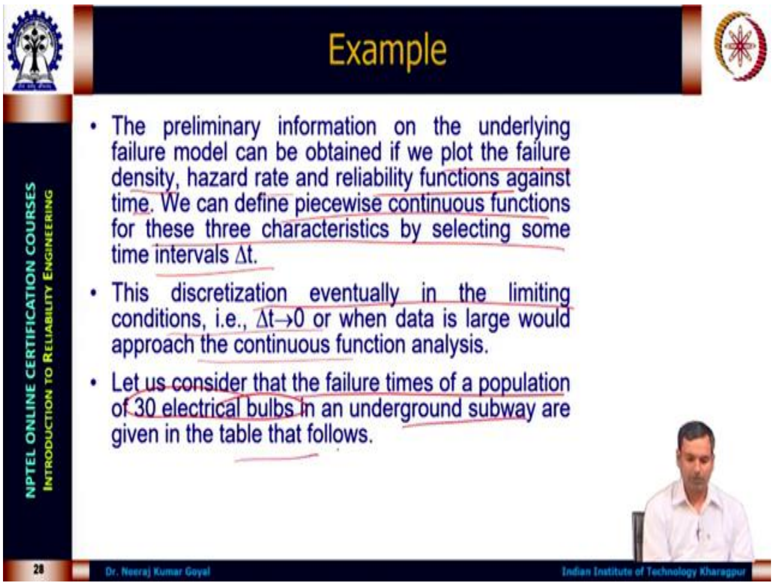
Introduction to Reliability Engineering
Professor Neeraj Kumar Goyal
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology, Kharagpur
Lecture: 28
Failure Data Analysis Non-Parametric Approach (Contd.)

(Refer Slide Time: 00:42)

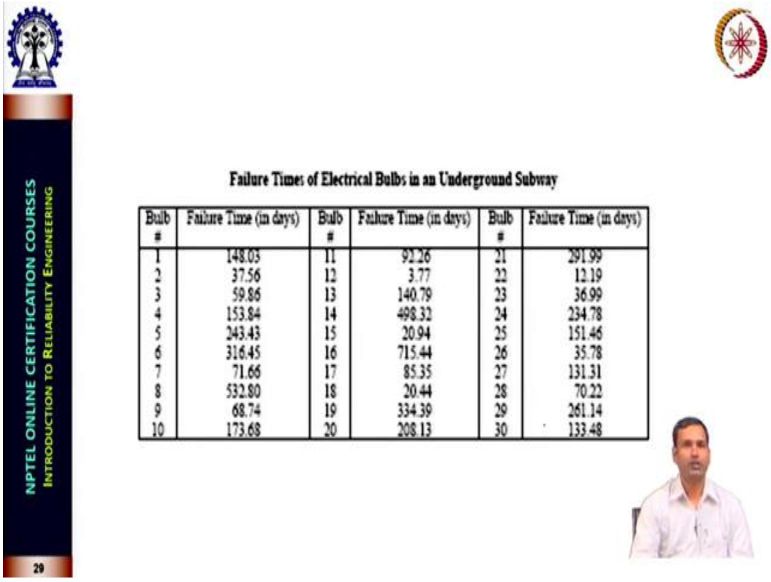


Hello everyone so we will continue our discussion where we left in previous class that is analysing the group failure data for evaluation of reliability and other important indices. So, as we discussed in previous class we derived these formulas now let us see that if we have the data then how we are able to calculate.

So, let us take an example that a preliminary information or underlying failure model can be obtained if we plot failure density, hazard rate this we have done already for the complete data we have time to failure was available reliability function against time. If we can define piecewise continuous functions for these three characteristics by selecting some time intervals delta t.

This discretization eventually in the limiting condition like delta t tending to 0 data is large would approach the continuous function analysis. Now let us say that we have the failure times for population of thirty electric bulbs and in underground subway which is for which we have these observations here.

(Refer Slide Time: 01:30)

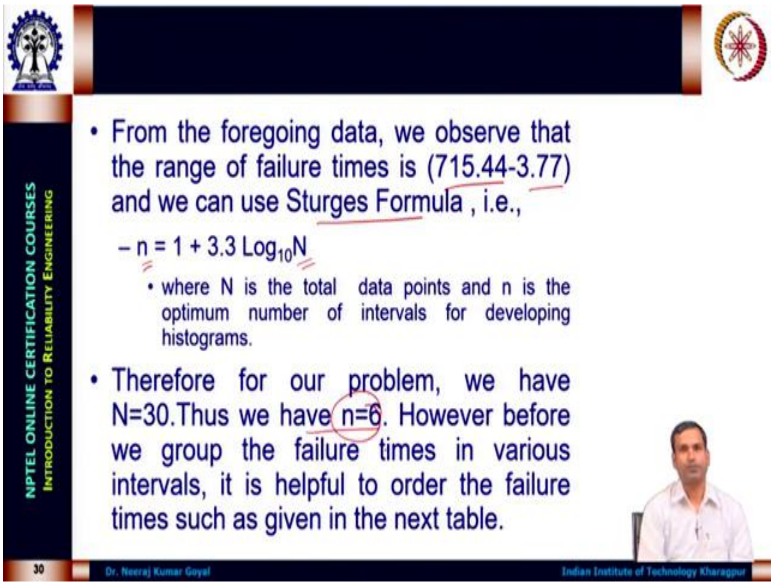


Failure Times of Electrical Bulbs in an Underground Subway

Bulb #	Failure Time (in days)	Bulb #	Failure Time (in days)	Bulb #	Failure Time (in days)
1	148.03	11	92.26	21	291.99
2	37.56	12	3.77	22	12.19
3	59.86	13	140.79	23	36.99
4	153.84	14	498.32	24	234.78
5	243.43	15	20.94	25	151.46
6	316.45	16	715.44	26	35.78
7	71.66	17	85.35	27	131.31
8	532.80	18	20.44	28	70.22
9	68.74	19	334.39	29	261.14
10	173.68	20	208.13	30	133.48

Now these observations which is given here I have plotted now I can do the analysis time to failure analysis also where like we discussed earlier I can arrange all these time to failure data into the increasing order I can do the same analysis find out the reliability unreliability and find out the Ft, Rt everything.

