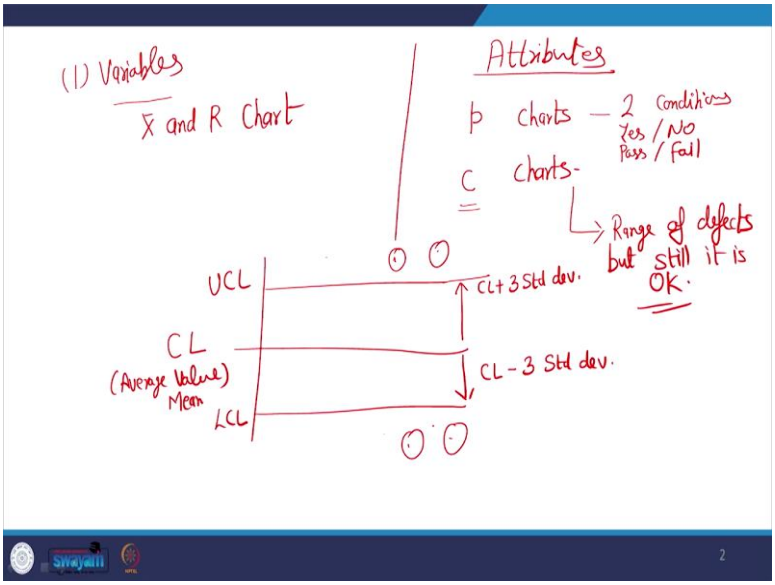
Production and Operation Management
Professor Rajat Agarwal
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture 42
Statistical Concepts in Quality Control- 4 (Run Test and Examples)

Welcome friends, in our last few sessions, we are discussing about quality control, particularly statistical quality control, when we are discussing statistical quality control, we discussed a diagram that either you can have acceptance sampling, which is more about appraisal and inspection or you can have process quality control.

In process quality control, we are trying to determine the quality during the production stage itself, so that we have minimum cost of quality, we do not add value to the defective pieces. And at the same time, we try to keep our processes within the limits of natural variation. And for that purpose, we discuss that quality control charts are the most used techniques.

(Refer Slide Time: 1:25)



And we discussed particularly two types of quality control charts, one charts which are used for the purpose of variables and then charts for attributes. For variables, we discussed about X bar and R chart and we discussed that depending upon how many variables we want to check in our process, as many number of X bar charts are required. For attributes, we discuss 2 types of chart p -charts and c –charts, p -charts we discussed that we are taking a sample and out of sample either a product is okay or defective, okay or defective.

So, you have only 2 conditions yes/no, pass/fail. So only these 2 conditions are there, but it is also possible that when you are passing you may pass with different grades, someone is passing with the 60 percent marks, another is passing with 50 percent mark and another is passing with 40 percent marks and someone is passing with distinction. So, there are possibilities that there is a product which is okay, but some defects are there, so it may fall in lower category of grades and a product is passing there are no defect is there, so it is passing with a higher grade.

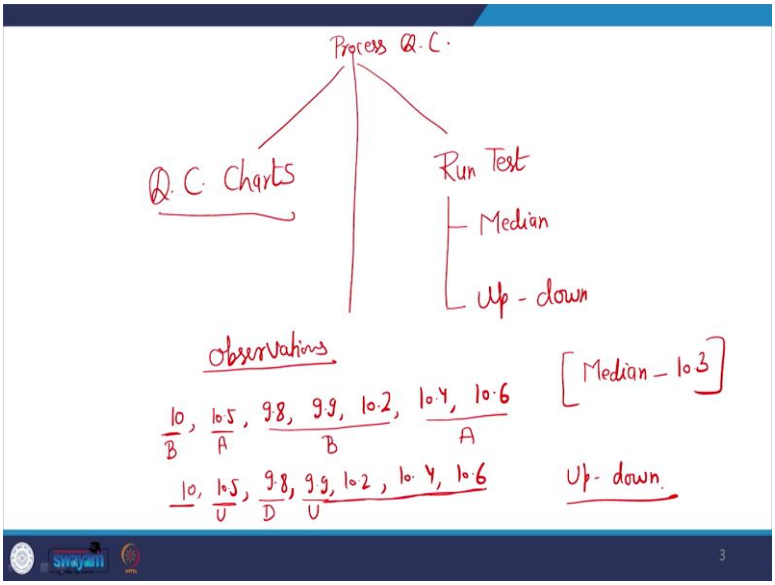
So when this type of system is there, that range of defects but is still it is okay, in that case, we go with c -charts that how many number of defects per product per unit are there on an average and there is a maximum limit and lower limit. If number of defects per piece is increasing beyond the UCL, then we stop the process, so like you are purchasing a wooden block in that wooden block, there may be some knots available in that, some wooden defects are there.

