# Introduction to Reliability Engineering

**Professor Neeraj Kumar Goyal**

**Subir Chowdhury School of Quality and Reliability**

**Indian Institutes of Technology, Kharagpur**

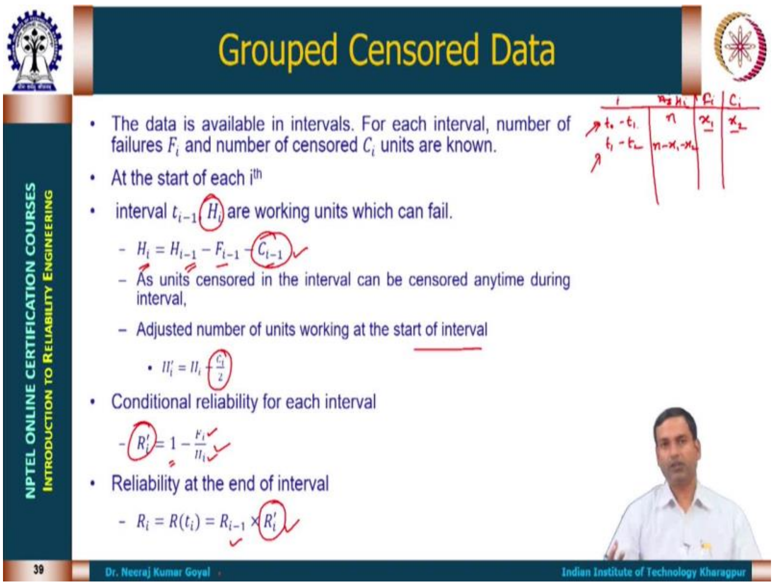
**Lecture: 29**

**Failure Data Analysis: Non- Parametric Approach (Contd.)**

Hello everyone, so we will continue our discussion about failure data analysis using non-parametric approach. In earlier classes we discussed about the failure data, which is generally time to failure data or which we have the group or interval data, then we discuss various methods that when we have the complete data how to analyze if we have the singly censored data or multiply censored data how to analyze.

One category which was the meaning was that if the data is grouped and data is multiplies censored and also we will discuss one more type of data that is static data where we have the fixed interval and for which we want to know the reliability.

(Refer Slide Time: 01:09)




So, whenever we say we are talking about group censored data that means, we have the intervals we have t 0 to t 1 then t 1 to t 2 and each interval we would like to know how much is a failure n i. So, rather than that, as we took earlier we can take the n s, n s mean so, at the start of interval how much for the unit so, working which we are writing as the H i here. So, if you write this H i and this is i.

So, for first interval that means, at t 0 how many units were working, so, that will be the initial number of units which you put on the test. So, that may be n, then how whatever failure have now we have the two categories here, we may have the failure here in this interval and we may have some censored unit. So, number of failures let us say is x, x 1 and number of censored let us say x 2 which we are writing as F and C.

Then what will happen for this second interval this will become n minus x 1 minus x 2, but what will happen number of units which have failed or censored will not be available to working for the next interval. So, what will happen at the start of each interval we will have reduction in unit because of failure or suspension this is what we are calling as the number of units which are working at the start of interval H i.

So, these will be H i minus 1 and from H i minus 1 we are reducing the number of failures in censored unit. Now, what happens that the censored units which you have removed from here, they might have been removed from anywhere in the interval they might have been removed from the start of the interval or they might have been removed at the end of the interval.

