Probability and Statistics

Prof. Dr. Somesh Kumar

Department of Mathematics

Indian Institute of Technology, Kharagpur

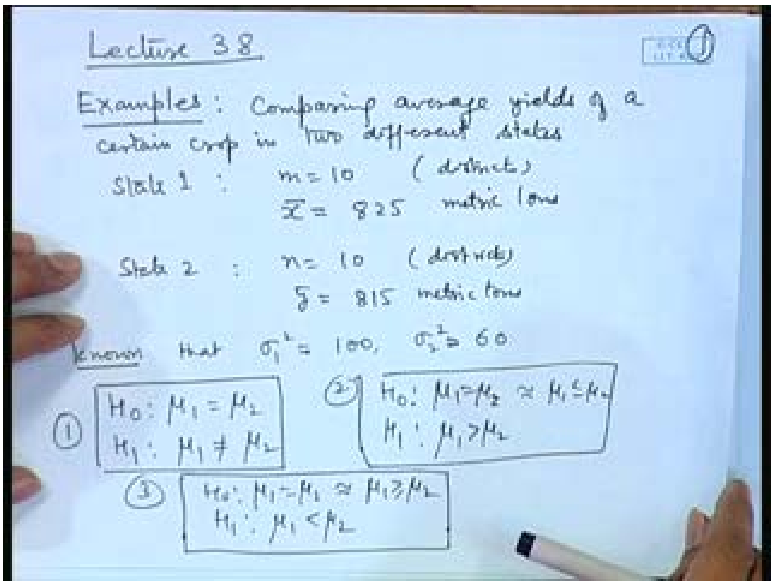
Module No. #01

Lecture No. #38

Testing of Hypothesis-VI

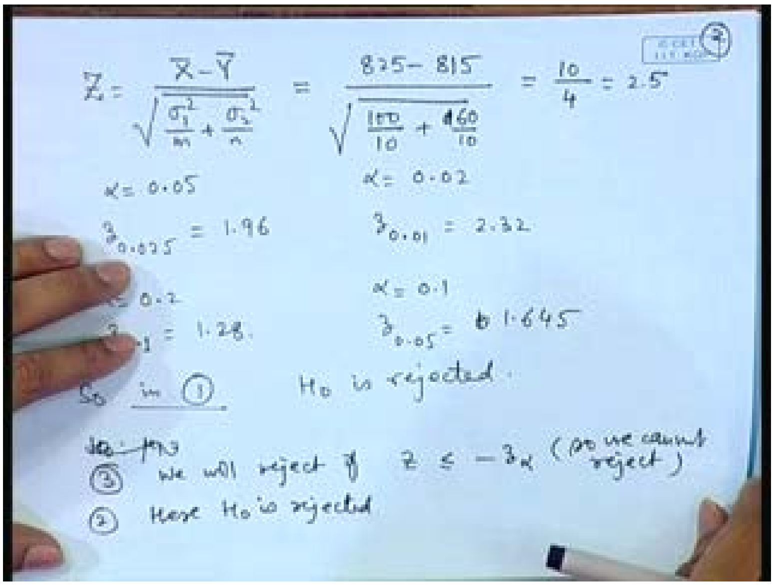
In the previous lecture, I have discussed various tests for testing about the equality or inequality of the means or variances of two normal populations. We also saw some special cases where the normality assumption may not be valid and therefore, we may take some approximate tests.Now, we will look at certain applications of these tests by various illustrations here. Suppose, we want to compare the average yield of a certain crop in two different states. Now, if we are looking at this, then we may have a random sample taken from the two states.

(Refer Slide Time: 01:17)



So, for example, you consider. So, we are looking at, say comparing, say average yields of a certain crop in two different states. So, in state one, we consider say ten districts.And on the basis of the random sample, we observed that the sample mean turns out to be 825, say metric tons.Similarly, in state two we took another random sample of tendistricts and the average yields turn out to be 815 metric tons. It is known that, the sigma 1 square is equal to say 100 and sigma 2 square is equal to 60. This data is known to us. So, we want to test say, whether mu 1 is equal to mu 2 against say H 1, mu 1 is not equal to mu 2.Now, onwards I will consider this hypothesis as 1, second hypothesis I will consider as H naught mu 1 is equal to mu 2 against H 1, mu 1 is greater than mu 2.And of course, this is equivalent to mu 1 less than or equal to mu 2 here. And the third one will be H naught mu 1 is equal to mu 2 which is equivalent to mu 1 greater than or equal to mu 2 against H 1 say mu 1 is less than mu 2.

(Refer Slide Time: 03:32)



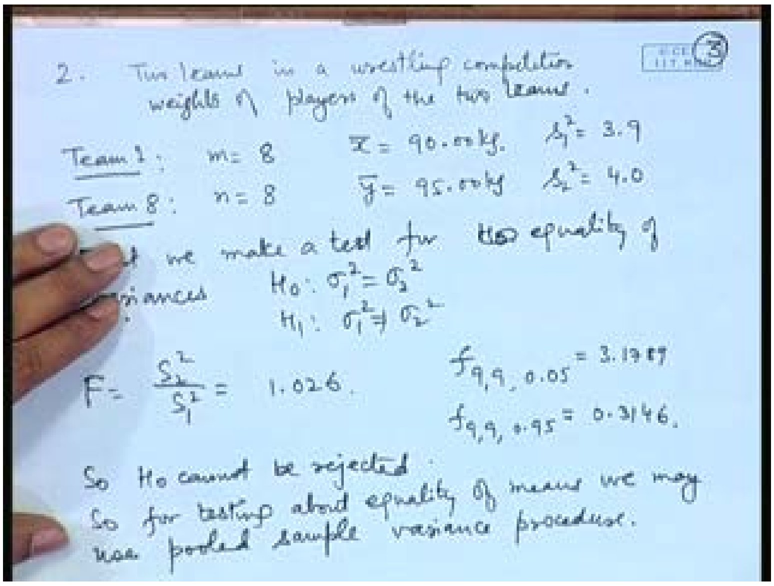
I will refer to the hypothesis 1, 2 and 3 to these hypothesis problems. So, here if the assumption that the variances are known is taken, then we will use the Z variable. That is Z test statistic, that was given by X bar minus Y bar divided by square root sigma 1 square by m plus sigma 2 square by n. Now, the values of X bar, Y bar, then sigma 1 square by 10 plus sigma 2 square by 10, now this value turns out to be simply 10 by 4; that is 2.5.If we are carrying out a test, say at level, say alpha is equal to 0.05,then I need to look at Z of 0.025, that is equal to 1.96. Suppose, I am having alpha is equal to 0.02, then I will have to see Z of 0.01 that is equal to 2.32.Suppose, I am saying alpha is equal to 0.2, then I will see 0.1, that is equal to 1.28.Suppose, I say alpha is equal to 0.1, then I will see 0.05, that is 1.64.You can see here, all of these values are smaller than this value.Therefore, at any practical level of significance,so, in testing problem 1, H naught is rejected.Of course, if we take much smaller alpha; that means, it may we may take say