So maybe in a wooden piece of 2 by 2 square feet, 3 knots are acceptable to you, because it cannot be without knots, it can be without it cannot be without some defects. So, 3 you are accepting, but if the number of defects increase to let us say 6 or 7, maybe and that case, you will say that it is not going to be acceptable, but if number of defects are 55 is still it is acceptable, but then you will not pay that much price for that.

So, when you are categorizing on the basis of number of defects, in that case c -chart is used. So, we have already discussed about X bar R charts, p -charts and c -charts. And in all these charts, you remember that we had this type of representation where we have a central line, the average value, the central line means average value or you can say the mean value, then the upper control limit and the lower control limit and between upper control limit you have CL plus 3 standard deviation and then for lower control limit, it is CL minus 3 standard deviation.

So with this if your, 3 plus minus plus minus 3 standard division are considered to be natural variation. So if your process if your end products are within the variation of natural limits, then you are going to accept it beyond this if some points are here, these are unacceptable points. So, all that we have already discussed with the help of examples in our previous sessions.

(Refer Slide Time: 6:13)



Now, there is another important thing that this is about process quality control and in process quality control, we discussed the use of quality control charts. But, one more important type of process quality control issue is there, whether your parameters whether the observations are showing any kind of trend or not.

So, for that purpose we apply run test, which is basically to determine some time with our naked eye, we will not be able to understand whether trend is there or not, this you remember in forecasting also we discussed that many a times simply by calculating the forecasting error, you will not be able to determine whether you are having any trend or not.

And for that purpose, we calculated tracking signal with the help of tracking signal, we are actually able to say whether trend is there or not. Similar to that, in this quality control, we have run test, where using the data of run test using the profile of run test, we will determine that some trend is there or not there.

Now, 2 types of run tests are there which popularly we use, one is based on median values and another is based on up and down values. So, these 2 types of run tests are there, median value is that you have some observations, let us say I am taking the observation of weight of a particular product and the weight of the products which I am observing like 10 grams, 10.5 grams, 9.8 grams, 9.9 grams, 10.2 grams, 10.4 grams, 10.6 gram and so on, these are the different observations for weight of a particular product.

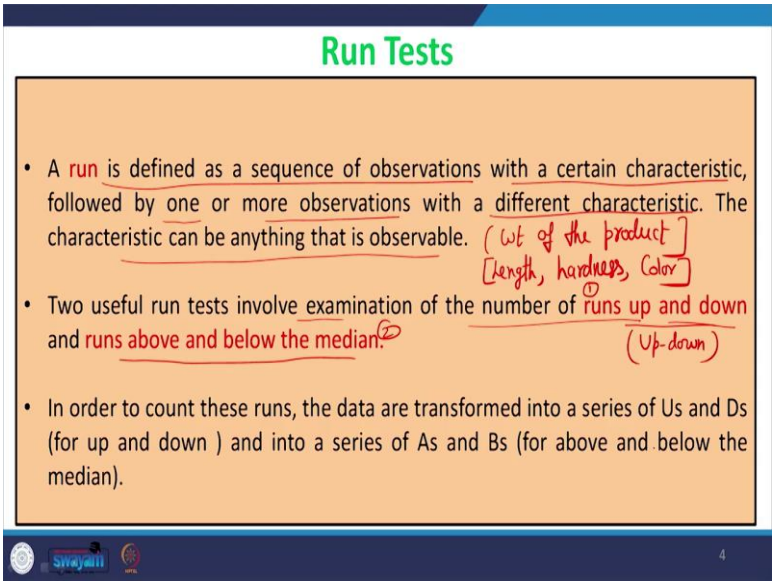
Now, the median weight for this particular product is 10.3. Now, there will be few observations which are above 2 median weight and there are a few observations which are