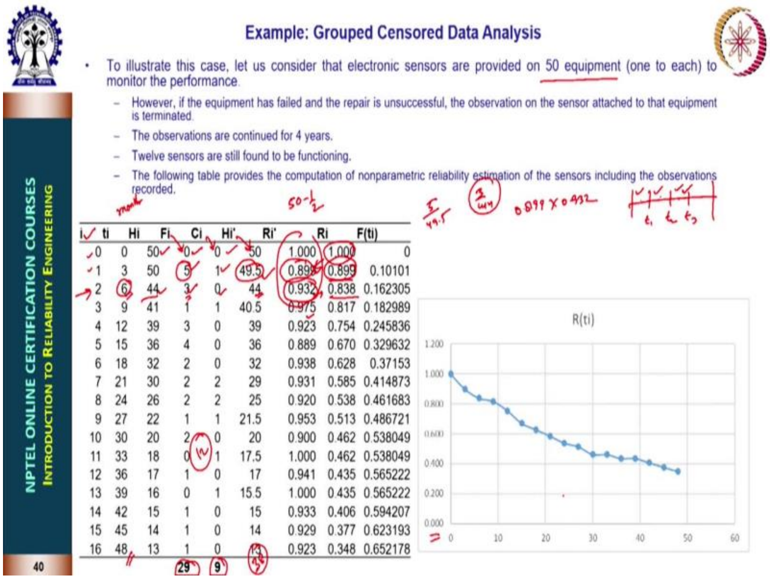
So, on an average the number of units which are working at the start of interval we try to consider that we have removed rather than c i minus 1 we will consider that all the i minus half of the half of the time they have been on an average working. So, number of units removed effectively we are taking a C i by 2 this is one type of approach another kind of approaches where we use we simply used our mean ranking method and we do not we use the same formula that you have considered here.

So, the next interval number of units removed will be the similar way. So, here the reliability and we are calculating this is nothing but 1 minus number of failures divided by number of units which are working at the start of the interval. So, this becomes our interval reliability that is the conditional reliability. So, it is the reliability for that particular interval only. So, here what is the probability that let us say first interval we are talking about, then it is giving the probability that in that interval, there is no failure or what is the probability of successful passing this particular interval.

So, if you want to know the continuous to have like if I want to know the reliability after 2 intervals, so, for that I have to multiply the reliability first interval and the second interval like we did the same in the case of ungrouped data censored data multiply censored data whenever we considered the found whenever failure happened we found our interval reliability and we multiplied that with the earlier known reliability, and that is how we considered it, so, same concept is applied here.

That whenever we are calculating here we are calculating the conditional reliability here for each interval and for current reliability, we have to multiply reliability for all earlier intervals, then only we get the reliability current interval. So the reliability at the end of each interval will be previous interval reliability multiply with the reliability at the end of previous interval multiply with the reliability or reliability of this interval, once we do that we will get the reliability of the at the end of the interval.

(Refer Slide Time: 05:10)




So, let us take one example here all these examples which I am presenting here, these all we have put up in using the Excel sheet also. So, those you may be able we may be able to share and you will be able to find those little later. Let us see how we have done it. So, like here i is the interval at number, so, that is 0 1. So, when I say 0 1 2 3, so, this becomes our interval number.

So, at the start how many units are working 50 units are working and so, here like in first interval that is from 0 to 0 here we are considering them no censoring no failure here. So, 0 failure and 0 censored this actually this values are right aligned. So, you have to read it according to this and H i dashes 50. So, here now, after one interval what will happen in this 0 to 1 interval we have the 5 failures and we have one censored.

So, because of this what will happen effective number of units which we had because of the loss of censored unit that will become 49.5 because 1 divided by 2 will be half. So, that will be 50 minus half. So, effectively we have 49.5 units working at this start. So, here for working for that interval.

Now, out of the 49.5, how many failures we have 5 failures. So, our reliability comes out to be for this interval is 5 divided by 49.5 this gives us this value. So, this because our earlier reliability was 1 if we multiply 0.899 and with this we get that 0.899 and here and failure probability will be nothing but 1 minus reliability.

Similarly, now, let us see the second interval for second interval that is from 1 to 2, we have past three more here like this is for four years. So, after four years means this is a month, so, 4 into 12 is 48 months. So, after 6 months, so, that means first 3 months, we got this 5 failure 1 suspension from 3 to 6 months, that means in the second quarter we got we had 44 units working because 5 were failed and 1 was removed the remaining units are only 44, 50 minus 6 is 44 out of 44 3 failed and 1 is, 0 is censored.

So, that means since no censoring is done, so 44 remain eligible for the complete period. So, out of 44 failure is 3. So, therefore, this interval how much is the reliability that is 3 out of 44 So, that becomes 3 divided by 44, 3 divided by 44 is 0.932. Now 0.932 is the reliability for second interval but for equipment to survive two intervals that means equipment to work up to 6 months, the reliability will become the equipment should be working for first interval as well as the second interval.