0.01 or 0.001, etcetera, then of course, this will not be rejected, but then it is not very reasonable to take such a low probability of the type one error. Because, that may reduce the power also.

So, if we are considering here, say H naught, say mu 1 is equal to mu two; that means, I am considering the problem say 3.If I am considering the third problem, in the third problem, we will reject if Z is less than or equal to minus Z alpha. But, this can never be true for any reasonable level of significance here. Because, Z is a positive value here. So, we cannot reject; that means, if we go by this third framework, then we were thinking that mu 1 is greater than or equal to mu 2 is cannot be rejected.
On other hand, if you look at 2, in the 2, we have to reject when Z is greater than or equal to some Z alpha. So, it is always, it will always be satisfied; that means, it is always rejected; that means, mu 1 greater than mu 2 is a strong conclusion. So, here H naught is rejected. So, here we cannot reject. So, we conclude that the average yield is higher in the state one.

Let us see this description carefully.What was our initial problem? Our initial problem is to compare the average yields in the 2 states. And we observe from the random sample of 10 districts each, that in the first state it is 825 metric tons, in the second state it is 815 metric tons. So, from the observed values, it is clear that, the first is having higher average yield, but statistically is it significant? So, for that we needed some other characteristics.In this particular case, the standard deviations are available. It is 10 and slightly less than 8respectively. So, since these standard deviations are not very large, the sample size is not very large, therefore, this hypothesis of equality is actually getting rejected at any reasonable level of significance. Since, this is rejected, that means, mu 1 is not equal to mu 2. Now, if mu 1 is not equal to mu 2, then we have suspicion that mu 1 is greater than mu 2 precisely from here. Because, X bar is greater than Y bar and therefore, the hypothesis of mu 1 less than mu 2 can; that means, in this favor we will never have a decision. And therefore, we consider the second hypothesis testing problem that is mu 1 less than or equal to mu 2 against mu 1 greater than mu 2 and here the decision is in the favor of H one; that means, we are rejecting H naught and this is a strong conclusion. And that therefore, we can say that it is significantly true; that means, the average yield in the state 1 is higher significantly than the state 2.

(Refer Slide Time: 09:46)

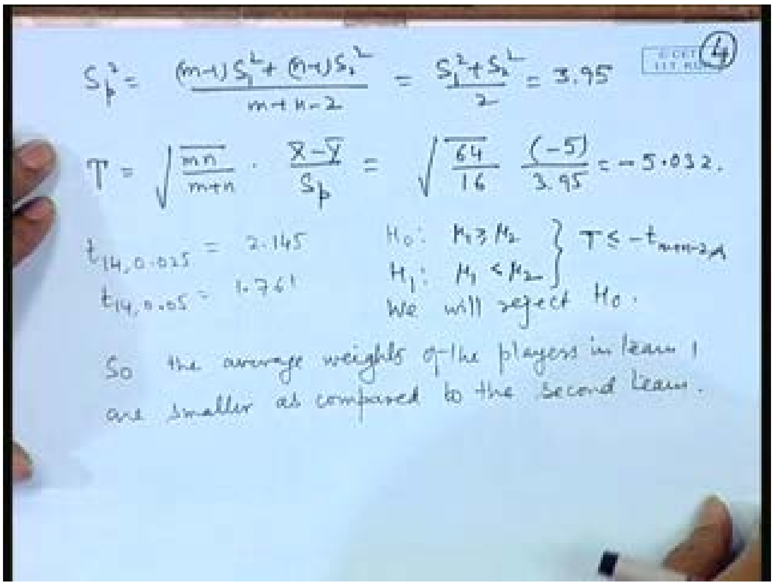


Let us take some more examples. Suppose, there are two teams in a say wrestling competition.And we want to see whether the average weights of the wrestlers are same in the two teams or one of them is having higher weights than the other one, because this is a crucial factor in the outcome of the actual match.So, we are having weights of players of the two teams. So, from team 18 wrestlers are randomly selected and their average weights are 90 kg. And the standard deviations are calculated from the sample and the S 1 square turns out to be 3.9.

From the second team, another set of 8 wrestlers were selected randomly, their average weights turn out to be 95 and S 2 square is turning out to be 4. Now, if you want to test whether the second team has a higher average weight than the first team, then the question comes up which procedure to be adopted. Because, when the variances are unknown, we have 2 different procedures rather 3 different procedures. Here,of course, the 2 samples are independent. So, the pairing is not required here. But, the variances equality plays a role here.

