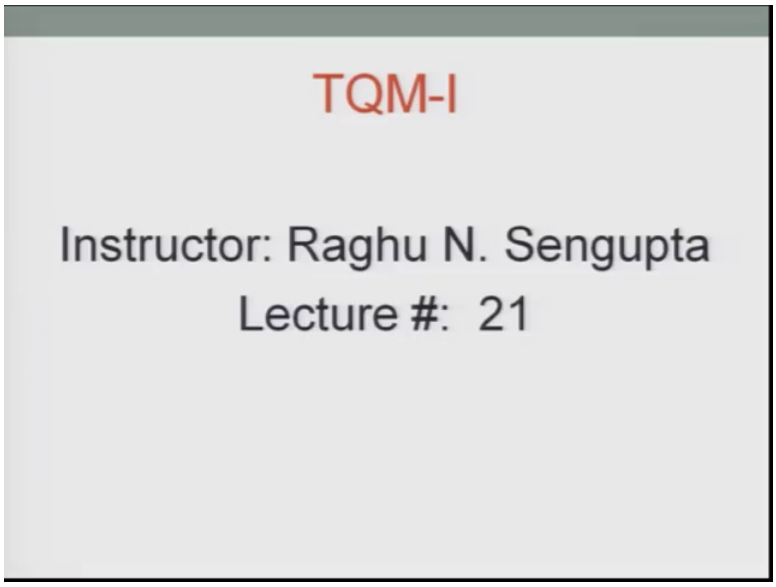
Total Quality Management - I
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 21
Basics of X bar and R chart

Very good morning, good afternoon and good evening my dear students I am Raghunandan Sengupta from IME department IIT, Kanpur.

(Refer Slide Time: 00:24)



So, this is the total quality management.

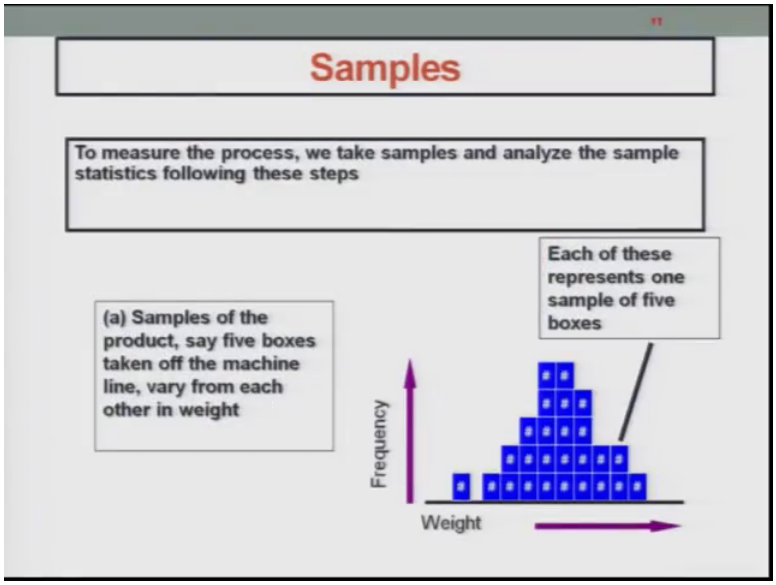
So, this is the total quality management one course and if you remember we had covered a lot in the last few lectures about OC curves LTPD concept AQL concept then military norms of statistical quality control, then one state sampling double samplings equation sampling, how the rules can be formulated depending on the values of n 1, n 2, d 1, d 2 c and all those values and then in the later part I do understand I did give a very brief background of r, because r is basically statistical packages which is free and it can be downloadable.

So, once somebody's acquainted with that many of the works either in quality control or in statistics in optimizations, in area of application in engineering anything you can safely use r, and that being free one which is exploding in a very positive manner with lot

of library functions been added people can definitely use that. I had just given a very brief preview of how to create columns, vectors and matrices and how to recall the numbers once you are aware data structures for that are should not be difficult.

