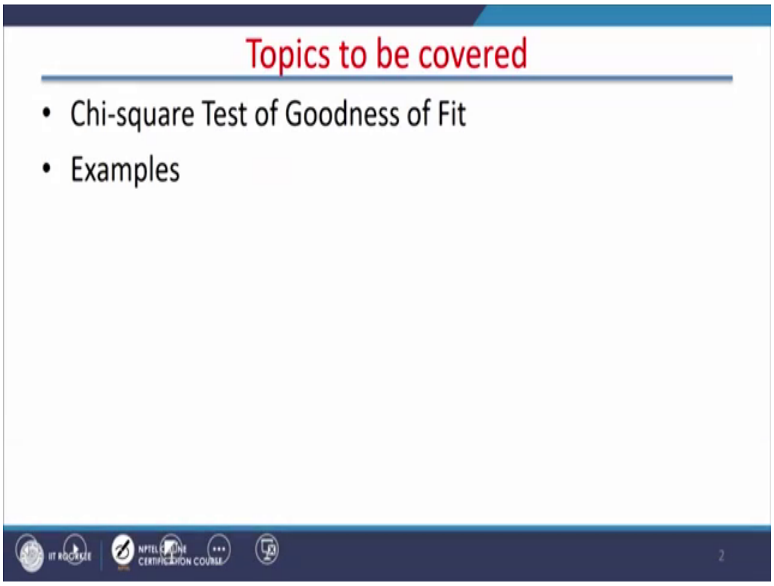
Mechanical Measurement Systems
Prof. Ravi Kumar
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee

Lecture – 09
Chi - square Test

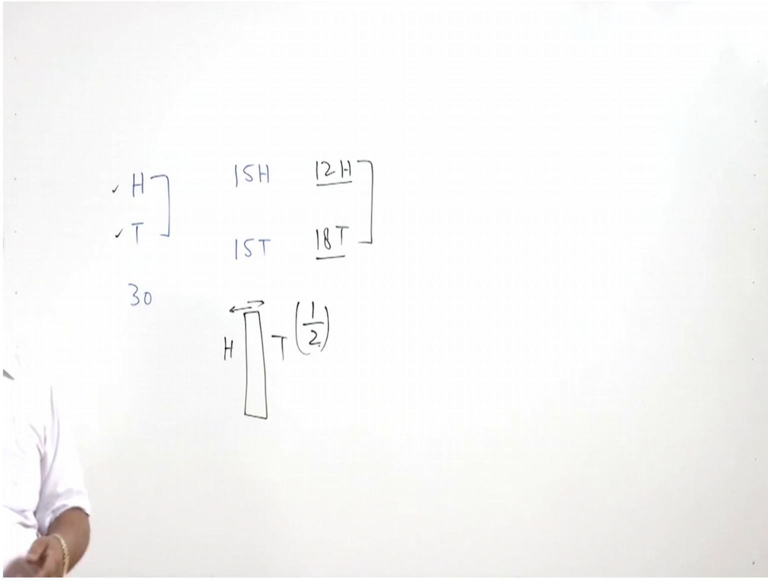
I welcome you all in this course on mechanical measurement systems. And today we will discuss the Chi-square test and we will solve certain example based on the chi-square test.

(Refer Slide Time: 00:29)



So, chi-square test is a test of goodness of it. I will give you an example suppose we toss a coin. So, there are two possibilities either we will get heads or we will get tails.

(Refer Slide Time: 00:45)

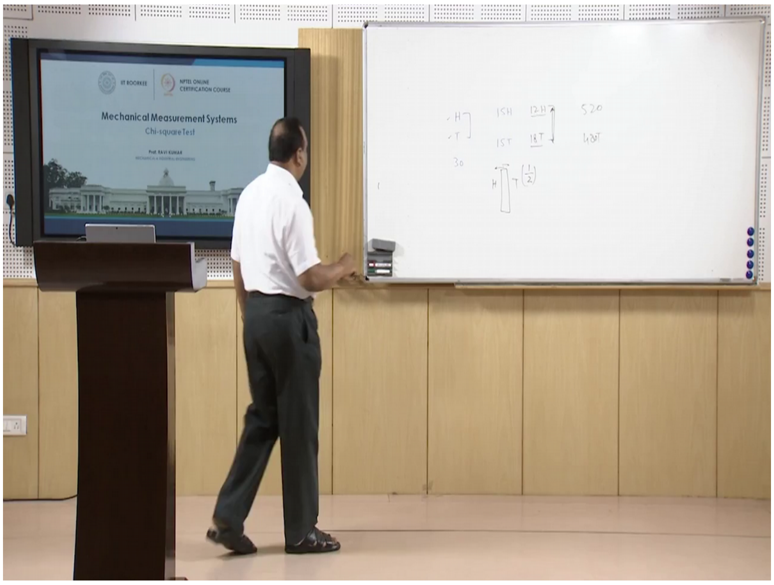


Suppose I toss coin for let us say 20 times or a 30 times if I toss coin for 30 times, ideally what should I get? I should get 15 heads and 15 tails because probability of getting heads is half and probability of getting tails is also half. So, half multiplied by 30, half multiplied by 30. But in actual practice what happens we will not get 15 heads we may get 12 heads and 1 tails. Now, the question is there any problem in the hypothesis or is there any problem with the coin.

