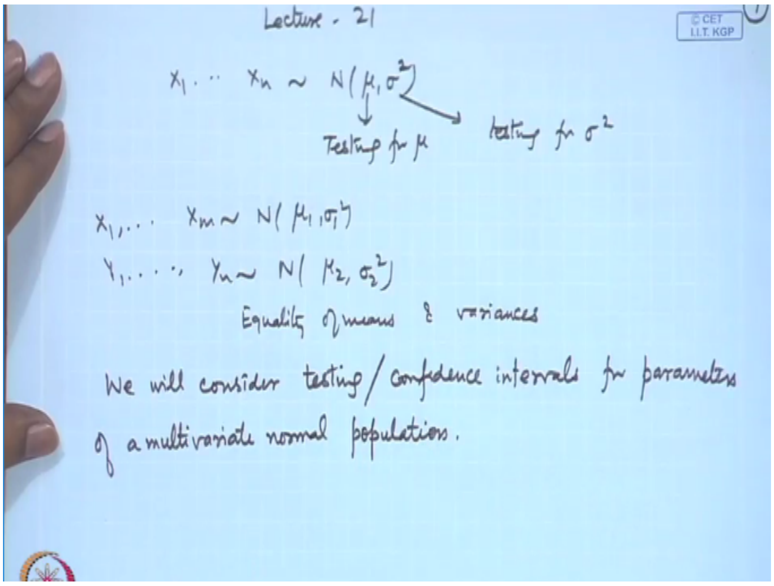
Statistical Methods for Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture – 21
Multivariate Analysis - VI

In the last lectures, I have introduced multivariate versions of the chi square distribution, which we call Wishart distribution, we also considered multivariate version of the student's t distribution, which we called Hotelling's T square distribution, we also see some other distribution such as non-central chi square, non-central t and non-central f and I showed a couple of applications where they arise.

(Refer Slide Time: 01:11)



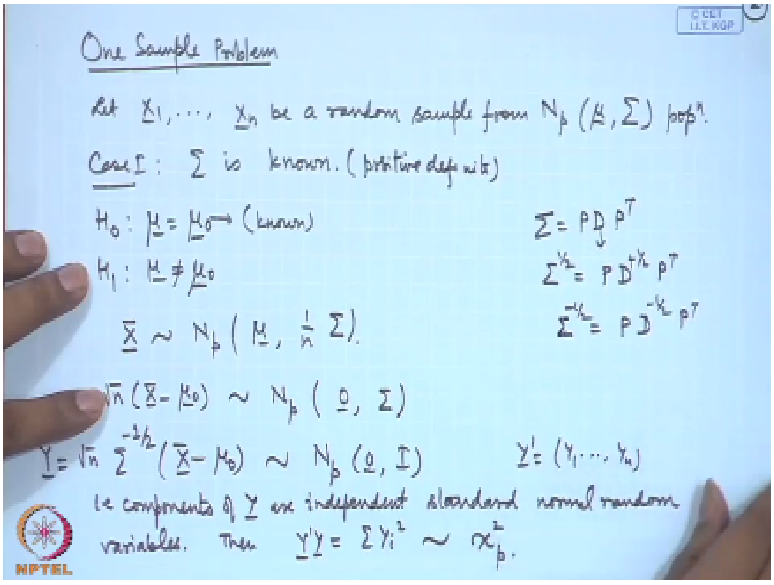
Basically, these distributions will be used when we consider testing in the multivariate normal population. If you remember the earlier lectures on the testing of hypothesis, we have introduced the testing for the parameters of a normal population. For example, testing for the mean like we have considered say, X1, X2, Xn a random sample from say normal Mu Sigma square.

Then, we have considered testing for Mu, we have also considered testing for variance, we also consider 2 sample problems that means we have say, X1, X2, Xm a random sample from normal Mu 1 sigma 1 square and Y1, Y2, Yn a random sample from another normal population say normal Mu 2 sigma 2 square. So, we have considered equality of means; equality of means and variances etc.

So, we have considered various testing situations for example, testing for Mu and sigma square is known, testing for Mu and sigma square is unknown, we have considered testing for sigma square, again when Mu is known or unknown. We also found the confidence intervals for these parameters in these situations. In the 2 sample problems, we considered testing for Mu 1 < or = Mu 2, Mu 1 is > Mu 2 etc., and similarly for sigma 1 square is = sigma 2 square, sigma 1 square < sigma 2 square etc.

We have seen that these tests are based on normal chi square t and f distributions. Now, in the multivariate situation, let us consider these. Of course, we also considered the confidence interval in all the situations and they were also dependent upon these distributions. So, now we consider the multivariate analogue of this testing and confidence interval problems. So, let us consider; we will consider testing and confidence interval for parameters of a multivariate normal population.

(Refer Slide Time: 03:40)



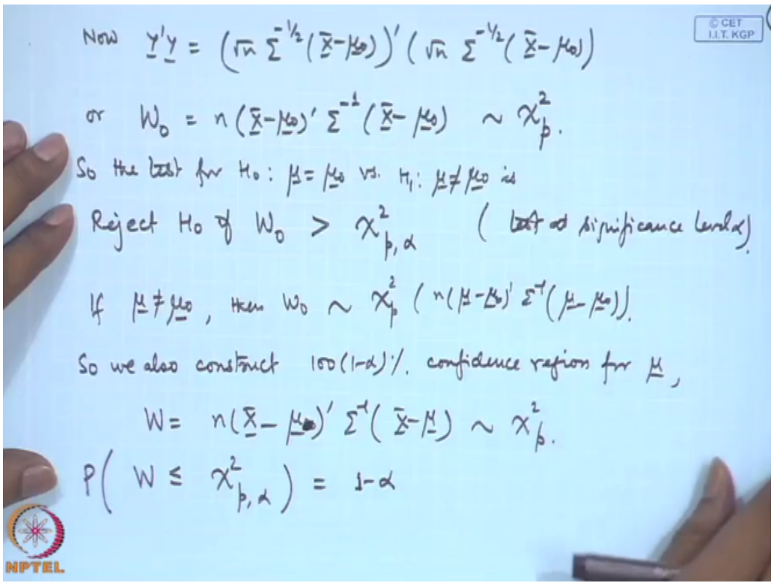
So, let me introduce 1 sample problem first. So, we will assume that we have a random sample; let X1, X2, Xn be a random sample from Np Mu Sigma population. So, let us consider sigma is known, so we can; let us consider say, testing for Mu, so we want to test whether the mean vector Mu is = a known vector Mu 0 against Mu is not = Mu 0, we consider the structure of the sufficient statistics here, X bar; X bar follows Np Mu 1/ n sigma.

So, based on this we can define root n X bar -; so for example, if I consider root n X bar - Mu 0 that will have Np 0 sigma and then if I consider a root n sigma to the power -1/2 X bar – Mu 0 that will have Np 0, I. Now, that means, we are assuming here sigma is known and positive definite, we are assuming it is positive definite, so that inverse is defined and we have already discussed in detail then that how to define sigma to the power -1/2 matrix.

That means we consider the spectral decomposition of sigma as P, D, P transpose, where D is a diagonal matrix and then we consider sigma to the power 1/2 as PD to the power minus 1/2 P transpose and sigma to the power -1/2 can now again we obtained as PD to the power -1/2 P transpose etc., so all these things can be determined for a positive definite matrix. Now, the components of this become independent standard normal random variables.

