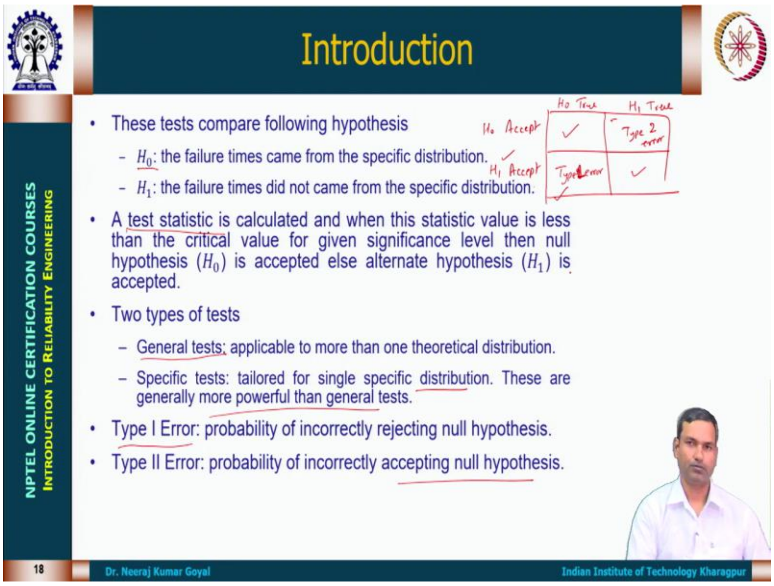
Introduction to Reliability Engineering
Professor Neeraj Kumar Goyal
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology, Kharagpur
Lecture 33
Goodness of Fit (GoF) Tests

Hello everyone. So, let us continue our discussion about the fitting of time to failure or time to repair data. So, in previous class, we discussed about how we can use MLE, maximum likelihood estimation. So, we discussed what are the steps we can follow, so that we are able to get the MLE estimate further from the data for any distribution. When we are using MLE, we can use the goodness of fit test to find out whether the data is belonging to a certain distribution or not. So, today we will start our discussion about the goodness of fit test. So, we will have brief discussions about it, we will not be going into very large details.

(Refer Slide Time: 1:08)



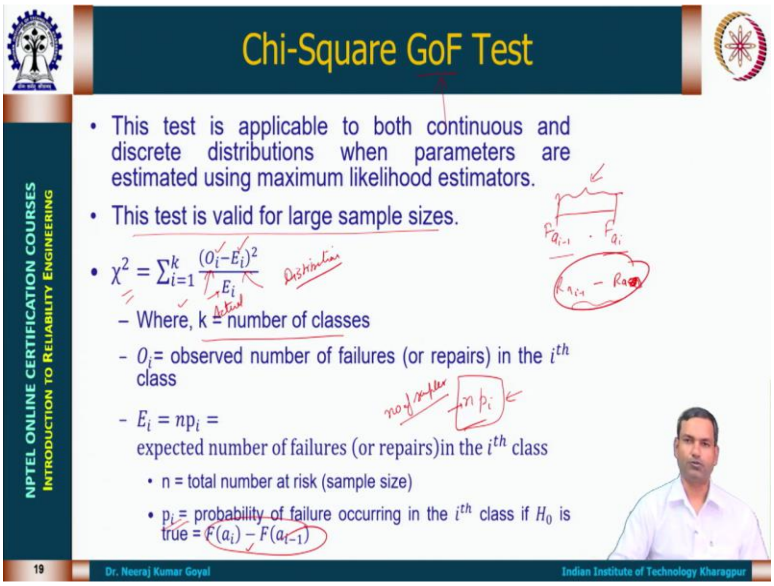
So, here as we see that for goodness of fit, we have the hypothesis testing. So, hypothesis is like we have the null hypothesis, and we have alternate hypothesis. Null hypothesis is represented as 0, and what is our null hypothesis or the other fact which we want to check, whether our assertion shown is true or not. So, the assertion here is that the time to fill a data or time to repair data which we have taken comes from a specific distribution, this specific distribution can be exponential, viable, normal log, normal whatever it is. So, we want to just check that whether this data belonging to this distribution or not, so that is our null hypothesis. So, effectively alternate hypothesis becomes the negative sentence of the same. That means, the failure times do not come from the specific distribution which we are choosing.

So, to do that what we do, we develop a test statistic. So, test statistic like chi-square and others. So, this gives a calculation using a statistic value, and this is statistic is another random variable, which follows a certain distribution. So, from the distribution we find out whether this statistic which you are calculating, whether this is falling within the designated interval or designated one sided interval or both sides interval. And if the statistic value is within the significance level or reason, then within the confidence reason, we will accept the null hypothesis. If it is outside the reason, we will accept the alternate hypothesis or we can say that, we will reject the null hypothesis in favour of alternate hypothesis. Because, we do not find the sufficient statistical proof in favour of null hypothesis.

So, here generally there can be 2 types of tests, one is the general test which we will be discussing that is like, so, chi-square is such test that which is applicable to not only one distribution, but it is applicable to most of the distributions. So, we can check based on the chi-square value whether this is belonging to a certain distribution or not. Then, there are specific tests which are specific to the distributions. So, these are generally more powerful to check a particular distribution, because they are specifically designed for checking those, they are specifically tailored for using for those kinds of applications. So, we can have that but, so, whenever we are making these hypothesis testing, so, with each hypothesis we have 2 types of error.

Generally, if we have the hypothesis here. So, we have a metric here. So, we can, we have H 0 accept, and another case is H 1 accept. So, there are 2 possible outcome, either we accept H 0 or we accept the H 1. But what is truth? Truth is H 0 is or truth is H 1. That means in actual if H 1 is true, H 0 is true or H 1 is true. So, if H 0 is true, and we accepted H 0, this is fine, right decision. If H 1 is true, and we accept H 1, then that is also right decision. But if we accept H 0, but it is not true, then this is called type II error, and this is called type I error. Type I error means, sorry typo I, this is type II error, this is type I error probability of incorrectly rejecting null hypothesis, that means we are accepting H 1, but while it is not true, and this is type II. So, that means if we are accepting H 1, but H 0 is true, that is type I error, and if you are accepting H 0, but H 1 is true in that case that is type II error. So, we generally if we want to balance one, another one will increase.

(Refer Slide Time: 5:51)



$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

So, let us discuss the chi-square goodness of fit test. So, goodness of it, I am writing as GoF here. So, chi-square test is like a global kind of application or it is applicable to almost all types of distribution, whether it is continuous distribution or discrete distribution or any other distribution. And though generally this is valid for large sample size, for goodness of fit, this test we require a little large sample of size, samples. If sample sizes is small, we may not be able to perform this, and we will see why that comes. So, generally what it does, it tries to divide the data into groups, and it is expected that these groups of all should be at least of size of 5 failures. So, if we do not have sufficient data, if we have only 10, 15 data points, then our groups formation will be very less, and we will not be able to get this chi-square statistic properly.

So, here chi-square statistic which we are getting is nothing but the summation of observed value of number of failures minus estimated value of number of failures in each interval. So, we divide the data into intervals, and for each interval we find out how much is the observed number of failures or observed number of repairs, and if I know the observed number of expected number of failures. Expected number of failures we calculate from the distribution. So, from distribution we calculate this, this we are getting from the actual observation, and that is square of this, so, error is squared, and again the divide from the error, that is the estimated value of the distribution. So, once we divide this, then this becomes kind of error.

So, definitely if error is high, chi-square value will be high, and we do not want high chi-square value. So, if chi, lesser the chi-square value, better is the fit, so, better is our applicability of the distribution. k is the number of classes, that is number of intervals, number of groups etcetera. So, this we already discussed, O is the observed number, E is the estimated number. How can we get the estimated number? To get the estimated number we can use the np i, what is n? n is the number of samples or number of device put on test, a number of devices, number of times you have tried to observe. So, out of total number of samples, within this interval, what is the failure probability that is our p i.

So, number of device put on test multiply by failure probability gives us the expected number of failure in the interval, this we know from the binomial distribution, by for binomial distribution expected number of failure is n into p i, where p is the failure probability. So, n is the total devices which are put on test. How can we get this failure probability for the interval? That is nothing but the CDF at higher, right side of the interval, and CDF at the right, left side of if I am interested in here. So, this is my a i, this is my a i minus 1, I want to know the probability of this. So, this will be f. So, subtraction of F a i minus F a i minus 1 will give me the what is the probability of this interval.

Similarly, if I want I can write as R a minus R a minus 1, a i minus 1. Because, whatever is the difference between the failure probability, same will be the, negative will be the difference in the reliability. Because reliability is a decreasing function, and failure probability is a increasing function. So, F a i minus F a minus 1 will give me a positive value, but for reliability, it will be the negative of that. So, we have to take R a i minus 1 and R a i, if you take that will be same as this one.