The coin may have certain bias I mean it is it may be a weighted coin, weighted coin means suppose this is a side view of the coin is not balanced right or it is weighted towards this side towards of let us say head side head side then always we will be getting tails, so probability of getting tails will be more. If it is weighted towards tails then probability of getting heads should be more. So, it should be a ideally it should be a weightless coin or unbiased coin.

So, now, we are getting these results they are the output of I mean is there any problem with the coin or I mean the coil is biased or unbiased right if the coil is unbiased then if you are getting this 12 H heads and 8 tails sorry if you are getting 12 heads and 18 tails. Is it following the same law of probability that probability of getting each one is half, because 30 sets of a experiments is not sufficient we should do at least 1000 or infinite number of experiments perhaps, if we do 1000 number of experiments we may get 520 heads and 480 tails.

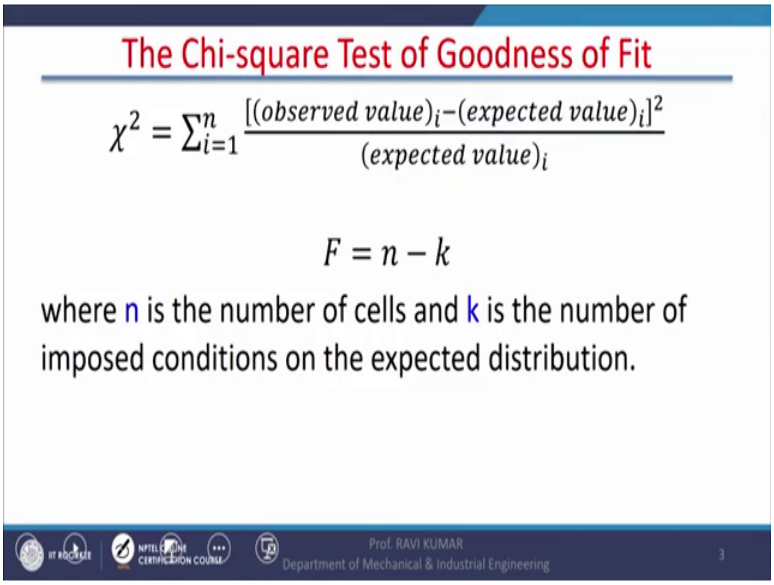
(Refer Slide Time: 02:54)



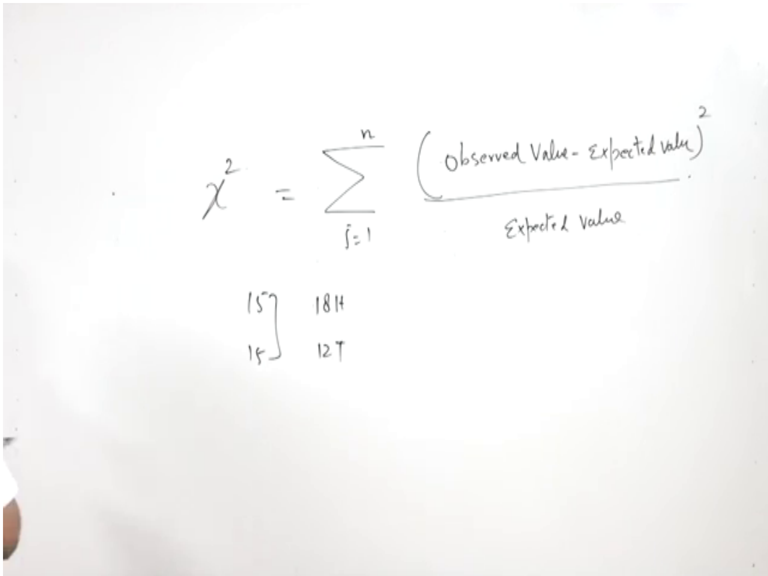
In that case also we not get a exact number of heads and tails. So, for conducting for having exact number of heads and tails we should conduct infinite number of experiments and that is not possible. I mean it is very time consuming and practically it is not possible to conduct infinite number of experiments.

So, if the coil is unbiased these observations do they verify this fact that we should have 15 heads or 50 or both the data we can say both the data setup for a both sets of data are following the same trend or not. That is the goodness of it and this is how now I will give you certain example then it will be very much clear to you. So, first of all we will start with the expression of chi-square.