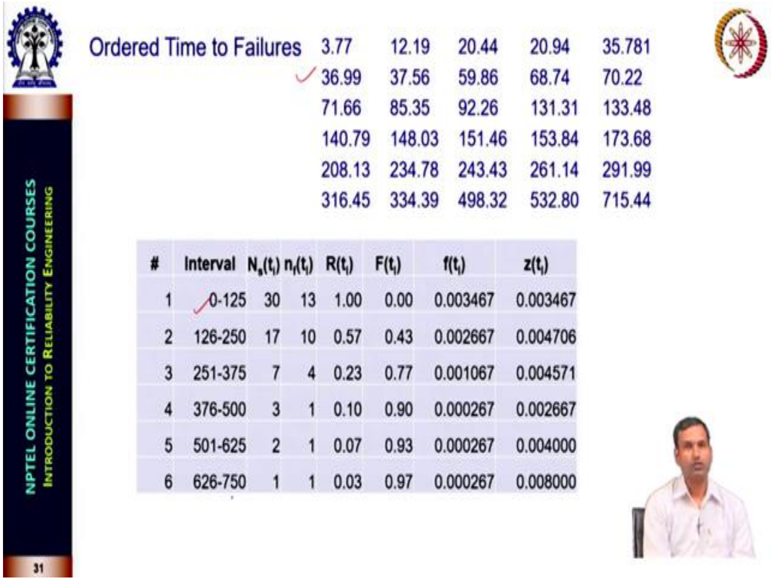
(Refer Slide Time: 01:58)



But we can also do one thing we can do the group analysis. So group analysis sometimes we are doing so that our problem is better understood sometimes that group did group analysis we are doing because data is in already in the group.

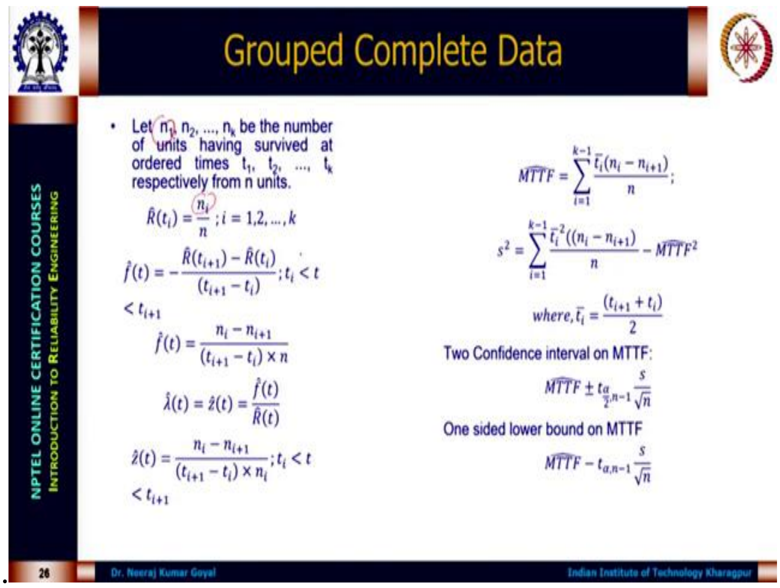
So, if data is not in group but we want to do the analysis in group what we can do we can use the Sturges formula. Sturges formula suggests is that how we can decide the interval so our minimum time is 3.77 and our maximum time is 715.44 from the data. So, how can we decide it, so total data points are 30, so when we have the 30 data points this n value when we use we it comes out somewhere around 6, so we can divide this interval into 6 intervals and then we can do the analysis.

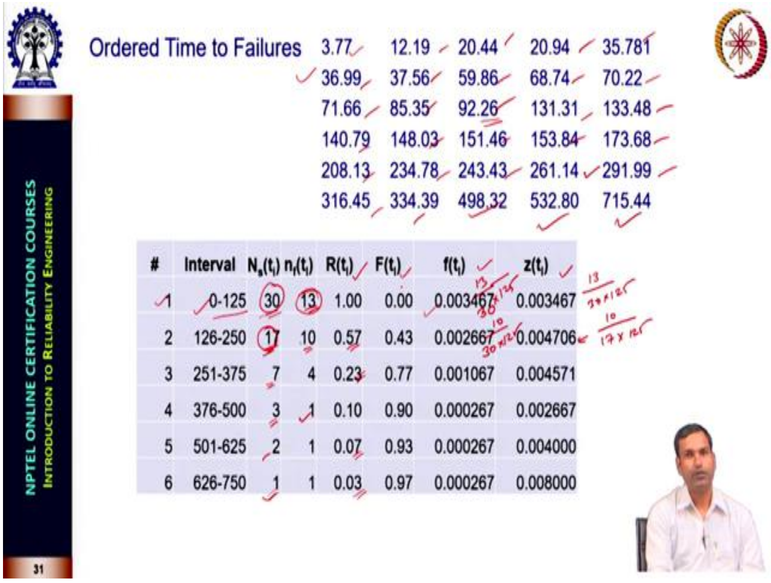
(Refer Slide Time: 02:48)



So, here this time to failure data which was there I have ordered it already and this is put up here, and this data I have taken in intervals. So 6 intervals are taken of 125 hours each. So 125 is here then 126 to 250, 251 to 375 like that, so we have the 6 intervals here.

(Refer Slide Time: 03:18)





Now in each interval as we discussed for the formulas like here we want to know at the start of interval how many units are working n1 n2 like that, so we will go like this. So, here as we see here for first interval at the start of interval all units are working 30, at the time t equal to 0, number of units surviving is 30. Now in time 0 to 125 how many failures are happened if you look at here this is this, this 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13. 13 failures have happened. So these 13 failures have happened in interval 0 to 125 so number of failures are 13.

