussion; we find n is also state, the plus value and the minus value are exactly as mentioned. So, they are plus l minus l; again it will depend on the level of confidence. So, this p hat would basically be given by the ratio of 1 by x double bar; x double bar is the actual average of the average minus a values. So, the a value which you already know and can be calculated and it can be utilized accordingly.

So, this calculations would be used for when the total number of events is given; which is h charts and this would be used for the case when the average number of events are given which is the h; that was the initially sorry that was the g chart this is the h chart. So, the control limits are based on an estimate of p; suppose that there are m sub groups.

So, we have as I mentioned and we discussed in the thirtieth lecture. So, we had subsamples, we take the first one, the second one third one, forth one and continue taking and we take m number of set of observations. And in each initially; if the sample size was fixed we took n observations in each set and that is what we are discussing. And later on we saw that we could change the sample size also in each subgroup.

(Refer Slide Time: 17:30)



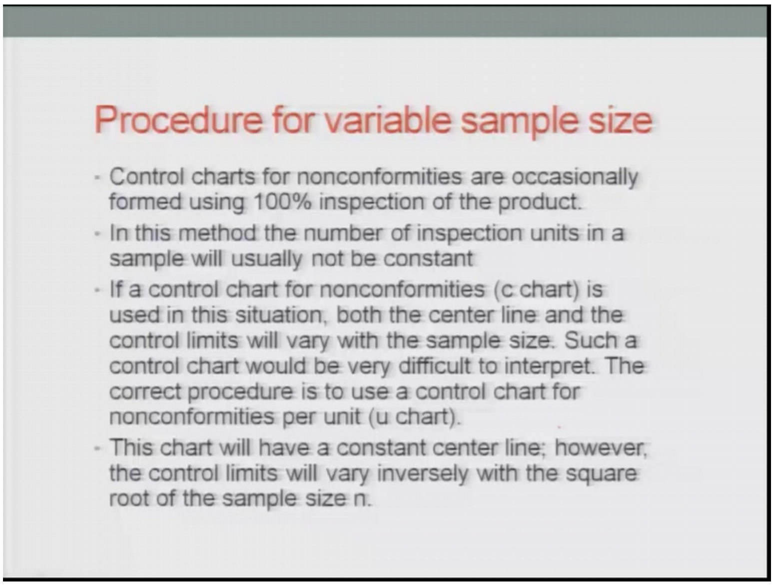
So, continuing the discussion suppose that there are m subgroups available each of size n. So, the size n is not changing and let the total number of events in each group b t 1, t 2. t 3 till t m. So, the average numbers of events per subgroup would be given by sum of t 1 to t m divided by m.

Now, if you want to find out the total number of events chart; which is the g chart and the average number of events chart; which is the h chart. The concept is exactly the same, now what you do is that let me use colours; which is easily discernible. So, this average basically comes as the central line; in this case and in this case it basically gets divide by n because I am talking about the average number of events charts.

In technically in that other example also it was basically divided by n. Now on the corresponding standard deviation for the g charts, where I will again highlight is given by this; not the l, l part is the level of confidence. Here in the other case the standard deviation is given by the circle part, here only you should remember this is been divided by square root of n. Now why square root of n? Because if you divide by n you will not find out the variances average; but then you want to find out the standard deviation, you should be divided by the square root of n; that is why.

So, it technically divide by n then later on in the next stage, when you find out the standard deviation; it will be divided by square root of n. So, again the same concept comes you would plus minus you add some plus minus values of l. So, those l would basically give you the level of confidence. So, here also; so, the control limits based on the estimate of p R found out and you have basically the g chart and the h chart which will give you the information as just stated in the slide.