(Refer Slide Time: 03:46)



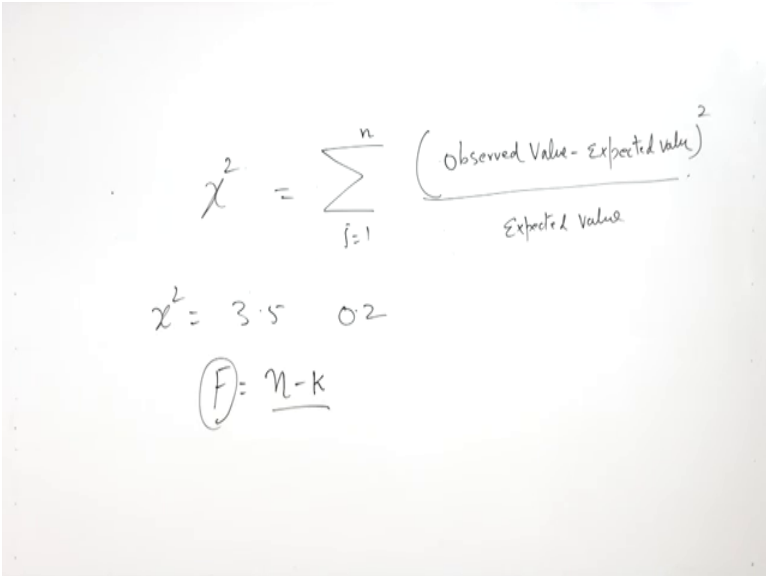
(Refer Slide Time: 03:59)



The value of chi-square can be calculated as the sum of observed value, observed value and a value minus expected value, expected value is what we get through the experiments sorry; this is the a value which we are expecting. In this case we are expecting 15 heads and 15 tails that is expected value.

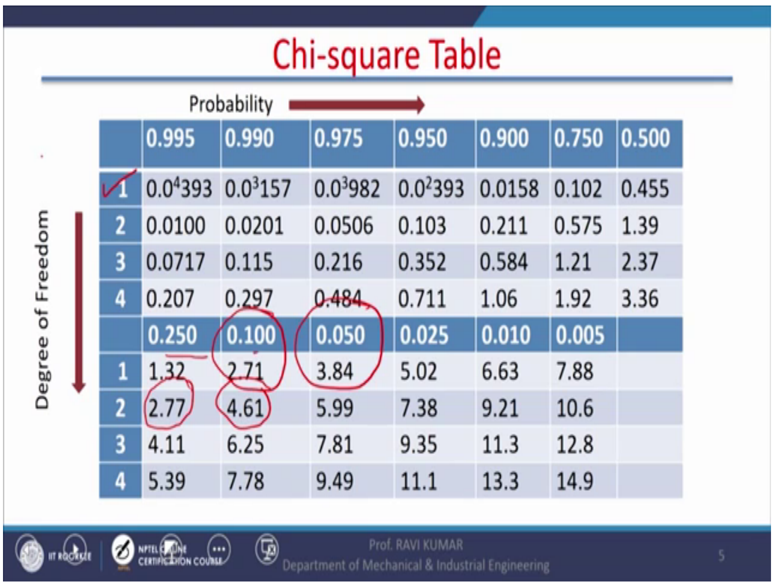
Observed value is a 18 heads and a 12 tails for example, out of 30 runs. So, expected value whole square divided by expected value and this sigma is of total sigma. This is i is equal to 1 to 10, right. Now, after calculating this chi-square value we may get certain figure let us say figure is 3.5, chi-square is 3.5 right or we may get 0.2 or any figure right.

(Refer Slide Time: 05:31)



After calculating this we will calculate the degree of freedom F, F is number of trials; suppose there are n number of trials. So, it is n minus restriction in the trial and this n minus k will be value of F. Then we have a table chi-square table I will show you on the screen.

(Refer Slide Time: 06:16)

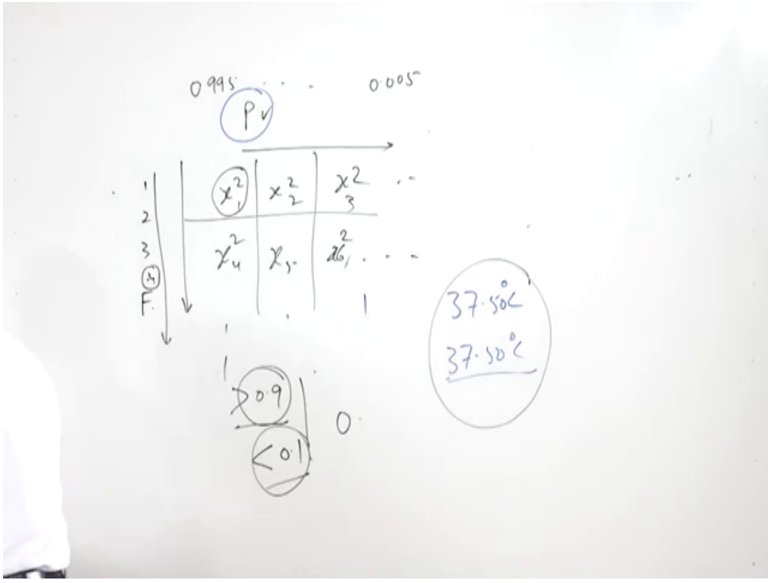


	0.995	0.990	0.975	0.950	0.900	0.750	0.500
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	0.102	0.455
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36

	0.250	0.100	0.050	0.025	0.010	0.005
1	1.32	2.71	3.84	5.02	6.63	7.88
2	2.77	4.61	5.99	7.38	9.21	10.6
3	4.11	6.25	7.81	9.35	11.3	12.8
4	5.39	7.78	9.49	11.1	13.3	14.9

This is the table where on one axis or one side of the table shows the variation of probability.