Now how many units are surviving at the next interval is 17 at the next interval only these 17 units are working out of these 17 units up to time 250 if we look at it then 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. 10 units are filled, so 10 units are filled then in next interval from 250 to 375 the remaining units are only 7, out of the 7, 4 failure occur 1, 2, 3, 4, and then again similarly 4 occur 7 remaining out of 7 1 failures in this interval, and from 500 to 625 1 failure occurs, then 626 to so 750 1 failure occurs.

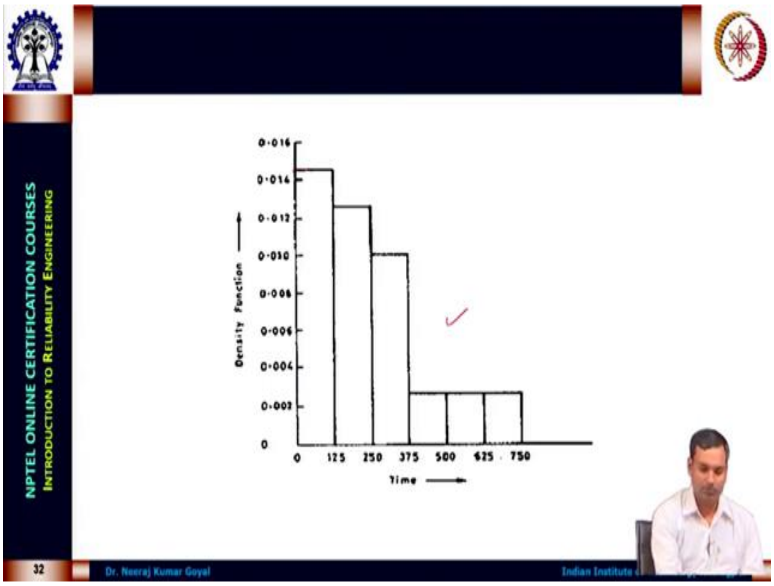
So, we have this number of survival data number of failures data now we can get the Rti. What is reliability at time ti at the start of the interval, so this Rti value is 1 because 30 out of 30 are working at the start of interval so 30 divided by 30 is 1.

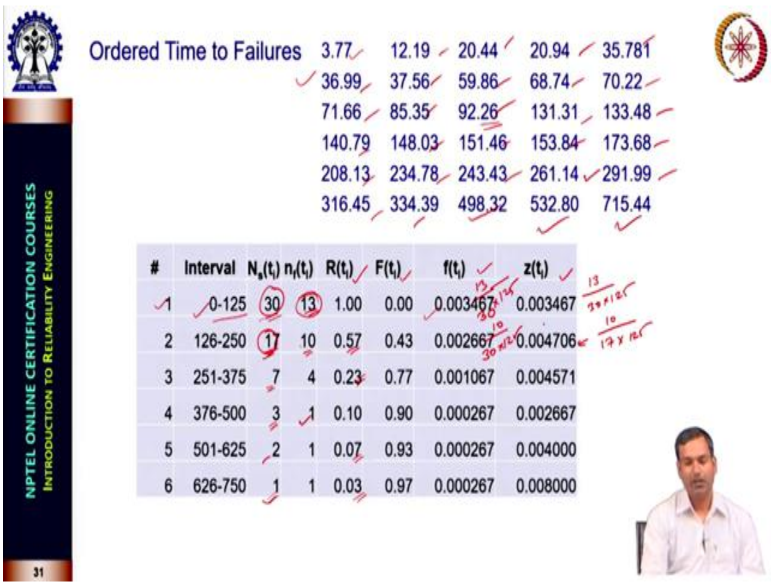
Here for second interval 17 out of 30 are working so 17 divided by 30 that gives me 0.57 is the reliability for second interval. Similarly for third interval 7 divided by 13, 0.23. Similarly we get 0.03, 0.07 lines all these values we are able to get.

Similarly, I can get the unreliability with as 1 minus reliability and ft is again change in reliability as we discussed here ft is number of failures in the interval divided by number of survivals at the start that means always 30, so this will be 13 divided by 30 that is probability of failure per unit time how much is the time each time interval is of 125.

So, 13 divided by 30 divided by 125 will give me this value similarly this value will be 10 failures out of starting 30 units multiply by 125, same way and zt would be I can get it two ways either I divide Ft with Rt or I can take number of failures divided by number of surviving units at the start that is 30 here into time period is 125 for this number of failures is 10 number of survivals at the start is 17. 17 into 125 like this way if we calculate we will be able to get the reliability, failure probability, ft and zt.

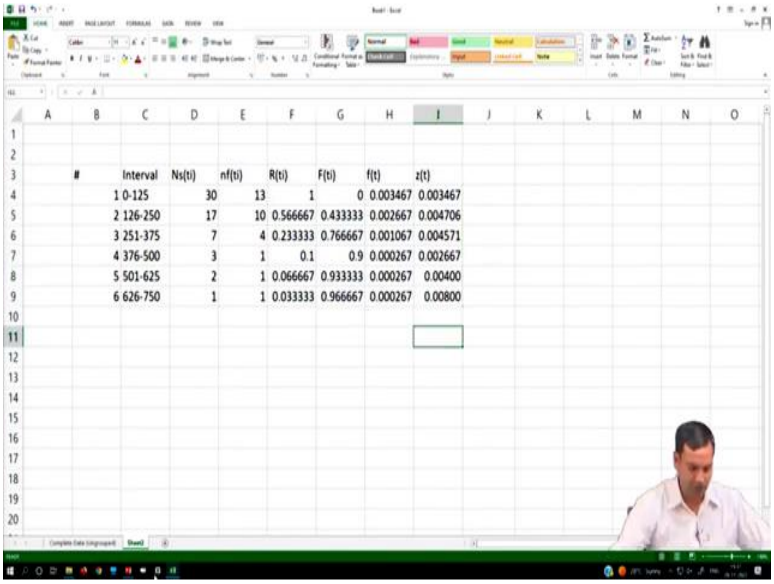
(Refer Slide Time: 06:50)





So, we are able to get all the parameters of our interest if I plot this then density function would like look like this that is of first value is for upon. As we discussed earlier zt values and ft values we will try to this is valid for the complete interval, so this will be for Interval same value, so we should try to plot a step function here. So first interval value around 0.14 something so that is, this may be a data from some other value 125 there is some error here. in this so we will not look at here we will see the same value which we have plotted here or we can check this also if you want let me just try to do this quickly.