So, Firstly, let us test about, first, we make a test for equality of variances; that means, we test H naught sigma 1 square is equal to sigma 2 square against H 1 sigma 1 square is not equal to sigma 2 square.Now, here you see that S 2 square is bigger than S 1 square. So, we make F as S 2 squares by S 1 square which is equal to 1.026. Now, if we look at the f value on 99 degrees of freedom at say 0.05, then it is 3.189.If we see f 99.95, that is equal to 0.3146. So, you see here, 1.026 is not greater than this value, it is not less than this value.So, H naught cannot be rejected. So, for testing about equality of means, we may use pooled sample variance procedure, that is S p square formula.

(Refer Slide Time: 13:29)

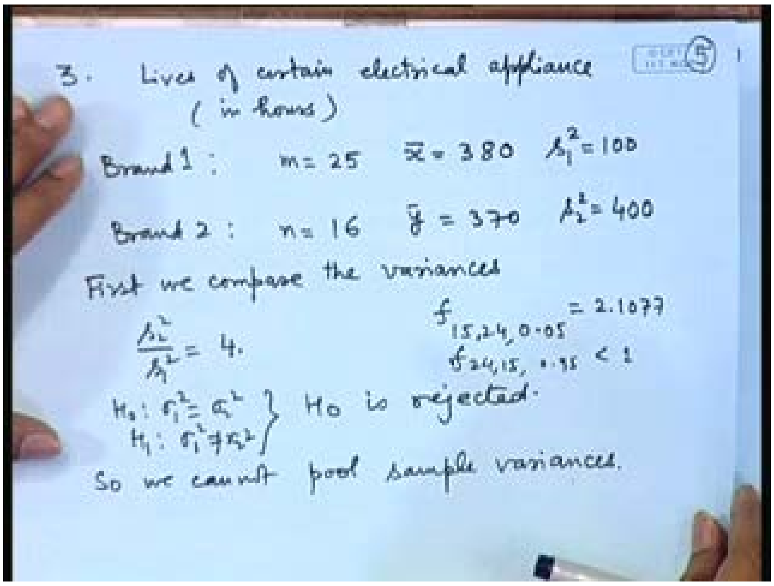


So, if we apply that, we need to calculate S p square. That is here, m minus 1, S 1 square plus n minus 1 S 2 square by m plus n minus 2.In this particular case, m and n are equal. Therefore, this becomes simply S 1 square plus S 2 square by 2, that is equal to 3.95. So, the formula for the pooled test statistic here, is square root of m n by m plus n X bar minus Y bar by S p.That is square root 64 divided by 16 and this is minus 5 divided by 3.95, that value turns out to be minus 5.032.If we see here, say t 14 at say 0.025 level, then it is 2.145 T 14 say 0.05, that is 1.761, etcetera. You can see here, the calculated value of t statistic here is minus 5.032.

If we are considering a test where our null and alternatives are given like mu 1 is greater than or equal to mu 2 against mu 1 less than mu 2. Then, here our rejection reason will be Z t is less than minus t m plus n minus 2 alpha. Since, minus 5 is smaller than minus 2 or minus 1, we will reject H naught. So, the average weights of the players in team 1 are smaller as compared to the second team.

So, here we will say it is significantly smaller. If we had reverse this hypothesis, suppose, I put mu 1 is less than or equal to mu 2 here and mu 1 greater than mu 2, then we cannot reject H naught so; that means, we are actually supporting mu 1 is less than or less than or equal to mu 2 here. And of course, the hypothesis of equality is also rejected here.Therefore, we should say that mu 1 is greater than mu 2, because equality is also ruled out. So, actually, this is what the point I wanted to emphasize. That, when we write the hypothesis, then we should not simply get a conclusion based on one hypothesis and say something, rather than we should analyze the closely related hypothesis also.For example, in this case, we are saying reject H naught. So, rejecting means mu 1 is less than mu 2.On the other hand, if we had framed it in the reverse way, then we cannot reject H naught but then, that would have meant mu 1 is less than or equal to mu 2. But actually, equal to mu 2 is also rejected here.Therefore, we conclude that mu 1 is significantly smaller than mu 2 in this case.

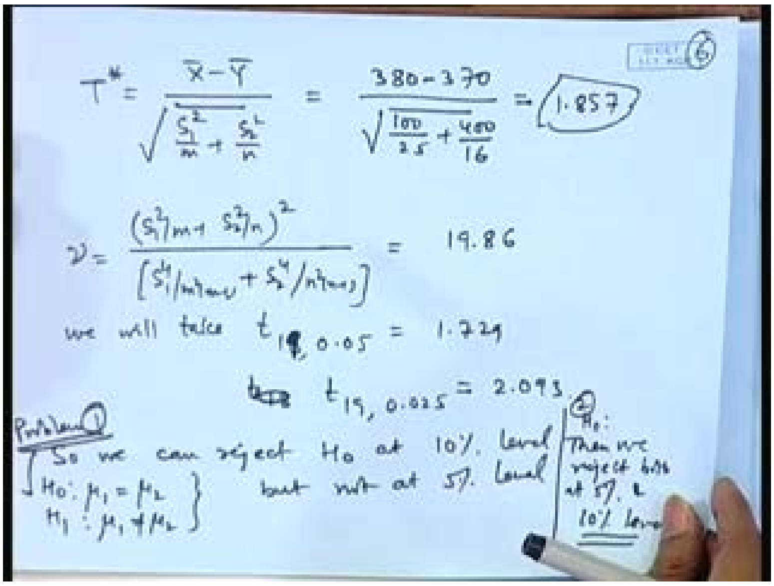
(Refer Slide Time: 17:20)



Let us consider, say lives of certain electrical appliance, it is in hours. So, we have 2 brands of appliances. We want to compare the average lives here. For the first brand, 25 units were selected at random and put on the test and their average lives were observed to be 380 hours with a sample variance of 100.