(Refer Slide Time: 10:08)

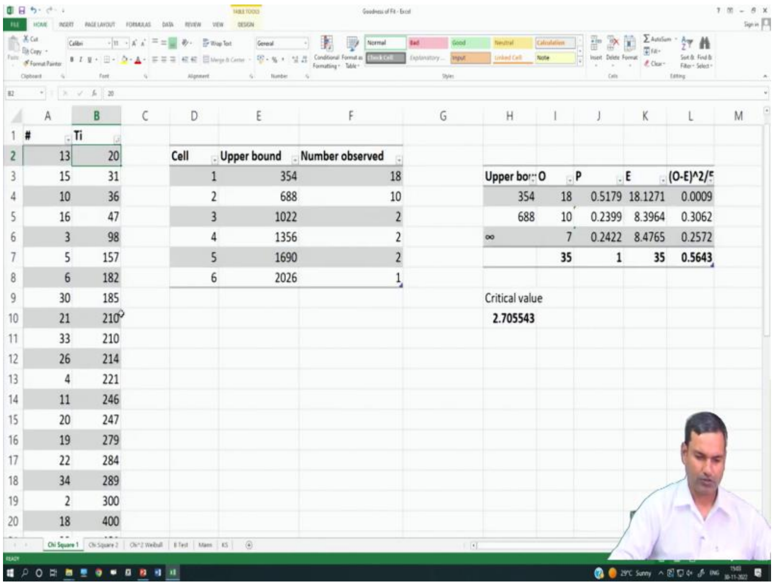
Example

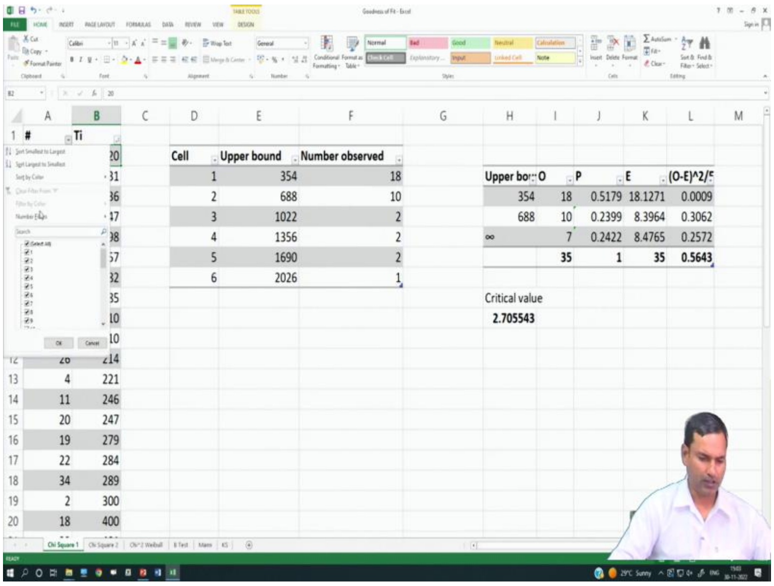
- 35 failure points are grouped into six cells. The MLE for parameter λ is 0.00206

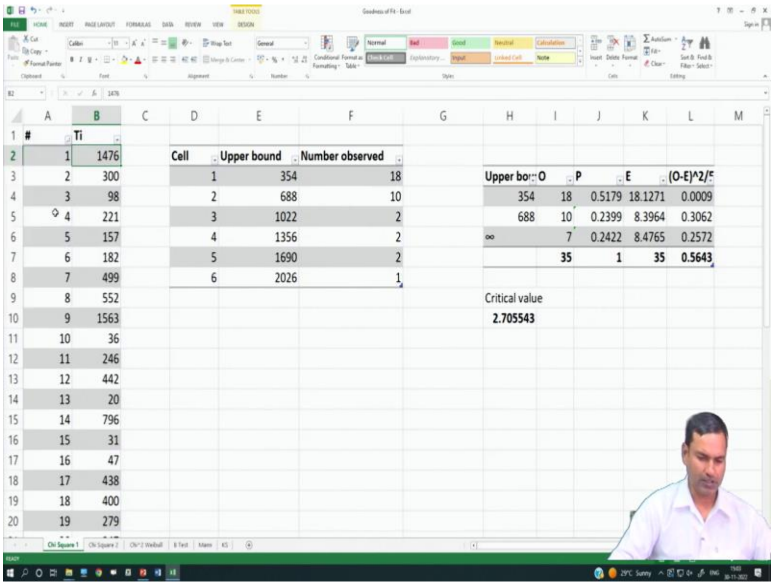
Cell	Upper bound	Number observed
1	354	18
2	688	10
3	1022	2
4	1356	2
5	1690	2
6	2026	1

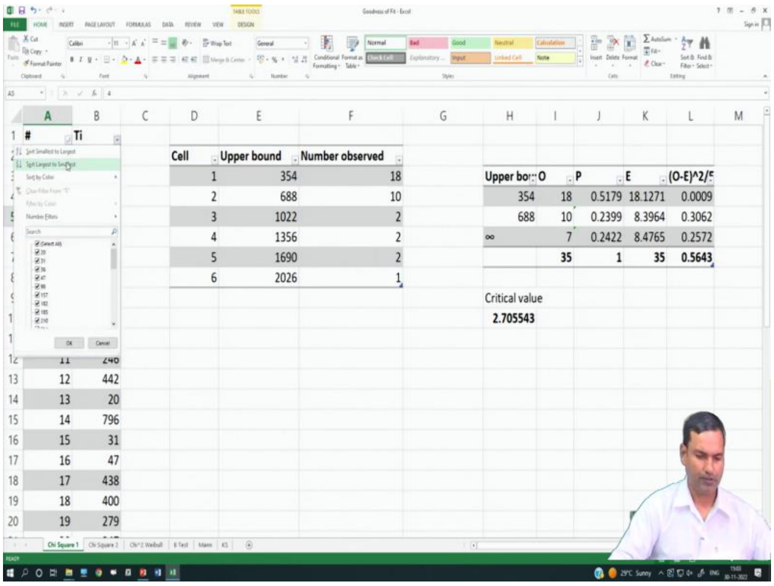
- The cells 3-6 are combine together so that each cell count is at least 5.
- As number of intervals =3 and number of parameters =1, Degree of freedom for Chi square distribution = 3-1-1 = 1
- Critical value for level of significance 0.1: $\chi^2_{0.1,1} = 2.71$
  - As statistic value 0.5643 is lower than critical value, the null hypothesis is accepted.

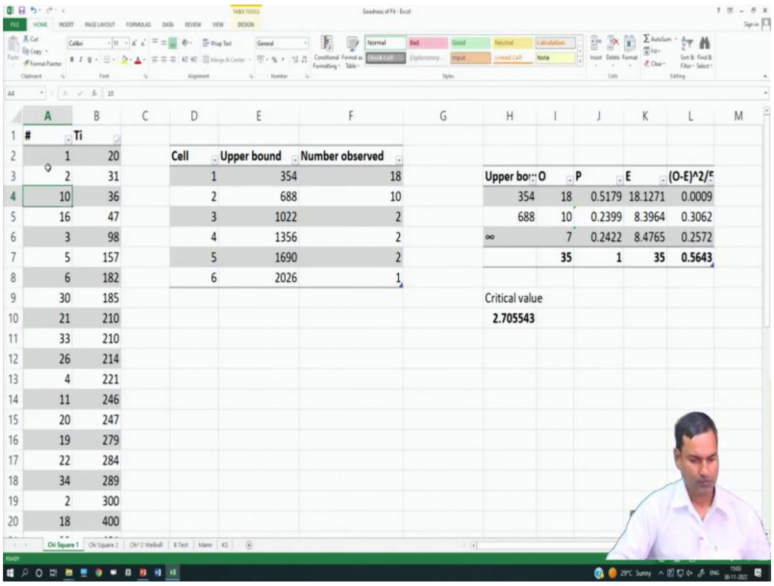
Upper bound	O	P	E	(O-E)^2/E
354	18	0.5179	18.1271	0.0009
688	10	0.2399	8.3964	0.3062
∞	7	0.2422	8.4765	0.2572
	35	1	35	0.5643