(Refer Slide Time: 08:00)

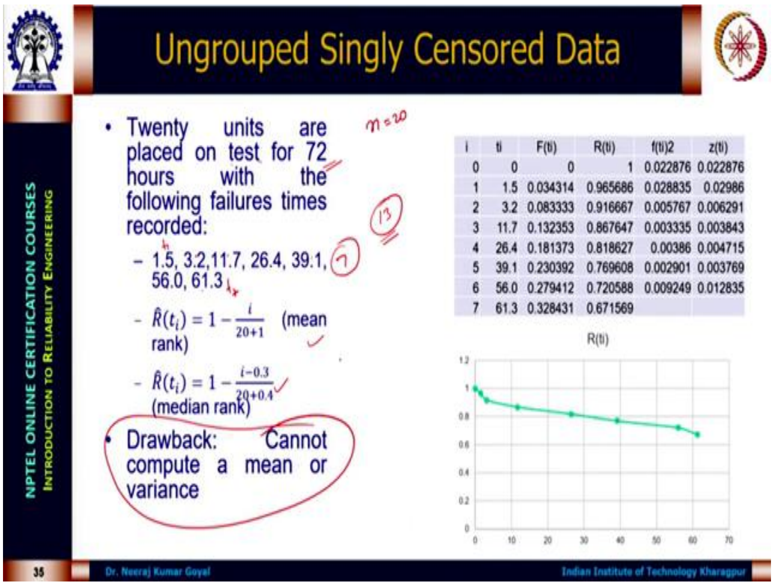


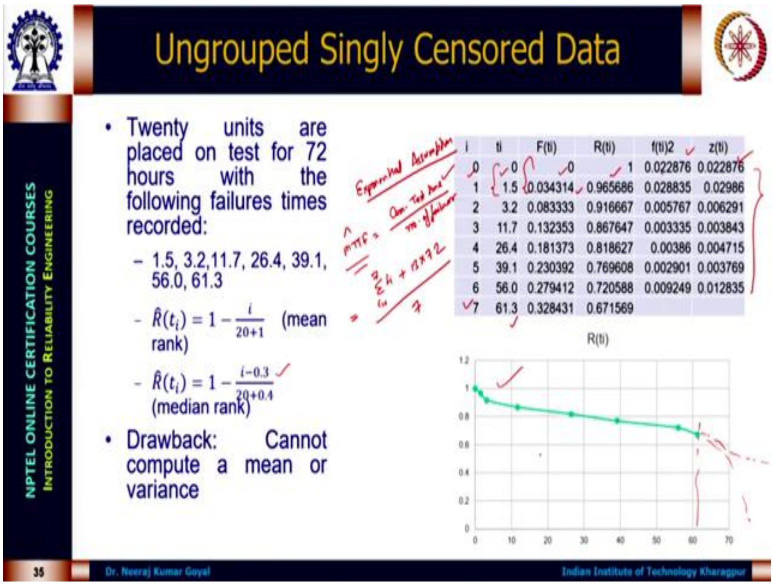
We have used the Excel sheet let me put one more Excel sheet here again I will do the same thing we have got interval we have got Fti. Now as we saw earlier that is equal to I will call the ft here. So ft is equal to this is equal to number of failures in the interval divided by units in at the start divided by interval is 125. So we are getting the same value maybe that figure is in some error.

And then zt value is number of failures in the interval divided by the number of units at the start and divided by interval, interval is 125. This value would be same as ft upon Rt if I show you that is ft divided by Rt same value will get with both the methods. So we are able to get this is coming because of let me just check Rt is equal to this divided by 30.

And this is actually what happened there was an approximation error, so that is why, it was looking like this. So because we have did the approximation of the digits, so that is why, now if we do not do approximation it would look exactly same. So, as we see here this will be kind of ft Rt. I think there is some scale problem otherwise it looks pattern would be same we see the same pattern for hazard rate Rt, ft would look like this.

(Refer Slide Time: 10:37)





Now if you look into other type of data we have looked into two types of data one is complete data ungrouped another is complete data grouped. Now let us see if we consider the singly censored data, singly censored data means mostly as you discussed it is the right censored data.

So, let us take an example for the same. Let us say we have the 20 units which is placed on test, so we have n equal to 20 and the test continues for 72 hours and these are the failure times recorded, so 1, 2, 3, 4, 5, 6, 7 failures have happened. So 7 failures happened in 72 hours but rest of the 13 failures have not happened.

So, these 13 failures are the censored data. So, if we use a mean ranking or median ranking we can get the because this is right side so whatever failure point is happening this is t1 and this is t7, so we are able to get the same way reliability value or unreliability value but the problem here would be we cannot calculate the mean on variance.

Because we do not know the how much is going to be the failure time for the rest of the units which are working so because of that we cannot calculate the mean, so for this kind of methods unless we fit the distribution we will not be able to calculate the mean. Sometimes a mean is calculated by assuming that there is exponential under exponential assumption.

In under assumption of exponential what happens we assume that failure rate is same, so if failure rate is same in that case we assume that given a population size whatever failure happens the number of probability of failure remains same for the rest of the failure in a given time.