And then in in the last lecture which was the twentieth one, we had the fag end I did start about on the concept of quality charts control charts and what was the essence or the difference between the variables in the attributes I did mention, and then will slowly go into x bar charts and r charts. So, if you remember at the last moment of the lecture I did mention that r or which is the range chart gives you some informations, which is in a way a type of variability or dispersions based on that we do different type of studies using quality control charts. So, with that we will start the 21st lecture.

(Refer Slide Time: 02:31)



So, when you are taking samples. So, in a big population you pick up samples and your main task is to basically understand the properties of the population using the characteristics of the sample check how close the characteristics of the samples are with the properties of the population and then make meaningful judgment about the overall population from the sample you consider and if you remember that I did mention in little bit details that cost structure is also important.

So obviously, it will mean that the more number of observations you take that means, increase the sample size it will have a negative impact on the total cost which means total cost would increase which is negative for you, but at the same time the efficiency

would increase; that means, this efficiency of prediction forecasting estimation would increase, but you have to make a compromise where to draw the line. So, continuing to that to measure the process we take samples, and analyze the samples or sample statistics following these steps.

So, what you do is there samples of the product say five boxes of the machine line are taking, they have a varying weight you basically find out the weight and try to find out the range between reach the weights vary or say for example, jam models are coming from the production line or say for example, some tyrod are being manufactured in the production line and they are coming out of production line you measure either the weight of the jam models or try to find out the length of the tyrod or they may be say for example, some heavy equipments coming out, and you want to measure what is the tensile strength of a particular part of that machine or you are manufacturing steel wires and you want to basically find out the young's modulus or some paint is coming you want to check what is the pigmentation level of pigmentation of this paint or see for example, some car is being manufacturer you want to basically test the brake horsepower capability are the car.

So, they can be different things or say for example, some coolant is being manufacture for high and manufacturing units they can use that coolant. Coolant means the; you are using CNC machines, you are using sophisticated machines and they are doing some mechanical work you want to cool down the overall process. So, use very sophisticated coolants you want to find out the viscosity of the coolant. So, all these things may be a part and parcel of your study considering the quality control checks need to be made.

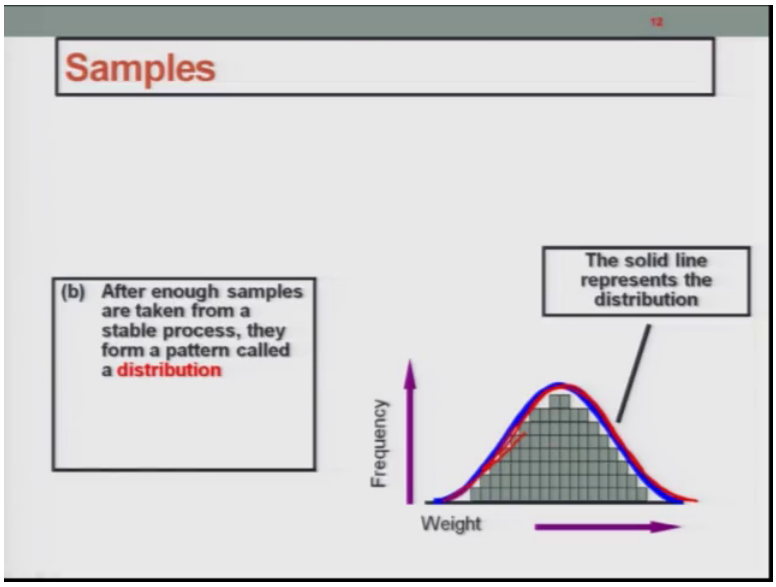
So, in this example again coming back to the example samples of the products, if five boxes taken out the machines lines with varying weights are checked and what you do is that you plot the weights along the x axis and the frequencies number of such boxes with which are exactly that a weight or within some range of weights are plotted along the y axis. So, say for example, that the graph looks like this as given in this slide which I will try to mark. So, this is the graph which we have.

So, each of these little boxes which are blue in color they represents one sample of five boxes. So, once you basically mark the frequencies, it will be say for example, for a certain weight which I am marking with this yellow highlighter would have say for

example, number of boxes which are 1, 2, 3, 4,5 five multiplied by 5. So, those 25 of these boxes would have a certain weight. In this case say for example, for a weight of let me mark the weight for ease of understanding say for example, this is 120 kgs.