For the second sample, 16 units were selected and theiraverage were observed to be 370 hours and the variances turn out to be 400.May be the second show has a smaller mean. Small, less number of smaller average life, but the variability is more. Once again, if we want to compare the means,firstly, let us compare the variances. So, first, we compare the variances. So, if we take the ratio S 2 square by S 1 square that is equal to 4. So, if we look at say f value on say 15,24,0.05; that is equal to 2.1077. So, naturally if I am considering, say the reverse of that, that is going to be smaller. So, H naught sigma 1 square is equal to sigma 2 square, against H 1 sigma 1 square is not equal to sigma 2 square. So, here H naught is rejected.Because, if I consider f 24, 15.95; that is going to be less than 1. So, naturally, this is falling into this region. So, H naught is rejected. So, we cannot pool sample variances. So, if we cannot pool, then we have to use that procedure which is given in the using smith satterthwaite procedure.

(Refer Slide Time: 20:02)

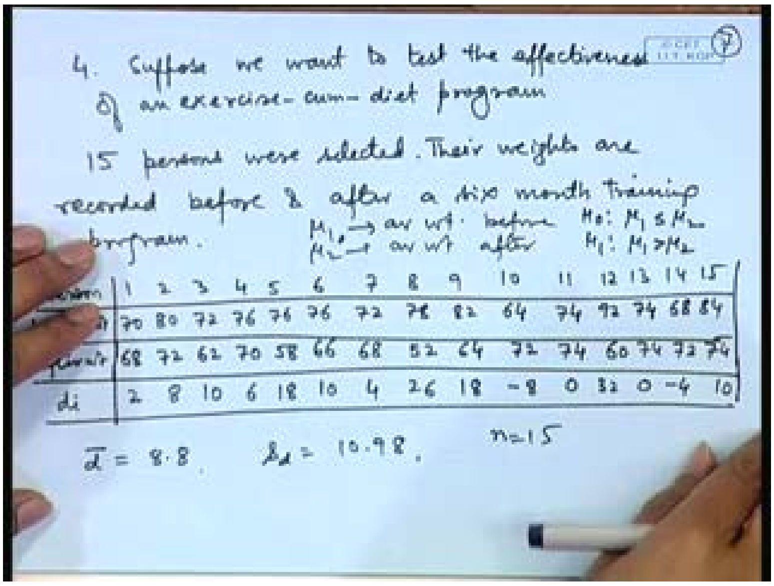


So, we have the T star as X bar minus Y bar divided by square root S 1 square by m plus S 2 square by n. So, that is equal to 380 minus 370 divided by square root of 100 divided by 25 plus 400 divided by 16.This value can be seen to be 1.857. Now, the calculated value of T star has to be compared with a T distribution on nu degrees of freedom. Now, nu was given by S 1 square by m plus S 2 square by n whole square divided by S 1 to the power 4 by m square into m minus 1 plus S 2 to the power 4 by n square into n minus 1. So, after substitution of the values given here, we obtain this value to be 19.86. So, we will take t 18. So, for example, at 0.05 levels, this value is 1.729.

Suppose I consider,sorry 19, t 19 at say 0.025, this is equal to 2.093, etcetera. You can easily see that, the value 1.857, this is bigger than this. But, smaller than this. So, we can reject H naught, suppose I am considering problem 1,testing problem 1,so, in that case, suppose I am considering H naught mu 1 is equal to mu 2 against say H 1 mu 1 is not equal to mu 2.If I am considering this, then we can reject H naught at 10 percent level, but not at 5 percent level. That at 5 percent level t alpha by 2 value will be 0.025, that is 2.093 which is larger than this.

However, if you are considering 1 side at test, that is H naught, say I am considering testing problem 2, in the testing problem 2 the rejection reason is t star is greater than or equal to t mu alpha. So, at 5percent level you will be rejecting but at 0.25 percent level you will not be rejecting. Then,we will be rejecting both at,then we reject both at 5percent and 10 percent levels. However, if you consider say 0.025 percent level, then you will not be rejecting here. So, this test is slightly more sensitive here.

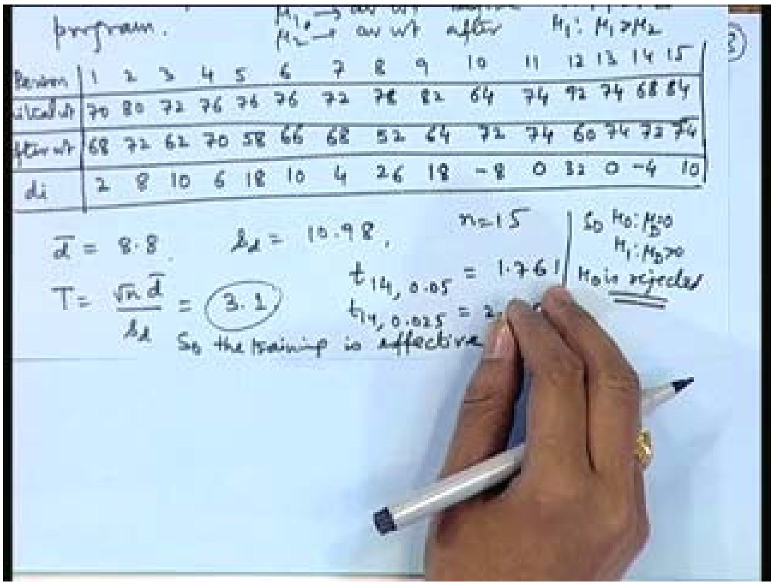
(Refer Slide Time: 23:44)



Let us consider one example, where the observations are paired. So, suppose we want to test the effectiveness of an exercise cum diet program. So, say 15 persons were selected. Their weights are recorded before and after a 6 month training program. So, we want to see whether the average weights have reduced. So, let us see the data is given in this form. Person. So, we have persons 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15. Initial weight is taken, to notice whether there is a significant difference later or not. The weights are 70 80 72 76 76 76 72 78 82 64 74 92 74 68 84. After the program is conducted, what is the weight then?68 72 62 70 52 66 68 52 64 72 74 60 74 72 and 74.