So, here because of this the for exponential assumption this MTTF becomes a cumulative test time divided by number of failures but as we see this is and only going to be valid when we assume the exponential this assumption many times people use this formula as the general formula but we should be show that data is exponential distribution because otherwise many times this MTTF would be very-very large and this large MTTF will suggest that life is more but this MTTF may not be the representative of the life.

Because what will happen if you place lot of units on test let us say 1000s units of test and then only observe 10 units of failure then what will happen for 990 units you will have the data which has been cumulated for the failure times, so MTTF will tend to show the larger value.

Because all units may not fail in similar pattern. So life limitation may be there but generally still it is valid many time because as we discussed earlier that is, if the device life is within the non-deterioration period or the constant failure rate period then this MTTF would be valid.

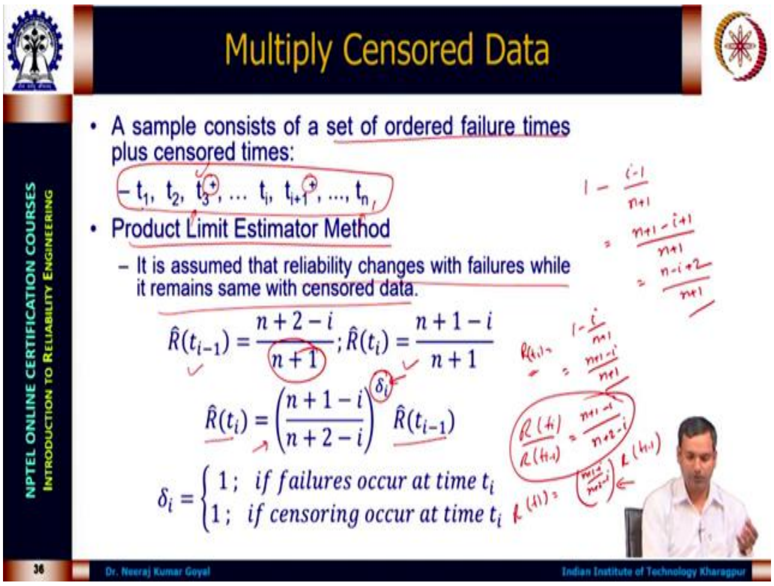
So, cumulative test time here like if I can take the summation of this time i equal to 1 to 7 and ti plus now for 13 units we have tested for 72 hours. So total cumulated time is how much the failure time which we recorded and the 13 units which have been tested for 72 hours but failure has not been observed. So, that is also cumulated and how many failures I have observed we have observed 7 failures. So, this formula when we use it will give us the MTTF but that is not the true MTTF that is the MTTF under exponential assumption

So, here as we see here I have done this exercise in excel, so we have 0 to 7 values, then ti values are there 1.5, 3.5, all these values I have put up here then Fti value as we have seen earlier we calculate in same way that is 0 and i minus 0.3 divided by 20 plus 0.4. Same way we calculate here and Rti value is 1 minus Fti and small fti as we discussed earlier that is difference in this divided by difference in this.

And zt is nothing but ft divided by Rt and that is how we get all these values. And Rt values if we plot, the plot looks like this, Rt is changing line. As we see here in complete data, it was reaching to 0 but since this is not a complete data, the censored data. So I know only up to here, I do not know how reliability will be, this may go like this also, this may go like this also, I do not know how will be the reliability behaving in the future because I do not have the data here.

But if I had put some distribution assumption then that would have suggested the future trend and because of that I would be able to get the MTTF and other values also. So, this singly censored data as we have seen we are able to analyse and we can use it for multiple purposes

(Refer Slide Time: 16:13)



We will continue our discussion with a more generic type of data which is multiply censored data so multiply censored data what we have we have ordered failure times this is all ordered data t1 is less than t2, t2 is less than t3 but here we are putting some signs plus here.

$$\hat{R}(t_{i-1}) = \frac{n+2-i}{n+1}; \hat{R}(t_i) = \frac{n+1-i}{n+1}$$

$$\hat{R}(t_i) = \left(\frac{n+1-i}{n+2-i}\right)^{\delta_i} \hat{R}(t_{i-1})$$

$$\delta_i = \begin{cases} 1; \text{ if failures occur at time } t_i \\ 0; \text{ if censoring occur at time } t_i \end{cases}$$

What does this plus mean, this plus means that data is censored. That means at t3 time the unit has been removed from the test but not due to the failure of concerned failure, so it has been removed either because it has failed in some other failure mode, it can be removed because some sort of other failure, structure failure has happened, or some sort of supporting device failure has happened which is not the failure concern.

So, because of that we do not know and we could not test it further and see we do not know at what time it will fail. So we are not able to know that time. Then, so this plus sign is for the data for the devices which have been removed from the test but not considered as the failure. So we may have different times we can have at the end also we let us say here it is looking like that nth failure is happening here but other devices are failing.

We may have some other way also that test may incomplete also may remove also. So, all types of data may be coming here so this general method can be used for that all the purposes. So, one popular method to analyse this kind of data is product limit estimator method. so in product limits estimator method is assume that reliability changes with failure while reliability does not change if there is a censored data.

So, what does it mean that reliability for ti minus 1 and ti, so, this is again that mean ranking formula when we use, so ti minus 1 means of i minus 1 failure we are talking about so that is 1 minus i minus 1 divided by n plus 1. So this will be equal to n plus 1 minus i plus 1 divided by n plus 1 this is equal to n minus i plus 2 divided by n plus 1.

