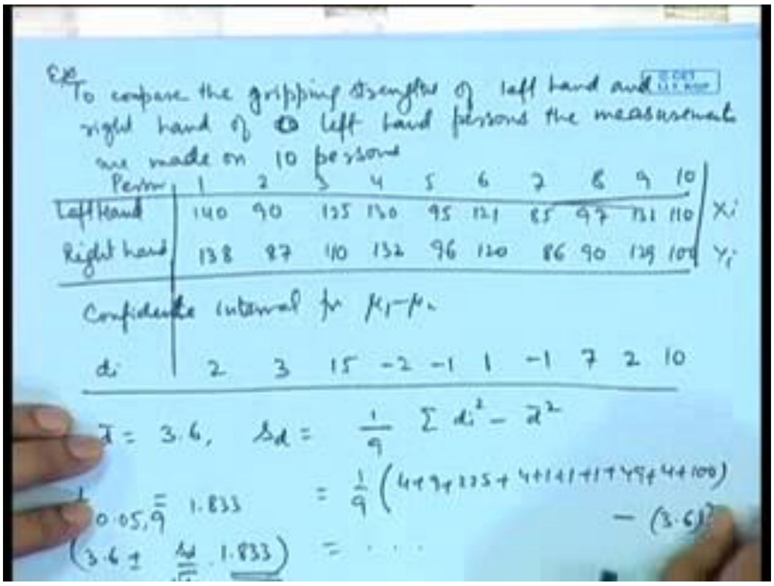
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 64
Confidence Intervals – IV

(Refer Slide Time: 00:41)

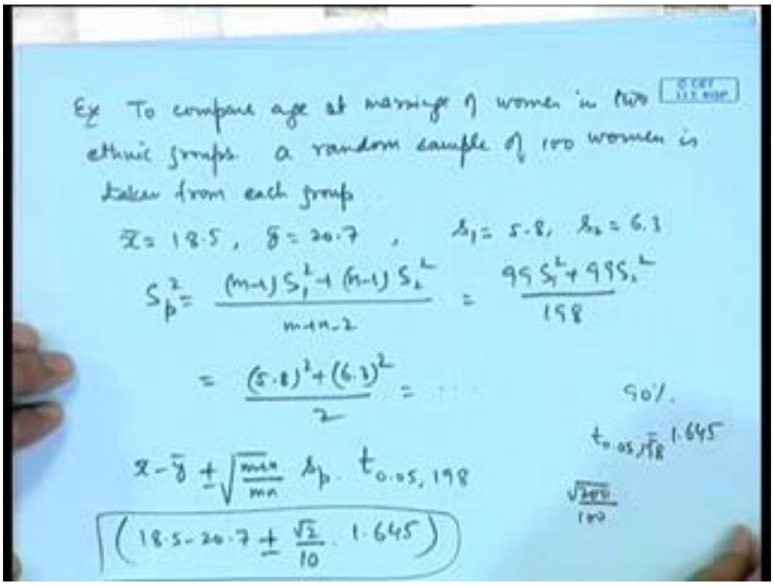


Let me take up some examples here to illustrate the situations. So, to compare the gripping a strengths of left hand and right hand of 10 handed of left handed persons the measurements are made on 10 persons and the data is observed. So, left hand and right hand, and we have of persons 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10. The gripping strengths are measured as 140, 90, 125, 130, 95, 121, 85, 97, 131, 110; for the right hand it is 138, 87, 110, 132, 96, 120, 86, 90, 129, 100.

So, we need the confidence interval for say mu 1 minus mu 2. Now observe here that this is the data related with the correlated observations. So, we will need here the means of; so let me call this as the first set. So, this is xi data this is yi data. So, we will look at di's the differences here. So, the differences here is 2 3 15 minus 2 minus 1 1 minus 1 7 2 and 10. So, we look at the d bar value here which is the mean of this that is 1 t, so 17, 24, 26, 36 that is 3.6. Similarly we calculate sd that will be equal to 1 by 9 sigma di square minus d bar square. So, once again it can be easily evaluated it is 4 plus 9 plus 225 plus 4 plus 1 plus 1 plus 1 plus 49 plus 4 plus 100 minus 3.6 square. So, this value can be evaluated.