(Refer Slide Time: 06:32)



So, probability in this direction, the increasing order of F in the this direction, decreasing order of probability in this direction and for all these values of p and k chi-square values are given and so on, in both the directions right. I have taken a part of this table it is shown on the screen right it is starting from 0.995 and it is going up to the probability value is starting from 0.995 and it is going up to 005. So, it is starting from 0.995, it is going up to 0.005. Likewise degree of freedom can increase in this point starting from 1 2 3 4 four this.

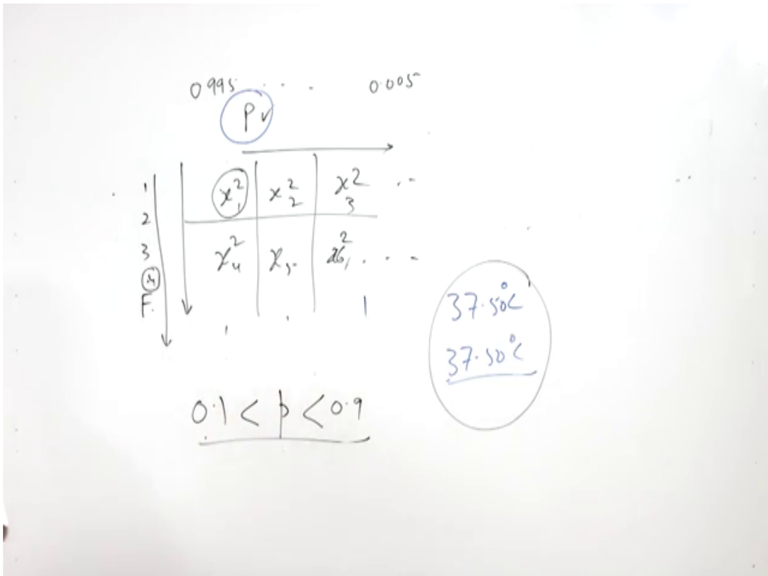
And we will have different chi-square value now for every case we have the value of chi-square we have already calculated, degree of freedom is also calculated. With the help of these two values, with the help of these two values we can comfortably calculate the probability of matching the data right ideally probability should be 1, exact matching. But those who are conducting the experiment they must be knowing when there is exact matching of the data with the expected data, then we have certain degree of doubt whether there is something wrong with the instrument.

Suppose I am expecting the outlet temperature of water from a heat exchanger I am expecting it should be 37.5 degree centigrade and thermocouple is also showing 30 or 37.50 degree C or thermocouple is also showing 37.50 degree centigrade right. Then I will have an element of doubt in my mind whether data are there is something wrong with the instrument then I will check. So, my confidence will be shaken in the data right.

So, in this test also with the probability of matching is greater than 0.9 then that is not considered to be acceptable matching of the data or agreement of the data with the expected one, because we always come under the it always comes under the clouds of doubt whether it is something wrong with the instrument it is giving me so, such an accurate reading or it is not working at all. So, an another extreme is when the data is less than 0.1, when the probability is less than 0.1 that is considered as a poor matching of a trend right.

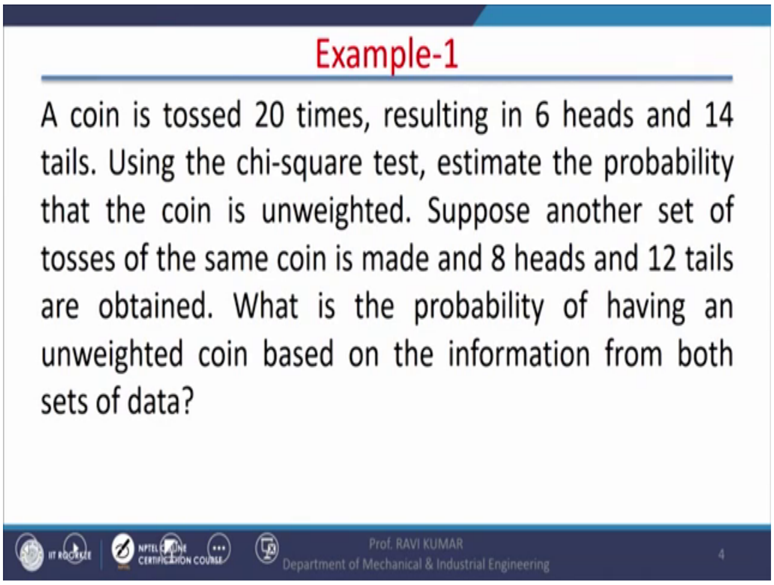
So, these if the values are falling greater than 0.9 and less than 0.1 it is not accepted, it is not accepted. So, the value of the probability should lie between these two point, it should lie it should be, I mean 0.1 or some probability and it should be less than 0.9.