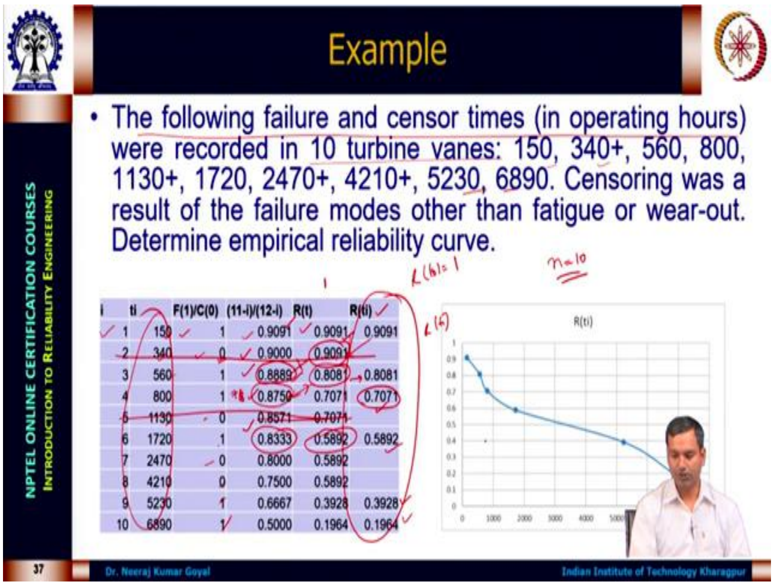
Similarly, if I take Rti, ti mean i failures so 1 minus i divided by n plus 1, so this will become n plus 1 minus i divided by n plus 1. So if you see here if I take the different ratio of this Rti divided by Rti minus 1 then this will be equal to n plus 1 minus i divided by n plus 2 minus I, n plus 1 and n plus 1 will get cancelled, this n plus 1, this n plus 1 will get cancelled.

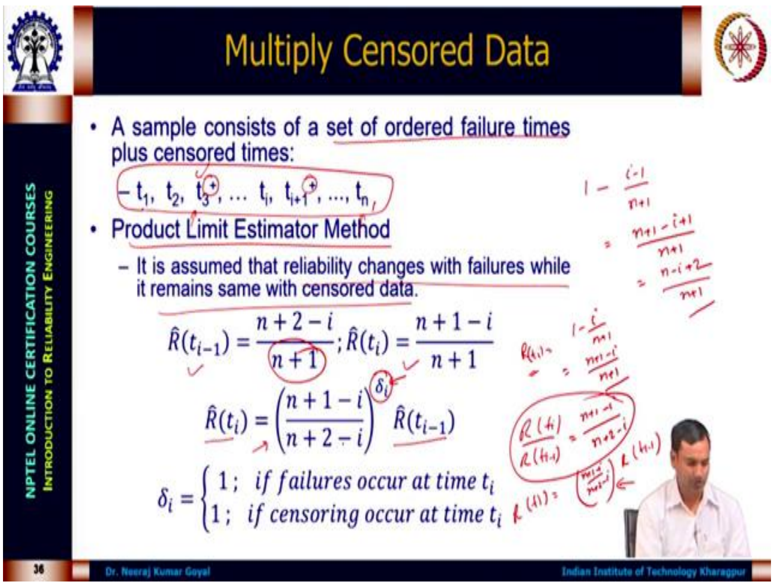
Now this so what I can say Rti is equal to this value n plus 1 minus i divided by n plus 2 minus i multiply by Rti minus 1. Now here I am using a power factor here that is delta I. So, this factor I will multiply, so what I am assuming that if it is a censored failure then reliability

Rti is same as Rti minus 1, but when it is censored when it is not censored then failure happens then this is to be multiplied with this factor.

So, delta i will be 1 if there is a failure and delta i will be 0 if there is no failure. So if delta is 0 then Rti will be equal to Rti minus 1 and then delta i is equal to 1 that means if failure happens for that point, then this value will be equal to, this value has to be multiplied by this factor.

(Refer Slide Time: 20:06)





Now let us see this with the data so let us say we take an example that we have the failure time and some censored time and for 10 turbine vanes. So, n is equal to 10 here, we have 10 vanes here and lack of 1 we have the failure data 150 is failure. So, for failure we are writing 1 for censored we are writing 0.

For 340 it is censored for 560 it is failure, 800 failure, 1130 censored, 1720 is failure, 2470 is censored, then again censored, then failure, failure, all these data we have put it here. Now how do we do the analysis as we have seen n plus 1 minus i and divided by n plus 2 minus i is the factor.

So, we will calculate this factor. So n is 10, so 11 minus i and n is 10, so that is 12 minus i. So 11 minus i divided by 12 minus i we will calculate here. i is what, this is, so this is 0.9091 that is 10 divided by 11, this will be 9 divided by 10, this will be 8 divided by 9, this will be 7 divided by 8. So, like this we are able to get this value.

But as we know that Rti value which we are calculating, Rti value will be calculated as Rt. So, Rt is this value for ti minus 1 as Rti, t0 is equal to 1, t equal to 0 is 1 for 0 failure 0, so previous value is 1 and if I take 0 i equal to 0, this value will be 1.

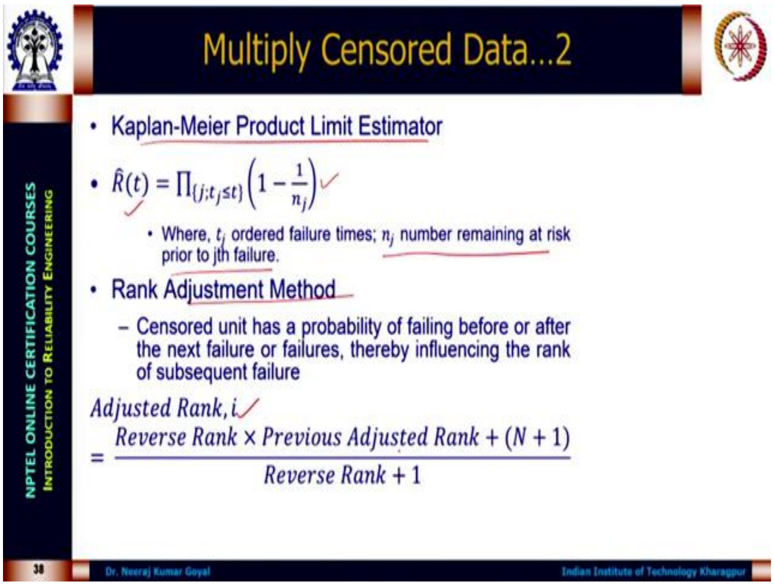
Now for this first interval reliability will be coming as it is 0.9091. Now this is a censored point, so for censored point the reliability remains same 0.9091, the same reliability which we have here that will remain same. Now for 560 points failure happened and the multiplying factor is 0.8889. So, what will happen this 0.9091 will get multiplied with this multiplication of 0.8889 with 0.9091 will give me 0.8081 this will be my new reliability then again failure has happened.