So, there are only one box say for example, this box this set of boxes are 119kgs. So, there would be. So, this each box is 5, 5 in number. So, this would be 5 in number this in this case it would be 5 into 2 in number. So, once you plot it you will basically have a. So, called histogram or the distribution or the weight of the boxes with respect to the weights and you can find out some characteristics of the overall sample as well as the population from this histogram which you have drawn.

(Refer Slide Time: 06:47).



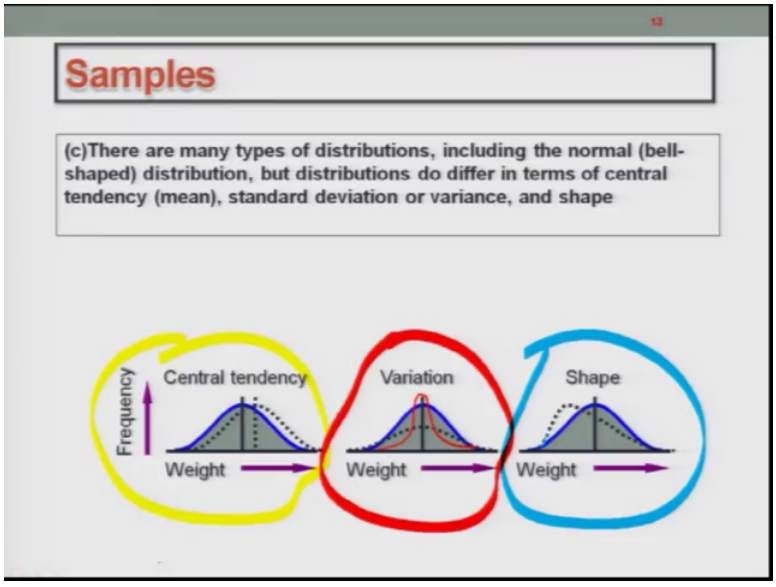
So, it means that after enough samples are taken from a stable process, stable means the variability all these things would be there, but they are variability which you cannot control there other sets of variability which you can control if, but if those variabilities go out of the control there is a problem for which you need all the statistical process control techniques. So, after enough samples are taken from the stable process they form a pattern which is called the distribution.

So, if you see the boxes going back to the last slide, if you basically join the midpoints of the of the histograms and join them and if the number of observations you are taking is quite large then in the long run using the concept of central limit theorem, you can

basically find out that the distribution would be almost normal as shown in this graph where I am basically plotting my this electronic pen.

So, this red line which shows which overlaps the blue one is basically a normal distribution and the histogram would basically give you that information that is really its true its normal. The solid line represents the distribution, now there are many types of distributions including the normal which is the bell shaped curve which I just mentioned few seconds back, but the distributions do differ in terms of this central tendency.

(Refer Slide Time: 08:06)



So, now, the word central tendency which will use will be respect to the mean or which is the center of gravity of the distribution, other central tendencies may be the mode may be the median, but will stick to the study of the mean as such for the statistical process control charts.

But distributions do differ in terms of the central tendency as I mentioned and obviously, thus the dispersions would be measured using the concept of standard deviation of variance and obviously, there would be a change in the shape also. So, when I am mentioning the word standard deviation or variance, do remember that standard deviation and variances with for any distribution is measured with respect to the mean only; obviously, you can formulate the problem when you want to understand the concept of standard deviation with respect to the median also, but for our discussion will only stick to the mean.

So, if you see the figure one. So, which I will try to highlight using yellow color, this has basically the central tendency which is and the central line. So, this is the mean median or the mode. So, this is a very unique characteristics or normal distribution, and the mean median mode is generated by bold black vertical line, now if the central tendency shifts to the right or to the left depending on in which direction the movement is, the dotted lines over which now the electronic pointer is pointing is basically now the new distribution and it has basically shifted from the bold line to the dotted vertical line.

So, again these are you are measuring the same thing which are the weights along the x axis and the frequencies or the relative frequencies or the chances are being basically be measured along the y axis. Now if the variability increases and the mean value remains the same, increasing means it becomes more flat; that means, more dispersed. So, if you look at the second diagram which I will try to highlight using the red color now. So, this would be the diagram where the variability is increasing, but the central tendency of the mean values same. If you see the vertical black line remains in the same spot the normal distribution has flattened up; that means, the number of observations I have become more dispersed on to the right or to the left of the mean value.