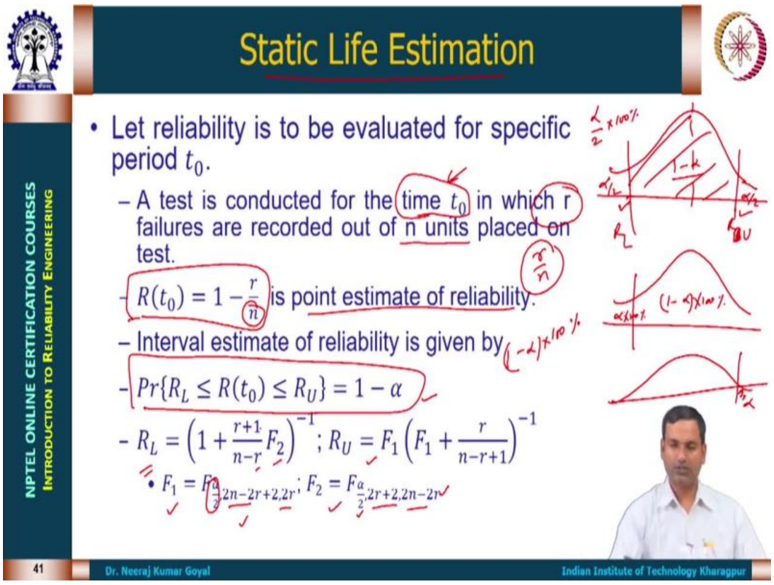
So, the reliability will become 0.899 into 0.932 this becomes our new reliability. So, after 2 intervals reliability is 0.838, same way we calculate for other intervals first interval reliability calculate then we multiply with the previous interval and reliability because as we know for equipment to survive, let us say this is t 1 this is t 2 this is t 3.

So, for equipment to continue working here it has to work here work here and work here. So, it has to be reliable here and reliable here and reliable here. So the reliability gets multiplied for later interval all earlier interval reliability has to be multiplied that means if we multiply all these, we will get this value.

Same way we have done this I have done this in Excel maybe we can share later. As you see here, we actually had done 50 equipment's working out of 50 29 are failed and 9 are censored. So that means 38 units, we have still 12 units of working here which are censored here, but they are right censored value. So which do not affect here, but what happens that we do not reach to the reliability new to 0 here about reliability curve. We are able to draw only up to the whatever values you are observing.

So but if we want to draw this curve further, either we can do the trend and we can draw this further by the curve fitting etcetera, but rather than doing that, we should be finding out the appropriate distribution for this data and we should use the distribution analysis so, that we are able to extend this curve and we know the full reliability distribution for the whole lifetime.

(Refer Slide Time: 10:20)




There is one type of the more type of data which we may be discussing that is static life estimation. So, this type of estimation is required for the equipment's like many times one shot devices are there are many times they work for a very short period of time. So, what happens here whenever we conduct the test, so, the test is conducted for a mission time.

This can be applicable for large operating devices also, but for large operating devices, the reliability is considered to be function of time. So, we would like to know, reliability at for a particular duration of the time, but here, we are more interested in let us say that short period devices are there, which work for a certain time t 0. So, this is kind of mission time. So, it is a mission that which it has to complete and that time for mission is t 0.

So, here what we will do, we will have multiple devices tested for this time, all devices and we do not know anywhere in between whatever failure happened or success happened, we come to know only at the end of the time whether that failure happened or this success happened. So, here when we put n units on the test, let us see r failures we observed.

So, for t 0 time at the start we put n units and at the time t 0 when tests completed, we found that r units have failed in that case of scenario, how can we find out the life or how can we find out the reliability. So, in that case reliability as we know. So, obviously, if we have the r failures out of n our unreliability's r divided by n, that is equal rank method and whatever we use so, all general sense r upon n will give you the unreliability.

So, here the reason is that the reliability will become 1 minus r divided by n. So, this becomes the point estimate of reliability. But point estimate of reliability can be misguiding here, the reason being that point has to be like one failure happened. So, one failure happened, maybe here we use, let us say 5 devices out of 5, 1 failure happened, but maybe we use 10 out of that also 1 failure happened.