That is components; let us call it say, Y; components of Y are independent standard normal random variables, then Y prime Y, suppose I am writing say, Y is = Y1, Y2, Yn that is Y transpose is the row vector, then Y prime Y that is sigma Yi square that will follow chi square distribution on p degrees of freedom. These will be p components because we are dealing with the p dimensional normal distribution.

(Refer Slide Time: 07:33)

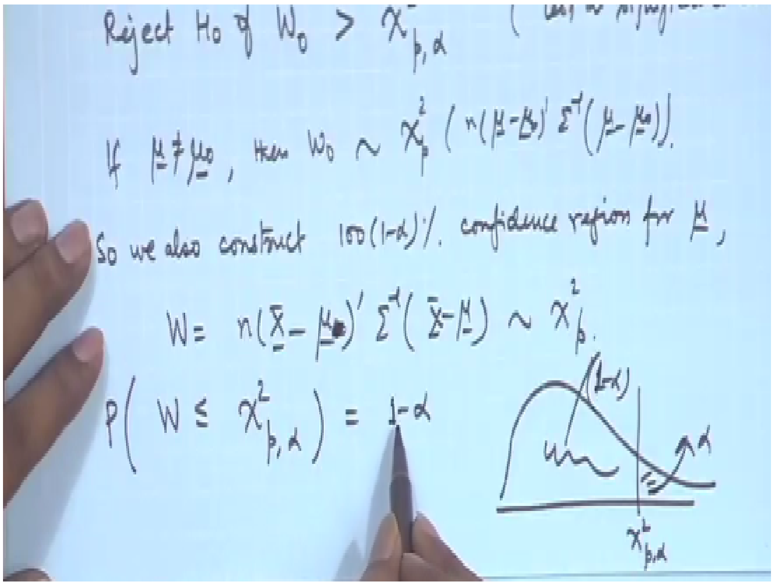


So, Y prime Y that is sigma Yi square will have a chi square distribution on p degrees of freedom, so if that is so; let us write Y prime Y. So, now what is Y prime Y? That will become = square root n sigma to the power -1/2 X bar - Mu 0 prime root n sigma to the power – 1/2 X bar - Mu 0 that is = n times X bar - Mu 0 prime sigma to the power - 1 X bar - Mu 0, then that will follow chi square distribution on p degree of freedom.

So, now based on this we can consider the test for Mu is = Mu 0, when Mu is = Mu 0, we are getting this, then Mu is = Mu 0, then we have this distribution. So, the test for H0 Mu is = Mu 0 against Mu is not = Mu 0 is reject H0, if this value; let us call it W0; W0 is > chi square p alpha at significance level alpha. Now, we can also consider based on this, see here what we are getting is that we are assuming Mu is = Mu 0.

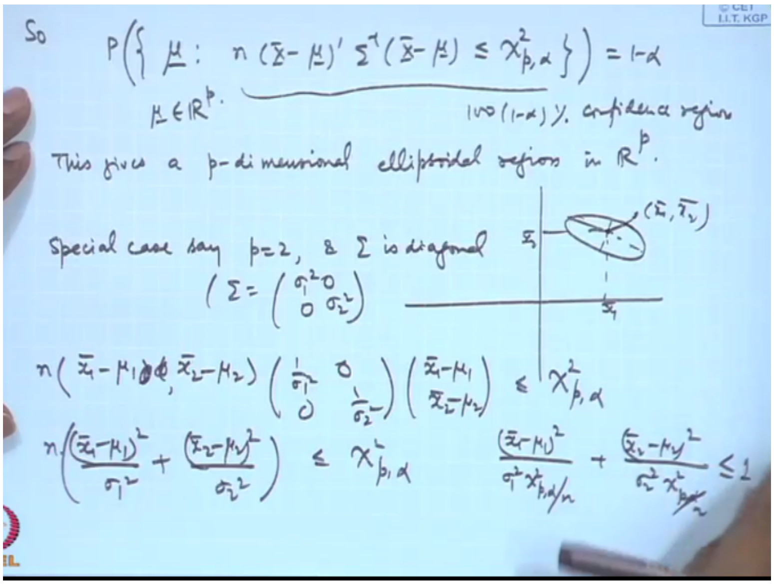
If Mu is not = MU 0 and then if I consider the distribution of W0, then that will be non-central chi square with non-centrality parameter Mu – Mu 0 prime sigma inverse Mu – Mu 0, so we can also construct 1 -; 100(1-alpha) % confidence region for Mu. If I consider say, W is = n X bar -; Mu 0 Mu 0, sigma inverse X bar – Mu, then that is having chi square p, so I can write probability of W < or = chi square p alpha that is = 1 – alpha.

(Refer Slide Time: 10:51)



So, because if I consider this region chi square p alpha, this is the point, this probability is alpha, so this probability is 1- alpha. So, if I consider this portion, then now here I consider the set of those Mu's for which this is satisfied.

(Refer Slide Time: 11:17)

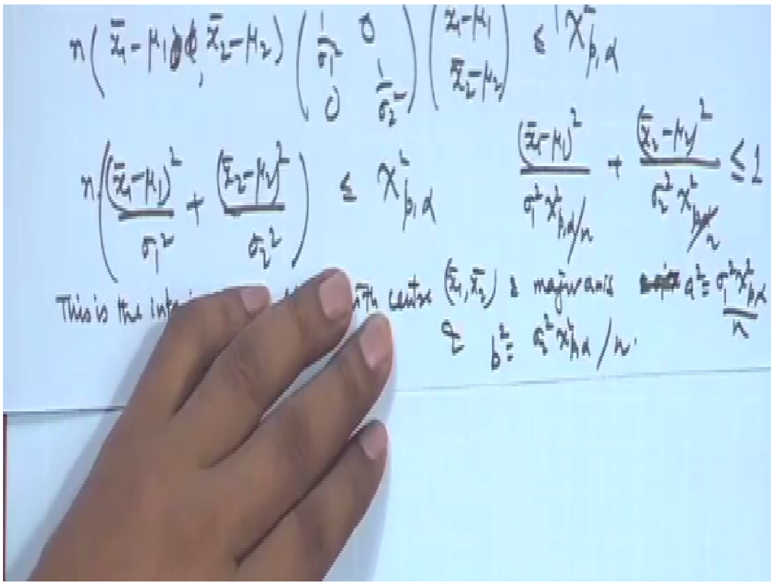


So, if I consider probability of the region, the set of all those Mu's for which n X bar - Mu transpose sigma inverse X bar - Mu is < or = chi square p alpha, then this is = 1- alpha, where Mu is vector in the p dimension. So, this gives a p dimensional ellipsoidal region in Rp, so this is called 100 (1-alpha) % confidence region. So, basically this is the interior and the boundary of the space. So, for example, if I consider 2 dimension, it may become something like this.

Suppose, this is my say, x1, x2 vector, so this is say; this point is say X1 bar, X2 bar and then you have this, so you are getting the components of this. Let us take say, special case say, p = 2 and sigma is a diagonal that is say, sigma is = sigma 1 square, sigma 2 square, then how this region will look like? This will become n X1 bar – Mu 1 X2 bar – Mu 2 1/ sigma 1 square, 1/ sigma 2 square, 0, 0, X1 bar - Mu 1, X2 bar – Mu 2 < or = chi square p alpha.