Now, we look at the value of t on; suppose we want a 90 percent confidence interval. So, we need 0.059 that is equal to 1.833. So, we get the confidence interval as 3.6 plus minus sd by root 10 into 1.833; that will be the confidence interval for the difference in the gripping strengths of the left hand and the right hand of the left handed persons.

(Refer Slide Time: 05:10)



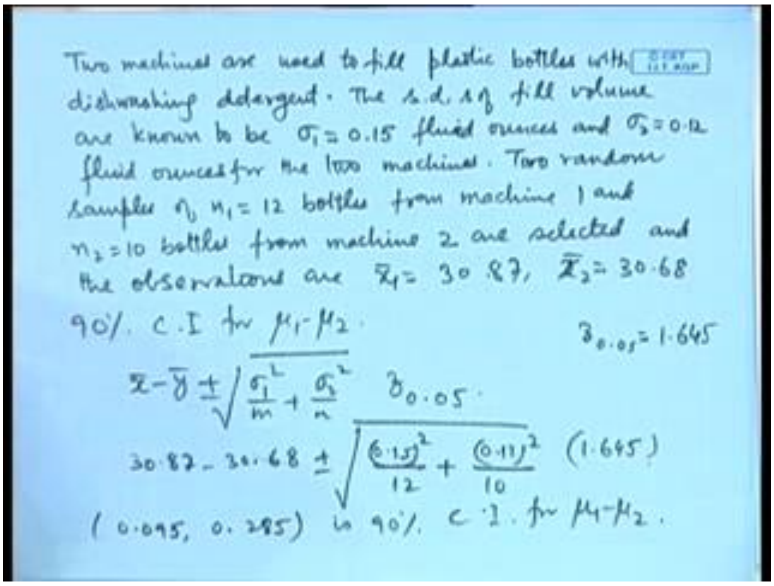
Let me take another example here: to compare age at marriage of women in two ethnic groups, a random sample of 100 women is taken and we observed that x bar is equal to 18.5 years, y bar is equal to 20.7 years and s 1 is equal to 5.8, s 2 is equal to 6.3 and we want say a confidence interval for this. So, we calculate here that we may use the model for S p square. So, S p square is equal to m minus 1 s 1 square plus n minus 1 s 2 square divided by m plus n minus 2 that is equal to 99 s 1 sqaure plus 99 s 2 square by 198 that is equal to 5.8 square plus 6.3 square by 2. So, this value can be evaluated.

In a similar way we have the confidence interval as x bar minus y bar minus root m plus n by m n that is and then s p into t alpha by 2; suppose I want 90 percent confidence interval so 0.05 and the degrees of freedom will be m plus n minus 2. So, this value we can see t 0.05 198; just almost as a normal distribution 1645. So, we substitute these values here 18.5 minus 20.7 minus this is 100 plus 100 that is 200 by 100. So, that is root

2 by 10 into 1.645. So, plus minus that gives the confidence interval for the difference in ages at marriage of women in two ethnic groups.

So, here we have you would the pooled formula. We may actually do a testing of hypothesis for sigma s 5 is equal to sigma 2 square and if sigma 1 square is equal to sigma 2 square is accepted then we may go for this formula.

(Refer Slide Time: 08:25)