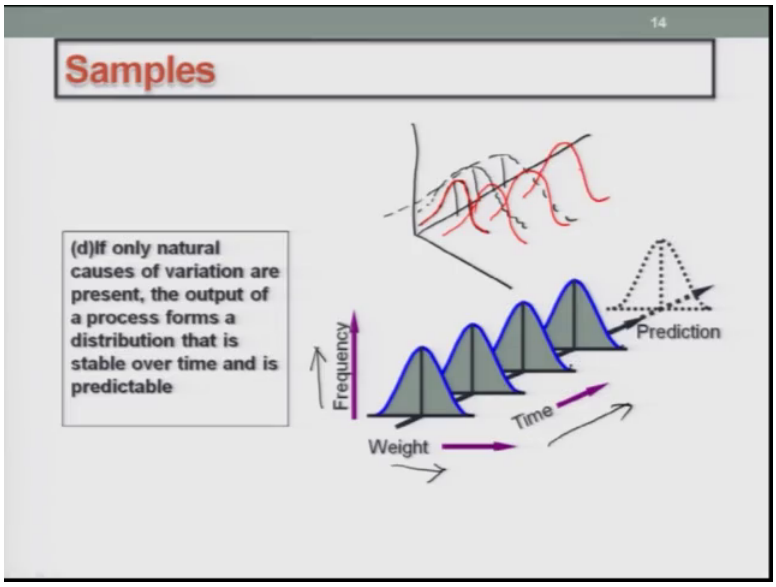
And if it is more dispersion is less that in that case the diagram would look like this. So, I will try my level best to draw it. So, this blue line would now become. So, this is the normal distribution which you have; then if I am going for the change of the shape. So, when you basically check the shape is changing it means, but the overall dilatation of the normal distribution I am talking from the point of the normal distribution would either shifted to the left or to the right.

So, if I am able to draw it for the right one let me try to do it accordingly. So, in this case where I am hovering my pen over the dotted one is shift to the left and if it shifts to the right the distribution would look like this. So obviously, it would mean that the mean values and the standard values would change for a change of the shape will come to that later on.

So, technically the variabilities of the samples or the change the word variability I am not using from the concept of variance, the change or the characteristics change of the sample which will have a some sort of information about the population parameters or the population characteristics, would come out from either the change of the central

tendency, change of the variability or change the shape of the distribution for this case we are considering the normal distribution to be true because it is easy for us to explain to each other how the changes occur and how they look like.

(Refer Slide Time: 12:16)



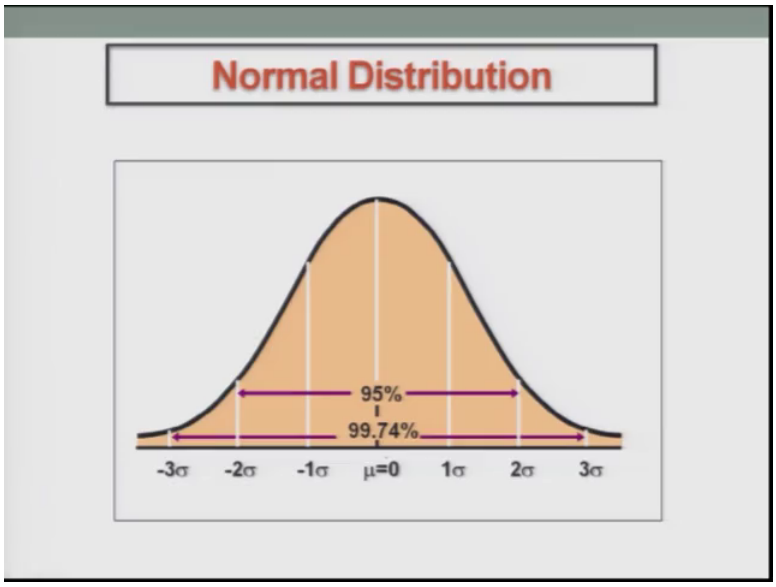
Now, if you are considering the samples, how would they look like on a time frame? So, only if only national causes are vary variations are present the output of the process forms a distribution that is stable over time. So, time measuring it at 9 ‘o’ clock in the morning and then again 10 ‘o’ clock in the morning, 11 ‘o’ clock in the morning; that means, I have readings at one hour interval or it may be may readings at on an interval scale of 7 days whatever it is.