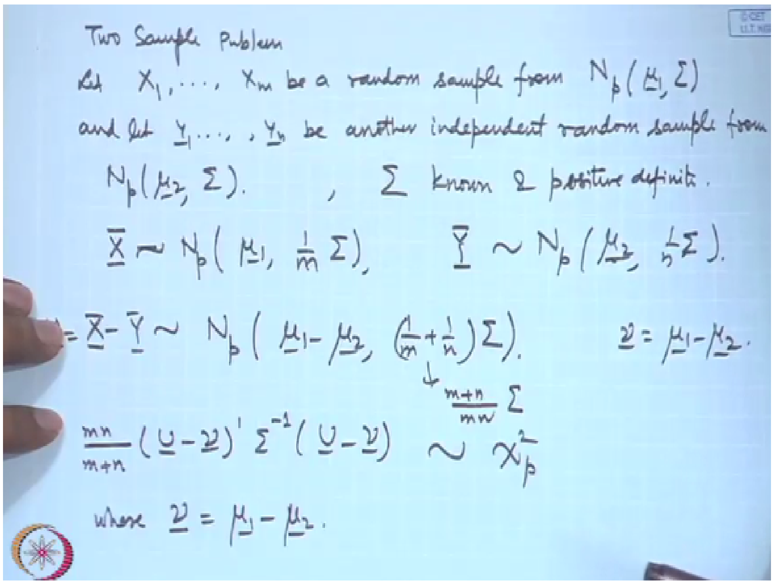
So, this quantity can be easily calculated, it is = n times; now if I multiply, I will get X1 bar - Mu 1 square/ sigma 1 square + X2 bar - Mu 2 square/ sigma 2 square < or = chi square p alpha, so this is can be easily seen that what is a ellipse here. Here, the centre is X1 bar, X2 bar and you are also getting, if I divide by this here, X1 bar - Mu 1 square/ sigma 1 square chi square p alpha divided by n + X2 bar - Mu 2 square/ sigma 2 square chi square p alpha/ n < or = 1.

(Refer Slide Time: 14:28)



So, this is the interior of the ellipse with centre X1 bar, X2 bar and major axis is = twice sigma 1 chi; so let us write, you are having, a square that is = sigma 1 square chi square p alpha/n and for minor axis, you are getting b, so b Square is here = sigma 2 square chi square p alpha/ n. So, you can easily plot the region and see how the ellipse will look like. So, we are able to solve this 1 sample problem, when the variance covariance matrix is assumed to be known.

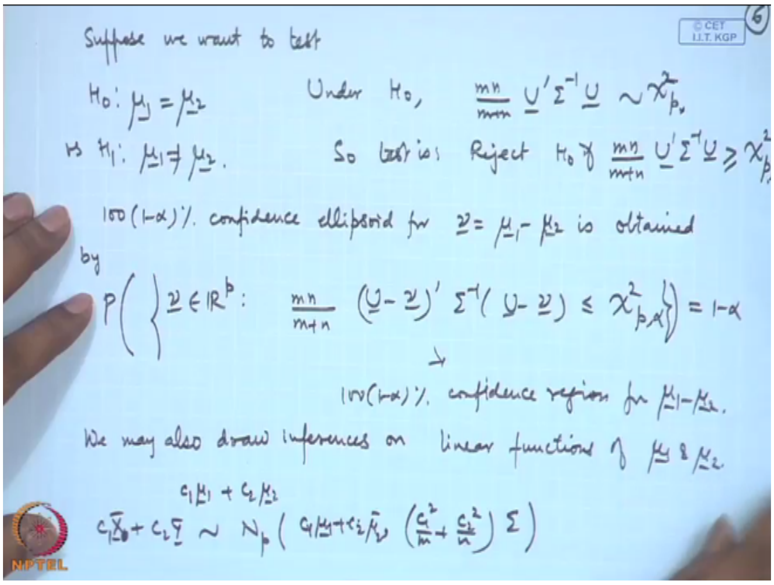
(Refer Slide Time: 15:58)



So, we are actually making use of the central chi square distribution. When the variance covariance matrix is known, we can also write down confidence region or the test for equality of means in the 2 population case or the 2 sample problem. So, let me consider 2 sample problem and so let us consider, let X1, X2, Xn be a random sample from Np Mu 1 sigma distribution.

And let, Y1, Y2, Yn be another independent random sample from Np Mu 2 Sigma population, here again I am assuming sigma is known and positive definite. Let us consider say, X bar, so that will have normal Np Mu 1, 1/m sigma, if I consider Y bar that will have Np Mu 2, 1/n sigma. So, let us work out the distribution theory here, X bar - Y bar that will be Np Mu 1 – Mu 2, 1/m + 1/n sigma, so we can call this Nu that is = Mu 1 – Mu 2.

(Refer Slide Time: 18:45)

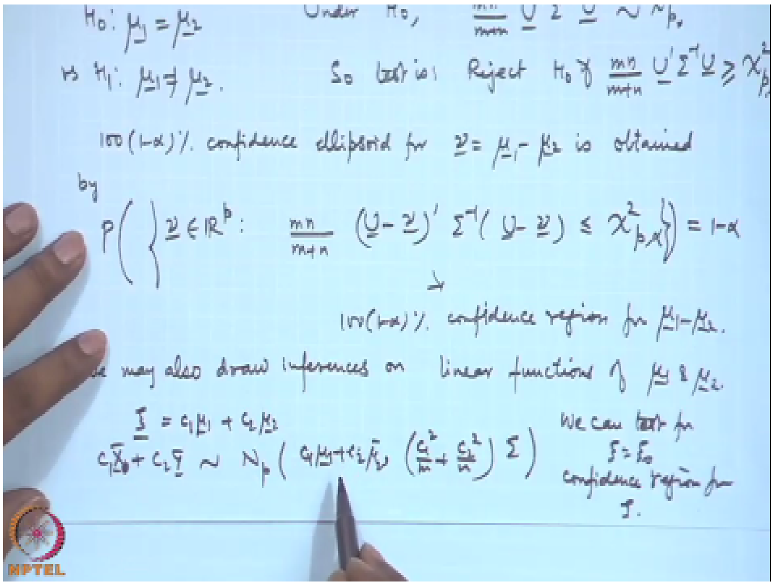


So, we can then write here, this is becoming m + n/mn, so we can write then, mn/ m + n, U -; say, Nu prime sigma inverse U - Nu that will have chi square distribution on p degrees of freedom, where I am defining this U is = the difference of X bar = Y bar and Nu is = Mu 1 –Mu 2. Therefore, this can be used for trying inference on Mu 1 – Mu 2. For example, if I want to do the testing, suppose we want to test say, H0; Mu 1 is = Mu 2 against say, H1 Mu 1 is not = Mu 2.

So, under H0, you will have mn/m + n, U - sigma inverse U; sorry, U prime sigma inverse U that will follow chi square p distribution. So, test is reject Ho, if this quantity mn /m + n U prime sigma inverse U is > or = chi square p alpha and we can also construct the confidence in region, 100(1- alpha) % confidence, again it will be ellipsoid only; ellipsoid for Nu is = Mu 1 – Mu 2 that will be; if I consider probability of say, Nu belonging to Rp mn / m + n U - Nu prime sigma inverse U – Nu < or = chi square p alpha, this is = 1- alpha.