We want to test, suppose I say mu 1 and mu 2. So, mu 1 is the average weight before and mu 2 is the average weight after the training program. So, we may be interested to test whether mu 1 is actually greater than mu 2; that means, we can put this as a strong hypothesis in the alternative, which actually we want to test. So, mu 1 is less than or equal to mu 2 against mu 1 is greater than mu 2. So, if we consider here the differences, d i, 2 8 10 6 18 10 4 26 18 minus 8 0 32 0 minus 4 10. So, let us calculate d bar, that is 8.8. S d is equal to 10.98. So, here n is equal to 15.

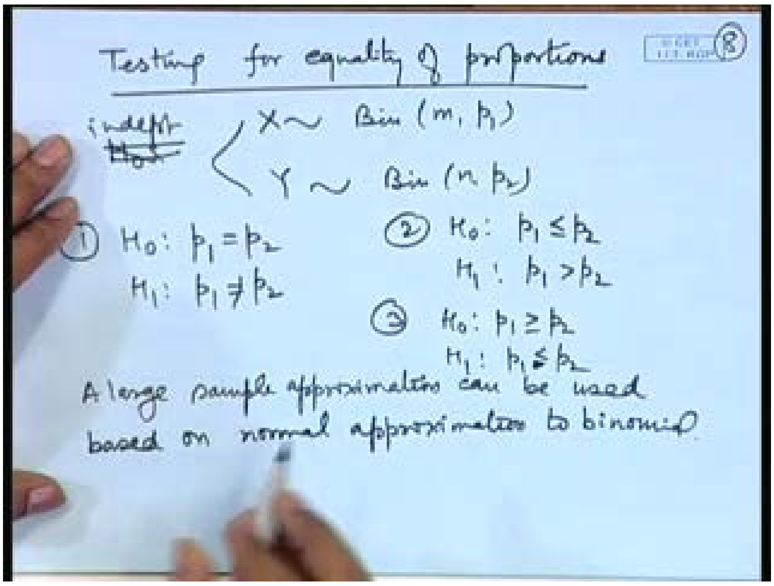
(Refer Slide Time: 27:57)



So, the value of the t statistic here, is root n d bar by s d, that turns out to be 3.1.If we are considering t distribution on 14degrees of freedom, then the values say 0.05, that is say 1.761 t 14 at say 0.025that is equal to 2. 145, etcetera. You can see that this value is significantly higher than these values. So, the hypothesis H naught mu is equal to 0 against mu d against mu d greater than 0. So, H naught is rejected; that means, we conclude that exercise program is effective. So, the training is effective. We are concluding that, there is a significant reduction in the weight after the 6 month training program.

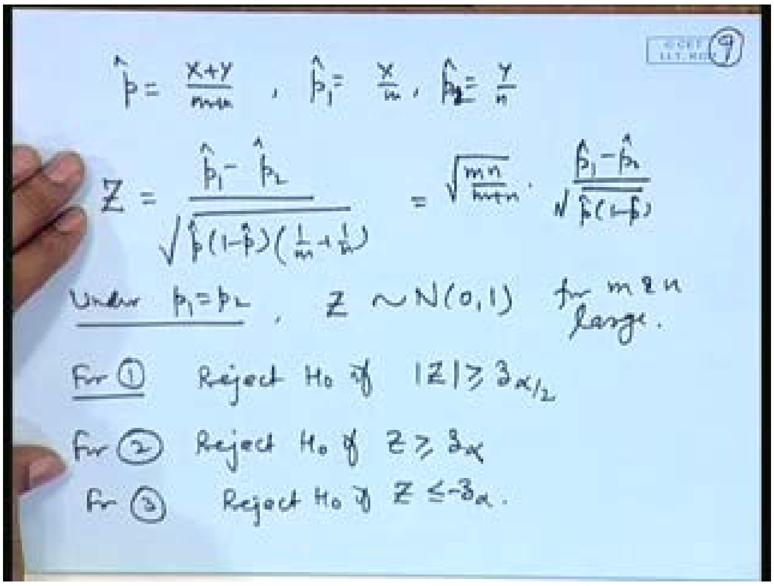
We also have the normal test, when we are comparing the proportions of the two binomial populations. So, for example, the data is recorded not in the numerical measure but in the characteristic form.For example, we want to see certain opinions, we want to see the effect of certain, say learning procedure. So, that may be, the result may be in the form of, that where a test is conducted. So, for example, a set of students start a certain instructional material, another set of students start another instructional material, a common test is conducted. We want to see how many passed in the first set and how many passed in the second? Is there a significant difference in the proportions? That means, we want to see whether the instructional material 1 is better or the instructional material 2 is better.

(Refer Slide Time: 30:11)



So, basically the situation is of the following form. Testing for equality of proportions. So, the general model is of the form X follows say binomial m p 1 and Y follows binomial n p 2.We assume these 2 samples to be independently taken here. So, we are interested in testing hypothesis of the form p 1 is equal to p 2 against, say p 1 is not equal to p 2.Say this is 1, we may test p 1 is less than or equal to p 2 against p 1 is greater than p 2 or say H naught p 1 is greater than or equal to p 2 against H 1 p 1 is less than p 2, etcetera. A large sample approximation can be used based on normal approximation to binomial.