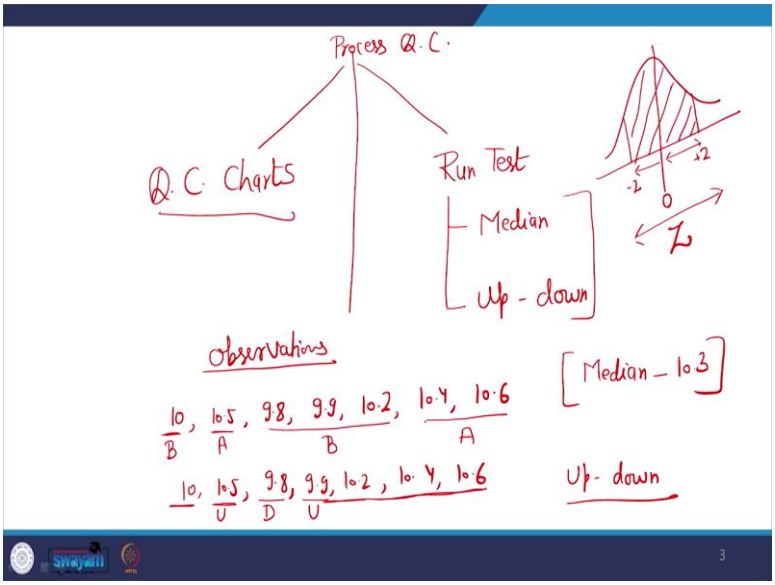
below this median weight and based on that I will decide, how many runs are there for an example, if you see this is below the median, these this is above, this is below all these are continuously below and all these are above.

So, that is how I will decide that how many runs are there. Now, we will do some statistical processing of this data to understand whether these runs are showing any kind of trend or not. Similarly, up and down is also there, again take the same data. Now, this is the comparison from the previous data, so there will not be any kind of notification for the first observation, but the second observation is compared with the previous observation.

So 10.5 is more than 10, so therefore, this is on the upside, but 9.8 is less than 10.5, so it is down then 9.9 is above to 9.8 it is up, then 10.2 is, so this is also up, 10.4 is also higher than 10.2 and 10.6 is also higher than 10.4. So, all these are under same line of you. So, these are the different runs with respect to up and down system. So, we have these two types of observations, which are known as run. Now, how do we do calculations with the help of these run data, so that we can determine whether any trend is there or not, that we will see now.

(Refer Slide Time: 11:15)





So a specific definition of run, so run is defined as a sequence of observations with a certain characteristic. So like in this previous numerical values, we have some characteristic of weight, where either they are continuously above to mean value median value or continuously lower to median value. So that is a type of observation with a certain characteristic followed by one or more observations with a different characteristic, the characteristics can be anything that is observable. So like in the previous example, the characteristic was weight of the product.

Similarly, you can take many other examples, length can be there, hardness can be there, color can be there. So there may be multiple type of characteristics, which you can observe and that can become your characteristics for deciding the run. Now two useful run tests involve examination of the number of runs up and down, so that is up down test and then this is runs above and below the median. So two types of systems are there, one is up and down second is median.

Now, in order to count these runs, the data are transformed into a series of Us and Ds for the up and down we have already seen that add into a series of As and Bs for above and below the median. So, how we are going to arrays that we will see with the help of some numerical data.

Then, once we have determined these number of As that how many above to median and how many below to median runs are there, how many U runs and how many D runs are there, so whatever type of system you are following, you have to determine that how many As, how many Bs, how many Us and how many Ds are there.

One more important thing is there like our X bar R chart, we also need to do both these tests simultaneously. It is quite possible that in one of the test, your data is coming within the range, but in the other test data may not come within the range. So, even if data is coming within the range for one test, and if it is not coming in the range for the other test, we will take the interpretation or we will finally conclude that data is going for a particular kind of trend. So it has to follow means data has to strictly come within the limits for both the test that is very important thing, so both median and up and down test normally we do simultaneously.

So now, this is about the calculation of number of runs for median test or for up and down test. After that, we need to determine that these are the actual numbers which we have just determined for that particular example. But what should be the expected number, what should be the expected number.