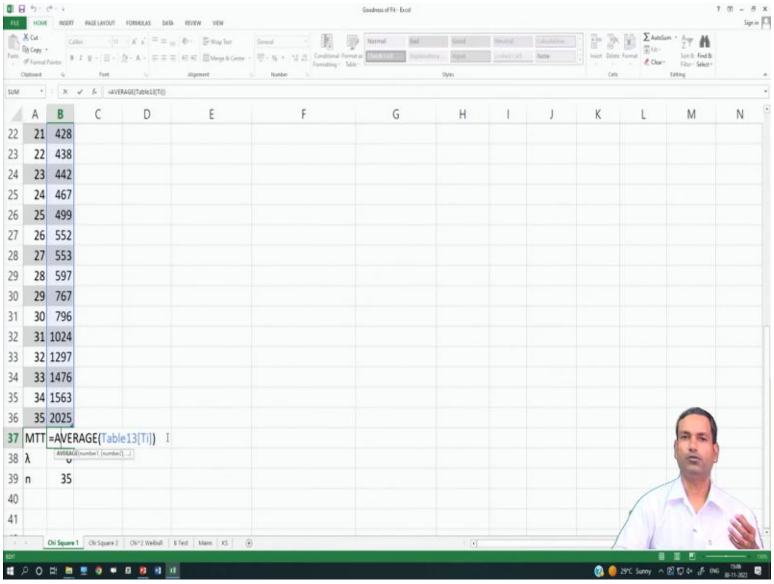




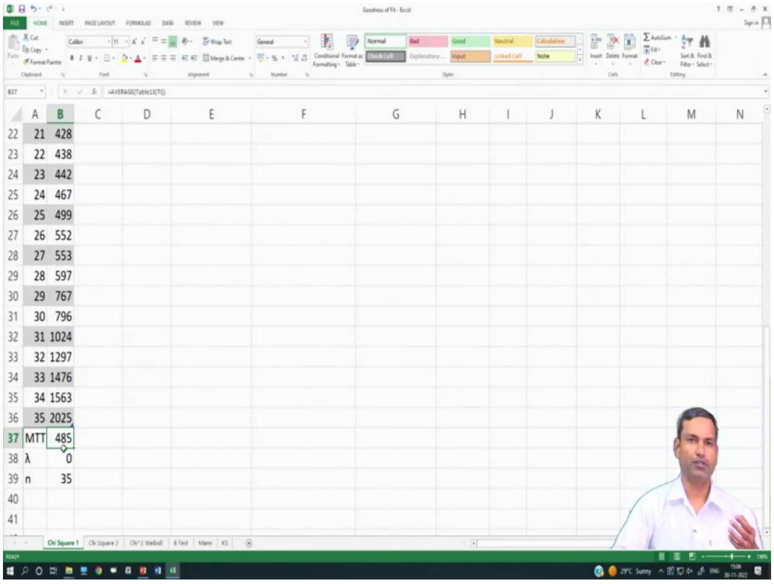

Goodness of Fit - Excel
Cell
Upper bound
Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bo O P E (O-E)^2/E
354 18 0.5179 18.1271 0.0009
688 10 0.2399 8.3964 0.3062
∞ 7 0.2422 8.4765 0.2572
35 1 35 0.5643
Critical value
2.705543
#
Ti
1 1476
2 300
3 98
4 221
5 157
6 182
7 499
8 552
9 1563
10 36
11 246
12 442
13 20
14 796
15 31
16 47
17 438
18 400
19 279
Chi Square 1


Goodness of Fit - Excel
Cell
Upper bound
Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bo O P E (O-E)^2/E
354 18 0.5179 18.1271 0.0009
688 10 0.2399 8.3964 0.3062
∞ 7 0.2422 8.4765 0.2572
35 1 35 0.5643
Critical value
2.705543
#
Ti
12 442
13 20
14 796
15 31
16 47
17 438
18 400
19 279
Chi Square 1

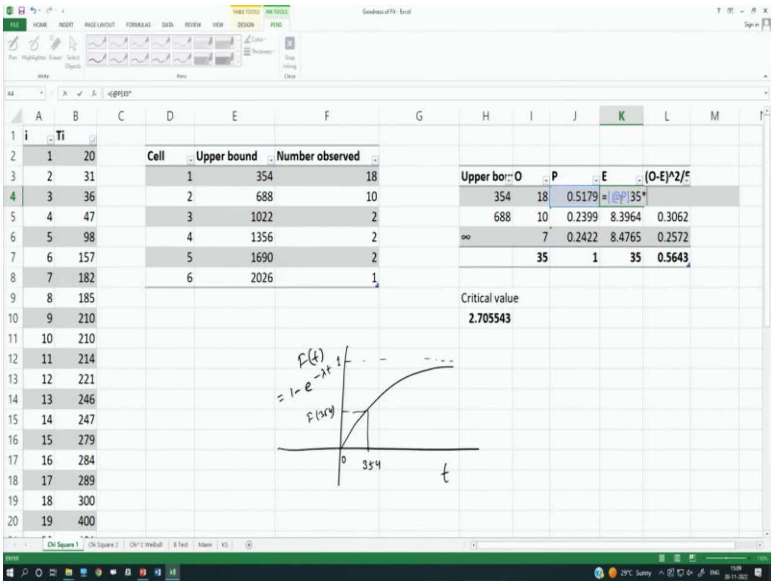

Goodness of Fit - Excel
Cell
Upper bound
Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bo O P E (O-E)^2/E
354 18 0.5179 18.1271 0.0009
688 10 0.2399 8.3964 0.3062
∞ 7 0.2422 8.4765 0.2572
35 1 35 0.5643
Critical value
2.705543
#
Ti
1 20
2 31
10 36
16 47
3 98
5 157
6 182
30 185
21 210
33 210
26 214
4 221
11 246
20 247
19 279
22 284
34 289
2 300
18 400
Chi Square 1



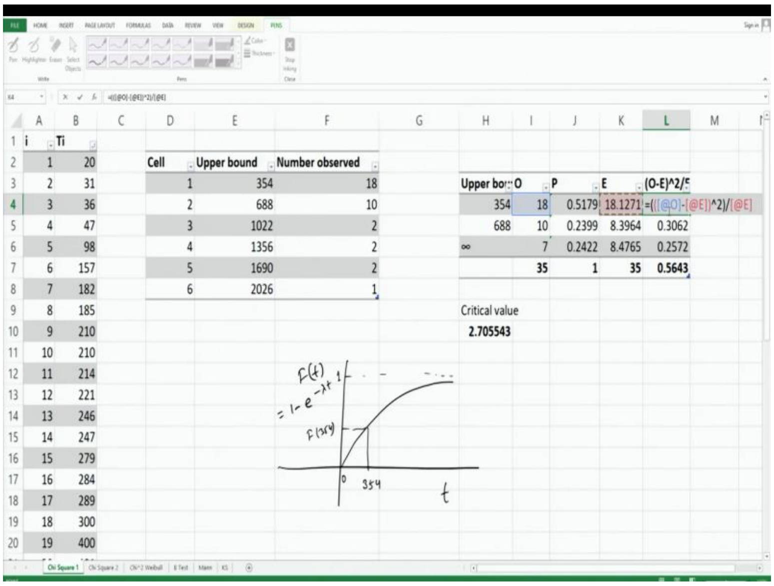
Goodness of Fit - Excel
=AVERAGE(Table13[Ti])
21 428
22 438
23 442
24 467
25 499
26 552
27 553
28 597
29 767
30 796
31 1024
32 1297
33 1476
34 1563
35 2025
MTT =AVERAGE(Table13[Ti])
λ
n 35
Chi Square 1

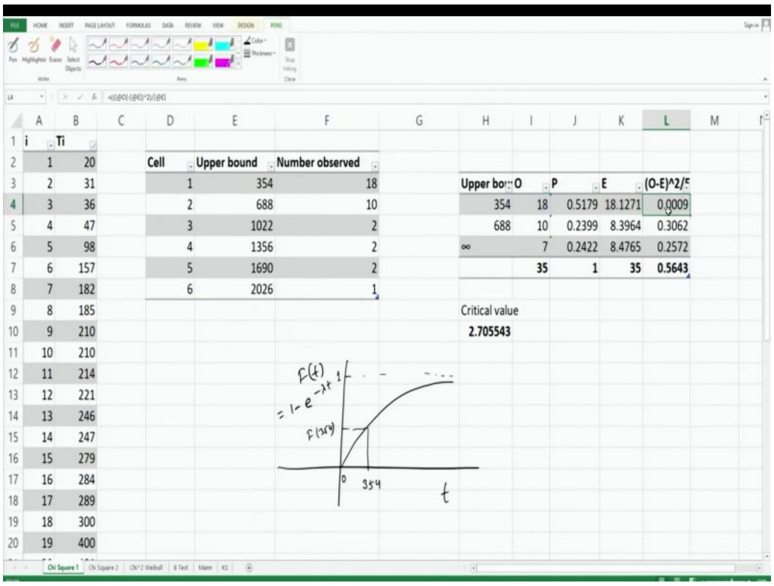


Goodness of Fit - Excel
=AVERAGE(Table13[Ti])
21 428
22 438
23 442
24 467
25 499
26 552
27 553
28 597
29 767
30 796
31 1024
32 1297
33 1476
34 1563
35 2025
MTT 485
λ 0
n 35
Chi Square 1