So, this estimate of reliability is dependent on the value of n how many samples we are using. So, when samples are less, this estimate can be biased depending on the outcome, because the sample which are chosen, if there are some equipment's, fault equipment's have come, so, it will make our reliability estimate bias.

So, it is useful if in such cases we are not only estimating reliability, but we are also estimating the lower bound on reliability and upper bound on reliability. So, we construct a confidence interval here. So, when we say 1 minus alpha percent into 100 percent confidence interval. So, essentially confidence interval concept is like this that we if we assume normal distribution, then we have the mean value here.

So, when we say confidence interval, then we have the lower bound here and we have the upper bound here, upper bound, so, when we have 1 minus alpha plus alpha here, so, probability of value falling in this region is 1 minus alpha. So, that means alpha by 2 belongs to this reason alpha by 2 belongs to this reason. So, we want to make sure find out the value this value this value if let us say this value will be representing to the alpha by 2 into 100 percent if we say alpha by 2 into 100 so, this will be alpha by 2 into 100 percentile.

So, we can say this, so, we can say 50 alpha percentile value while this one will be 100 minus alpha by 2. So, that will be because only alpha by 2 is covered here. So, that is right said is alpha by 2, so, this will become 100 minus alpha by 2. So, if my alpha is let us say, 0.1 that means 90 percent confidence interval I am talking about then 5 percent to 5 percent here, so, this becomes fifth percentile, this becomes 95th percentile value. So, this percentile value which we are getting, we can use it.

So, here the for this kind of scenario F distribution comes into the picture. So, this is statistically given that R L value can be evaluated as 1 plus r plus 1 divided by n minus 1 into

F 2 inverse whole inverse so, that is 1 divided by 1 plus r plus 1 upon n minus 1 r into F 2. Similarly, upper reliability estimate can be achieved by F 1 into F 1 plus r upon n minus r plus 1 inverse that we can write as F 1 divided by F 1 plus r upon n minus r plus 1.

So, this R U and R L gives us the estimate this will be R L this will be R U. Most of the time while F 1 is and F 2 we have to get from the F distribution, F distribution generally has the 2 parameters one is this 2 n minus 2r. So, parameter one is this second parameter is this and the probability for which we have to calculate is alpha by 2.

Similarly, for alpha by 2 then we have the degree of freedom 2 degrees of freedom which we use here for this second F 2 case the degrees of freedom is 2 r plus 2 and 2 n minus 2r and 2n minus 2r. So, for these degrees of freedom we can get this F 1 and F 2 and we can use in this formula and once we use this the in this formula we get the lower estimate and upper estimate on reliability.

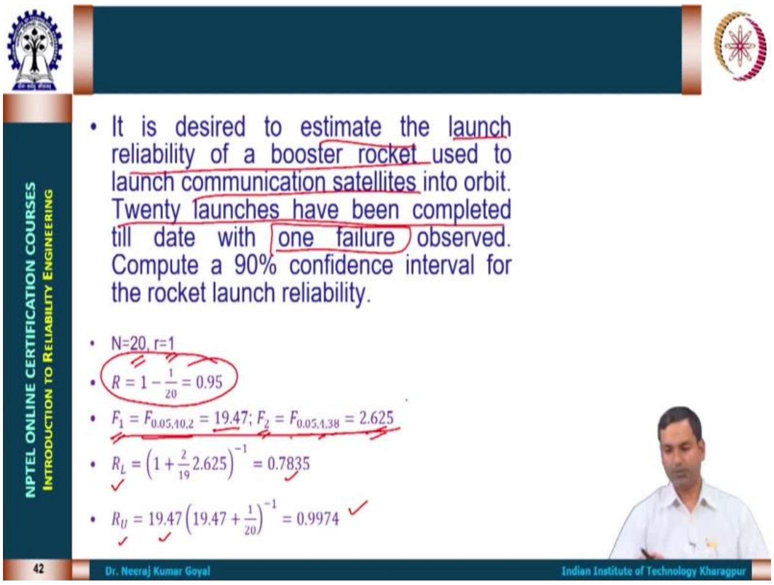
So, we get the range of reliability in which we expect that the average reliability will fall for the given 1 minus alpha into 100 percent confidence. So, here generally for the reliability we are more interested in lower estimate, a lower bound, because we want to know how much minimum reliability we are going to achieve. So, generally, whenever we are talking about good things, positive things, we want at least how much we will get.