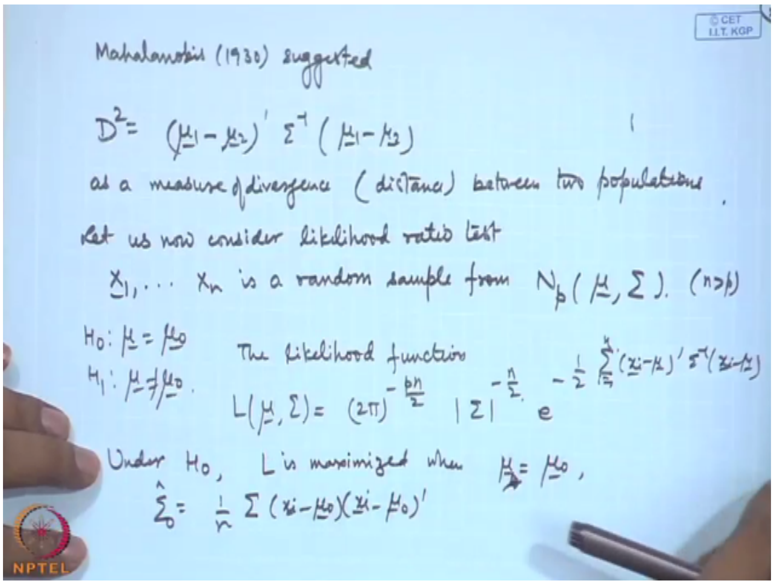
If I consider this region in the p dimensional Euclidean space, so this is 100(1-alpha) % confidence region for Mu1 – Mu2. One can actually also write for some linear combinations also, we may also draw inferences on linear functions of Mu1 and Mu2. For example, I consider say, some c1 Mu1 + c2 Mu2, then I can consider say, c1 X1 bar; c1 X bar + c2 y bar, then that will have Np and we can write down the distribution here.

(Refer Slide Time: 22:29)



We can also consider linear combination of the components of Mu1 and Mu2 that also we can consider, so for example here it will become c1 Mu1 + c2 mu2 and here I will get c1 square/m + c2 square/ n sigma. So, based on this, again we can construct test and confidence interval for c1 Mu1 + c2 Mu2, suppose I call it Xi, so we can test for Xi is = Xi 0 or we can find confidence intervals or confidence region for Xi.

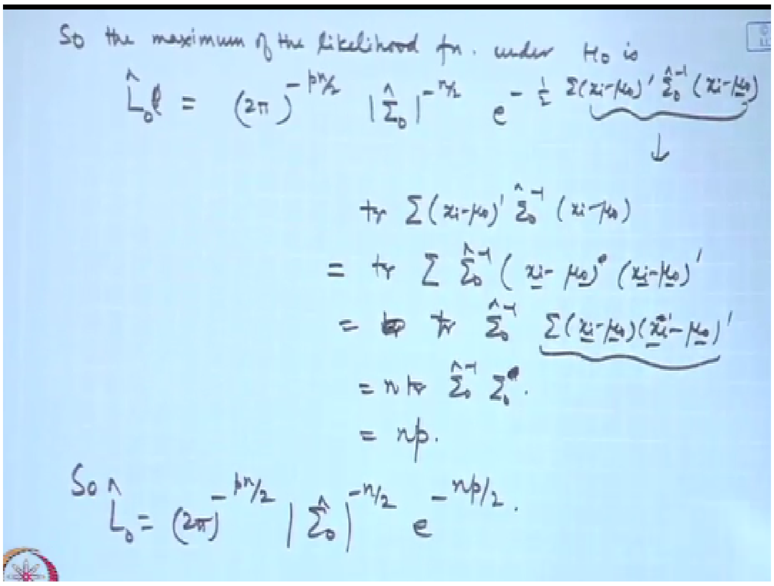
(Refer Slide Time: 23:15)



So, again it will be in the terms of chi square p distribution that is the central chi square distribution that we will be getting here. Actually this idea for making use of X bar – Mu, this term actually the initial ideas are hidden in the Mahalanobis D square statistic. So, let me just mention that thing. He suggested using D square that is Mu1 – Mu2 prime sigma inverse Mu1 – Mu2 as a measure of; he called it divergence or distance basically between 2 populations.

Let us also consider the general situations here say, likelihood ratio test, so again X1, X2, Xn is a random sample from Np Mu sigma and of course, we assume as usual n is > p and we are considering Mu is = Mu0 against Mu is not = Mu0. The likelihood ratio criteria involve the likelihood function, so we calculate the likelihood function here; 2 pi to the power – pn/2 determinant of sigma to the power –n/2, e to the power minu1/2 sigma xi - Mu prime sigma inverse xi – Mu.

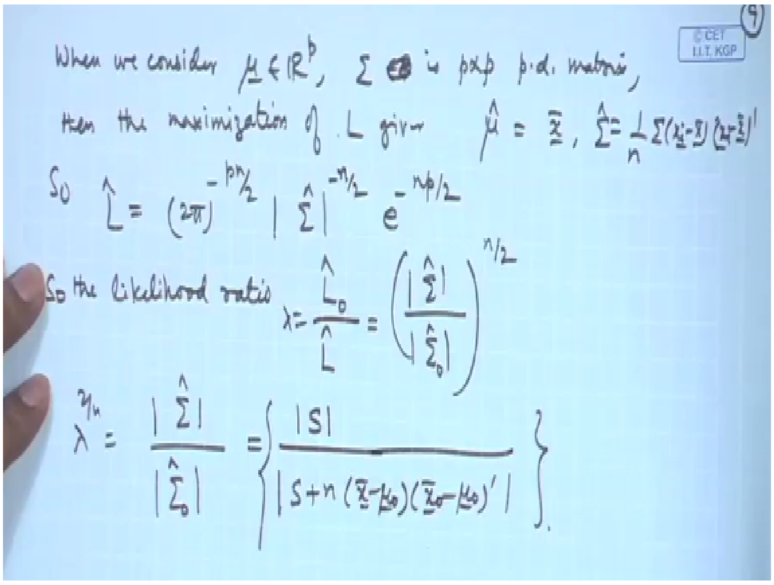
(Refer Slide Time: 26:10)



Now, under H0, this L is maximised, then Mu is = Mu0 because under H0 Mu is = Mu0 and sigma will be considered as; so let us put Mu0 here and sigma 0 that will be = 1/n sigma xi – Mu0, xi – Mu0 prime. So, the maximum of the likelihood function under H0 is; let us call it L head 0 that is = 2 pi to the power – pn/2 determinant of sigma 0 head to the power –n/2 and if I consider this term here; e to the power -1/2 sigma xi – Mu0 prime, sigma 0 head inverse xi – Mu0.

Now, if you look at this term, this is actually a scalar term; this term is a scalar, so we can also consider it as trace of this term. Now, this I can write as trace of sigma, sigma 0 head inverse xi – Mu0 prime xi – Mu0; xi – Mu0, xi – Mu0 prime. Now, this sigma I can take inside, so it becomes trace of sigma 0 head inverse sigma xi – Mu0, xi – Mu0 transpose but this is nothing but sigma 0, so this is becoming sigma 0 head inverse, sigma 0 head * n.