So, if I look at the distribution what I will check now it this diagram is more on a 3 dimensional scale, you are measuring the weights along the x axis which is this one, let me change the color to black. So, I am measuring weight along this direction the frequency or the relative frequency along the so called y direction, and if I am going one on the time frame it would be shifting more towards the back. So, say for example, this gives where the pointer is marking is at 9 ‘o’ clock in the morning, then 10 ‘o’ clock in the morning, 11 ‘o’ clock in the morning 12 ‘o’ clock in the morning, I keep measuring the information. So, if there is no variability and output from process forms a distribution which is stable over time, it will look exactly the like this. Stable means both mean

values both shape both dispersions are not changing and if at all if the dispersions are changing. So, they would I will try to draw on a scale of three dimension.

So, if the variability is changing. So, in one case the distribution like this, if the variability now increases may further increases. So, the mean values always remains at the same point, but the variability is increasing or decreasing. In cases the mean values are changing let me draw it on the same graph using a different color. So, you consider the first one is the black one, if the dispersion remains the same but the vary the central tendency changes. So, it become like in the second time. So, consider the variability remains the same. So, that is what I have tried to drawn using my artistic skills, but if the variability also changes along with the central tendency then it will shift as well as be more expanded and contracted depending on the dispersion increase or decrease.

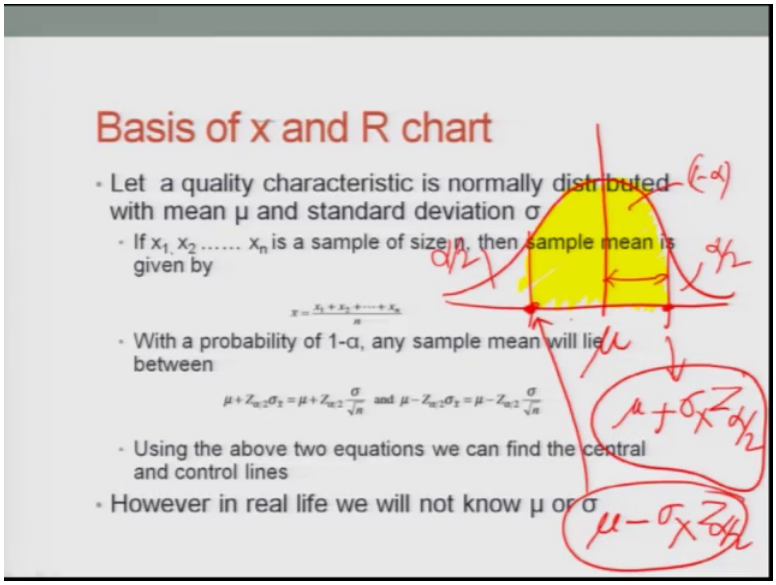
(Refer Slide Time: 14:55)



So, we all are aware of the very well-known normal distribution. So, in the normal distribution what we do is that for any normal case we first convert into a z distribution which is to each standard normal distribution, the mean value is 0 which is marked here. So, again I will use the color as black for ease of representation. So, this is the mean value and the dispersion which you have would be either plus minus 1 sigma on to the right or the left plus minus 2 sigma on to the right or the left plus minus 3 sigma on to the right on our left.

So, if it is plus minus 1 sigma then the overall dispersion is where I am marking the arrow and the total area would be about 67 percentage it is enough value it can be 67 point something we have to check the table. If the dispersion is plus minus 2 sigma then the overall coverage probability would be 95 and if the dispersion is plus minus 3 sigma then the coverage probability is about 99 percentage as shown in this diagram.