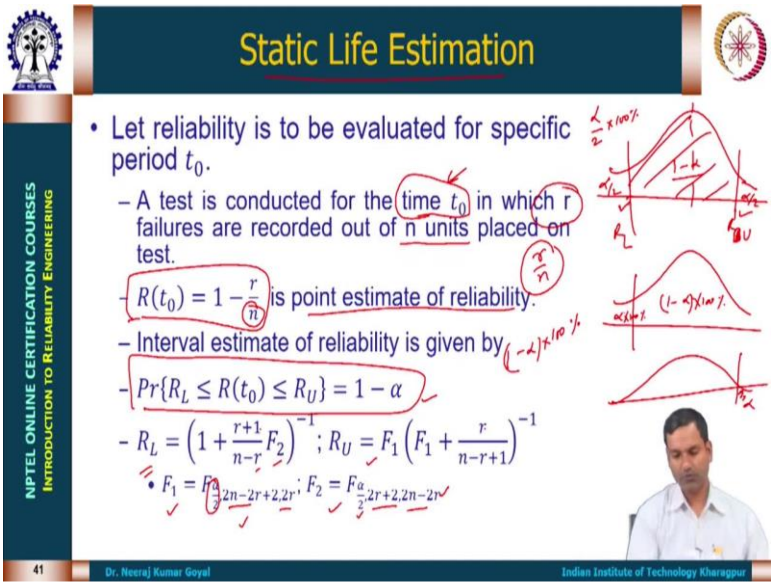
Similarly, when we talk about so, we are interested in lower bound. Similarly, when we talk about the bad things like risk failure probability, we talk about the upper bound because we wanted at max how much that is, so, we are interested in upper bound. Many times here we are constructing the two sided bound many times we may be interested in one sided bound, then we are interested in one sided bond that is a lower bound that is alpha percent on left side and 1 minus alpha percent on 1 minus alpha into 100 percent on the right hand side, alpha into 100 here.

So, rather than alpha by 2 we will be using alpha. So, same formula may be applicable, but when we say one sided only in that case, we are not taking alpha by 2 alpha by 2 we are taking alpha on the left hand side or on the right hand side for lower bound on left hand side and for right hand, for upper bound, we will be using the same on the right hand side which is leaving alpha into the right hand side.

So, using these formulas is general and applicable to almost all distribution in all cases we can use it. Let us see the example for this.

(Refer Slide Time: 17:35)




$$\Pr\left\{R_L \le R\left(t_0\right) \le R_U\right\} = 1-\alpha$$

$$R_L = \left(1+\frac{r+1}{n-r}F_2\right)^{-1}; R_U = F_1\left(F_1+\frac{r}{n-r+1}\right)^{-1}$$

$$- F_1 = F_{\frac{\alpha}{2},2n-2r+2,2r}; F_2 = F_{\frac{\alpha}{2},2r+2,2n-2r}$$

Let us say that we want to estimate the launch reliability of a booster rocket. So, generally these kind of formulas are much more applicable to rockets, missiles, guns, or we can say bullets or we can say the different kind of ammunitions, because they are once fired then job is over. So, one shot devices you may have or you may have some mission oriented devices which goes and complete the mission. So, like satellites etc.

So, here satellites generally work for a longer period. So, you may be interested in time based reliability, but for rocket you will be interested in this static reliability. Now, here in this case, the rocket is used to launch the communication satellite into the orbit. Now, this 20 launches have been completed. So, for that launch 20 launch has been completed. And out of 20 one time the launch was unsuccessful, so that one failure is observed.

So, how much is this reliability for this rocket, someone wants to know how much is that laptop for this rocket with a 90 percent confidence interval. So, we can apply whatever we have done in previous slide that N is 20 r is 1. So, our estimate of reliability is 0.95 but as we know that number of samples are only 20. So, it is still it is a good size, but still if you take larger size the value may change.