(Refer Slide Time: 19:27)



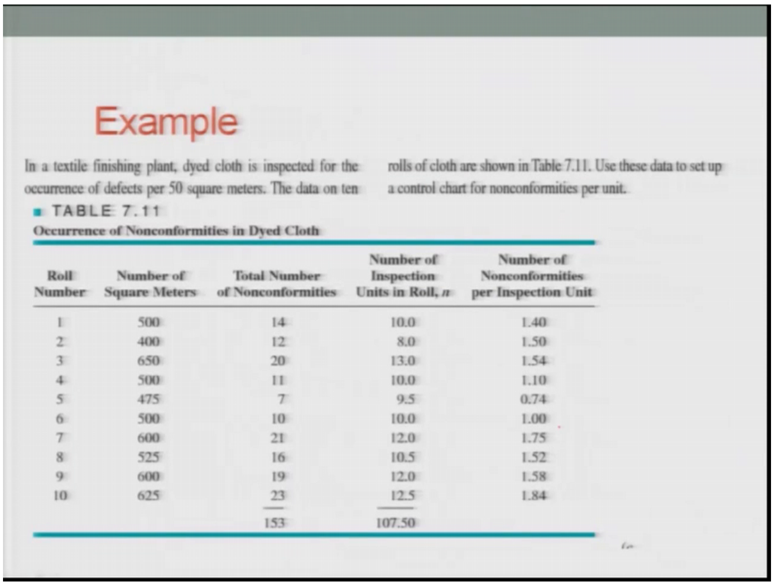
Now, procedures for variable sample size. So, control charts for nonconformities are occasionally formed using 100 percent inspection of the products. In this method, the number of inspection unit in a sample would usually not be a constant. So, if a control charts for nonconformities which is basically the c chart is used in this situation; both the central line and the control line which will vary with the sample size.

So, as the sample size changes; obviously, the central line and the upper control and the lower control will change. Because in this case, the upper control and lower control would basically be dependent on the central line based on which we will need to recalculated.

Such a control charts would be very difficult to interpret because they are changing; which means the central line is changing in the upper control lower control values are changing. So, what we want to basically compare against the central line that would also be very difficult for us to infer. The correct procedure is to use a control charge for nonconformities which is per unit; which is the u charts. This charts will have a constant central line; however, the control limits will vary inversely with the square root of the sample size.

So, generally the u charts will have a central line which is fixed, but the upper control and the lower control values would basically change depending on so, called the sample size. Because that would be divided by n in the case of variance and once you take the standard division, it would become basically divided by square root of n. So, you have to understand it technically and do the calculations as necessary.

(Refer Slide Time: 21:01)

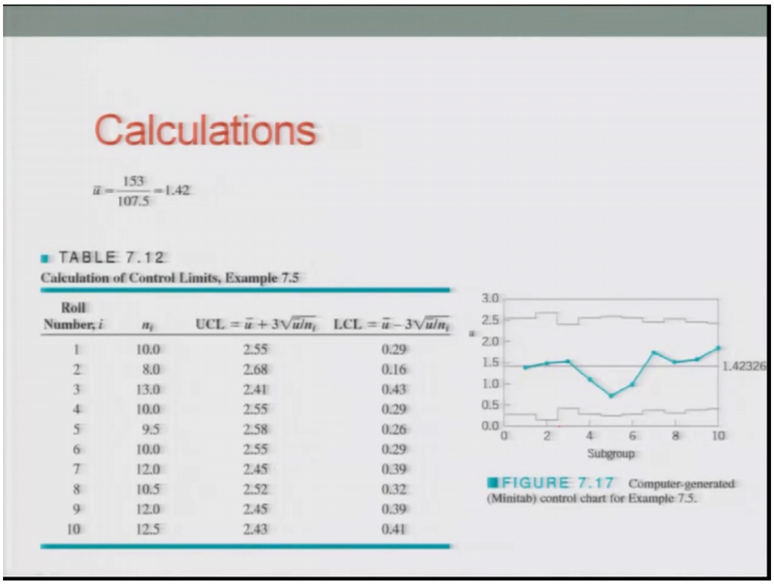


Roll Number	Number of Square Meters	Total Number of Nonconformities	Number of Inspection Units in Roll, n	Number of Nonconformities per Inspection Unit
1	500	14	10.0	1.40
2	400	12	8.0	1.50
3	650	20	13.0	1.54
4	500	11	10.0	1.10
5	475	7	9.5	0.74
6	500	10	10.0	1.00
7	600	21	12.0	1.75
8	525	16	10.5	1.52
9	600	19	12.0	1.58
10	625	23	12.5	1.84
		153	107.50	