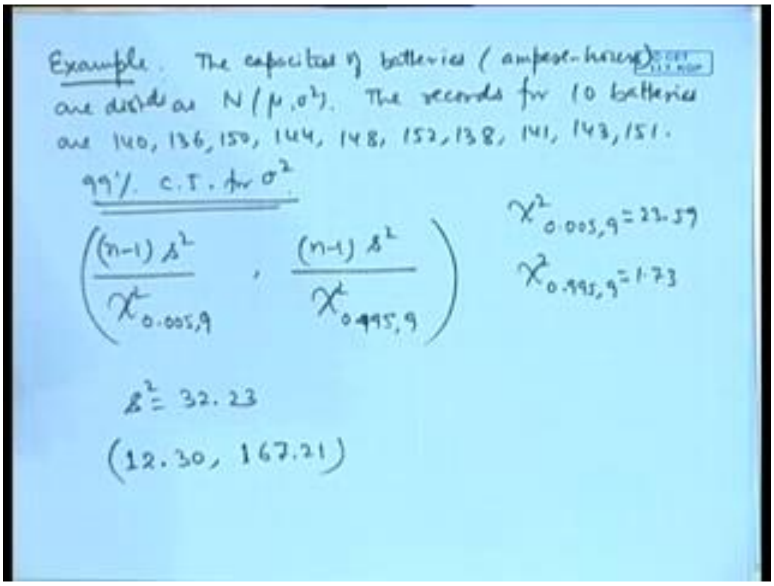


Let me take another example here: two machines are used to fill plastic bottles with dishwashing detergent the standard deviations of fill volume are known to be sigma 1 is equal to 0.15 fluid ounces and sigma 2 is equal to 0.12 fluid ounces, for the two machines. Now two random samples of n 1 is equal to 12 bottles from machine 1 and n 2 is equal to 10 bottles from machine 2 are selected. And the observations are: x 1 bar is equal to 30.87, x bar is equal to 30.68. So, find 95 percent confidence interval for mu 1 minus mu 2.

So here we can see; we can look at the confidence interval as x bar minus y bar plus minus square root sigma 1 square by m plus sigma 2 square by n z 0.05. Now z 0.05 we can see from the tables of normal distribution it is 1.645. So, this interval becomes 30.87 minus 30.68 plus minus square root. Now 1 sigma square is 0.15 square by 12 plus sigma 2 square is 0.12 square by n, n is 10 multiplied by 1.645. So, after simplification this value is turn out to be 0.095 to 0.285. So, this is 90 percent confidence interval for the mean difference that is mu 1 minus mu 2.

So, here the variances were known sigma 1 square and sigma 2 square so we have adopted a procedure were the formula for known variances is utilized here.

(Refer Slide Time: 12:19)



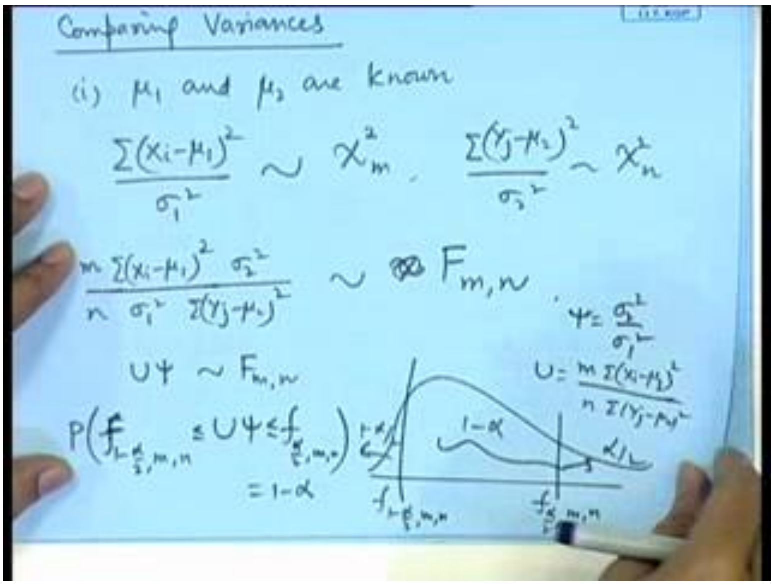
Let me take another example here: the capacities of batteries, so these are measured in say ampere hours they are distributed as normal mu sigma square. The records for 10 batteries are say 140, 136, 150, 144, 148, 152, 138, 141, 143, 151; we want 99 percent confidence interval for sigma square.

So, now here we will make use of the fact that mu is unknown. So, if mu is unknown then the formula for confidence interval for sigma square is based on chi square on n minus 1 degrees of freedom. The formula is n minus 1 s square by chi square, so 0.005 n is 10 so this is 9 to n minus 1 s square by chi square 0.095 9; so 995.

So, this values we see from the tables of the chi square distribution that is chi square 0.005 9 it is 23.59 and chi square 0.995 on 9 degrees of freedom is 1.73. So, s square we calculate here it is turning out to 32.23. So, after substitution of these values the confidence interval turns out to be 12.30 to 167.21, which is pretty large confidence interval but that will be there because we are considering for sigma square and the variability of the original sample itself is large; this is s square is 32.23 here.