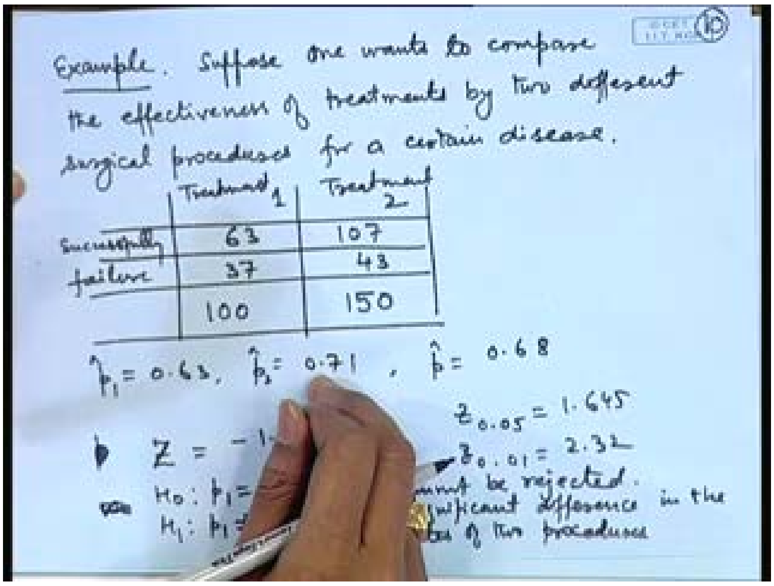
(Refer Slide Time: 31:50)



So, we may calculate here, p as head X plus Y by m plus n p 1 head is equal to X by m p 2 head is equal to Y by n. So, this is the first proportion, this is the second proportion, sample proportion. This is the pooled proportion. So, we construct a statistic Z as p 1 head minus p 2 head divided by square root p head into 1 minus p head 1 by m plus 1 by n. Which is actually equal to root of m n by m plus n p 1 head minus p 2 head divided by root p head into 1 minus p head.Under the assumption that p 1 is equal to p 2 Z is approximately normal 0 1 for m and n large. So, we may make a test based on this. For 1 reject H naught, if modulus Z is greater than or equal to Z alpha by 2.For 2, it will be reject H naught, if Z is greater than or equal to Z alpha. For 3 reject H naught, if Z is less than or equal to minus Z alpha.

So, this is an approximate test. If the assumption that m and n are large is not true, in that case, we may have to go for an exact procedure. But, that procedure will make use of the distribution which is calculated from the binomial. So, under p 1 equal to p 2, the distribution of X plus Y is again binomial and one can make use of the distribution of X given X plus Y which is hyper geometric. And there is a test procedure for that, but we are not going to discuss that, in this particular course here.

(Refer Slide Time: 34:15)



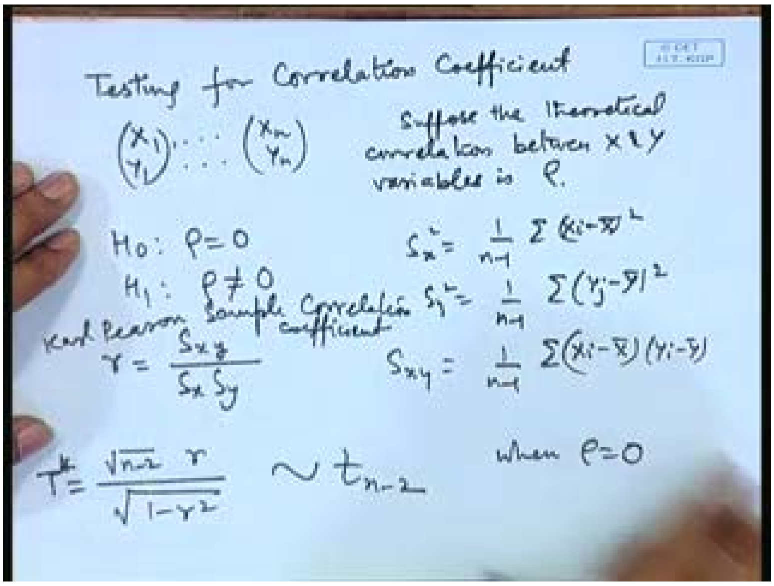
Let me give an application of this here. Suppose, one wants to compare the effectiveness of treatments by two different surgical procedures for a certain disease. So, for this a set of patients is taken. On one set of patients, one surgical procedure is adopted and we observe the proportion of success. So, let us make the data in this particular fashion. Suppose, 100 patients are there on which treatment procedure 1 is adopted. We see how many are successfully treated and how many of them are failures. Suppose, in out of 100 here, it turns out that 63 are successfully treated, 37 are unsuccessful.Whereas, using the treatment procedure 2, suppose 100 patient, 150 patients were given this procedure. Out of that, 107 were successfully treated and 43 are not successfully treated; that means, on them the surgical procedure did not yield any positive result.

Let us look at the proportions here. p 1 head is 0.63, p 2 head is equal to 0.71, and p head is equal to 0.68. So, if we calculate the Z statistic here, that is p 1 head minus p 2 head divided by root of p head into 1 minus p head into root m n by m plus n, this value turns out to be minus 1.33. If we are considering any reasonable level of significance, say Z 0.05; that is 1.65. Suppose, I take Z 0.01, that is 2.32, etcetera. Then, you can see that, if we consider say hypothesis H naught p 1 is equal to p 2 against, say H 1 p 1 is not equal to p 2. Then, at level of significance, say 10 percent, 2 percent, etcetera.H naught cannot be rejected, because the absolute value of Z; that is 1.33 is smaller than these values. So, that means, there is no significant difference in the two, in the success rate of the two surgical procedures. So, no significant difference in the success rates of two procedures.

Although, from here it looks that the second procedure is more effective but statistically speaking, there is no significant difference here.

We have seen here, that sometimes the assumption of the correlated observations is required, as in the case of exercise program, etcetera. But, we have seen the use of t paired t test should be done under certain checks; that means, we should use it with care.For example, we are adopting a paired t test procedure but the value of the correlation is say 0. In that case sigma 1 square plus sigma 2 square will be the variance of X i minus y i; that means, we have unnecessarily used, reduce over degrees of freedom here.We are using only n minus 1 degrees of freedom. Consequently, our power of the test will reduce. So, it is not advisable to go for a pair t test here; that means, in such a case it will be a reasonable option, firstly, to check whether the correlation is 0 or not in the given data set. So, fortunately, we can find actually a test for the correlation coefficient being significantly different from 0 or not.