So, let us consider an example here in the textile finishing plant; dyed cloth is inspected for the occurrence of defects per 50 square meters; the data on 10 rolls of cloths are shown. So, basically it goes from roll which are given by number 1 to 10; number of square meters is given starting from 500 to 625, total number of nonconformities in each role is given. So, the first one in 14, second is 12; second last is 19 and 23.

So, the number of inspections units in roll whatever the roll number is basically n is given as 10, 8, 13 and the last two values are 12 and 12.5. So, the number of nonconformities per inspection unit are found out. So, based on that you will try to basically recalculate and draw the charts such that we will give you a much better perspective where the process and how the process is going.

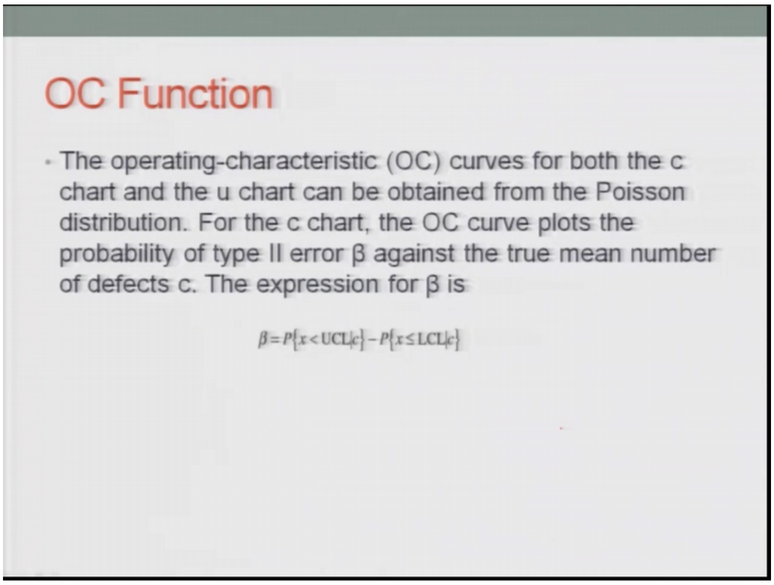
(Refer Slide Time: 21:55)



So, based on that we do the calculations and find out u bar as 1.42; in the other case for calculation the control charts. So, ni's are given, so these are the sample size which are changing from set of observations to observations. So, the first one was basically the ni's where given; depending on the values of the calculations. So, once you find out the upper control limit, which is basically u hat plus 3 of into multiplied by the term; which gave us the standard deviation and minus value would again be minus 3 into that term which gives us the standard deviations.

So, if you draw that; so, the chart which we have in front of us which is in figure 7.17; gives you the subgroup number which are noted along the X axis. And along the Y axis; you have basically u and the central line would basically be u bar and the fluctuations would be basically found out using so, called change values of the; or the recalculate values of the standard deviation.

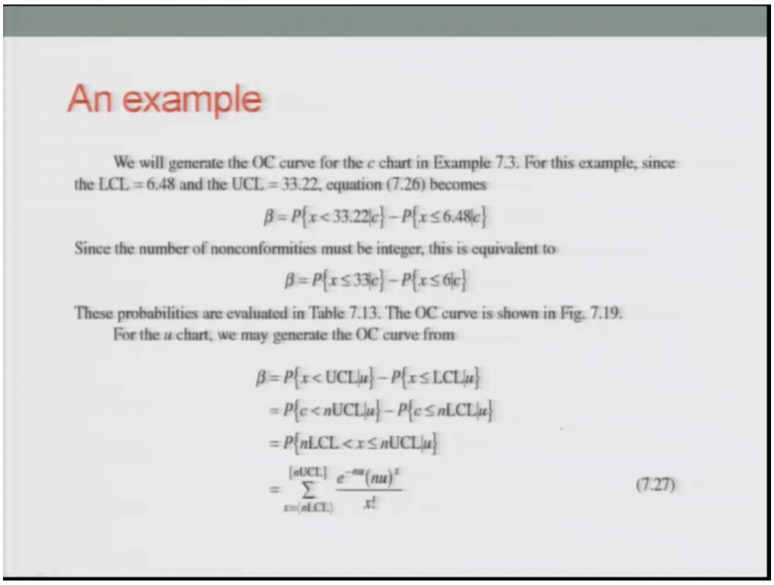
(Refer Slide Time: 23:05)