If we reduce the confidence level, suppose we make it 90 percent then this will be shrinking. Since we have made a very a high confidence level that is why the confidence interval is very large which look likely in practical also.

(Refer Slide Time: 15:38)

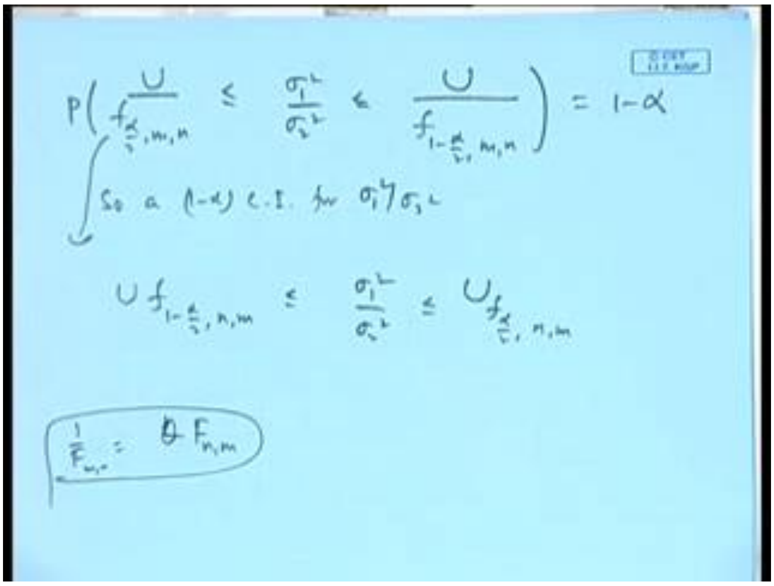


Next we look at the confidence intervals for variances, so comparing variances. Again we have two cases that is mu 1 and mu 2 are known, if mu 1 and mu 2 are known then we make use of sigma xi minus mu 1 square by sigma 1 square following chi square distributions on m degrees of freedom and sigma y j minus mu 2 square by sigma 2 square follows chi square distribution on n degrees of freedom. So, if you take the ratios here sigma xi minus mu 1 square by sigma 1 square pi m, so that is m here divided by sigma y j minus mu 2 square by sigma 2 square so that will come in the numerator divided by n. That will have chi square f distribution on m and n degrees of freedom.

So, if we look at this quantity if mu 1 and mu 2 are known then here the ratio sigma 2 square by sigma 1 square is coming. Let us denote it by say psi that is sigma 2 square by sigma 1 square. So, we are having and let me use the notation say U as m sigma xi minus mu 1 square divided by n sigma y j minus mu 2 square. So if we look at this one, then we are having U psi following f distribution on m n degrees of freedom. So, if we make use of the tables of f distributions that is f on m and n degrees of freedom here and f 1 minus alpha by 2 on m n degrees of freedom this is alpha by 2 and this is 1 minus alpha by 2, so this is 1 minus alpha.

So, probability of f 1 minus alpha by 2 m n less than or equal to U psi less than or equal to f alpha by 2 m n that is equal to 1 minus alpha.

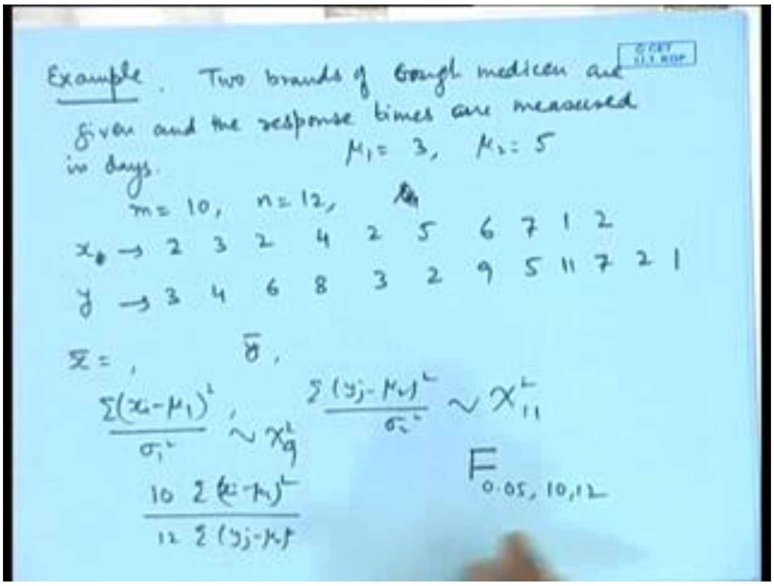
(Refer Slide Time: 18:24)