If we are close to expected number, it means there is no trend, but as we move away from the expected number, chances of some kind of trend comes into the existence. So now the second part of the calculation is to determine the expected number because this whole idea of run test, basically follow the normal distribution.

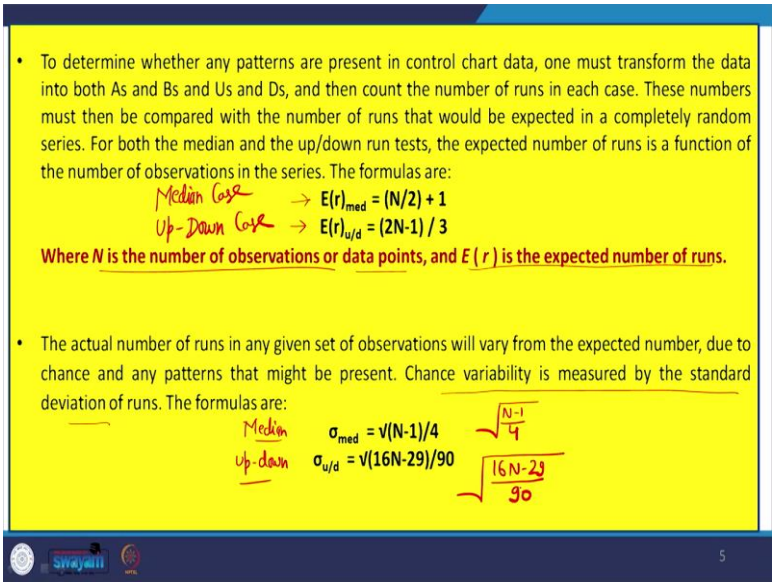
So, your expected number and the difference in the actual number will give you some Z value. And now, if that Z value is within your expected limits, let us see, if I take a Z value within plus minus 2, this is plus 2 Z, this is minus 2 Z. So if the Z value, which is calculated on the basis of difference between the actual values of R run, Rs run, and the expected value of R that is the expected R, if those values give you the Z value, which is between minus 2 to plus 2 that will say that our data is not showing any kind of trend, it is expected, it is acceptable data, if the observed trend observed number of run and the expected number of run are equal, in that case, the value of Z will come 0.

And then it means you are here and that means, that will mean that your number of runs are exactly matching with the theoretical ones, but that is a very rare situation, but just for the sake of our understanding that what we are going to do, I am giving you this part before it actually going to come in our session.

So, we will determine the this is the Z value. So we are going to determine the Z value on the basis of observations and expectations, and that that value should fall within our pre specified limits, so in some cases you may have minus 2 to plus 2 or you may have minus 3 to plus 3,

so it is up to you that what is that predefined limits of acceptance and on the basis of that, how much trend how much variations you are able to accept.

(Refer Slide Time: 17:36)



So, for determining the values of expected run for the median case, this is for the median case and this is for the up and down, let me right here, this is median case; this is up, down case. So, E(r), E(r) represents expected values of runs. In case of median it is N by 2 plus 1, in case of up and down, it is to N minus 1 divided by 3. So these are the number of expected runs in a particular situation.

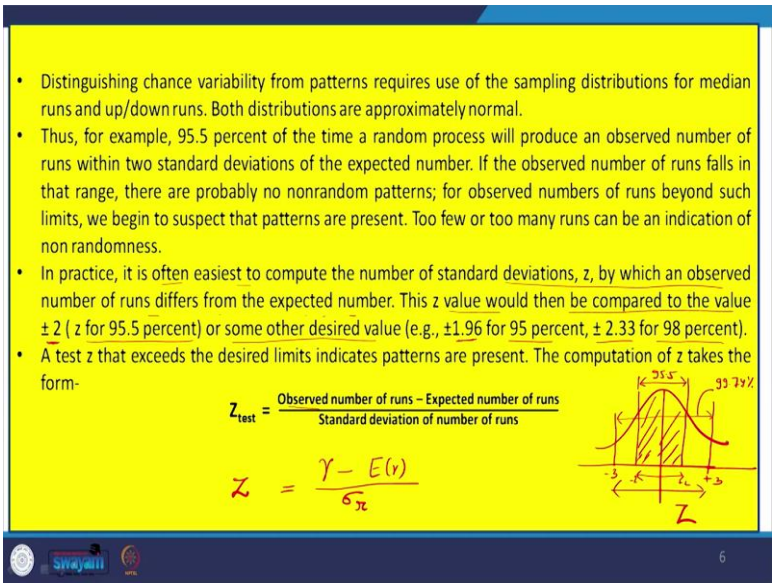
So, N is the number of total observations you are going to have, N is the number of observations or data points and E(r) is representing the expected number of runs, then, we have to determine, because, whatever variations are happening because of chance, we are going to accept that, but whatever variations are happening, because of some assignable reasons, we are not willing to accept them, that is the underlying assumption of entire statistical quality control, all random things are acceptable, but all non random things are not acceptable.