(Refer Slide Time: 16:06)



Now, the basis of x bar charts and r charts, x bars would be with respect to the sample mean r would be respect to the range which basically gives an information with respect to the dispersion. Later quality characteristics be normally distributed with the mean value of mu and a standard deviation of sigma from the population, if you are picking up observations. So, what happens is then along the day at 9 'o' clock in the morning you pick up say for example, n number of observations. So, n can be 30, 40, 50 whatever it is. So, these x 1 to x n is a sample size of n and what you want to find out is whether the central tendency or the mean value is changing.

So, consider this example again going back to the example which I gave that jam bottles are coming out from the factory, the actual weight of the jam bottle should be 250 grams and there has been some complaints from the distributors that on an average thus the quantum of jam which is there in in each bottle is less than 250. So obviously, there would be some problem for which this is the occurred and you want to check whether that is due to some assignable causes or un assignable causes. Causes can be both at our

end where either the humidity has change or whether the machine capability has change the machine is not having the same amount of flow of jams which is happening which is being automatically filled into the bottles, maybe the workmanship has change we do not know what it is or there may be some on a assignable causes whatever there we have to find it out.

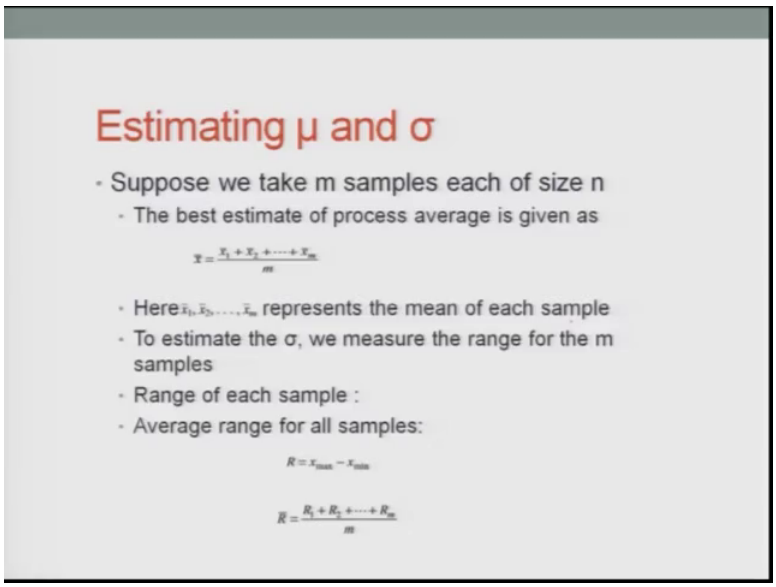
So, what do we do is that we take n number of jam bottles being produced in that factory at 9 'o' clock in the morning. So, those are x one to x n that is what is mentioned and the second bullet point which is if x 1 x 2 till x n is the sample of size n then I want to find out that what is the average size of the jam bottle with respect to the weight. So, that is the sample mean is given by sum of all the weights divided by n. Now if we if with the probability 1 minus alpha any sample mean lies between that range so obviously, that would become like this. So, let me go back to the diagram itself, I am using the red color for better understanding this is the real line where you measure the weights this is the normal distribution this is the mean value.

So, the mean value what we have mentioned is basically the mean value from the sample and you want to compare it with respect to the population. So, what we actually need is these values. So, if this is mu these values are the dispersions which are occurring, hence the right hand limit value and the left hand values would be calculated according to. So, if I go on to the right it is increasing. So, it will be mu plus sigma x which is the standard deviation multiplied by z alpha by 2. The reason is that if the total coverage probability this part, if this total area this value is 1 minus alpha hence these values on the right or the left of these limits would be alpha by 2 and alpha by 2 because is equally dispersed.

So, the z value which you are taking here would be plus z alpha by 2 and the z value which you are taking here for the standard normal deviate would be minus z alpha by 2 minus being because it is on the left hand side. So, now, if you want to find out the probability that any sample mean will lie between that, what we have to do is that find out the limits. So, this right hand would be plus and the left hand which I am circling now would be mu minus sigma x z alpha by 2. So, this goes here. So, based on that you can calculate that; what is the coverage probability and whether that that overall sample observation set of observations which you have taken for x 1 to x n really meets that criteria.