Goodness of Fit - Excel
i Ti
1 20
2 31
3 36
4 47
5 98
6 157
7 182
8 185
9 210
10 210
11 214
12 221
13 246
14 247
15 279
16 284
17 289
18 300
19 400
Cell Upper bound Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bound O P E (O-E)^2/E
354 18 0.5179 =[@P]35*
688 10 0.2399 8.3964 0.3062
∞ 7 0.2422 8.4765 0.2572
35 1 35 0.5643
Critical value
2.705543
F(t)
= 1- e^-λt
F(354)
354
t
Chi Square 1





So, let us try to see one example here, we are focusing more on exponential distribution because in practically when you use for your cloud project etcetera, exponential distribution is the one which is primarily used. So, let us say we have certain data, this data is as shown here in Excel sheet. So, this is my data, here we have the, actually it was unarranged way, it was like this, then I arranged it for T i, that is smallest to largest, and then if I put this 1, 2, 3 then this becomes my i, this I can say is the i. Now here so, how many failures we have here, we have the 35 failures, and for 35 failures, we have the time to failure data.

Now, I want to fit this data to the exponential distribution, and I want to my problem is, what I want to know, I want to know that whether this fitting the exponential distribution is good or not, whether exponential distribution is applicable for this data or not. So, for that we actually have, if we were dividing as per the Sturgis formula, we actually are supposed to get the 6 possible classes. So, because that is but here what happens, if we use the 6 possible classes, then my classes 3, class 4, class 5, I only have 2-2 failures, and class 6 only 1 failure. But as we discussed earlier that we need minimum 5 data points for using the chi-square distribution in each interval.

So, that is why what we do rather than using all these separately, we will use them all together. That means, I will say that my intervals are like first interval is 354, second is 688 and third interval is infinity, that means from 60 all values larger than 688 that will be counted here. So, with that what will happen my values will be 2 plus, 2 plus, 2 plus, 1. So, 7 observations will find my file, fall in this region.

So, my first interval that is 0 to 354 has the 18 values here, if you see here, that 354 is up to here 300, then from 19 to our 10 value, so that means 19 to 28 are falling in the second region, that is second region is up to 688 because this is higher than 688. So, this becomes the next region, and higher than 688 is all falling in next interval, that is my third interval. So, here 18 times seven becomes my the number of observations in each interval.

I could not use this though in development of doing the initial analysis, I was supposed to divide it in these classes, but these classes are no longer applicable. So I have to further group the data. Now, what is the probability? If I want to know the probability for this. So to calculate the property first let us see with MLE how do we calculate the lambda.

So as we know what is the lambda here, if I want I can calculate the MTTF here, MTTF is because it is all the failure data, there is no sensor data. Since there is no sensor data, MTTF is nothing but the average, that is a summation of time to failure divided by n. So this is nothing but if you see this is nothing but the average, average of time to failure data. This gives me the MTTF, and what is lambda? Lambda is equal to 1 divided by MTTF.

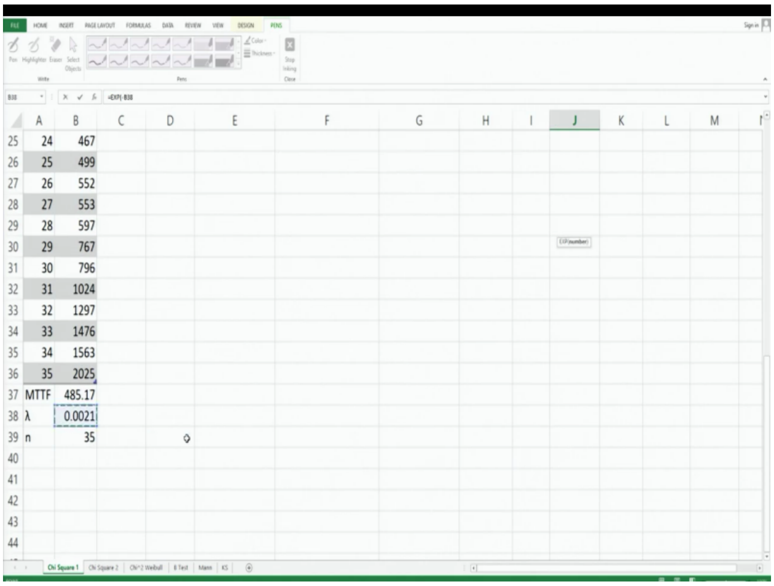
So same formula which we have developed earlier, we have used here, the only difference here is that the formula we have seen for sensor data also, but here is no censoring. So that is why the, we only, we do not have to add the sensor value here. All the failure time we add that will give, and take the average that will give us the MTTF, and inverse of the MTTF will give us the lambda. So, we get the lambda here and number of sample is 35.

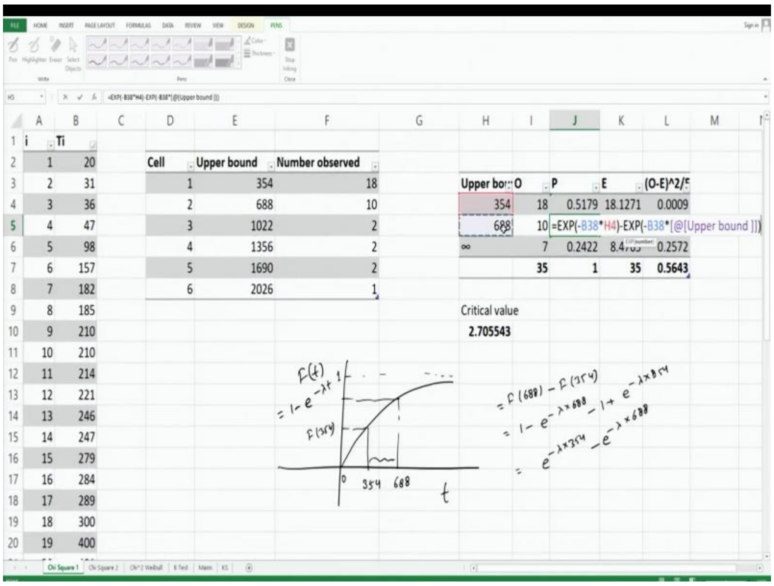
So, I want to now know the probability that my distribution is, what is the probability that my values or time will be 0 to 354. So, we know that I will use a pen here, we know that if I have this data t versus capital F t, what is capital F t in this case? That is 1 minus e to the power minus lambda t. So, capital F t will be starting from 0 and it may be increasing further, sorry not like this, but with somethings like this, it will be touching 1 at the time t equal to infinity.

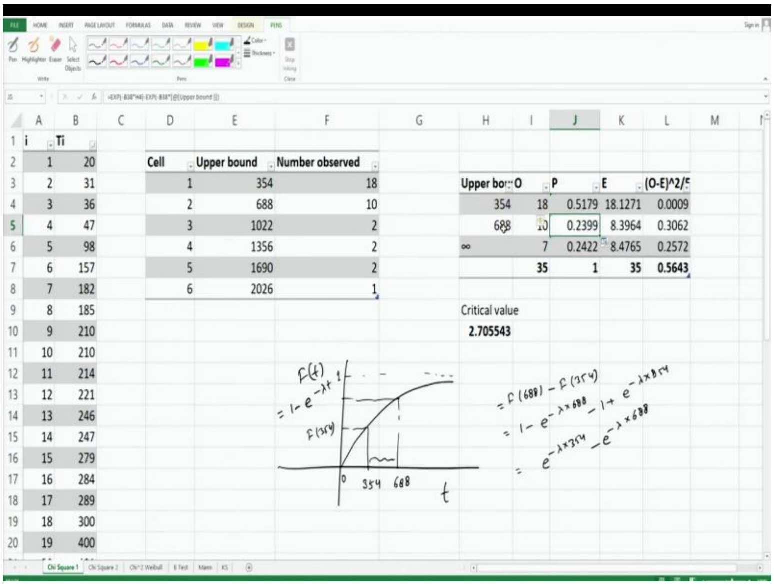
Now, I want to know that at 354, what is the probability that values are lying in this region that means, I want to know F at 354, that is the probability that my value will be from 0 to 354. So, how can I calculate here? This value I can calculate as 1 minus exponential of minus lambda, lambda value we have calculated here, lambda into t is how much, 354 default.