So random sometime it is mentioned as chance, sometime it is mentioned by accidents. So all these are the names for the randomness, but wherever cause is possible, wherever assignment is possible, that is not random, because you know the reason of that particular thing and therefore, you have to remove the reason. And chance variability, that is the randomness is measured with the help of a standard deviation of runs and for that purpose again for the

median case and then this is for up and down case, we will calculate the standard deviation of runs.

So in this case, it is under root for better clarity let me write like that. And for this it is like 16 N minus 29 divided by 90. So, these are the formula for determining the standard deviation of runs for median and up and down case respectively.

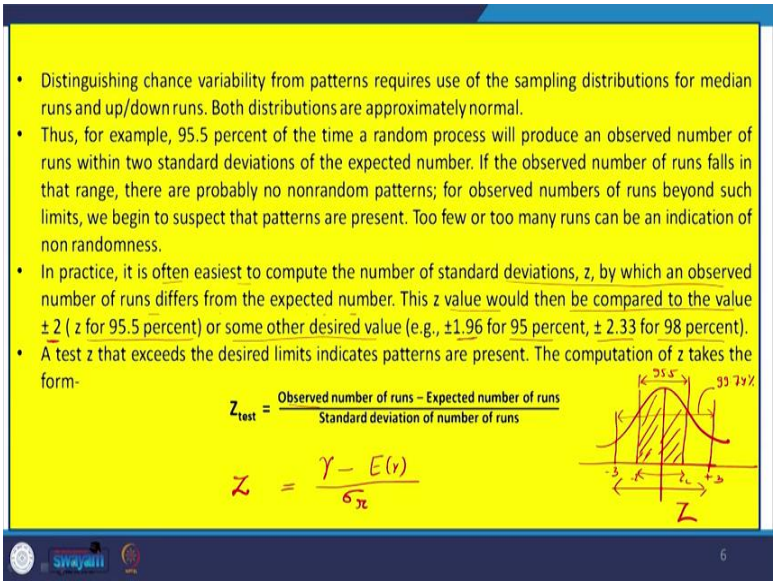
(Refer Slide Time: 20:08)

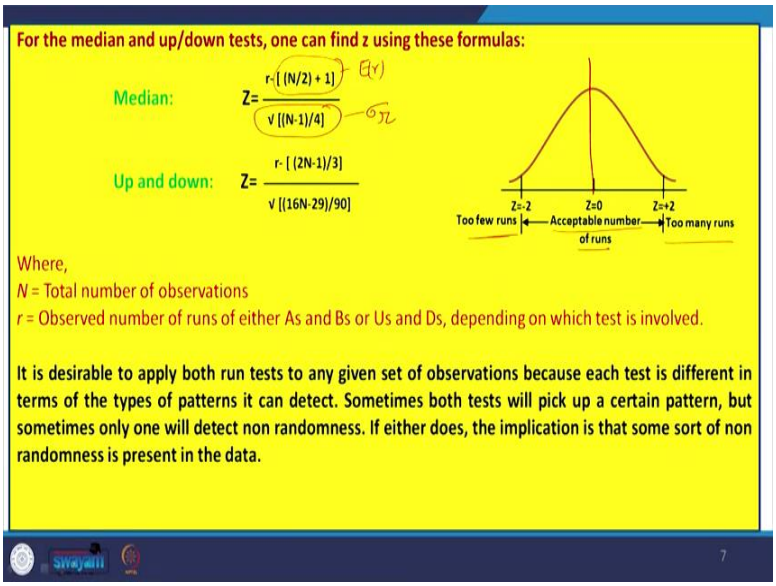


Then, since we are going to apply a Z test for because we have assumed that it is going to behave in a normal distribution curve and for knowing the value of our variation, we will actually normalize the entire variation on this Z scale. And therefore, we are interested in determining the Z value for a particular test.