So, using the above two equations we can find out the central and the control line. Now remember one thing in the central and the control line would give you the overall average value and the plus and minus deviations which may happen due to the change of variability which may be unintentional or intentional whatever the case is. However, in real case we know that the actual values of mean and sigmas are not known to me because if they are known; obviously, we would have solved the problems accordingly. So, what do we do? If mean value of the population is not known we take the records or from the sample find out the sample mean estimate and use at the becks proxy. In the similar case if sigma how the population standard deviation is not known we take the help of the sample and do it accordingly.

(Refer Slide Time: 21:47)



So, now first we need to estimate mu and sigma suppose we take m samples. So, what we are doing is that the first chunk we take of n observations that is the first sample again say for example, at 10 ‘o’ clock in the morning we take another chunk of n observations that is the second sample, again in the third time at 11 ‘o’ clock in the morning we take n number of observations that is third sample and if we continue doing it m number of times. So, that is what is written, suppose we take m samples each of size n. So, the best estimate of the process average would be given by this.

So, what we will basically have that in the first part which is 9 ‘o’ clock in the morning will have the first sample average, which would be given by x 1 bar. X 1 bar technically

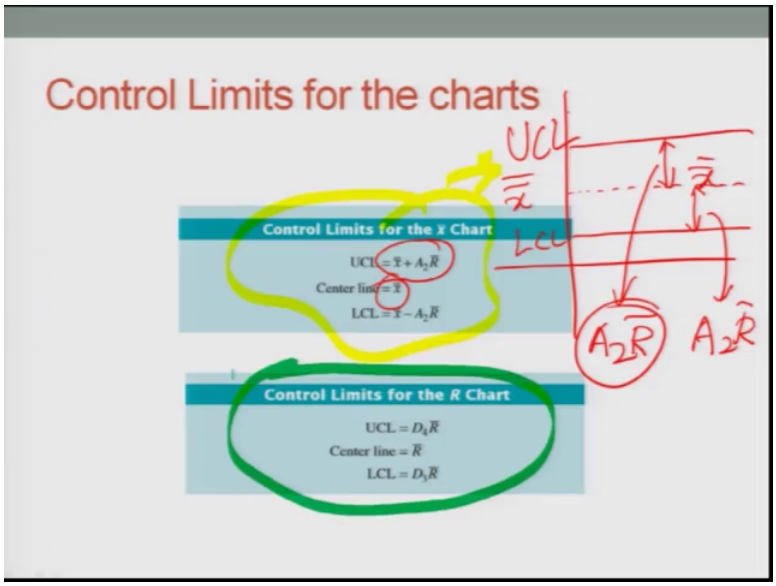
would be x 11 plus dot dot plus x 1 n divided by n. Now you should remember what the suffix 11 is, suffix first one is the sample number and the second suffix is basically the observations. So, if you are picking up the third sample and you have the fourth observation. So, it would be denoted by x 34, similarly you will find on the averages which is x 2 bar x 3bar till the last one which is x m bar. So, adding all of them and divided by m would basically be in the mean of the or the averages or the averages that should be in the long run be exactly equal to the sample mean mu.

So, the best estimate of the process average which is the mu would be given technically by the sample averages that is given by x double bar because you are taking two averages that is why its double bar. So, that is basically sum of x 1 bar plus x 2 bar dot dot till x m bar divided by m. So, hence here x 1 bar, x 2 bar x m bar represents the mean of each sample now to estimate sigma which is the standard deviation we measure the range for the m samples. So, the ranges are given by the formula. So, what we need to find out is the difference between the maximum on the minimum. So, that is what is given here where I am circling.

And once you have the difference between the maximum the minimum will find out the average dispersion using the average range that would be r 1 which is r suffix one would be basically be the dispersion in in for this first sample r 2 bar would be the dispersion for the second sample similarly r m would be the dispersion for the last sample. So, you add them up divide by m. So, what we have is now is the sample average which is the average of the averages and we basically have the average of the range.

So, what we want to do is that consider this average of the averages as the best proxy of this population mean, which is sigma which is mu and considering this range average to be in a way be the best proxy about the dispersion and we continue solving the problem using the graph likewise. So, what we do is now here.