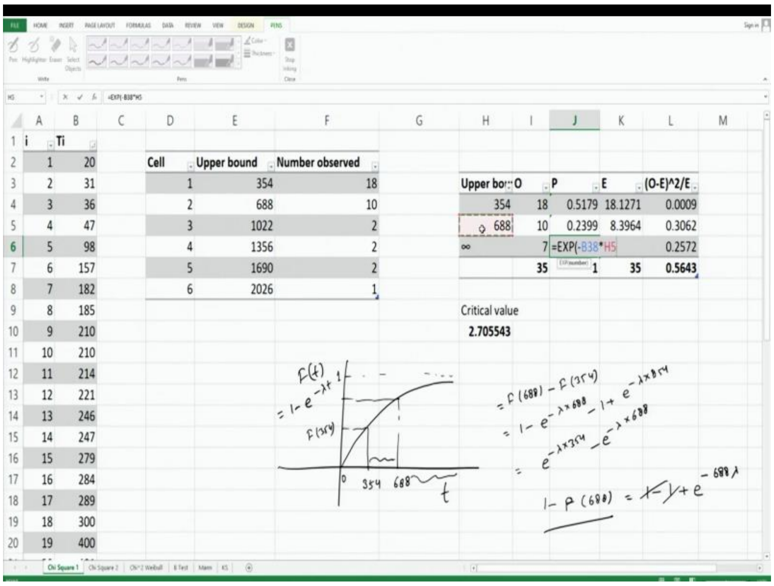
So, this becomes my probability and how many total samples I put, 35. So, if I multiply 35 to this is equal to this multiply by 35 I get 18 failures, expected failures are 18, and how many I observed? I observed also 18. So, if I want to calculate that formula which I have O minus O y minus E i, whole square divided by E. So, O is18, E is 18.12 and divided by E, if I use this is equal to O minus E whole square divided by E, same formula we have used I get the error here.

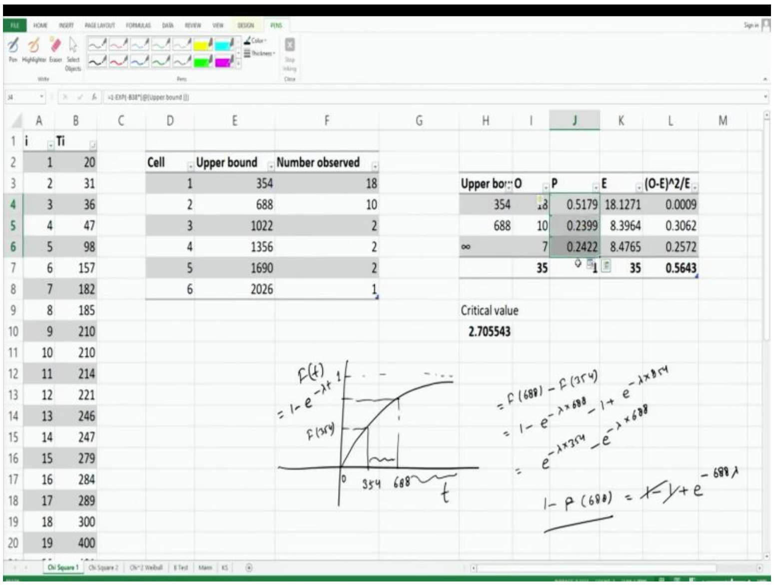
(Refer Slide Time: 18:04)

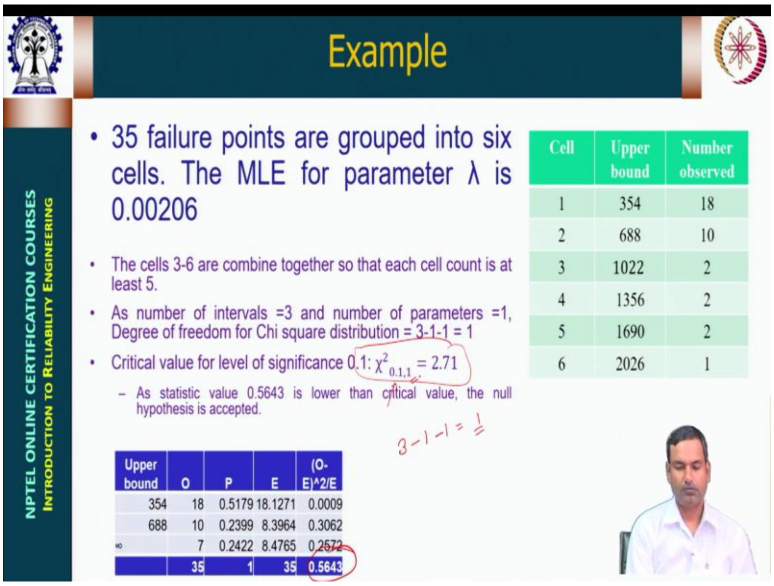



=EXP(-B38*H4)-EXP(-B38*[@[Upper bound]])
Cell
Upper bound
Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bound
O
P
E
(O-E)^2/E
354 18 0.5179 18.1271 0.0009
688 10 =EXP(-B38*H4)-EXP(-B38*[@[Upper bound]])
∞ 7 0.2422 0.2572
35 1 35 0.5643
Critical value
2.705543
Chi Square 1
Chi Square 2
Chi^2 Weibull
Main
KS


Cell
Upper bound
Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bound
O
P
E
(O-E)^2/E
354 18 0.5179 18.1271 0.0009
688 10 0.2399 8.3964 0.3062
∞ 7 0.2422 8.4765 0.2572
35 1 35 0.5643
Critical value
2.705543
Chi Square 1
Chi Square 2
Chi^2 Weibull
Main
KS


=EXP(-B38*H5
Cell
Upper bound
Number observed
1 354 18
2 688 10
3 1022 2
4 1356 2
5 1690 2
6 2026 1
Upper bound
O
P
E
(O-E)^2/E
354 18 0.5179 18.1271 0.0009
688 10 0.2399 8.3964 0.3062
∞ 7 =EXP(-B38*H5 0.2572
35 1 35 0.5643
Critical value
2.705543
Chi Square 1
Chi Square 2
Chi^2 Weibull
Main
KS





Now, let us see the second one, for second one is I am talking about here the range for in this range that is 354 to 688. So, within this range, I want to know the failure probability. So, with this range if I want to know the failure probability, that will be equal to F 688 minus F 354 that will be the probability that value will be in this range. So, this probability I can get like 1 minus e to the power minus lambda into 688 minus 1 plus e to the power minus lambda into 354. Sorry 354. So or I can say e to the power minus lambda into 354, minus e to the power minus lambda into 688.

So, if I use the same thing here so, I various ways I can do, see this value is already available with me, 1 minus e to the power minus this is already available. So, one way I can do is I can calculate the F at this value 688, and subtract the F at 354. That way I can do or I can calculate like this directly that is equal to exponential of minus lambda, lambda is this value, minus lambda into. Now, I have to take the previous value, 354 minus exponential of minus lambda into the current range, that is 688. If I do that my value will be here.

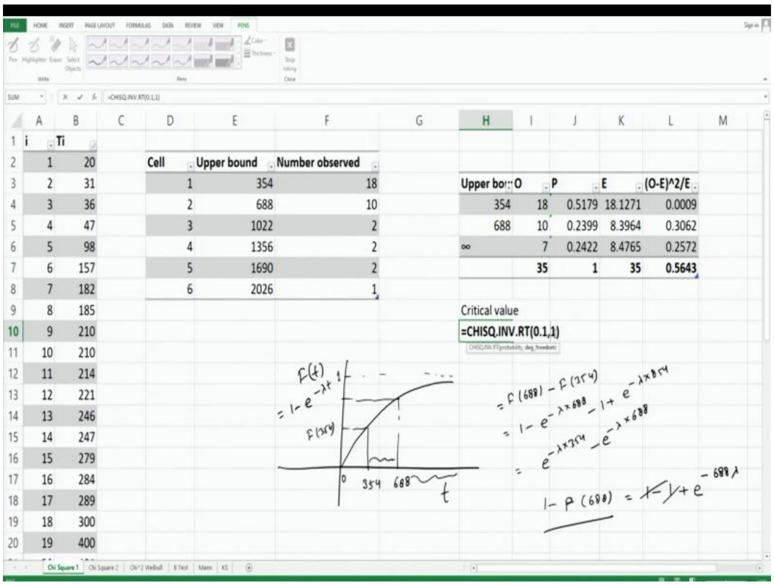
So, my 2.399 becomes the value, and how much will be the expected failure? Expected failure is total 35 units of put on test multiply by quality of failure, I will get the expected failure, that is it. And how much will be the contribution of this towards error, that will be O minus E squared divided by E. So, same thing I will get it here.