So, we can write probability of U, so divided by f of alpha by 2 m n less than or equal to sigma 1 square that is 1 by psi it becomes sigma 1 square by sigma 2 square less than or equal to U divided by f 1 minus alpha by 2 m n; that is equal to 1 minus alpha. So, we have a 1 minus alpha confidence interval for sigma 1 square by sigma 2 square. This can also be write as U f 1 minus alpha by 2 n m to U f alpha by 2 n m by using the ratio or you can say reciprocal property of the f distribution. Because we know that 1 by f m n is 1 by is equal to f m n, so this property can be utilized here.

Let me give one example here for confidence interval for the ratios.

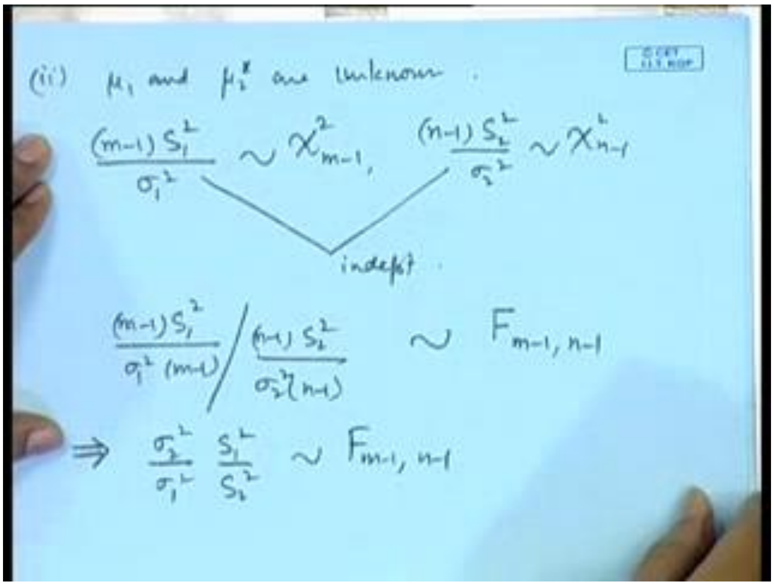
(Refer Slide Time: 20:40)



So, two brands of say cough medicine are given and the response times are measured in days. So, here we are having the data m is equal to say 10, n is equal to 12, s 1 and we are getting the observations as x 1 is equal to say 2 3 2 4 2 5 6; so 3 7 and then 1 2. So, we have 10 data here and for y we have the data say 3 4 6 8 3 2 9 5 11 7 2 1. Now based on this we calculate x bar, y bar, and we calculate sigma xi minus mu 1. So, it is given that mu 1 is say 3 and mu 2 is equal to 5. So, if we are looking at sigma xi minus mu 1 square and sigma y j minus mu 2 square then that will follow chi square on 9 and this divided by sigma 2 square will follow chi square on 11 degrees of freedom.

So, we can construct 10 sigma xi minus mu 1 square by 12 sigma y j minus mu 2 square. And then we need to look at the tables of F on say 0.05 10 and 12 degrees of freedom.

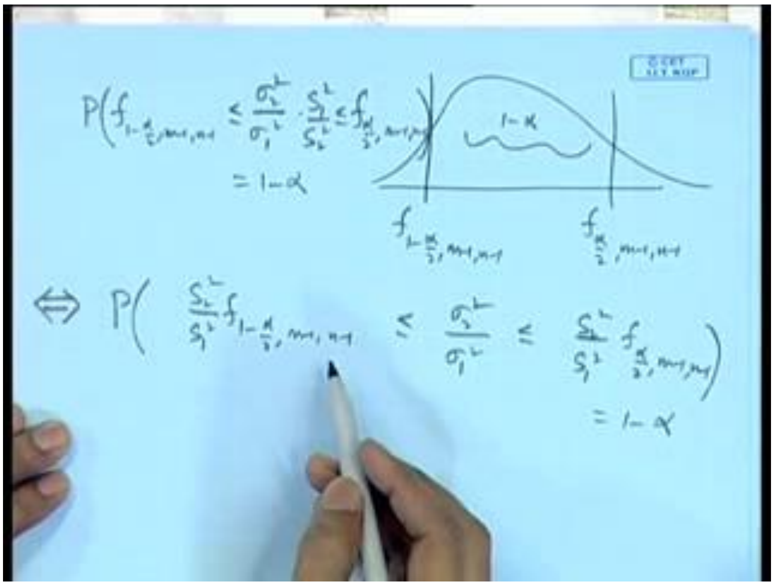
(Refer Slide Time: 23:18)