So, again this value will be multiplied with 0.8081 and this will be my new reliability. Again censored point, so I do not bother about censored point even the censored point if I do not list it here it will not make much difference. Then for again failure happened here the factor is this, this will be multiplied with 0.7071 and this will give me 0.5892 so same way I am getting the reliability values here.

This reliability values becomes the Rti value, so Rti value will plot against ti, so this is my ti, and this is my Rti. And when I plot this reliability function is looking like this I am plotting only 1, 2, 3, 4, 5, 6 points, other points which are not calculated is not plotted here. So that you can see that how reliability is changing with time.

So, with this example we are able to see that how reliability changes or we are able to estimate reliability whenever and same once we know the reliability and with time then change in reliability per unit change in time, if we take that will give you the density function, negative of that. And if you take the divide by again Rti if you do density function divided by Rti, you will get again the failure rate function. So all the functions which other functions same formula same table you use you will be able to get it.

(Refer Slide Time: 24:16)

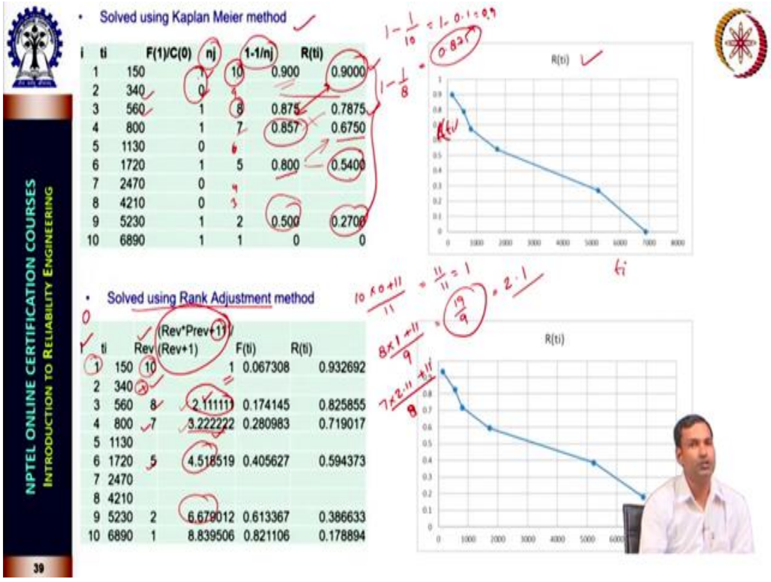


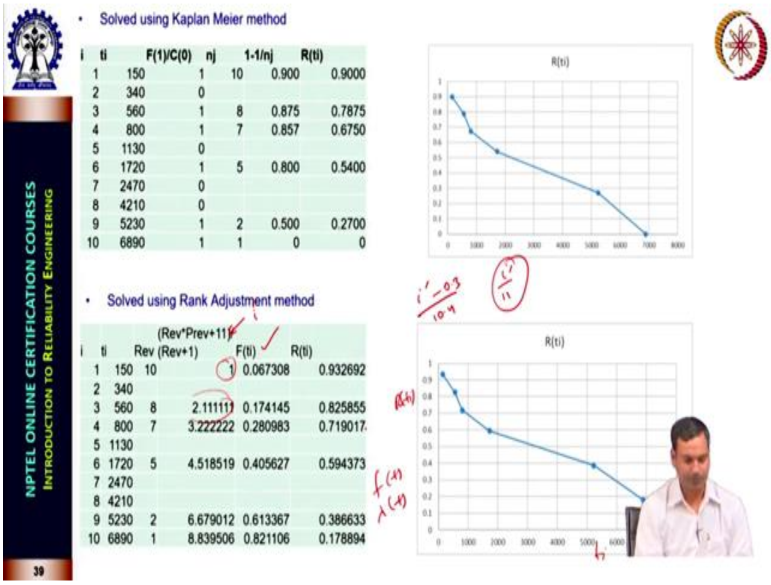
Now this analysis can also be done with other methods so there are two more popular methods one is Kaplan Meier Product Limit Estimator. So product limit estimator which we used here it has this assumption and these assumptions were later on changed this they told that you can do different ways. So here the reliability is calculated as 1 minus 1 upon nj.

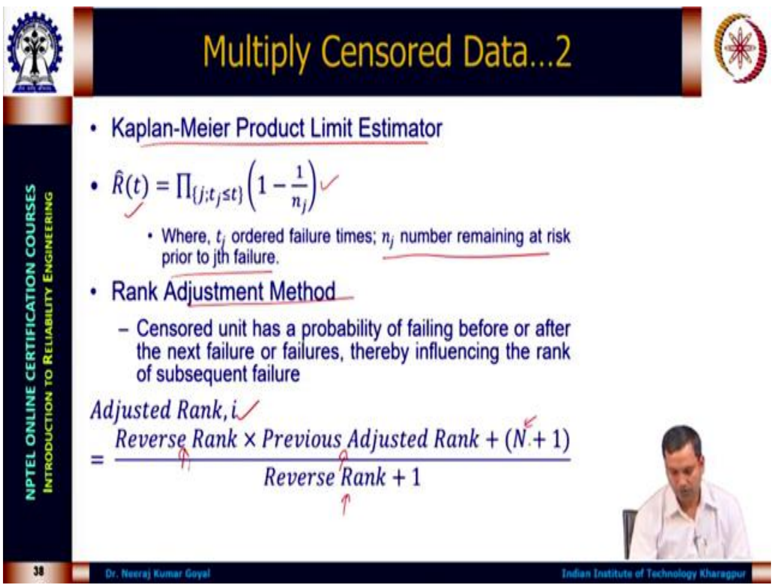
So, here tj is the order time and failure times and nj the number of items remaining at risk at prior to jth failure, so that means at that start of the interval how many units are working. So, the units before this which have been failed or which have been censored all those units are removed, only remaining units which have been working has are counted here.