So, operating characteristic also both the c and the u charts can be obtained from the Poisson probability distribution; for the c chart the OC curve plots the probability of type II errors; against the true mean values, the expression of value of beta is given. That given c, whatever the value c is that you want to find out the probability of x being less than the upper control or x being less than the LCL and basically find out the difference of this probability and concealment on what is the value of beta.

So, let us consider an example; we will generator OC curves from the c chart as given in sample; example 7.3; for this examples since lower control limit is 6.48; upper control limit is 33.22; the equation would basically yield the value of beta.

(Refer Slide Time: 23:42)



Which is now difference of two terms; what are the those terms? One is the probability that x is less than 33.22 given a certain value of c and the other value which will basically subtract from this initial value is the probability that x is less than equal to 6.48; given again the value of c is known.

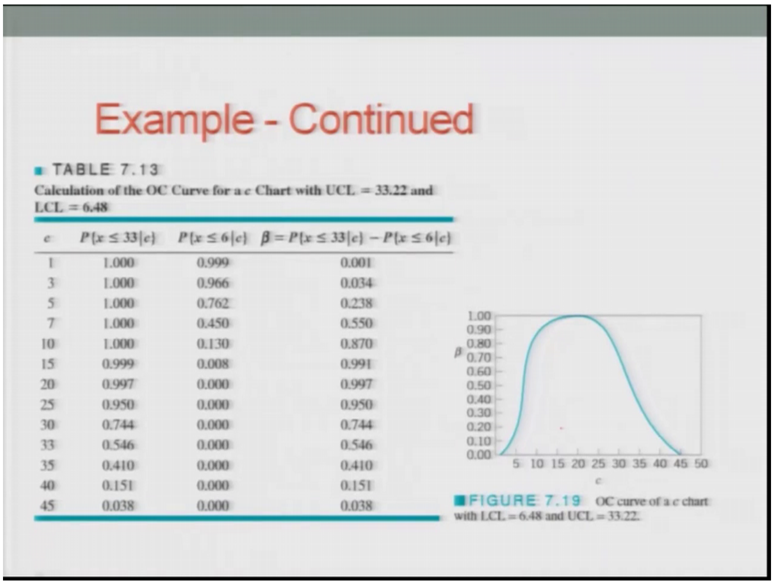
So, c is the number of nonconformities must be integers; this is equivalent to basically trying to find out. So, 33.22 becomes 33 and 6.48 basically becomes 6 and you do the calculations to find out. These probabilities are evaluated in and given table 7.13, the OC curve is shown in figure 7.19. For the u charts; we generate the OC curves from the following. So, say for example, we have the beta value; beta value is given by x being less than equal to UCL; given some of value of u and the x being less than equal to LCL; given sum value of u.

So, based on that again you will basically multiply the value of n on both on the left hand side and the right hand side and you get the so called numbers; not proportions now. They would be n numbers; so, you basically put the numbers and you find out what is the control limit for the value of x, which is basically in between n into the lower control limit. Lower control limit is basically coming from the proportions. So, proportions multiplied by n would give you the numbers.

So; obviously, they may not be in exactly integers; you have to basically find out the numbers in decimals and then basically extrapolate them according to whatever your decision is. So, say for example, if 32.13 then; obviously, I will be tempted to take the 33 number of observations; so, which is n.