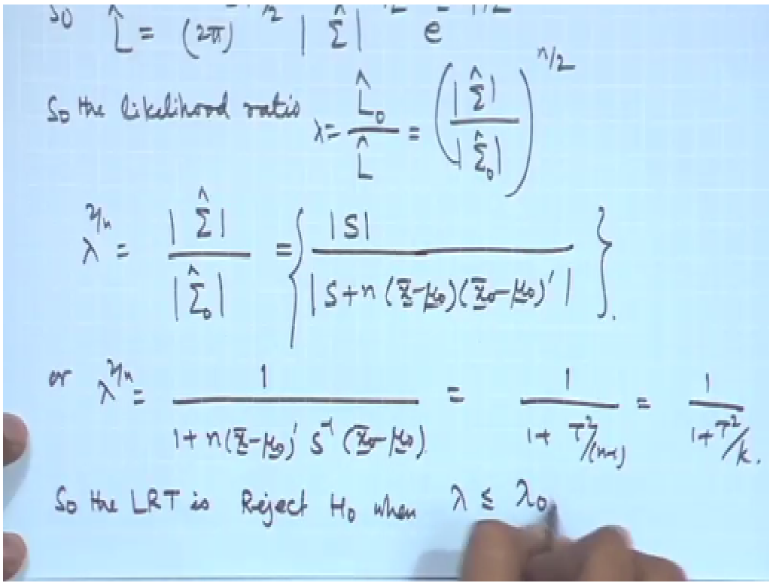
(Refer Slide Time: 28:25)



So, this is becoming np, so L0 head is becoming 2 pi to the power –pn/2 determinant of sigma head to the power –n/2 e to the power –np/2, Now, under the full space, then we consider Mu belonging to Rp and sigma is a p/p positive definite matrix, then the maximization of L gives Mu head is = X bar and sigma head is = 1/ n sigma xi - x bar, xi – x bar prime. Once again, if I put L head that is = 2 pi to the power –pn/2 determinant of sigma to the power; sigma head to the power -1/2, e to the power –np/2.

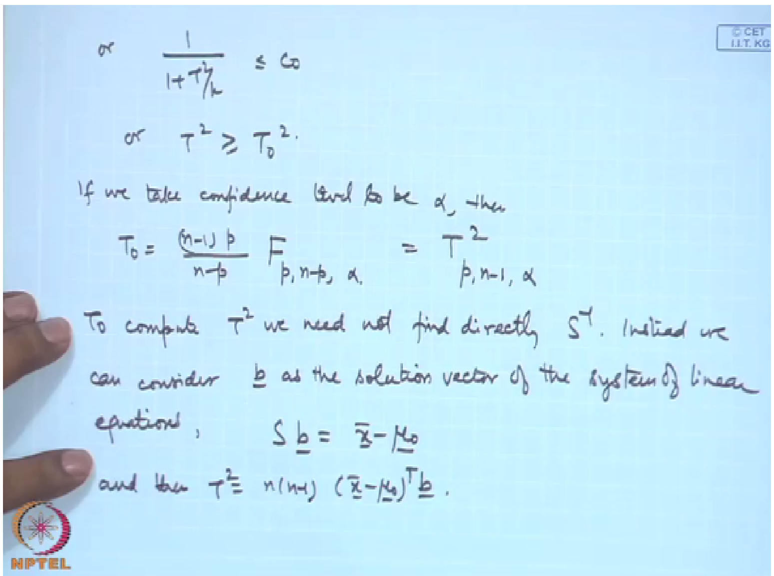
So, this will be same part here, so the likelihood ratio that we consider that is L0 head/ L head that is =; so if you look at these terms here, L0 head and L head, then these 2 things are common; these two terms are common, so this will get cancelled out, you are left with only determinant of sigma 0 head and determinant of sigma head and this is -1/2, I am sorry; this is –n/2 here, so this will be –n/2 and this will become –n/2.

(Refer Slide Time: 31:00)



So, this is now I am getting determinant of sigma head/ determinant of sigma 0 head to the power n/2, let us call it say, lambda. So, lambda to the power 2/ n that is = determinant of sigma head divided by determinant of sigma. Now, this is nothing but S here; divided by S + n times x bar – Mu0, x0 – Mu0 prime or we can consider lambda to the power 2/n, this will become = 1/ 1+ n x bar – Mu0 prime S inverse x0 – Mu0, which is nothing but 1/ 1 + T square/ n - 1 that is 1/ 1 + T square/ k.

(Refer Slide Time: 31:58)

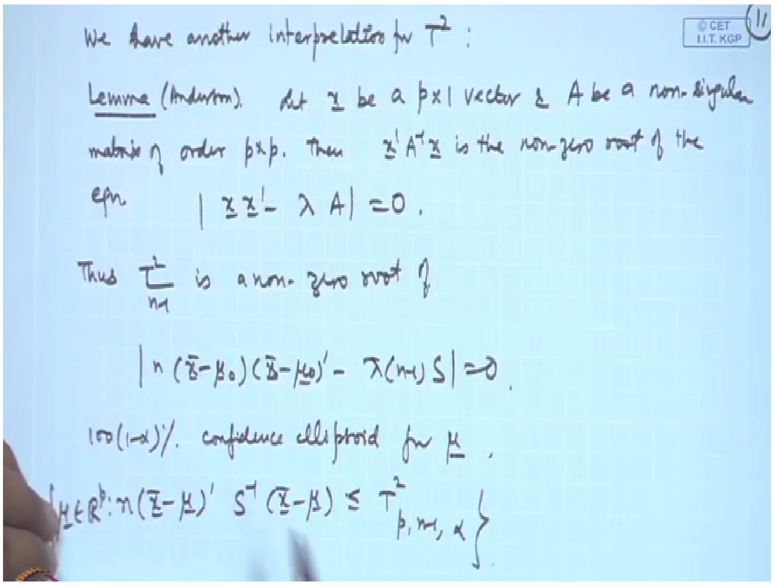


This T square/ k, this term I introduced in the last class, which is coming from the Hotelling's T square distribution, so the likelihood ratio test is reject H0, when lambda is < or = some lambda 0, 1/1 + T square/ k < or = some c0 or we can say, T square > or = some T0 square. If we take confidence level to be alpha then, T0 we can choose to be n – 1 p/ n – p, F on p, n - p alpha, this value actually we will call the percentage point of the Hotelling's T square distribution.

Here, 1 computational problem is there that is; if the data is given to you, you need to evaluate this T square here that is nx bar – Mu0 prime S inverse X0 – Mu; this involves the evaluation of the inverse of S, which may be quite complicated. For example, if you have p is = 4 or p is = 5, then this is quite complicated exercise but one can actually do it by using numerical techniques, you consider it as a solution of the simultaneous linear equations.

Let me just present a method here. To compute T square, we need not find directly S inverse, instead we can consider b as the solution vector of the system of linear equations that is Sb is = X bar – Mu0 and then T square is nothing but n * n - 1 X bar – Mu0 transpose b. So, one can use some numerical technique like Gauss elimination backward etc., all those things can be used for solving this system.

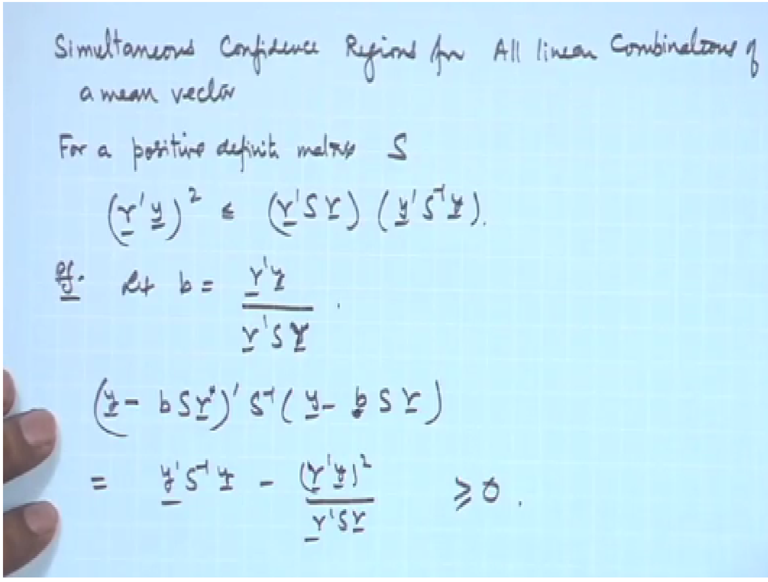
(Refer Slide Time: 34:42)



There is another interpretation to this, here we have another interpretation for T square, let me firstly state a Lemma, which is from Anderson, let x be a p/1 vector and A be a non-singular matrix of order p/p, then x prime A inverse x is the nonzero root of the equation x, x prime – lambda, A is = 0. If I use this, then I can say that T square/ n -1 is a nonzero root of n, x bar – Mu0 x bar – Mu0 transpose - lambda n – 1 S is = 0.