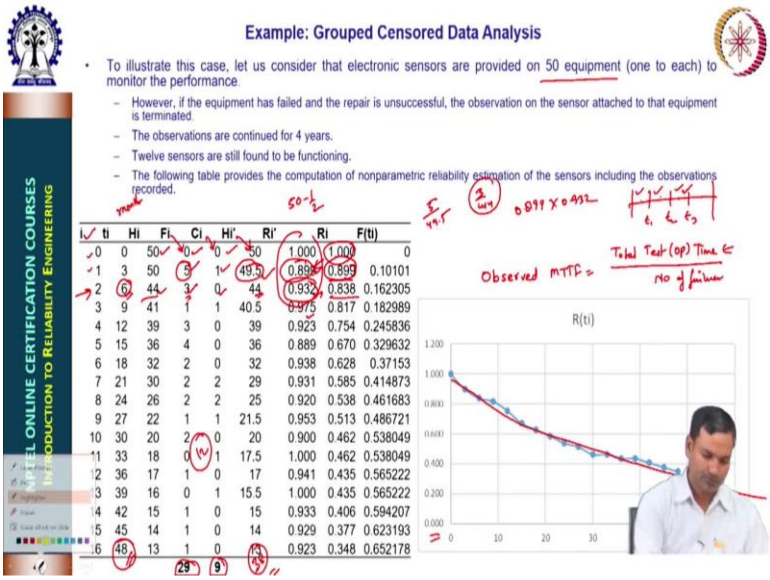
So, how much is uncertainty about our estimation or because of that, how much is going to be the lower bound and upper or if there is no uncertainty of our estimate value would have been good. So, because of uncertainty, we are trying to find out the lower bound on reliability. So, for lower and upper bound, we need 2 values F 1 and F 2, F 1 and F 2 we have to get from the F distribution.

So from F distribution when we choose these parameters, 0.05 is the probability 40 and 2 are the two degree of freedom. Then we get to 19.47 and 4.05 4 and 38 degrees of freedom we get the 2.625. So based on that we calculate reliability R L, R L is 1 plus 2 divided by as you see here, r is 1, so, 1 plus 1 2 n minus r, n is 20 r is 1. So, 2 divided by 90 and multiply by F 1 so, multiply by F 2. So, F 2 is 2.625.

And then we took 1 upon of that inverse of that, we get the value 0.7835. Similarly, R U is calculated by F 1 into F 1 plus r upon n minus r plus 1. So, 1 R is 1 n is 20 So, 20 minus 1 plus 1 is 20 so, this will be 1 by 20 whole raised to the power minus 1 and we get the upper reliability is 0.9974.

Now these values if we want we can see in excel also we can get it otherwise we have the F distribution tables from there we can get this F distribution values. So, here as you have seen that we are able to calculate various system reliability when by using the nonparametric approaches, but as we discussed earlier nonparametric approaches like we presented earlier.

(Refer Slide Time: 21:07)




We do not get the full scope of reliability right we get a limited scope, because here we know what happens for the reliability for 48 months, but we do not know the reliability for 50 months or we do not know the reliability for 60 months because, that data we do not have. So, unless we do some sort of trend fitting here, we do certain things here, if you do some trend fitting and we are able to extend this line, then we will know if assuming that the same trend will follow.

We will be knowing the reliability for 60 also here, we will know the reliability of 50 also here 55 also here and other times, but to do that, we have to do certain model fitting because model is going to fit that trend to this data. Without using a model we cannot extend or interpolate or extrapolate our results.