So, Z value for a particular test is observed from number of runs, that is the actual number of runs which we have calculated by that sample data, the expected number of runs which is coming from the formula in the previous slide, we discuss these formulas. So, that is the expected number of runs and in the denominator, these standard deviation comes will come.

(Refer Slide Time: 20:58)





So observed number of runs that is r minus E(r) that is the expected number of runs divided by the standard deviation of r. So, that becomes your formula for the calculation of Z value. And if you want to have a very specific formula for these, then for the median, this is r and this term is actually E(r) and this term is sigma r. Similarly, for up and down r minus E(r) divided by a sigma r. So, this is the final formula for the calculation of Z.

Now if you read this theory, it says that in practice, it is often easiest to compute the number of standard deviations Z by which an observed number of runs differs from the expected number. This Z value would then be compared to the value like plus minus 2 that is Z for 95.5 percent or some other desired value like plus minus 1.96 for 95 percent or plus minus 2.33 for 98 percent.

So depending upon like, if I want to include variations up to 99.74 if I want to include variations up to 99.74, I may take the value of Z plus minus 3. So depending upon how much variations I am including in my discussion, I will take a value of Z because these area these area under this curve it may be from here to here minus 2 plus 2, then from here to here, it can be minus 3 plus 3, when it is minus 2 to plus 2, it is taking care of 95.5 percent area under this curve.

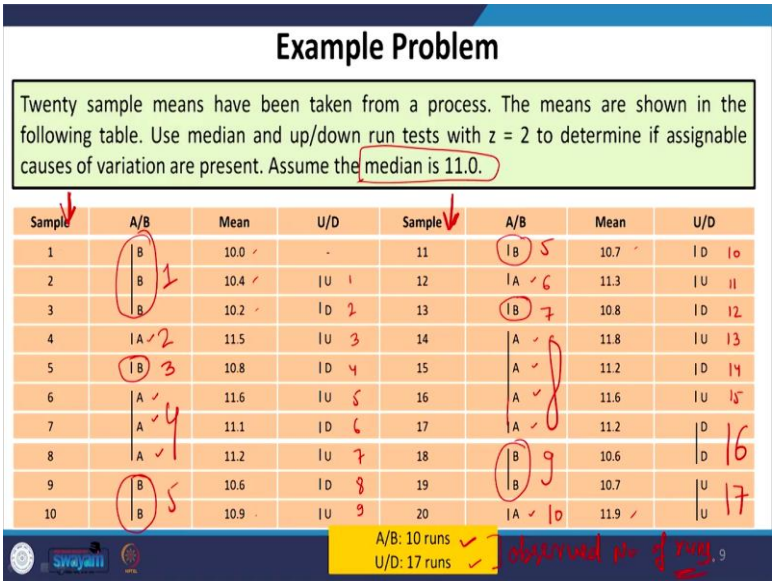
When it is from minus 3 to plus 3, it is taking care of 99.74 percent of area. So, how much variations you are ready to accept in your this distribution on the basis of that you are going to have a desired value of Z and that value of Z will decide whether the trend is there or not. So, now we have calculated the values of Z for two tests, median and up and down test and that is going to have this kind of normal distribution, where this is the central value of that is the mean value where Z is 0, in case observed number of N and the expected number of runs are equal, then the value of Z will come 0.

But we have already understood now, that there will be some difference between the actual number and the expected number. And therefore, there will be some difference. So how much difference is acceptable, so that is written as acceptable number of runs, that how many difference of runs are possible and this means, if you are coming to this side Z is coming around minus 2 that means, there are 2 less number of runs and if it is coming towards this side, that means there are too many numbers of runs.

So neither we want too many numbers, nor we want too less numbers, we want a kind of balance between the number of runs. If too many runs are there, that means a highly randomized thing, there is no stability in the process and if it is very less number of runs that means, there is going to be a particular kind of trend, because all values are either As or all values are either Bs all values are either Us or all values are either Ds.

So, we want a kind of mixture of these things, so this is the interpretation of this particular graph. Now, we understand that what is the median test and up and down test.