(Refer Slide Time: 09:58)



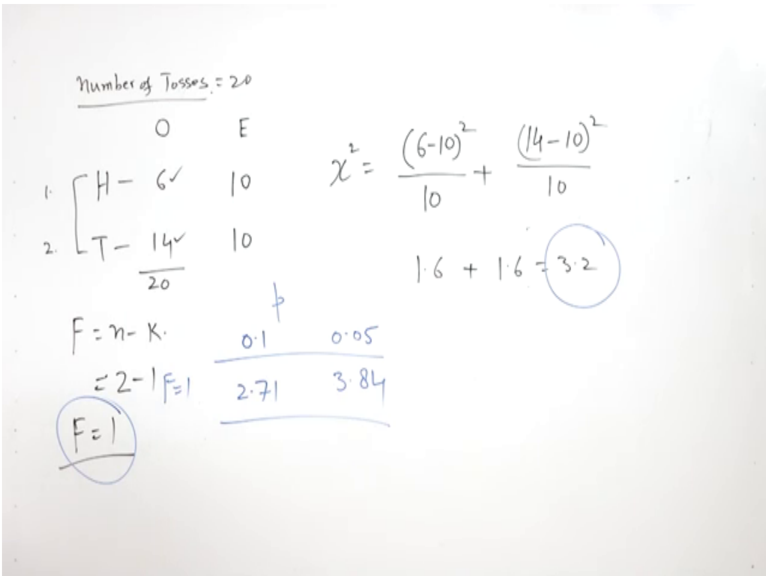
So, it should lie between 0.1 and 0.9. I will take certain example that will make concept clear about this chi-square test.

(Refer Slide Time: 10:16)



Let us take one example a coin is tossed 20 times, a coin is tossed 20 times resulting in 6 heads and 14 tails.

(Refer Slide Time: 10:49)



So, n is 20 rights and resulting in 6 heads and 14 tails. So, heads are 6 tails are 14. Using the chi-square test estimate the probability that coin is un weighted. There is no unbiased in the coin, so in order to prove that we will have to make sure that this trend this is the same trend as which is for the h is equal to 10 and tails is equal to 10 right then we can say there is no problem with the coin and it is following the same trend if it is following the chi-square test.

Now, here this is observed value and this is expected value now I have to calculate the value of chi-square. So, here the chi-square is going to be, first observed value minus expected value, so observed value is 6 expected value is 10 whole square divided by expected value that is 10 plus again observed value minus expected value. So, observed value is fourteen expected value is 10 whole square divided by again 10 this is the value of chi-square for this experiment right.

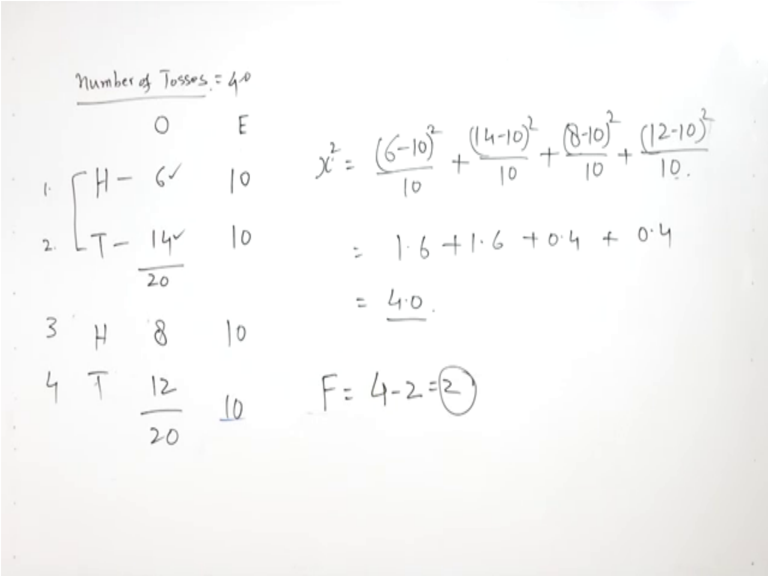
So, in order to calculate that we will have to just 4 square divided by 10 so, it is going to be 1.6 plus this is also 1.6 is equal to 3.2. So, value of the chi-square is 3.2 right. Now, what about F? F is n minus k, here n is 20 sorry not 20, this is the 2, either heads this is serial number 1 and this is serial number 2 these are the number of tosses this is not n this is number of tosses. So, this is the sum of this 20, right.