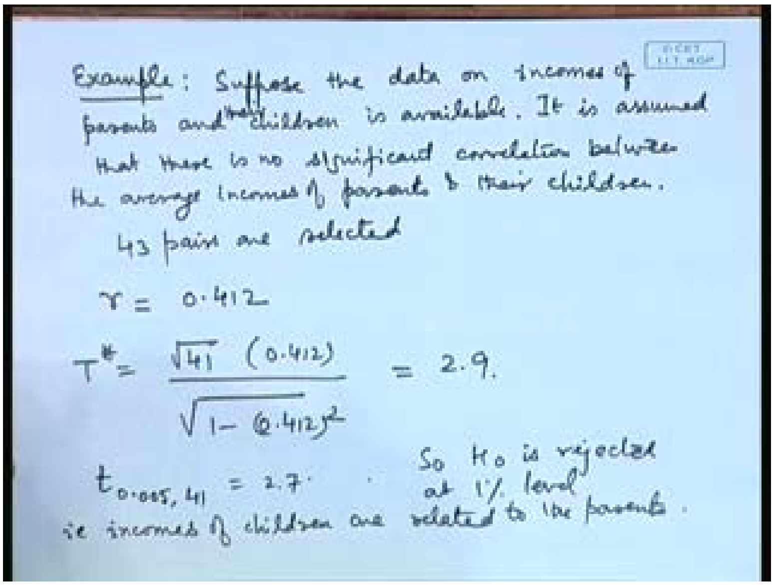
(Refer Slide Time: 39:40)



So, testing for correlation coefficient. So, we have the data X 1 X 2 X n, Y 1 Y 2 Y n. So, in the two data sets, we consider the sample correlation coefficient and the population correlation coefficient. Suppose the theoretical correlation between X and Y variables is rho. So, we want to test whether rho is equal to 0against, say rho is not equal to 0. For this, we calculate the sample correlation coefficient that is r, that is equal to S x y divided by S x S y, where this S x square is 1 by n minus 1 sigma x i minus x bar whole square S y square is 1 by n minus 1 sigma y j minus y bar whole square and S x y is equal to 1 by n minus 1 sigma x i minus x bar into y i minus y bar. That is the sample, standard sample variances for the 2 samples and this is the sample covariance. So, based on this, we define the Karl Pearson sample correlation coefficient.

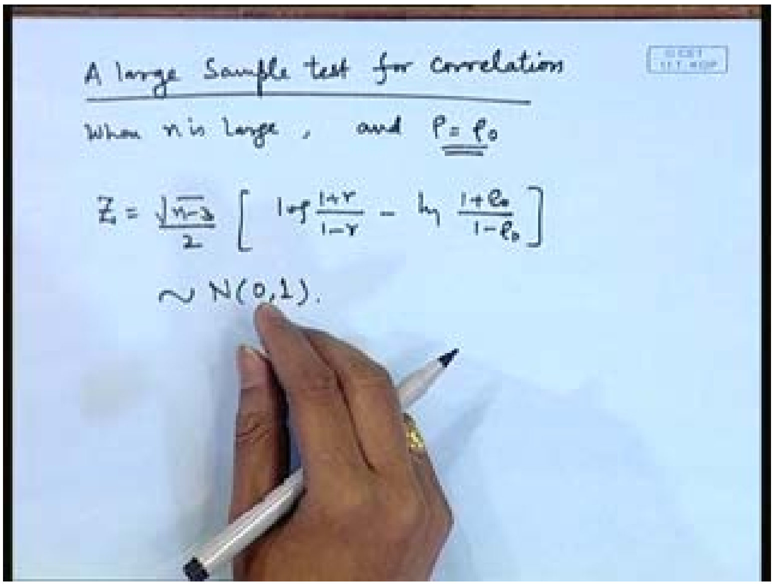
Then, it has been observed that square root n minus 2 into r divided by root 1 minus r square. This is having a t distribution on n minus 2 degrees of freedom when rho is equal to 0 let me call it T star. So, one can make use of this for testing the significance of correlation; that means, whether there is a significant correlation between the 2 variables or not.

(Refer Slide Time: 42:14)



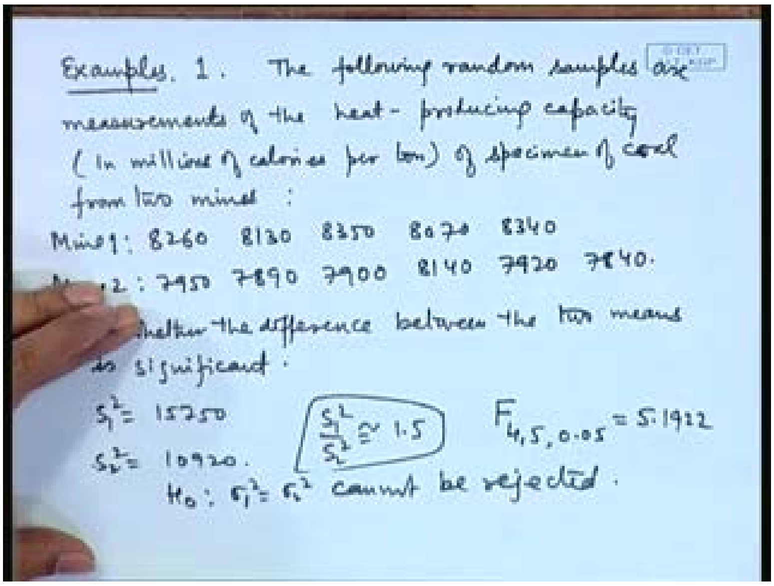
Let us consider one example here. Suppose the data on incomes of parents and children, their children is available. It is assumed that there is no significant correlation between the average incomes of parents and their children. So, 43 pairs are selected and there sample correlation was calculated, which turned out to be 0.412. So, from here, this T star value, that is square root 41 into 0.412 divided by square root 1 minus 0.412 square is calculated. This value turns out to be 2.9. Suppose, we are considering say t, on say 0.005 at 41, this value is 2.7. Now, you can see here that this is extremely small level of significance, we are taking. So, H naught is rejected at any reasonable level of significance.This is actually 1 percent level. So, at 1 percent level itself, this is rejected; that means, the income incomes of children are related to the parents; that means, higher income parents, their children will also tend to earn higher incomes and lower income parents, their children will have lower incomes here.