Similarly, the 1 - alpha confidence region; so here we will have, 100(1-alpha) % confidence ellipsoid for Mu, this will be nx bar - Mu prime S inverse x bar - Mu < or = T square pn - 1 alpha. The set of all the Mu's in Rp satisfying this condition, this set is the confidence region for Mu here. As we have given the interpretation earlier, this is ellipsoid in the higher dimensional space. One can also actually find out, as I mentioned a little earlier that we can consider linear combinations of vectors.

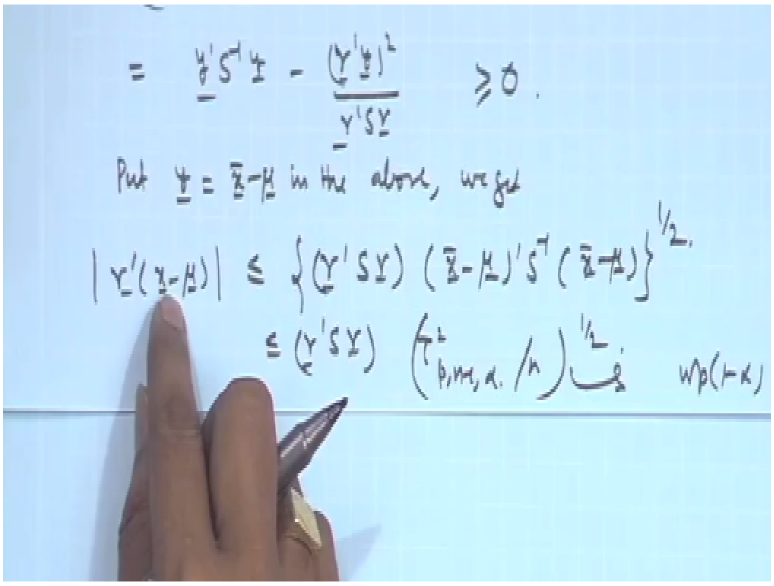
(Refer Slide Time: 37:45)



So, for example here I mentioned that we can consider c1 Mu1 + c2 Mu2 etc, we can consider more than 1 also that means, we can consider simultaneous confidence intervals for all linear combinations of a mean vector. So, that also we can give; let me just briefly mention about that also here. For all linear combinations of a mean vector, we first have the following result that for a positive definite matrix S, gamma prime y square is < or = gamma prime S gamma, y prime S inverse y.

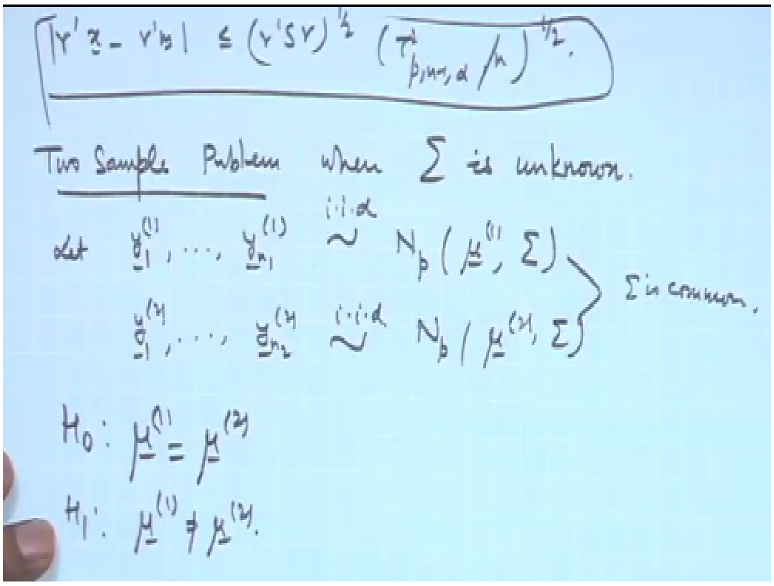
Let us look at the proof here, let us consider say, b to be gamma prime y divided by gamma prime Sy; S gamma. Now, if I consider y - bS gamma prime S inverse y - bS gamma, this is a scalar here. Now, if I expand this, I can get it as = y prime S inverse y - gamma prime y square/ gamma prime S gamma, now this is > or = o, so this gives the result here. This is basically you can consider as a generalization of the Cauchy Schwarz in equality to higher dimensions.

(Refer Slide Time: 39:39)



So, if I substitute say, y is = X bar – Mu, then we get gamma prime x - Mu is < or = gamma prime S gamma x bar Mu S inverse x bar - Mu to the power 1/2 but if we use the distribution of this, then this is nothing but gamma prime S gamma T square pn - 1 alpha divided by n to the power 1/2, this is true with probability 1 – alpha, so we are getting that all linear combinations here, they will satisfy simultaneous inequalities of the form.

(Refer Slide Time: 40:46)



That is gamma prime x - gamma prime m < or = gamma prime S gamma to the power 1/2 T square pn -1 alpha divided by n to the power 1/2, so simultaneously these are satisfied here. So, we have considered 1 sample problem for Mu, when Sigma is known and also we have resolved the problem, when sigma is unknown. We have considered the 2 sample problem, when sigma is known.

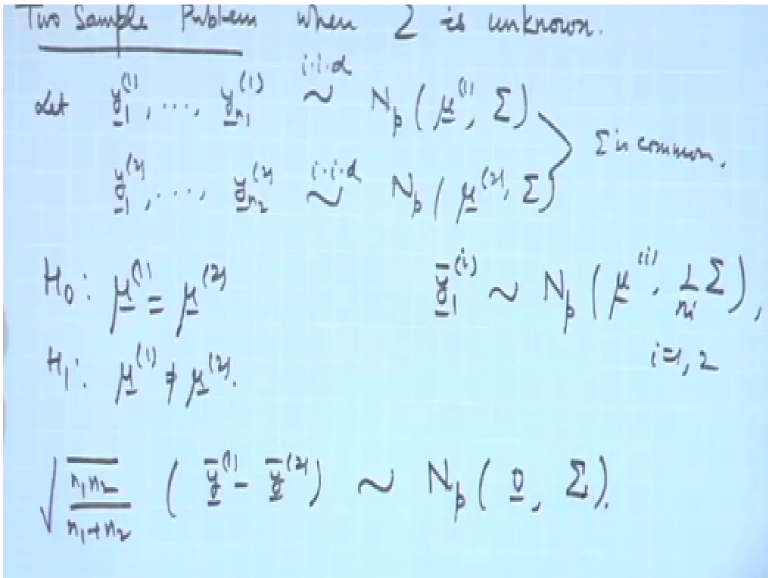
Now, let us consider 2 sample problem, when sigma is unknown, here again as before we have 2 cases; 1 case in which the variance covariance matrix is considered to be common and in another case, we will consider it to be uncommon and the procedures will be different as you had seen in the case of univariate problem. So, let us consider the 2 sample problem, then sigma is; now when sigma is unknown, so we are actually considering; I am just a little bit modifying the notations here.