So, n is 2 either heads or tails. Now number of restriction, restriction is 1 because we do not have the number of n greater than 2 numbers are restricted. So, restriction is 1, so F is equal to 1. Now, for F is equal to 1 and chi-square is equal to 3.2. I will have to look into the probability table. So, I have taken a part of the probability table here this table is available in all the books right. So, complete table you can take from there, but I have taken part of the table here, right here.

So, F is equal to 1. So, for degree of freedom 1, this is F 1 and chi-square 3.2. So, when it is 3.2 it will lie between 0.1 and 0.05, 0.1 and 0.05, it means it is less than 0.1 right. So, I will write on the blackboard also for 0.1, 0.1 the value of chi-square is 2.71 and for 0.05 the value of chi-square is a 3.84, for F is equal to 1, this is value of p, this is value of chi-square. So, definitely it is lying between 0.1 and 0.05.

So, we are not very confident about this data. So, we cannot comfortably say that when we are getting in a toss 6 heads and 14 tails the coin is unweighted or there is no bias in the coin, coin may have a the certain bias. So, let us move back to the statement of the problem. Suppose another set of the tosses or the same coin is made and 8 heads and 12 tails, another set we have now this is in this set it is proved that the we are not very comfortable with the outcome of this set because the probability is less than 0.1 right. So, what we will do? We will conduct another test. So, another test was conducted and in another test we are getting 3 and 4, heads and tails, right.

(Refer Slide Time: 16:09)



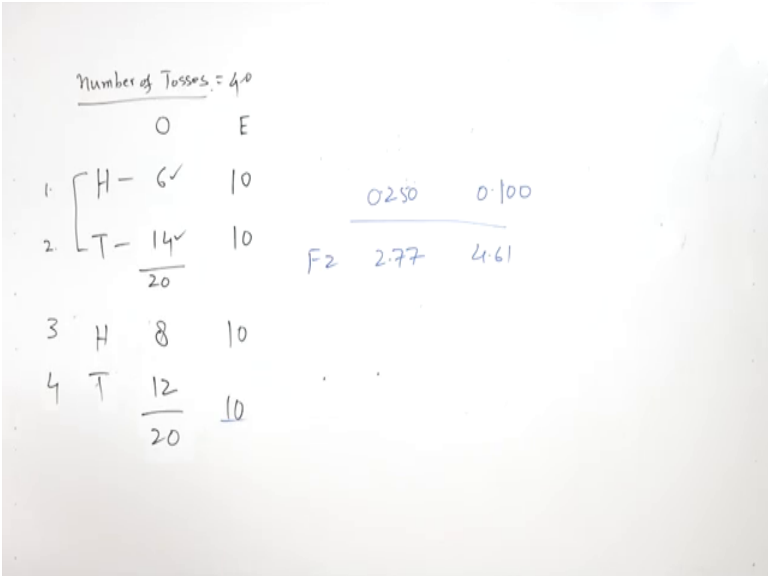
And then we are getting 8 and 12, 8 heads and 12 tails and these are to 20. So, total 40 experiments. We have number of tosses now they have increased to 40 right, now for this again we will do the chi-square test because once we toss 20 times we got 6 heads and 4, 14 tails again we tossed 20 times we got 8 heads and 12 tails.

Now, let us see what happens? In this case here it is expected value is 10 and 10. Now, in this case also we will calculate the value of chi-square. Now, again it is observed value minus expected value divided by observed value minus expected value whole square divided by expected value. So, observed value 6 minus 10 whole square divided by 10 plus 14 minus 10 whole square divided by 10 this we have already done in the previous case.

Now, again 8 minus 10 whole square divided by 10 plus 12 minus 10 whole square divided by 10 right. Now, here we have already this calculated this as 1.6 plus 1.6 plus 0.4, and this is also 0.4 and finally, we will get 4.0. This is the value of chi-square.

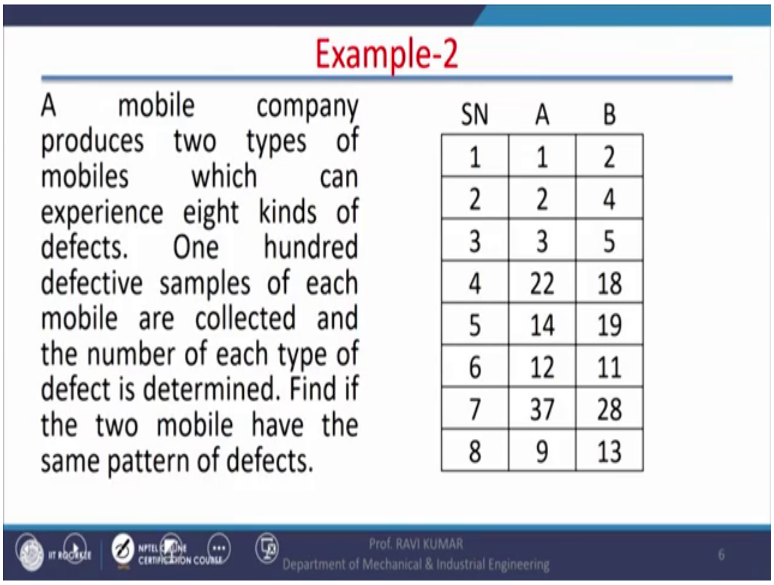
Now, what about the degree of freedom F? Now, in degree of freedom we have conducted 2 tosses. So, that will also add one more restriction. So, now, we have 4, n is equal to 4 and there are two restrictions. So, degree of freedom is 2. Now, if you again look at the chart, now we again look at the chart degree of freedom 2 and chi-square is equal to 4 then it comes between 2.77 and 4.61 it is between 0.25 and 0.1. So, if degree of freedom is 2, when degree of freedom is 2, when degree of freedom is 2, then probability 0.250 and 0.100, for this the value of chi-square is 2.77 and 4.61. So, the value of chi-square is about 0.1.