(Refer Slide Time: 45:05)



Now, this test is again based on the normality assumption. Sometimes the normality assumption may not be valid.In that case, there is a large sample test for correlation coefficient.When n is large and say rho is equal to rho naught. So, we may not test that rho is equal to zero but some arbitrary value rho naught,then we construct Z.That is equal to root n minus 3 by 2 log of 1 plus r divided by 1 minus r minus log 1 plus rho naught divided by 1 minus rho naught.Then, this has approximately normal 0 1 distribution. So, consequently if n is large, then we may test about the correlation being equal to any arbitrary value.Of course, if rho naught is 0here, then this term will vanish and we will have only this particular term. So, this is an approximate normal test here. And this is not based on the assumption of normality for the initial samples, that is xi's and y i's here.

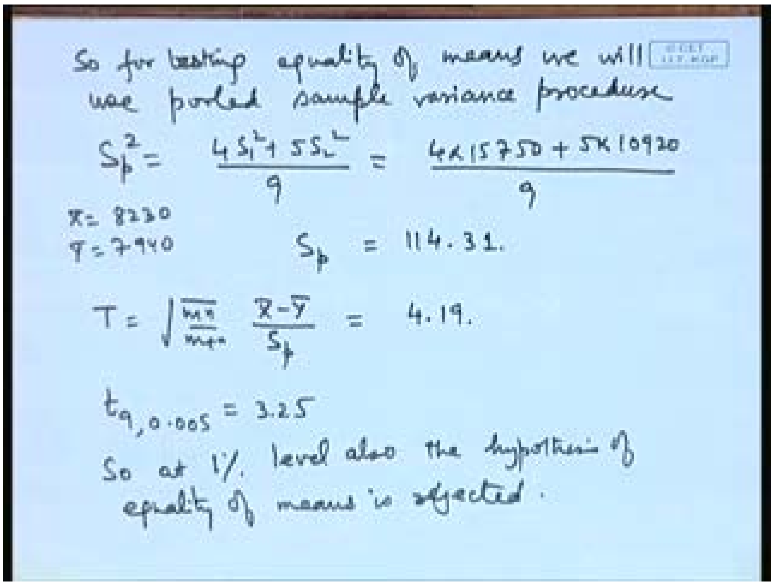
(Refer Slide Time: 47:10)



Let us take a few more examples here, of the test that we have discussed today. So, let us take one example based on, say normal distributions here. The following random samples are measurements of the heat producing capacity in million of millions of calories per ton of specimen of coal from two mines. So, in mine 1, there will be 5 measurements are taken. The values are 82608130835080708340. In the mine 2, 6 observations were taken, 79507890790081407920and 7840. Test whether the difference between the 2 means is significant. So, here 2 samples are available to us and we are having 5 and 6 observations respectively from the 2 samples.

So, what we do here? We have to check whether mu 1 is equal to mu 2 or mu 1 is not equal to mu 2. But, once again here, to test this hypothesis we will have to test about the variances also. So, whether the variances are same. So, here we see that, if we consider S 1 square here, S 1 square is equal to 15250, etcetera. S 2 square is equal to 10920. So, if we consider the ratio here, S 1 square by S 2 square, that is approximately 1.5. So, if I am looking at the F value on, say 1 2 3 4 5. So, 4 and 5degrees of freedom.Let us see one example here, from the tables of the F distribution and say at point 05level, if we are seeing F 45. So, at 0.05 the value is 5.1922. So, H naught sigma 1 square is equal to sigma 2 square cannot be rejected. Now, if this cannot be rejected, then for the equality of means, we will go for the pooled sample variance procedure.

(Refer Slide Time: 51:02)



So, for testing equality of means, we will use pooled sample variance procedure. So, if you seeS p square, that is equal to 4 S 1 square plus 5 S 2 square by 9, that turns out to be, so, that is equal to, that is a huge value here.4 into 15750 plus 5 into 10920 by 9. So, that is equal to some value. So, S p is taken to be square root of that. That is 114.31. So, if we calculate the T variable here, that is root m n by m plus n x bar minus y bar by S p. Then, first thing is we observe here, that x bar is equal to 8230 and y bar is equal to 7940. So, this value turns out to be after calculation 4.19. So, if we see the t values at say 9 degrees of freedom, then even at 0.005, this value is 3.25. So, at 1 percent level also the hypothesis of equality of means is rejected.

So, we conclude that in the 2 mines, the measurements of the heat producing capacity are significantly different, because it may be due to difference in the type of the coal that is available, it may be due to the type of the mine that you are having. May be in one of the mines, you have a very low levelroofs and various kind of parameters which may be operating in those mines there. So, it may be due to that.

So, to sum up, if we are comparing the means of two normal populations, the first thing is we have to look at is that, what type of variances are there? If the variances are known then, we have one type of procedure. If the variances are unknown then we firstly, test whether the variances are same or not. If they are same then, we go for a pooled sample variance procedure. If they are not same then, we go for a different procedure which is an approximate test. We also see the correlation. If the correlation is present then we may go for a pairing. If we do not have the correlation, then we may go for independent samples. So, before adopting any procedure one has to carefully examine the problem and then, choose the appropriate test.

We have also seen the effect of choosing the null and alternative hypothesis. As I already mentioned, since we are controlling the probability of type 1 error, therefore, it is always reasonable to put the stronger hypothesis or you can say the conviction in which, we have more thatas an alternative hypothesis. Because, rejection of the hypothesis is strong conclusion whereas, acceptance of the hypothesis becomes a weaker conclusion. Simply, because of the reason that we are actually controlling the probability of type 1 error. In the forth coming lecture, I will be discussing the chi square test for goodness of fit or testing for the independence.