Applied Multivariate Analysis

Prof. Amit Mitra
Prof. Sharmishtha Mitra
Department of Mathematics and Statistics

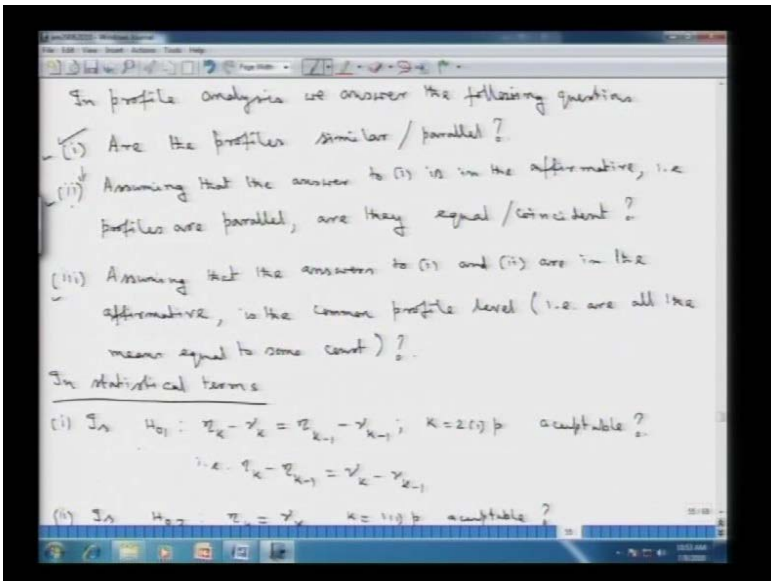
Indian Institute of Technology, Kanpur

Lecture No. # 17

Profile Analysis – II

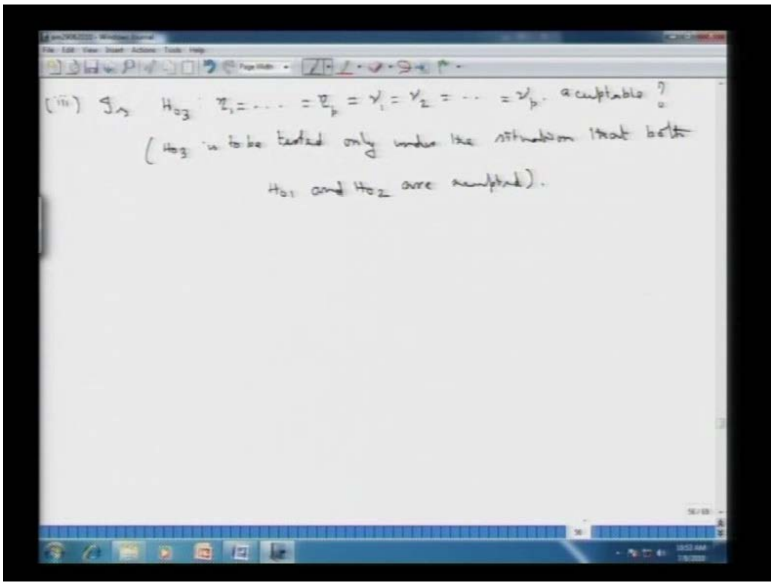
In the last lecture, we had introduced the profile analysis for this multivariate data.

(Refer Slide Time: 00:25)



And we had posed the following questions, one, two and three, and we had discussed in detail, how we actually go through this these answering these questions, sequentially. And we had also put these questions that are of interest in profile analysis, in terms of statistical hypothesis testing.

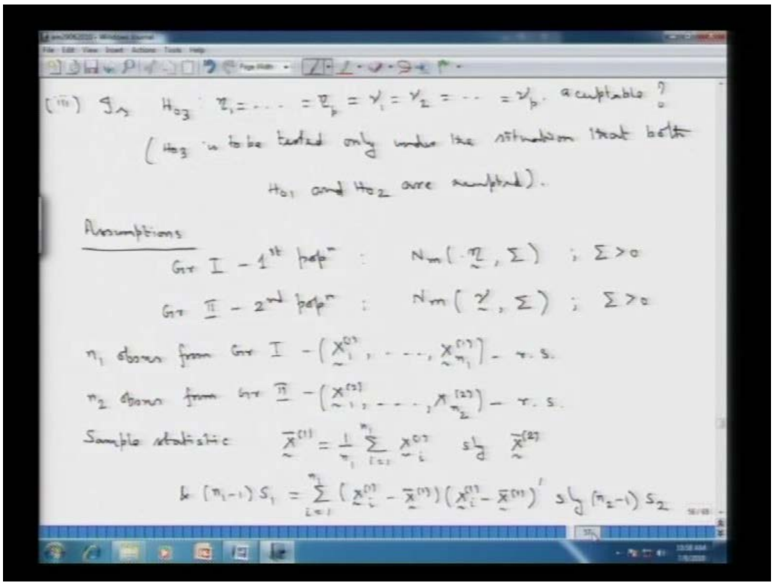
(Refer Slide Time: 00:50)



So, these questions were translated in statistical terms, and we had this H naught 1, the first hypothesis, H naught 2 the second hypothesis, H naught 3 the third hypothesis; that is to be tested, in order to answer the questions, associated with the profile analysis. And we had also said the what is the sequencing actually, when we are talking about this three sets of hypothesis, H naught 1, H naught 2, and H naught 3.