(Refer Slide Time: 25:56)



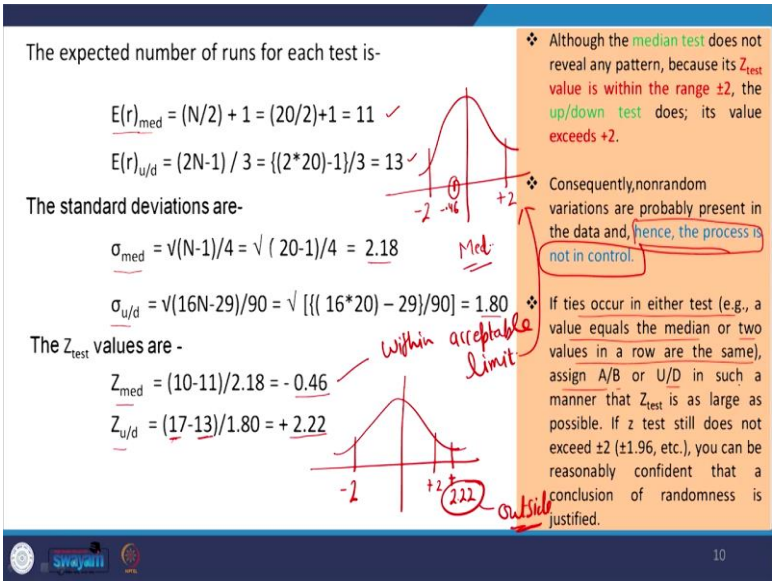
Sample	A/B	Mean	U/D	Sample	A/B	Mean	U/D
1	B	10.0	-	11	B	10.7	D
2	B	10.4	U	12	A	11.3	U
3	B	10.2	D	13	B	10.8	D
4	A	11.5	U	14	A	11.8	U
5	B	10.8	D	15	A	11.2	D
6	A	11.6	U	16	A	11.6	U
7	A	11.1	D	17	A	11.2	D
8	A	11.2	U	18	B	10.6	D
9	B	10.6	D	19	B	10.7	U
10	B	10.9	U	20	A	11.9	U

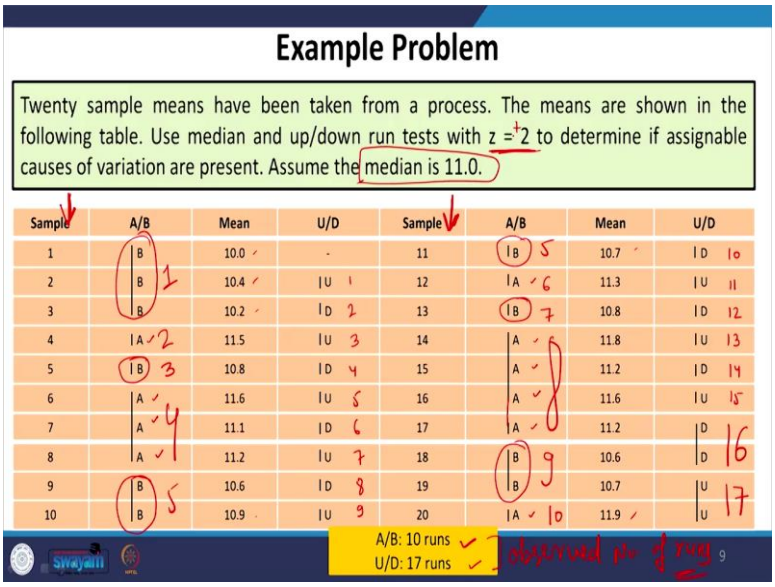
Now let us apply this, on some numerical data. And for that purpose we have a data of 20 sample means, now these are the different sample, 1 to 10 and 11 to 20, so these are the sample, and in that sample, the values are 10, 10.4 like that these values are up to 10.7 to 11.9. Now, in this particular case, the median value is given as 11 median value is given as 11. So, you first see, on the basis of this, that 10 is less than 11, 10.4 is less than 11, 10.2 is less than 11, so all these are Bs then 11.5 is more than 11, so it is A above 10.8 is less than 11 it is below that is B, 11.6, 11.1, 11.2 all these are above A above median.