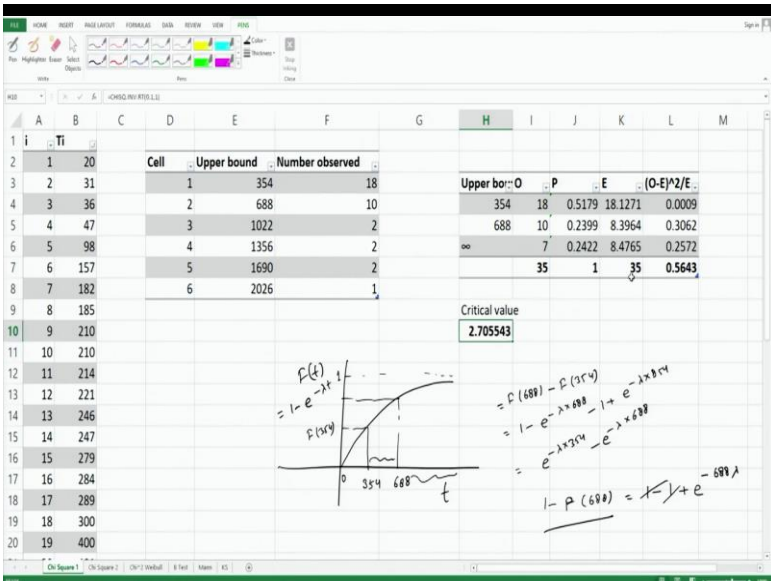
Now, in last region is 7 failures, for last region if you see that means everything above 688. So, that means, total probability will be from 688 onwards to infinity, at infinity we have probability 1. So, 1 minus F at 688, if you do we will get the this probability, and this probability is 1 minus 1 F 68 is e to the power, 1 minus e to the power minus 688 into lambda. So, that means 1 and 1 will get cancelled so, I can use this probability as the this will be equal to nothing but there is another way that I subtract earlier probabilities from 1, I will get this.

Because total has to be 1 or this will be same as exponential or minus lambda into our previous value 688, and summation of the probabilities will always be 1, because total probability 1, we have divided into few intervals. So, expected number of failures will also be total 35, which is divided into different intervals, and how much is the O minus E square divided by E, we are getting for all and this becomes a summation of all these values. So, this becomes our chi-square statistic. So, our chi-square statistic here come out to be 0.5643. So, this 0.5643 is the chi-square statistic which we have calculated.

Now, what is our criteria this chi-square value should be less than our critical value, and how much is of a critical value? To calculate critical value, this has one parameter so, we have 3 intervals. So, 3 minus 1, 1 is lost already and because this is having one parameters, so, for one parameter one another degree of freedom is lost. So, we have total one degree of freedom. So, we have to get the chi-square value for 0.1 is the critical level. So, that means, critical level 0.1 is we have the 90 percent probability or we can say right hand side 0.1, 10 percent, we are leading on the right hand side. So, here 0.1 and our this value is 1.

(Refer Slide Time: 23:13)





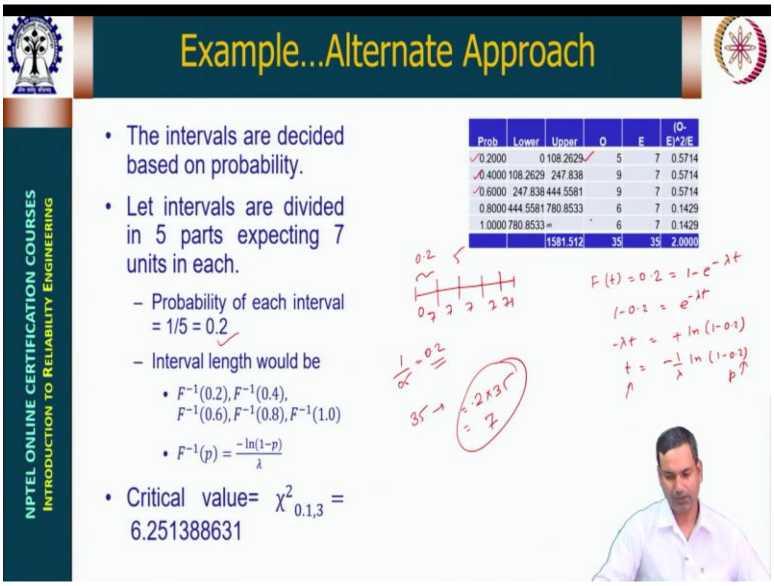
Example

- 35 failure points are grouped into six cells. The MLE for parameter λ is 0.00206
  - The cells 3-6 are combine together so that each cell count is at least 5.
  - As number of intervals =3 and number of parameters =1, Degree of freedom for Chi square distribution = 3-1-1 = 1
  - Critical value for level of significance 0.1: $\chi^2_{0.1,1} = 2.71$
    - As statistic value 0.5643 is lower than critical value, the null hypothesis is accepted.

Cell	Upper bound	Number observed
1	354	18
2	688	10
3	1022	2
4	1356	2
5	1690	2
6	2026	1

Upper bound	O	P	E	(O-E)^2/E
354	18	0.5179	18.1271	0.0009
688	10	0.2399	8.3964	0.3062
∞	7	0.2422	8.4765	0.2572
	35	1	35	0.5643

NPTEL ONLINE CERTIFICATION COURSES
INTRODUCTION TO RELIABILITY ENGINEERING



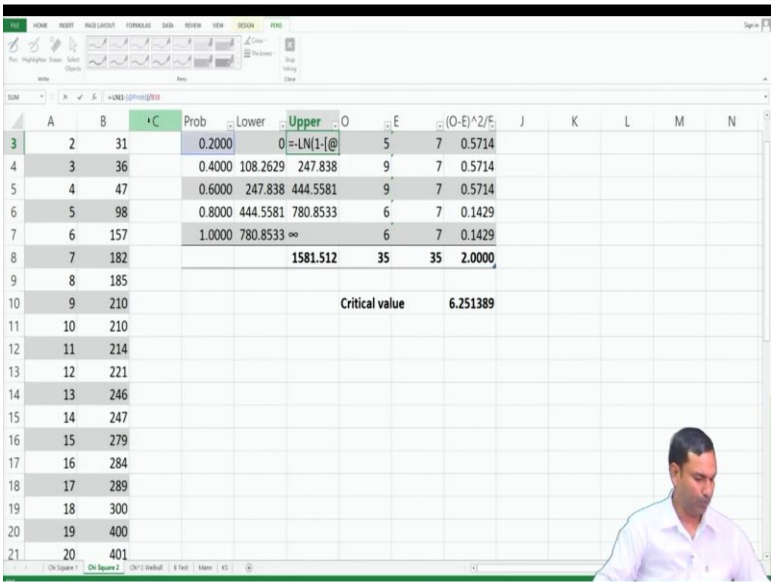
So, this also we can get from Excel sheet. So, critical value if you want to know, what is this critical value? Critical value is equal to chi-square inverse value towards right hand side because we are leaving 0.1 data towards right. That means 0.1 and what is the degree of freedom? Degree of freedom is 1. So, we are able to get the critical value is 2.7. So, 2.7 is higher than the 0.5643. And what is the significance level here that is the 0.1 or we can say the confidence bound is 90 percent confidence bond. So, here we are this value is higher than 0.5643. So, we can say that our null hypothesis is accepted. So, null hypothesis get accepted here and out chi-square values.

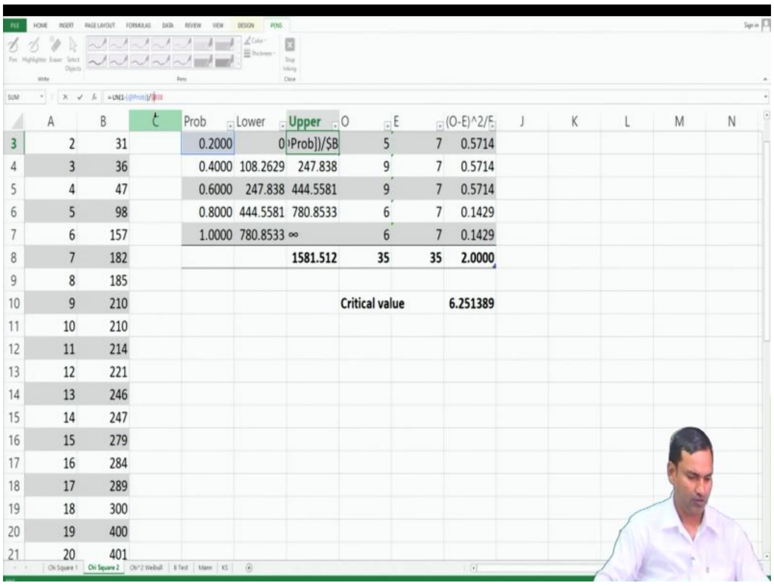
So, this is the procedure which we can follow to get this value. There is another way as we have seen here, because of this grouping of the data, what happened because here intervals initially were of same size 354 into 2, into 3, into 4 like that. So, because of equal intervals, what was happening number of failures, we are having different in different cases, but our criteria is that we want to have the minimum 5 failures in each region. So, to do that, we have to divide this into a different bin size. So, what we can have we can have these intervals which are not uniform, but in every interval we will try to have the similar number of failures. So, let us see if we had the 35 failures here, if you divide the 35 failures into 5 different intervals, then 7-7 failures will supposed to be there in the each interval.