Now, another situation may occur when mu 1 and mu 2 are unknown. If mu 1 mu 2 are unknown then we will be not able to make use of the formula that we derived earlier because they are in the confident interval mu 1 and mu 2 are actually appearing. So, what we do we make use of s 1 square and s 2 square, we have m minus 1 s 1 square follows chi square distribution on m minus 1 degrees of freedom and n minus 1 s 2 square by sigma 2 square follows chi square distribution on n minus 1 degree of freedom.

Furthermore, these two random variables are independent. So, we can make use of the ratios m minus 1 s 1 square by sigma 1 square divided by m minus 1 divided by n minus 1 s 2 square by sigma 2 square into n minus 1 that will follow f distribution on m minus 1 n minus 1 degree of freedom; which is reducing to sigma 2 square by sigma 1 square s 1 square by s 2 square this follows f distribution on m minus 1 n minus 1 degrees of freedom.

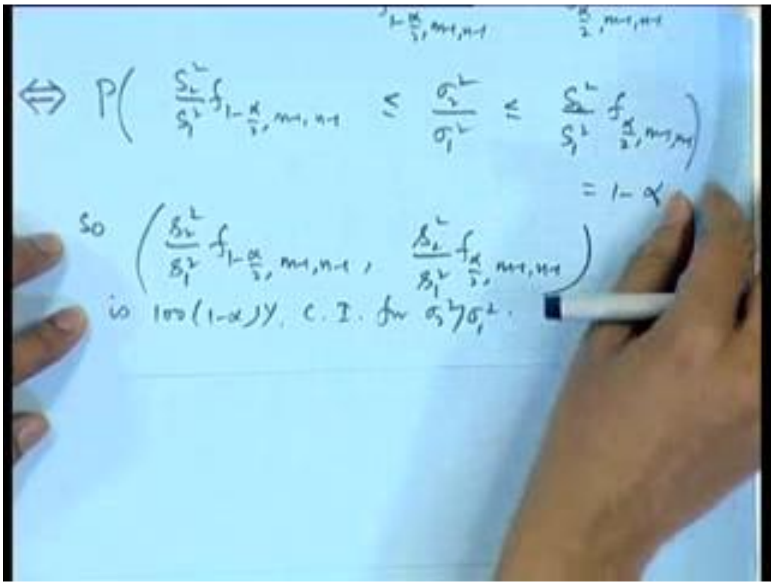
(Refer Slide Time: 24:40)



So, making the use of distribution of f that is we have f alpha by 2 m minus 1 n minus 1 and f 1 minus alpha by 2 m minus 1 n minus 1, intermediate probability 1 minus alpha. So, probability that f 1 minus alpha by 2 m minus 1 n minus 1 is less than are equal to sigma 2 square by sigma 1 square s 1 square by s 2 square is less than are equal to f alpha by 2 m minus 1 n minus 1 that is equal to 1 minus alpha. So, we make use of this and adjust the coefficient as probability that s 2 square by s 1 square f 1 minus alpha by 2 m minus 1 and n minus 1 less than are equal to sigma 2 square by sigma 1 square less than are equal to s 2 square by s 1 square f alpha by 2 m minus 1 n minus 1 that is equal to 1 minus alpha.