So, based on that you find out the sum of this value of the OC curve values; the OC curve would basically the integration values of e to the power n into u. So, u value was also there, now the n values coming because there is the total sample size you have taken. So, based on that you find out what is the value of beta. So, example let us continue.

(Refer Slide Time: 26:08)

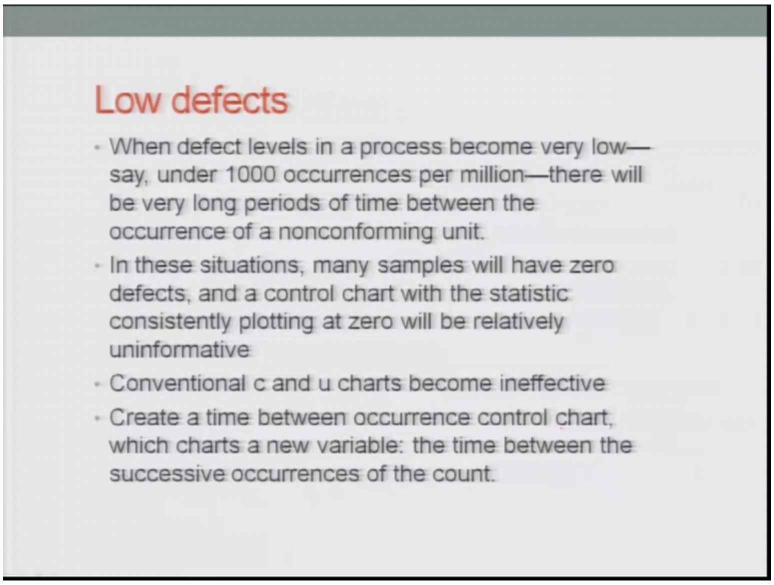


c	$P\{x \le 33 \mid c\}$	$P\{x \le 6 \mid c\}$	$\beta = P\{x \le 33 \mid c\} - P\{x \le 6 \mid c\}$
1	1.000	0.999	0.001
3	1.000	0.966	0.034
5	1.000	0.762	0.238
7	1.000	0.450	0.550
10	1.000	0.130	0.870
15	0.999	0.008	0.991
20	0.997	0.000	0.997
25	0.950	0.000	0.950
30	0.744	0.000	0.744
33	0.546	0.000	0.546
35	0.410	0.000	0.410
40	0.151	0.000	0.151
45	0.038	0.000	0.038

Calculate the OC curves for c charts with upper control limit as 33.22 and the lower control limit as 6.48. Column 1 gives you the c value as 1, 3, 5, 7 till the last; second last one is 40 then 45; then the probability that x is less than equal to 33; given some certain values of c; whatever it is it basically ranges from 1 to 0.038 depending on where c is. And then you can find out the probability that x would be less than equal to 6. So, this is considering the lower control limit.

So, these probabilities are given as 99.9 percent till about 0. So, there difference between that in decimal system will basically give you the fourth column which is the value of beta. So, the OC curve which c chart with LCL is 6.48 and UCL as 33.22 is basically shown; you can basically solve it using the same concept as I discussed.

(Refer Slide Time: 27:11)



So, low defects we are interested to find out; what is the reason and what are the causes for defects and how they can be made low. In general technically we will not want any defect, but that is practically not possible. So, we will have defects and how you overcome that, so which was saw that using different top charts and tables. When defect level in a process becomes very low; say for example, under 1000 occurrences per million.

So, they would be very long periods of time between the occurrences and the non conforming unit. In this situations, many samples we will have 0 defects and the control charts with static consistently plotting at 0 will be relatively uninformative. So, it will not give you any information and conventional c charts and u charts; basically in those cases would become defective.

So, create time between the occurrences of control charts which charts a new variable and the time between the successive occurrences of these events are noted out, such that somebody can basically take some collective decision based on that.

So, with this I will end the thirty first lecture and continue discussion the charts further on from thirty second and so on and so forth.

Thank you. Have a nice day.