So, if you want to do that, what will happen here that total probability from 0 to 1, we have total probabilities from 0 to 1, which we have to divide in 5 intervals. So, that means, each interval will have a probability of 0.2, 1 divided by 5, will be equal to 0.2 that means, we have to divide in 20 percent of 0.2 probability intervals, then only we can divide the data into 5 equal numbers. So, we have the 35 data points. So, in each interval 7-7 points are supposed to appear, because probability is same for all 0.2 probability. So, 0.2 into 35 will give me 7. So, we are able to get the 7 failures, expected failures in the each interval, but actual failures may be different, actual failures we have to get from the data.

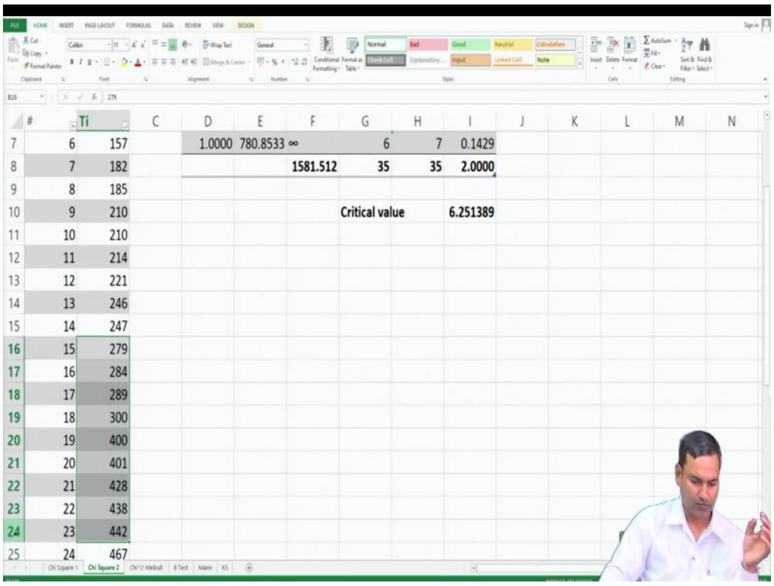
So, what we do here we will divide the probability of interval into cells. So, that way we will get the intervals first, and in each interval then we will see how many failures are lying actually, and once we see that, then we will compare the same. So, here probability is 0.2, 0.4. So, we get the corresponding time. So, if I want to know the time let us say for we know F of t is let us say 0.2 and that is equal to 1 minus e to the power minus lambda into t. So, I want to know the t. So, t will be equal to 1 minus 0.2 will be equal to e to the power minus lambda t. So, lambda t, minus lambda t would be equal to minus ln of 1 minus 0.2 and t will be equal to. So, this is plus and t will be equal to minus 1 upon lambda ln of 1 minus 0.2. So, for 0.2 I will get similarly for 0.4 I want to get, for 0.4, also I will get. So, for all the values when I change here, that is probability value, I will get the corresponding upper bound on the interval. So, here if you see that interval comes out to be this I have done already in Excel I will show you.

(Refer Slide Time: 27:38)

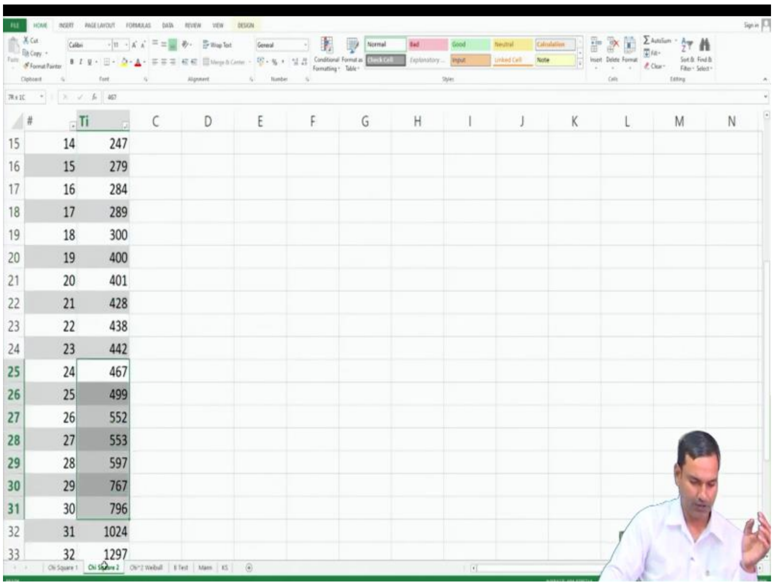




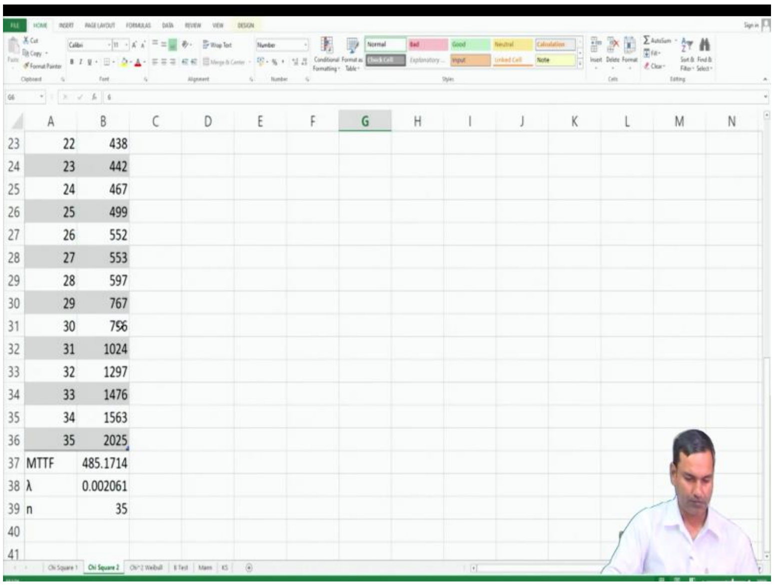
	A	B	C	Prob	Lower	Upper	O	E	(O-E)^2/E
3	2	31		0.2000	0	Prob])/$B	5	7	0.5714
4	3	36		0.4000	108.2629	247.838	9	7	0.5714
5	4	47		0.6000	247.838	444.5581	9	7	0.5714
6	5	98		0.8000	444.5581	780.8533	6	7	0.1429
7	6	157		1.0000	780.8533	∞	6	7	0.1429
8	7	182				1581.512	35	35	2.0000
9	8	185							
10	9	210				Critical value			6.251389
11	10	210							
12	11	214							
13	12	221							
14	13	246							
15	14	247							
16	15	279							
17	16	284							
18	17	289							
19	18	300							
20	19	400							
21	20	401							



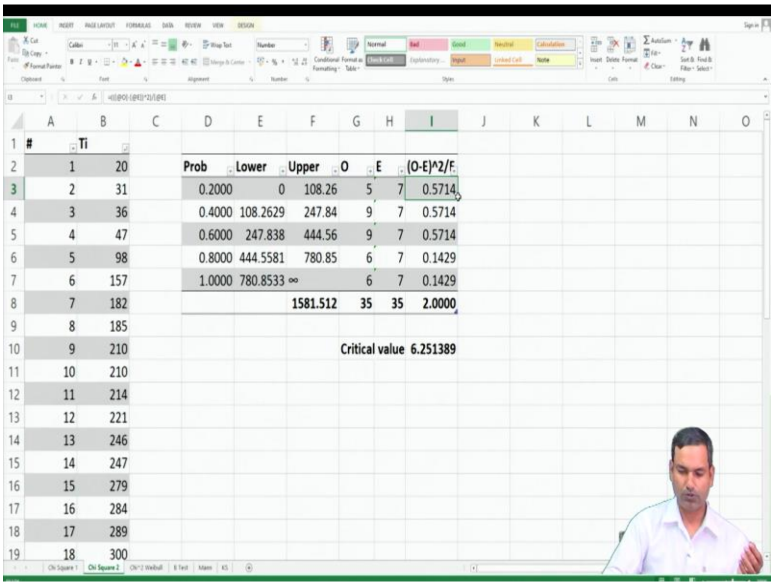
	#	Ti	C	D	E	F	G	H	I
7	6	157		1.0000	780.8533	∞	6	7	0.1429
8	7	182				1581.512	35	35	2.0000
9	8	185							
10	9	210				Critical value			6.251389
11	10	210							
12	11	214							
13	12	221							
14	13	246							
15	14	247							
16	15	279							
17	16	284							
18	17	289							
19	18	300							
20	19	400							
21	20	401							
22	21	428							
23	22	438							
24	23	442							
25	24	467							