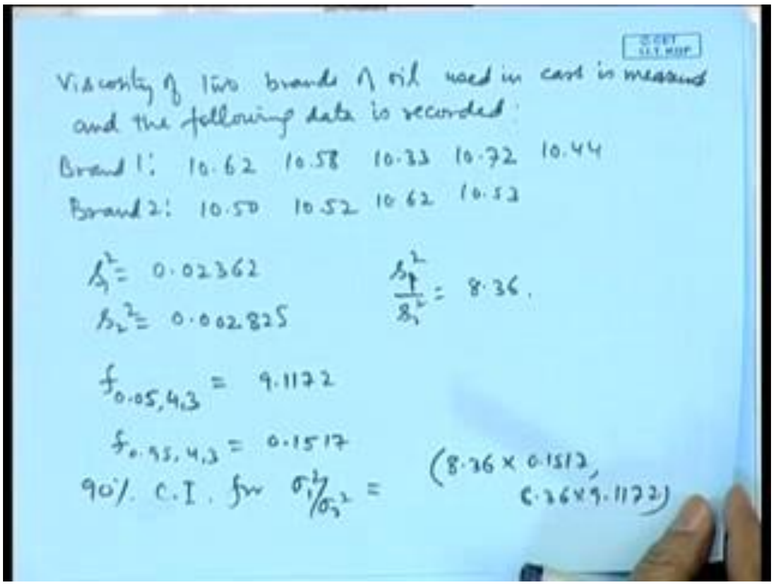
So, we are getting 101 minus alpha percent confidence interval for sigma 2 square by sigma 1 square. We can reverse if we want for sigma 1 square by sigma 2 square then we inter change the roles here we put sigma s 1 square by s 2 square and the degrees of freedom will get reverse it will become n minus 1 m minus 1.

(Refer Slide Time: 26:14)



We give one example here. So, s 2 square by s 1 square f 1 minus alpha by 2 m minus 1 n minus 1 to s 2 square by s 1 square f alpha by 2 m minus 1 n minus 1 is 100 1 minus alpha percent confidence interval for sigma 2 square by sigma 1 square.

(Refer Slide Time: 26:38)



So, say viscosity of two brands of oil used in car is measured and the following data is recorded. So, from brand 1 you have 10.62 10.58 10.33 10.72 10.44; for brand 2 it is 10.50 10.52 10.62 10.53. Suppose we want a confident interval for sigma 2 square by sigma 1 square. So, we will calculate the values here s 1 square s 2 square. So, s 1 square turns out be 0.02362, s 2 square is equal to 0.002825. You can see here there is a 10 times difference here. So, the f value that s 2 square by or you can say s 1 square by s 2 square will be equal to 8.36.

So, if we look at the f value on 0.05 say 1 2 3 4 5, so 4 three degrees of freedom that is equal to 9.1172 and f value 0.9543 that is equal to 0.1517. So, a 90 percent confident interval for sigma 1 square by sigma 2 square that will be equal to 8.36 into 0.1517 to 8.36 into 9.1172. So, this is the confidence interval for the ratio of the variances here.

So, in a given practical situation we need to analyze that what is the model that will be applicable and accordingly we make use of the formulae. So for example, when we are looking at the confidence interval for mu 1 minus mu 2 then we worry about that what is the status of the variances. If the variances are known then we have some formula, if that is based on the z that is normal distribution. If we have variance is unknown but equal then we have a formula which is based on a t distribution based on the pooling of the concept, pooling the variances. And if we have variances to be completely unknown then in that case we have another approximate t distribution formula and we make use of that.

On the other hand if the data is correlated then we make use of pairing and pair t formula is used. Similarly when we are worry about the confidence interval for the sigma 1 square and sigma 2 square then we look at the knowledge about the means. If the means are known then we have a formula based on f distributions on the total degree of freedom m and n. If the means are unknown then we have another formula which is based on s 1 square and s 2 square and the degree of freedom are slightly reduce to m minus 1 n minus 1.

Now, these formulae are quite standard because they are making use of sampling distributions from the normal populations. When we do not have normal populations then in that case we may have to look for appropriate sampling distribution; for example if we are dealing with uniform distribution, if we are dealing with from exponential populations, then we look at from the description that what is the sufficient statistics from where we find out the pivoting quantity if we are able to derive the sampling distribution of that.

So, the techniques for that and also for the propulsions are available and one can work out various formulae for confidence interval from other populations as well. So, that is part of another course that is statistical inference that will be doing later on.