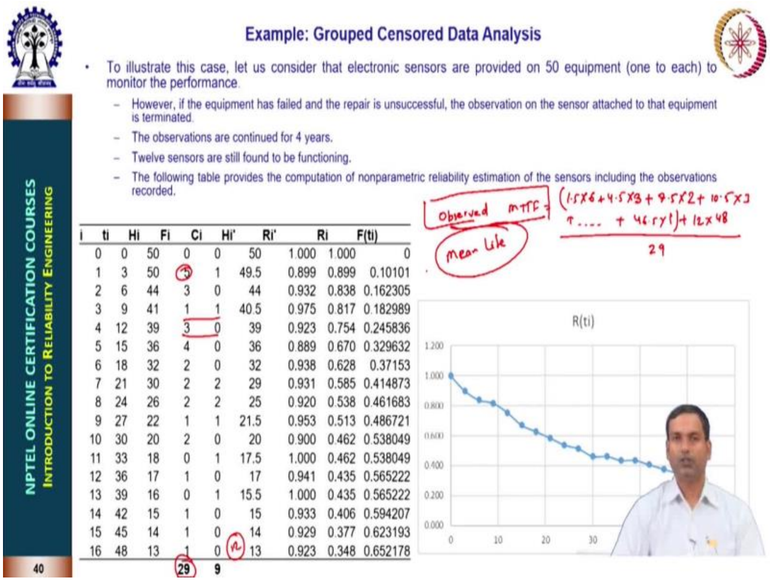
So, here our next topic, which we will discuss for this purposes that how to do this estimation using the parametric methods, where we will mostly be discussing about the 4 distributions. So, the distributions which are made mostly used in reliability or the exponential distribution viable distribution and 2 more distributions, which may be used on may not be directly to the time to failure data, but for repair data many times log normal is used for time to failure data also log normal distribution is quite widely used.

Sometimes normal distribution is also used for the purpose of stress data, strain data that kind of applications whenever we involved then normal distribution also comes into the picture. So, we will discuss in our next class how we can use these distributions fit the data to these distributions. And once we fit the data to these distributions, how can we evaluate the parameters of reliability and they are able to know more elaborate information and we are able to know MTTF for like here also we are not able to calculate MTTF because we do not have all the failures.

MTTF is only we are able to calculate under the assumption that if he assumed the exponential distribution by default here, then only we can calculate the MTTF, which is we are calling his observed MTTF, which as we discussed in previous class, observed MTTF observed MTTF as we are calculating is total time or test time we can say or total operation time we can say divided by number of failures.

Where the count all the time which has been spent as working. So we can see total working time either by the failed unit or by the censored unit. So like here, if we say for this 38 failure, so we have 12 failures which are remaining is still so for if we here want to calculate this time, then for this time, how do we calculate this sorry just one second.

(Refer Slide Time: 24:43)




So like if you want to calculate MTTF here, then first I have to assume that if I do not assume distribution, I cannot calculate MTTF here, but normal normally we assume the exponential distribution because exponential distribution we are assuming that time to failure is constant. So in that case we are able to calculate MTTF here. So MTTF if you want to calculate then we know that on an average how much time.

Like if you see these 5 units, these 5 units plus 1. So, on an average if you take they might have felt any view so, we can take the average time. So, that means 1.5 into 6 so 1.5 units have 1.5 months for 6, then again for this interval that means up to 6 from 0 to 6 right? So, which so, that becomes 4.5 so, 4.5 on an average midpoint multiply with how many device? 3 devices worked. Plus, similarly, we can take 7.5 into for 7.5 two devices worked then 10.5 10.5 three devices work. Similarly, we calculate all.

So, last value would be that up to 46.5. In 46.5, how many devices worked one device right into 1 plus this is the failure time this time is recorded for the failure devices or the censored devices in between, but there is a sensor devices 12 sensor devices here which have been working here not failed and which is continuing to work. So, for these 12 devices they have worked for 48 months, so, 48 months will become the total time and divided by total number of failures, how many failures we observed we observed only 29 failures.

So, this becomes our observed MTTF. But we have to be cautious when we are using this, this observed MTTF may not mean the life, mean life, because we do not have the failure data for all the equipment. So, if the components later on start degrading and fail faster, because of if a component continues to follow exponential distribution for whole life this will be valid, but that generally does not happen as you have seen earlier. Components do degrade and because of the degradation later on the failure rate may rise.

So because of that a higher failure rate or the degradation, this observed MTTF may not be valid. So in that case, we have to be cautious unless we have all the failure data. Calculating MTTF may be tricky part. So if we observe if we use exponential distribution for this, we should be cautious here that we should not miss understood this with mean life. This is the MTTF under the consideration that system is following the constant failure rate or exponential distribution.

So we will continue our discussion and we will next time discuss the failure data analysis using the parametric models. Thank you.