	#	Ti
15	14	247
16	15	279
17	16	284
18	17	289
19	18	300
20	19	400
21	20	401
22	21	428
23	22	438
24	23	442
25	24	467
26	25	499
27	26	552
28	27	553
29	28	597
30	29	767
31	30	796
32	31	1024
33	32	1297



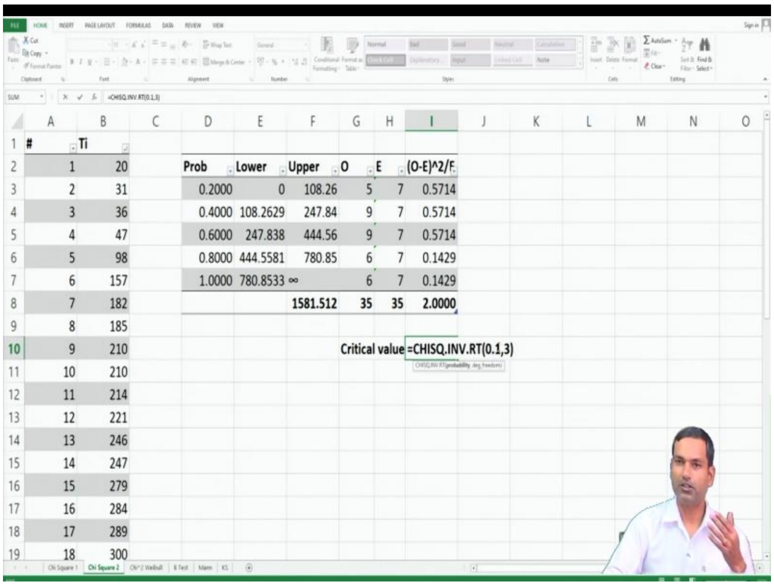
	A	B
23	22	438
24	23	442
25	24	467
26	25	499
27	26	552
28	27	553
29	28	597
30	29	767
31	30	796
32	31	1024
33	32	1297
34	33	1476
35	34	1563
36	35	2025
37	MTTF	485.1714
38	λ	0.002061
39	n	35
40		
41		



	A	B
1	#	Ti
2	1	20
3	2	31
4	3	36
5	4	47
6	5	98
7	6	157
8	7	182
9	8	185
10	9	210
11	10	210
12	11	214
13	12	221
14	13	246
15	14	247
16	15	279
17	16	284
18	17	289
19	18	300

Prob	Lower	Upper	O	E	(O-E)^2/E
0.2000	0	108.26	5	7	0.5714
0.4000	108.2629	247.84	9	7	0.5714
0.6000	247.838	444.56	9	7	0.5714
0.8000	444.5581	780.85	6	7	0.1429
1.0000	780.8533	∞	6	7	0.1429
		1581.512	35	35	2.0000

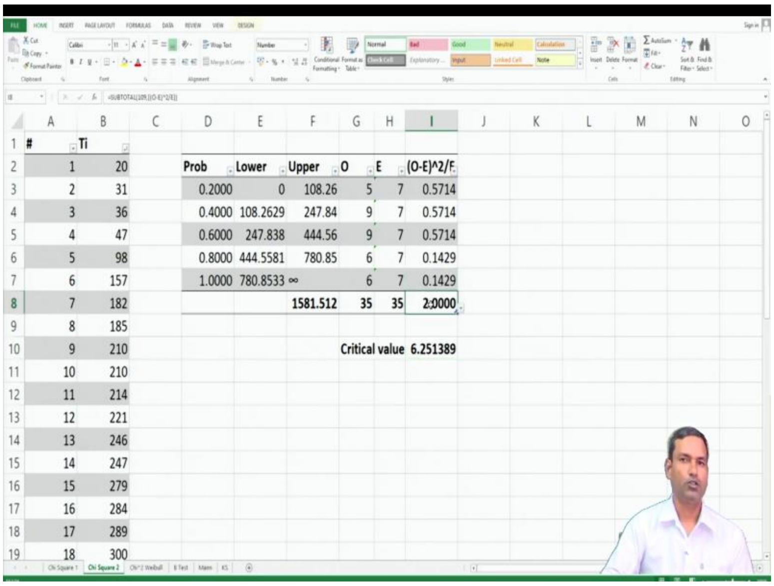
Critical value 6.251389



	A	B
1	#	Ti
2	1	20
3	2	31
4	3	36
5	4	47
6	5	98
7	6	157
8	7	182
9	8	185
10	9	210
11	10	210
12	11	214
13	12	221
14	13	246
15	14	247
16	15	279
17	16	284
18	17	289
19	18	300

Prob	Lower	Upper	O	E	(O-E)^2/E
0.2000	0	108.26	5	7	0.5714
0.4000	108.2629	247.84	9	7	0.5714
0.6000	247.838	444.56	9	7	0.5714
0.8000	444.5581	780.85	6	7	0.1429
1.0000	780.8533	∞	6	7	0.1429
		1581.512	35	35	2.0000

Critical value =CHISQ.INV.RT(0.1,3)



So, here like let us say if we use the same approach in another way, so, we have the probability 0.2. So, for 0.2 how much will be my time at which the probability is 0.2, the time is minus ln of 1 minus probability, and divided by lambda. How much is lambda, lambda is same as what we calculated earlier. I am using dollar signs here so, that this does not change and I get the same value. So, as you see here that we get the 108 here, I will use same number decimal point here. So, 108.26 is the point at which I will have the 20 percent probability and 40 percent probability will be at the point 247, 60 percent as per my fitted distribution, because I have fitted to exponential distribution. So, at 108.26, the probability cumulative for this 0.2. At 247.82 cumulative probability is 0.4, same way.

So, I get this upper bound on the intervals and lower bound is nothing but the previous value 0 to 108 fault is 2, 0.2. 102 to 247 point again total property will be 0.2, cumulative probability will be 0.4. We are able to get this data. Now, let us see that for up to 108 how much data is falling. So, we will see here, 5 data points are falling in the up to 108. Then from 108 to 247 108 to 247. Like up to here. I have the 9 failures here, 7 to 15. Then 247 to 444 about 257 to 444, 444 means up to here. That means of 16 to 24 that is again 9 data, if you see count is 9. Then 444 to 780 that means from here to 780, 780 is up to here 767, sorry. So, that is again my 6 data points. So, I have the 6 data points, and then all above 780 that means, here again I have the 6 data points, so, these are my actual observations, and how much is the observations which I have in each interval that is the 0.2 into 35 that is 7.

So, expected values, expected observations are near same, because I have divided the data as per the distribution. So, expected observation supposed to be here is 7-7 in each interval, but actual observations which I have is here like 5, 9. So, now, the formula is same O minus E squared divided by E, same is calculated O5 minus 7 divided by O5-7 square, that is 2 square 4 divided by 7.

So, same way we are able to calculate this here, once you get it here, we can get the total summation of this. So, this becomes our chi-square statistic, whatever we have used earlier same chi-square statistic, but there is a difference now, the difference here is I have the 5 intervals, here earlier I was having only 3 intervals here.

So, since 5 intervals are here my degrees of freedom will be 5, 1 is lost in estimation, then 1 is lost in the single parameter estimation, so, 2 degrees are lost. So, that means, I have the 3 degrees of freedom. So, I will get the chi-square value inverse value for 10 percent confidence or 10 percent that is significance level, and the number of degrees of freedom is 5 minus 1 minus 1 that is 3. So, we get this, so, my critical value is 6.25. So, this value, 2 value is quite low than critical value. So, that means we can say that my hypothesis that this data follows the exponential distribution is accepted. Or I can assume that this data is following the exponential distribution. We will stop here; we will continue our discussion in this direction with more distributions and tests. Thank you.