Then similarly, these are below, it is also below 11.3 is more than that it is above 10.8 is less than 11.8, 11.2, 11.6 all these are more than median A again it is more than median and then 10.6, 10.7, it is less, so these are 1 run and 11.9, it is again higher than that. So, let us see how many A and B runs are there 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 let me join this line because this is one single run, so that there is not a confusion, so this is 1, 2, 3, 4 this is 5, this is 6, this is 7, this is 8, this is 9 and this is 10. So you have AB runs, which are equal to 10.

And similarly now, we will calculate U and D runs for first there is no mentioned because we have to take this as our starting point. So, there is no U and D value for the first but 10.4 is more than 10. So, here U will come 10.2 is less than 10.4, so D will come, 11.5 is more than 10.2, U will come 10.8 is less than 11.5, so D will come and so on we have U and D values. Now, you see, it is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17. So total 17 U and D runs are there. So, these are the observed number of runs.

(Refer Slide Time: 29:48)





Sample	A/B	Mean	U/D	Sample	A/B	Mean	U/D
1	B	10.0	-	11	B	10.7	D
2	B	10.4	U	12	A	11.3	U
3	B	10.2	D	13	B	10.8	D
4	A	11.5	U	14	A	11.8	U
5	B	10.8	D	15	A	11.2	D
6	A	11.6	U	16	A	11.6	U
7	A	11.1	D	17	A	11.2	D
8	A	11.2	U	18	B	10.6	D
9	B	10.6	D	19	B	10.7	U
10	B	10.9	U	20	A	11.9	U

And now with the help of those formula, we will calculate the expected number of runs. So, the expected number of runs with the help of formula for the median case N by 2 plus 1, so 20 observations divided by 2 plus 1 that is the 11 runs. And for up and down using the formula, which is already discussed 30, so these are the expected number of runs for different types of test.

Now, we will calculate the median values again with the help of the formula which we have discussed. So median value see, standard value is standard deviation for median case is coming 2.18 18 standard deviation for up and down is coming 1.80 and then using these values, the Z values for median is our observed value that is 10 minus 11 divided by a

standard deviation. So, it is coming minus 0.46 and for up and down the 17 is the observed value 13 is the expected value divided by a standard deviation 1.80 so it is coming plus 2.22.

Now let us see, the plotting of these Z values on our x axis and we need to have 2 curves here. So one curve because our expected values should have within plus minus 2. The question says that Z should be plus minus 2, Z value is plus minus 2. And here we have 2 curves of Z, plus minus 2. Now, in one curve which is let us say for the median, the value of Z is coming, minus 0.46 somewhere here, this is minus 0.46.

So you can understand this is within acceptable limit as evident here, but this plus 2.22 it will come somewhere here, this point now this is outside of acceptable limit. So, since this point is outside the acceptable limit, we are not going to accept this particular point because it is beyond plus 2 value.

So there are too many observations too many runs, because we want a balance between few runs and many runs. And since, this 2.22 is towards the positive side, it means that there are many runs in case of U and D, up and down and actually 17 runs are there, 30s or expected runs, but we have actually the 17 runs, a standard deviation of 1.80 is acceptable means 13, 1.80 you may have at a maximum up to 16 runs that will give you the value of Z within plus 2, but 17 runs are giving a too much fluctuation in your data point. So, here we can say that this particular process is not in control.

So, we have concluded that this particular process is not in control and we have to therefore, decide that we are we need to re-check this process because it is going beyond the acceptable limit some time if ties occur in either test, for example, a value equals the median, or 2 values in a row are the same, then assign A and B or U and D in such a case in a manner that Z test is as large as possible.

If Z test is still does not exceed plus minus 2, you can be reasonably confident that the conclusion of randomness is justified. Sometime it is possible that in this particular example, we did not we did not have that kind of data, where the value is exactly on the mean value or two continuous values are same.

So, whether you are going to take them as a U or as D. So, in that particular case, we should be taking the values of those runs in such a manner that it should actually help in increasing the value of Z, so that you can be doubly sure you means on your own, you should be able to justify that there is enough randomness in your sample if that is being justified then you can

take that particular process as a process, which is giving you statistically correct results. So with this, we come to the end of this run test discussion. And with this, we are able to complete the discussions related to various process control in the quality control using statistics. Thank you very much.