(Refer Slide Time: 19:20)



So, now we can definitely say that when we are tossing coin 2 times 20 times, 2 times in each time 20 times then values we are getting 6, 14 and 8 12 for heads and tails respectively. In that case the coin is unbiased and it is following the same trend or this is confirmed we are confirmed these data are confirming this data. So, this is the test which is very important those who are conducting experiments or doing experimental analysis of the data.

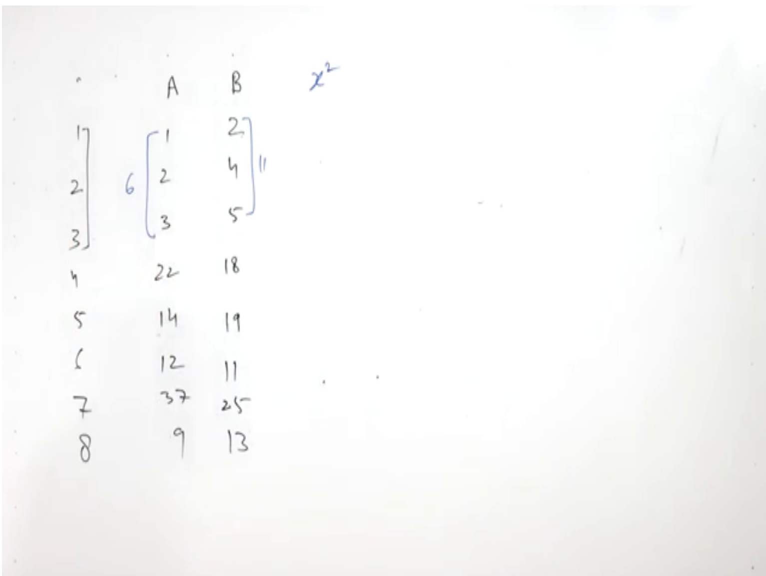
(Refer Slide Time: 20:24)



Now, I can take another example right. This is a mobile company example.

A mobile company produces two types of mobiles which can experience 8 kind of defects. Now, this is very interesting. Now, we knows here in the tossing of the coin we were restricted to only two things, heads and tails. Now, there is a mobile company it is producing two mobiles and they have identified 8 type of defects 1 2 3 4 5 6 7 8 right, and 100 defective sample of each mobile are collected.

(Refer Slide Time: 20:52)



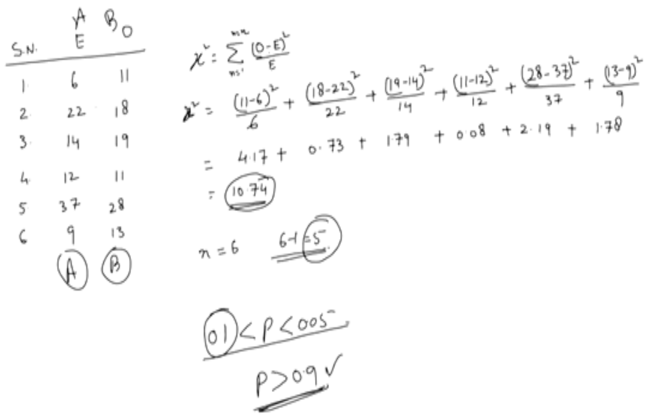
So, they have two types of mobiles model A and model B. So, the company has model A and model B, and they have taken sample of each, 100 sample of A and 100 samples of B. And the figures are 1 2 3 22 14 12 37 9 and here it is 2 4 5 18 19 11 28 and 13 right. So, this is a mobile, this is mobile A and type of defect one is 1, type 2 defect 2 mobiles are having type 2 defect, let us say 14 mobiles are having type 5 defects, 37 mobiles are having type 7 defects this is a trend, right.

Now, B, type model B it has 2 mobiles of defect 1, 18 mobiles of defect 4, 25 mobiles of defect 7. Now, the issue is both these data are following the same trend, this we have to issue right and in order to find that let us do one thing. Let us assume this is expected value expected output and we will try to fit the do the chi-square of this data referring this data in ideal case we assume that this will also have the value of 1 2 3 22 14 12 37 and 9, but this is not happening. In another model these figures are varying.