So, let us consider say, we write in terms of y itself just because we will consider a generalization to higher dimensions also and that means multi sample also, so if I consider y1,1 and so on, yn1, so this is a random sample from Np Mu1 sigma, so these are independent and identically distributed and similarly I consider y1,2 and so on, yn2, 2, this is a random sample from Mu2 sigma.

So, these 2 are same, sigma is common but unknown and now as before we will be testing equality of the mean vectors. Let me mention here that we have considered the problems in which the testing problem is about the equality or not, in the case of univariate we had seen other kind of testing problems also like Mu1 < or = Mu2 or Mu1 > Mu2 etc., but here I am not giving those procedures here.

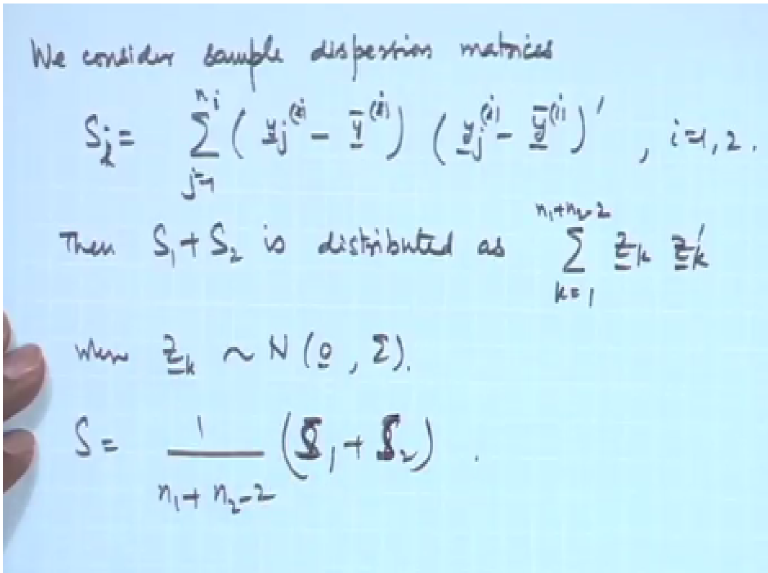
In fact, if we consider inequalities, then there can be various cases for example, first component may be less, second component may be more, third component may be equal, so you can have various kind of hypothesis testing problems. Some of the popular ones are like ordered alternatives we call, in which the concept of isotonic regression is used, so there are many research papers currently available on that topic both for the known and unknown variance cases.

(Refer Slide Time: 44:30)



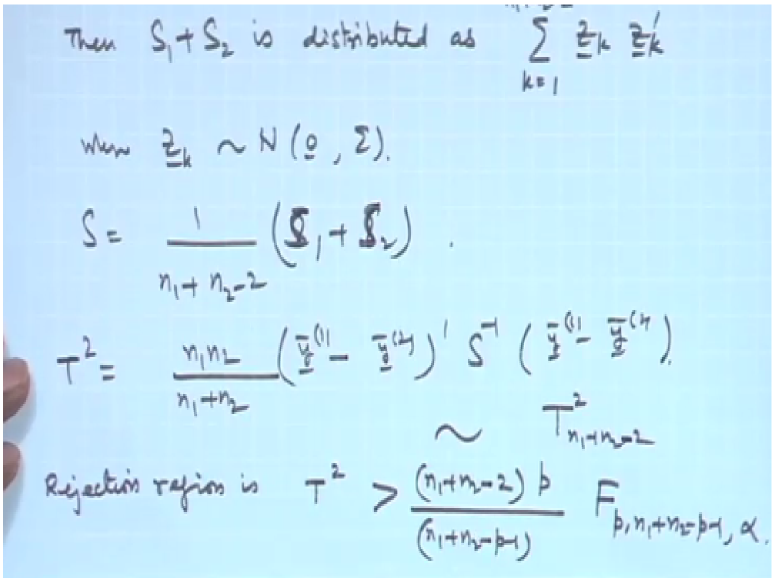
In this particular course, we will discuss only the basic ones that means the equality concept is being tested here. So, we will construct the Hotelling's T square here, so let us consider the sample mean vectors; y1 bar that will be Np Mu i 1/ni sigma, for i is = 1, 2. So, if I consider the difference square root n1 n2/n1 + n2 y bar 1 - y bar 2 that will follow Np 0, sigma. Now, in this case sigma is unknown, so we make use of S, now.

(Refer Slide Time: 45:27)



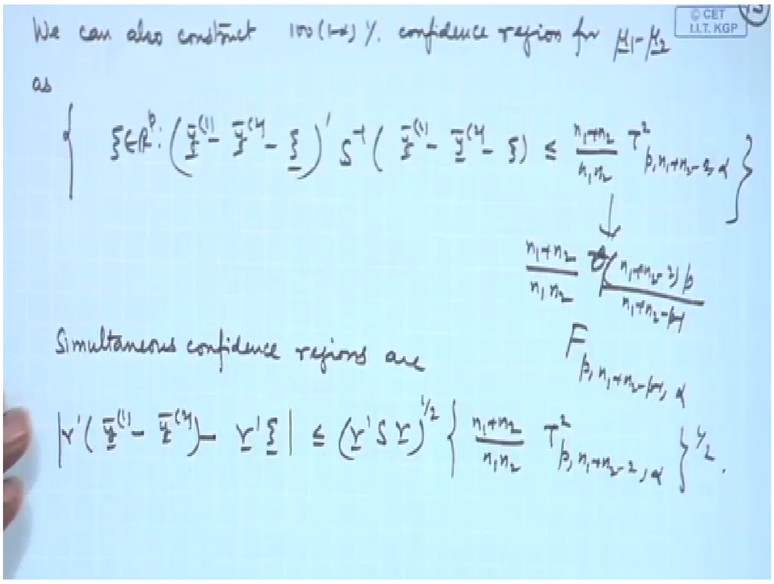
And for S, we have 2 things like from the first sample, we will get the variance covariance matrix as S1 and from the second, I will get variance covariance matrix as S2 and then we will consider pooling of that, so let us define this. We consider sample dispersion matrices, so that is S1 that is = sigma yj1 - y bar1, yj1 - y bar1 transpose, j is = 1 to n1, so if I put here ni and here I put i and this is I can call i; i is = 1, 2.

(Refer Slide Time: 47:08)



And we can consider S1 + S2, then this is having the same distribution, sigma zk, zk transpose, k = 1 to n1 + n2 -2, where zk is normal 0, sigma, so we define S is = 1/n1 + n2 - 2 sigma 1 + sigma 2; sorry S1 + S2, then based on this we can define the Hotelling's T square that is n1 n2/ n1 + n2 y bar1 - y bar2 transpose S inverse y bar1 - y bar2. Then, this has a Hotelling's T square distribution on n1 + n2 - 2 degrees of freedom.