We will see this then rank adjustment method, the same thing it is supposed to propose that we can calculate it by using the rank adjustment method in which we try to use the adjusted rank. So it is assumed that whenever you remove a failure because of removal of failure what has happened, the rank which was 1, 2, 3 we were taking for failure times that gets modified and we try to get the new adjusted rank which is little higher than the rank, because the failure could have happened the removed device could have failed in the interval also.

(Refer Slide Time: 25:45)







So, let us see how do we do this so if we take the Kaplan Meier method so same data which we had earlier same data I have taken here now in this case we calculate the nj, nj is the number of device working at the start of the interval, so at the start of first interval the number of device working for 10.

Now what happened out of 10, 1 unit is failed and 1 unit is under 1 unit has gone for it has been censored. So 2 units have been removed from the testing further and at this for this failure the units which are working is 2 removed, so only 8 units are working. Again 1 failed, so 7 units are working, in a way if I say it is a reverse ranking.

In a way if I say that this is 1 to 10 and this is 10 to 1, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1. So this reverse ranking is denoting the number of units which are working at the start of the interval. Now reliability calculation is simple, unreliability calculation is 1 minus 1 divided by nj whenever failure happened like or we can say this is my reliability calculation.

That means 1 minus failure happen out of 1 failure happened here in each interval. So here 1 failure happened out of 10 units. So 1 minus 0.1 that is 0.9. This is my reliability at the end of this interval. Similarly, for this time I will not calculate reliability, the censored time reliability is not calculated, like earlier also we did not calculate.

For 560 number of device working is 8, so 1 minus 1 failure out of 8 devices. So that would be 0.875. Now this 0.875 is the reliability for this interval only 1 failure for this, but actually 2 failures have happened, so for earlier failure the reliability induction is 0.9. So effective reliability is that for up to here that means it should survive this period also, this period also. So reliability 0.875 will be multiplied with 0.9 and that will be giving you the reliability. So reliability in 560 hours is 0.7875

The reliability in this interval second failure interval that is 0.875 but gravity has to survive first interval also second interval also, so reliability will gets multiplied that is the formula we have seen all these values are to be multiplied whenever failure happens. For seventh again the value is 0.857 that is 1 minus 1 divided by 7 and what is Rti,

Rti is this value multiplied by 0.7875 multiplied by 0.875 that will give me the 0.6750. 1 minus 1 by 0.5 is 0.8 then we get this 0.8 multiplied by this will give me this value. Similarly we get this value, and this is giving us reliability versus time, ti versus Rti. we are able to plot and we can see that how Rti is changing with ti.

Similarly, if we use the rank adjustment method, it ranks adjustment method again, it is a similar way that we calculate this is my rank this is my reverse rank. So the rank adjustment method gives me the adjusted rank so for adjusted rank first we calculate the reverse into previous, this is my reverse ranking, and this is my previous reverse ranking. Plus, the formula is reverse rank multiply by previous adjusted rank plus n plus 1 divided by reverse rank plus 1. So n here is 10, so 10 plus 1 is 11.

So, that is why I am writing 11, this is my reverse rank and previous ranking previous adjust rank is 0 because i equal to 0. So here I am writing this adjustment this is my new adjusted rank so for i equal to 0 previous adjustment adjust rank is 0. So what will happen, this will become 10 into 0 plus 11 divided by or reverse rank plus 1 11, so effectively I will get 11 divided by 11 which will be equal to 1. So, my first rank comes out to be 1 only.

Now the second rank, because here what happens this is the failure, this is a censored unit, no failure has happened, so I will not do anything here, then I will do the second failure happened at 560. So, I will use this 8, so 8 is my reverse rank, so but I will do 8 and how much is my previous adjusted rank that is my 1 plus 11 divided by reverse rank plus 1 9, so that will be 19 divided by 9.

So, this becomes my new rank and this will do is 2.1 something. So, this gives me the new rank similarly for this I will get the new rank that is 7 into previous adjust rank is this, that is 2.11 plus 11 divided by 8, this gives me this 3.22 as you see here this is my first failure this is my second this is my third failure, so third failure is not rank 3 this is little higher than rank 3.

Because the failures which we have removed earlier that they could have failed and because of that there is a little higher weightage to the rank, so it is not 3, third it maybe 3.22 failure. Similarly, this fourth failure is 4.51 because more devices which have been removed could have failed during this period and because of that rank has become higher.

Rest of the process is same once we have the rank then if we have the rank that that this becomes my now i value, new adjusted rank. So let us say i dash, so i dash if I want to calculate Fti that will be i minus 0.3 divided by n plus 0.4 that is 10.4 if I am using the median rank.

If I am using mean ranking I will use i dash divided by 11. So depending on the this rank I will use to 1, 2.1 divided by the same formula which we have developed and seen earlier same formula will be used for calculating the Fti. We get the Fti, Rti will be one minus Fti we can plot ti versus Rti or Fti and you will get the change in reliability versus time similarly we can get small ft can get lambda t using the same table if we extend to calculate we will be able to get it.

So, we have discussed various methods which using which we are able to get this analysis and we are able to get this data. We will discuss the same thing in more detail, for one more type of data which is remaining that is the grouped censored data. If you have the group censored data, how we can do the analysis that is also again simpler way of doing that and calculating the reliability, ft, rt, lambda t, etc. We will stop here today, we will continue our discussion in next class. Thank you.