So, we take this as a reference value or expected output and this is the actual output or this is we can say expected value and this is observed value, fine. So, in order to do that first of all we should remember, I forgot to tell you in chi-square test the frequency of I mean it should not be less than 5, it should not be less than 5. So, what we will do? We can club these 3 defects, we can club these 3 and then we club these 3 defects this becomes 6 and this becomes 11. Just a minute or I have, yes, it becomes 11.

Now, after this we will conduct the chi-square test. Now, this is the expected value and this is a observed values, in order to find the value of chi-square we will take sum of the observed value minus expected value whole square divided by expected value.

(Refer Slide Time: 24:13)

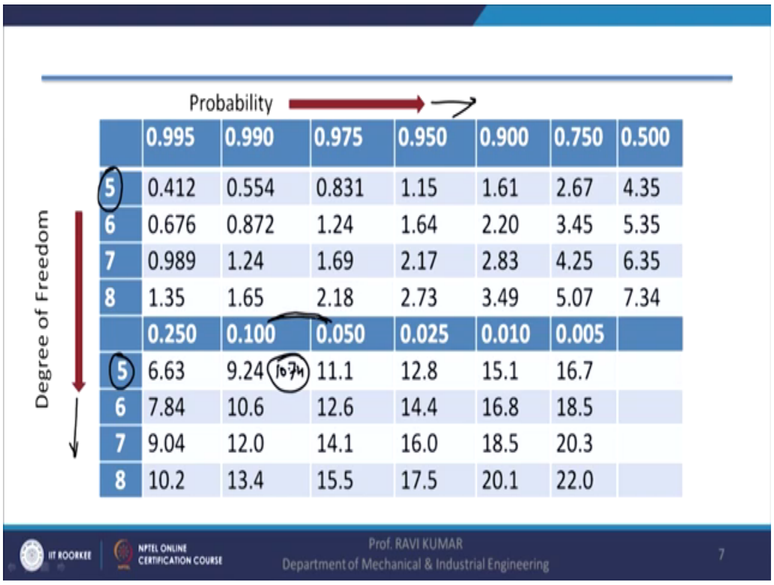


Now, here in this case the value of chi-square is observed value 11 minus expected value 6 whole square or we can take 6 minus 11 as well there is no difference divided by expected value that is 22 plus next 18 minus 22 whole square divided by 22.

Next is 19 minus 14 whole square divided by 14, next 11 minus 12 whole square divided by 12 plus 28 minus 37, whole square divided by 37 plus 13 minus 9 whole square divided by 9. Now, if we further simplify this we will be getting 11 minus 6 whole square divided by 22, this is not 22, this is 6, this is 6. So, chi-square is equal to 4.17 plus 0.73 plus 1.79 plus 0.08 plus 2.19 plus 1.78 and the sum of all these is equal to 10.74.

Now, the value of chi-square is 10.74, and degree of freedom degree of freedom is there are 6,. So, n is equal to 6, and degree of freedom is 6 minus 1 is equal to 5. Now, for 5 degree of freedom and chi-square value 10 0.74 we look at the table, in this table for 10.74 value of chi-square because on this axis this side the probability is shown, this side degree of freedom is shown. So, we will go for 5 degree of freedom, so 5 degree of freedom the value of chi-square 10.74.

(Refer Slide Time: 27:08)



	0.995	0.990	0.975	0.950	0.900	0.750	0.500
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35
8	1.35	1.65	2.18	2.73	3.49	5.07	7.34
	0.250	0.100	0.050	0.025	0.010	0.005	
5	6.63	9.24	11.1	12.8	15.1	16.7	
6	7.84	10.6	12.6	14.4	16.8	18.5	
7	9.04	12.0	14.1	16.0	18.5	20.3	
8	10.2	13.4	15.5	17.5	20.1	22.0	

So, 10.74 will lie somewhere here and it will lie between 0.1 to 0.05. So, the value of probability the p is greater than 0.1 and it is less than 0.05. Now, for this value of probability because the agreement is poor, the agreement is poor, so we cannot conclude from these data that the same type of defect pattern has been followed in both the cases A and B. So, this is A and this is B. So, the defect pattern is not same in both the cases. So, in chi-square test if the probability is less than or sorry is greater than 0.9 it reflects the agreement is too good. So, too good is also not good in the in the case of chi-square test.

So, if the probability is greater than 0.9 or if it is less than 0.1 in that case we conclude that the same trend has not been followed. So, in this question because the probability the in probability is less than 0.1, 0.1 and it is coming around it is between 0.1 to 0.05. So, we can definitely say that the 2 mobiles do not have both the, so these mobiles do not have the same pattern. So, we can comfortably conclude that the defect pattern in both the values is not same. That is all for today.

Thank you very much.