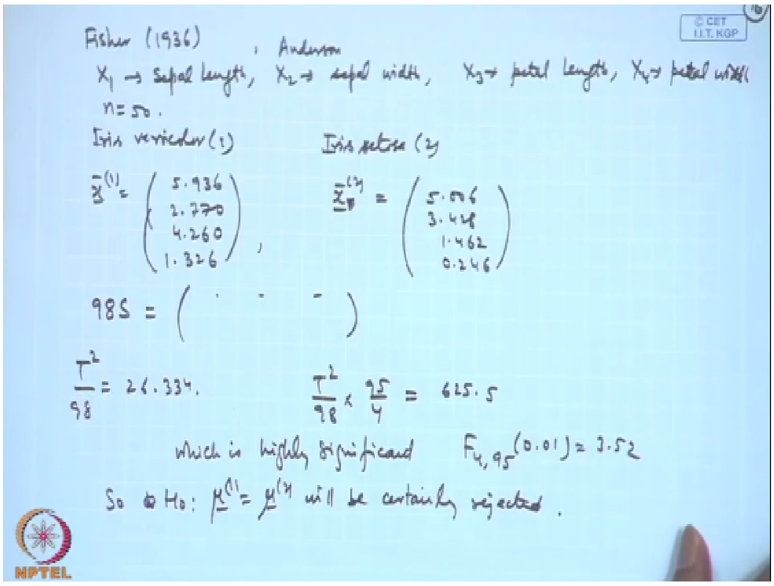
(Refer Slide Time: 48:27)



So, if we consider based on the representation in terms of F, so the rejection region is T square > n1 + n2 - 2 * p/ n1 + n2 – p – 1, Fp n1 + n2 - p -1 alpha, so this is level of significance level here will be alpha for this. We can make use of this for constructing the confidence region also for Mu1 – Mu2, we can also construct; it is the set y1 bar – y2 bar – some Xi S inverse y1 bar – y2bar - Xi < or = n1 + n2/ n1 * T square pn1 + n2 - 2 alpha.

The set of all p dimensional vectors, which satisfy this, so this is the 100(1-alpha) % confidence ellipsoid for Mu1 – Mu2. We can also write the; of course, this term you can see that this is also equal to n1 + n2/ n1n2 T square * n1 + n2 - 2 p divided by n1 + n2 - p – 1, Fpn1 + n2 – p - 1 alpha, so one can evaluate this using the tables of the F distribution. Similarly, the simultaneous confidence intervals can be written, gamma prime - gamma prime Xi < or = gamma prime S gamma to the power 1/2 n1 + n2/n1 n2 T square p n1 + n2 - 2 alpha to the power 1/2.

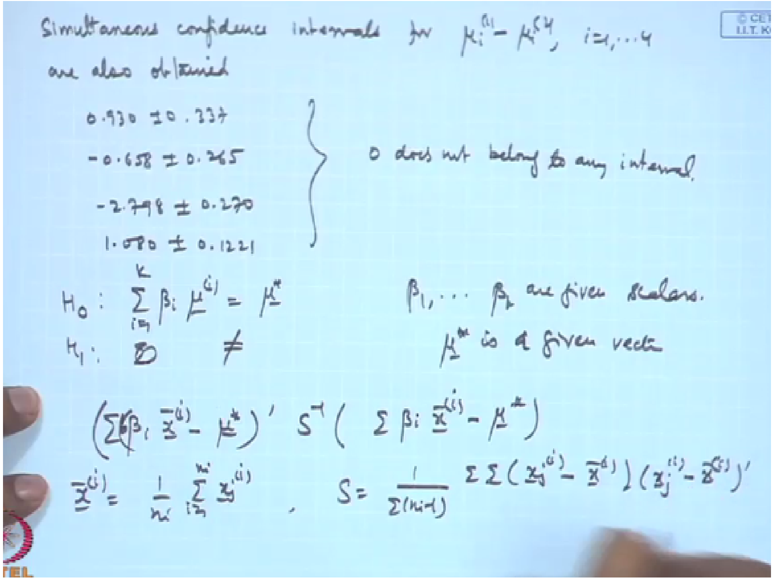
(Refer Slide Time: 51:02)



One of the classical examples is given in the Fisher's paper in 1936, in which he considered the 4 variables as sepal length, x2 as the sepal width, x3 as the petal length, and x4 as the petal width and this data I have taken from the book of Anderson and 50 observations were taken on 2 populations; 1 is iris versicolor and another is the iris setosa, the summarized data is x bar1 = 5.936, 2.770, 4.260, 1.326, this is the mean vector; sample mean vector based on 50 observations on the iris versicolor trees.

And the x2 bar vector that is on the 50 random observations taken on iris setosa trees; 5.006, 3.428, 1.462, 0.246 and n1 + n2 - 2 is 98S, so that is given here, I am not writing it here. So, T square/ 98 value turned out to be 26.334, so if we consider T square/98 * 95/4 is = 625.5, which is highly significant, if I take F4, 95 at say, 0.01 etc., that is 3.52 only, so naturally; so H0 that is Mu1 is = Mu2 will be certainly rejected.

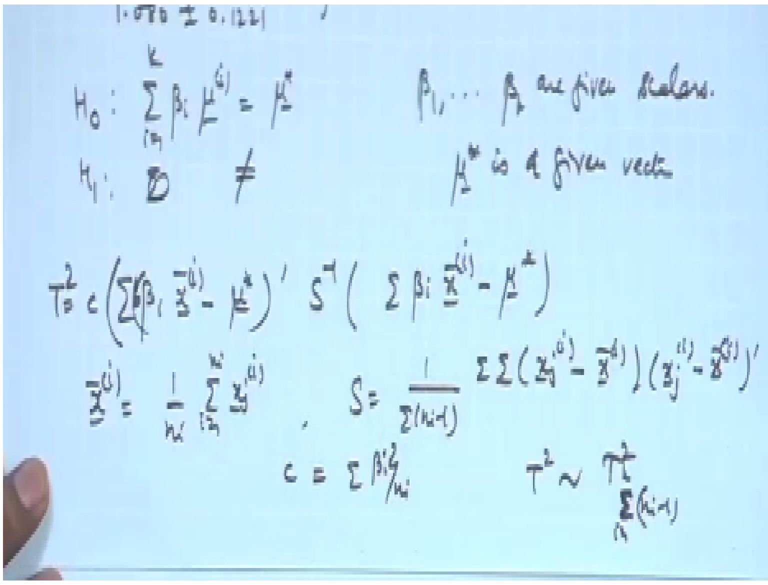
(Refer Slide Time: 53:44)



Here, the simultaneous confidence intervals have also been obtained; simultaneous confidence intervals for Mu i 1 – Mu i 2 for I is = 1 to 4, they are also obtained, so it is something like 0.930 +- 0.337, -0.658 +- 0.265, -2. 798 +- 0.270, 1.080 +- 0.1221, you can see that 0 does not belong to any interval in fact, this is quite different from 0, this is quite different from zero, this maybe is little bit closer to 0.

So, naturally you can say that the means of the 2 populations are quite different. As I mentioned that one may consider linear combinations also for example, I may consider testing H0 sigma beta i Mui is = Mu; I is = 1 to k against H1 sigma; so that is not equal, where beta1, beta2, beta k are given scalars and Mu; let me say Mu star, this is a given vector, then we can construct the statistic sigma beta i x bar i - mu star prime S inverse sigma beta i x bar i - Mu star, where x bar i is actually 1/ ni sigma xj i; i is = 1 to ni and S is 1/ sigma ni - 1 xj i - x bar i, xj i - x bar i transpose.

(Refer Slide Time: 56:38)



And c is sigma beta i square/ ni, then T Square will follow T square distribution, on sigma ni - 1 degrees of freedom that is Hotelling's T square here, so we can consider rejecting H0 when this value is > T square sigma ni – k here. In the next lecture, I will consider a problem which is based on symmetry; we will also consider the case when sigma 1 and sigma 2 are not assumed to be known.

Now, this case is again like in the case of univariate, we had only approximate procedures, in the multivariate case however, exact procedures can be constructed but then there is a compromise like we may have to ignore some of the observations, so I will be discussing in detail this problem in the following lecture here.