(Refer Slide Time: 25:06)



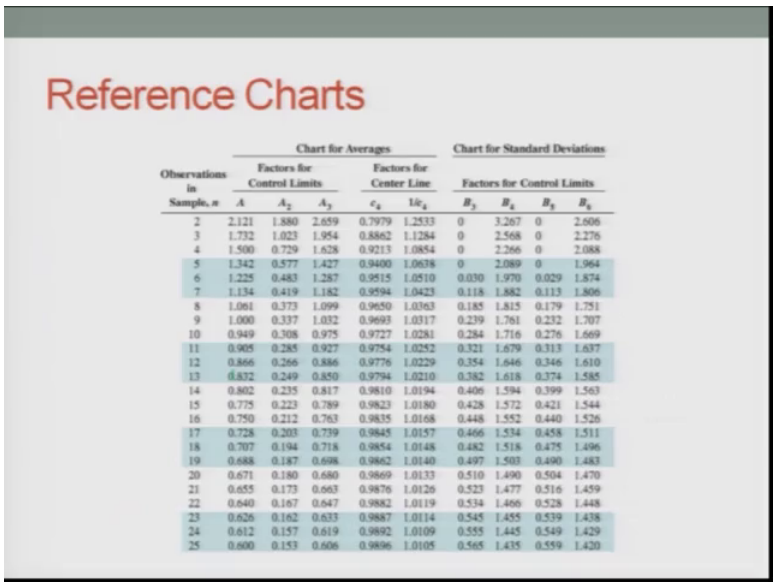
Now, we need to central set up the control limits in the graph. So, the control limits r the central line would be the average of the averages, which would be x double bar where I am circling my electronic pen. The upper control limit would be given by the overall positive movement on to the upper side of x double bar that would be given by x double bar plus A2R bar. So, A2 would be a positive quantum. So, that will basically be plus 1 plus 2 plus 3 depending on the efficiency which we want for our observations.

And the lower control limit would be given by x double bar minus that total quantum of movement which you have taken on the positive side. So, if you consider the graph would come with into a few seconds. So, if you consider the graph as like this. So, what you only have the central value which will be the x bar bar and they would be upper control limit and the lower control limit. So, these upper control limit differences from the central line and the lower control limit differences for the central line are given I am only giving the quantum of the value plus minus would come because it one is more one is less, this would be given by A2R bar this would also be given by A2R bar. So, with one x bar bar you add up this value and go to the upper control limit and I am 2 x bar you subtract the a this is not bar sorry. So, remind mistake. So, this is A2 into R bar. So, that part has to be subtracted from x bar bar and you get the lower control limit.

Now, the first part which I will highlight; so, this set of calculations which are showing in the graph on to the right which I have just drawn using my electronic pointer. So, that

gives the control limit for the x bar charts, and the lower portion which will mark highlight with the different color this is the control limit for the r charts. So, the r charts; obviously, range technically should be 0 so; obviously, the central line technically in the long run should be 0, but they would be some range. So, the central line is given by r bar and the positive and negative fluctuations would be given by D4 R 3 R bar and D3R bar depending on what the values of D 4 and D3 are predefined defined way based on the sample size which you are taking which is n, and also the number of observe and number of such times the sample you are taking which is m.

(Refer Slide Time: 27:55)



Observations in Sample, n	Chart for Averages: Factors for Control Limits A	A_2	A_3	Chart for Averages: Factors for Center Line c_4	$1/c_4$	Chart for Standard Deviations: Factors for Control Limits B_3	B_4	B_5	B_6
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420

So, these reference charts are definitely use. So, in the first column you have the observations which is the samples is n, the factor for the control limits which are for A which is technically for the x bars are given as in the second third and fourth column with the heading of A A2 and A 3, the factors for the central line which would be based on which you will multiply the central line is given by C 4 and the reciprocal of C 4 and the charts for the standard deviations or the ranges are given by the values of B with some suffix which has B3, B 4, B 5 in B 6. So, with this chart you can utilize these concept of the of the process control charts and try to collect observation such that you will get much more meaningful observations and studies can be done based on the information which is being supplied from the sample based on which you will predict about the population, will come to that in more details later on and;

Thank you very much for your attention have a nice day, bye.