H naught 2 is to be tested only if we have H naught 1 accepted, H naught 3 is to be tested only, if a both H naught 1 and H naught 2 in that order are accepted, and hence we go to the third hypothesis testing. Now, what we will see today is we look at how to perform these tests H naught 1, H naught 2 and H naught 3 using (( )) T square statistic, we will also see some numerical examples of some actual profiles; and then perform the profile analysis questions that we have in mind. In order to see, how these questions are answered for some practical data sets.

(Refer Slide Time: 01:54)



So, let us move forward to looking at testing of H naught 1 first, but before that we will have to have the following assumptions, so we make the following assumptions for testing, the underlying assumptions is that this group I, which is the first population, so this is the first population. Let us have that been characterized by a multivariate normal, m dimensional, say I will write that as normal multivariate normal m, we had denoted that by perhaps eta yesterday.

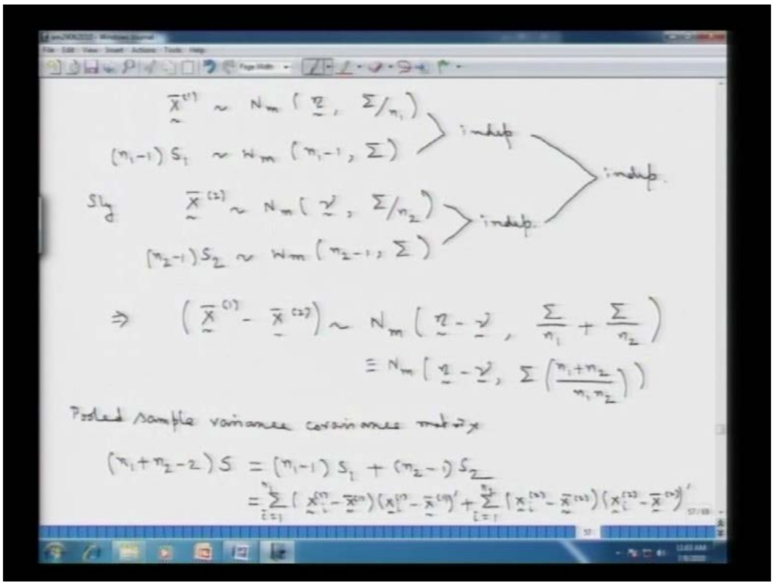
So, this eta vector and a covariance matrix sigma, where the sigma is positive definite and the group II, the second population is also a multivariate normal distribution with mean vector, as nu vector and sigma as its variance covariance matrix. So, we keep this sigma in both the populations to be same, so we have two populations group I, group II, which differ in their mean vector component, the variance covariance component remaining the same.

So, we have n 1 observations n 1 observations from group I, the random variables let then be denoted by x 1 1, x n 1 1 and n 2 observations are taken from group II. So, we will have this random sample coming from the second group here, so these are denoted by x 1 2, x n 2 2. So, these two are the two random samples, so this is the random sample set and this is also the random sample set from the two respective populations.

Now, these are forming an independent set of random vectors, and so will be these forming set of independent random vectors, and these two also are the two random samples, from two groups and hence they are also going to be independent. Now, the sample statistic that, what we have derived from these two is x 1 bar let me denote that by x 1 bar, in order to signify that it is basically coming from the first population. So, this is on upon n 1 summation i equal to 1 to up to n 1 and then we will have this x i 1.

Similarly, what we will be having is x 2 bar vector, which is the sample mean vector coming from the second population, and we will have n 1 minus 1 S 1, this is S 1 is the sample variance covariance matrix, when it is based on the random sample coming from the first population. So, this is equal to summation i equal to 1 to up to n 1 x i 1 minus x bar 1 vector this multiplied by x i 1 vector this minus x bar 1 vector its transpose. Similarly, we will have n 2 minus 1 times S 2, which is the sum of squares and cross product matrix similarly, from the second population.

(Refer Slide Time: 05:34)



Now, from the distribution theory, what we know are the following quantities, that this x 1 bar this follows a multivariate normal m dimension with a mean vector as eta vector, the underlined mean vector of that population. And this sigma divided by n 1 from the first sample itself, we will be also be having n 1 minus 1 S 1, this to follow a wishart m n 1 minus 1 sigma, the two are independent because, they are based on that random sample from the first population.

Similarly, from the second population, we will have these characteristics the statistic x 2 bar this will follow a multivariate normal with new vector, as its mean vector and the sigma from n 2, as its variance covariance matrix. And we will also be having n 2 minus 1 S 2 this would follow a wishart distribution m n 2 minus 1, and the variance covariance matrix same as the previous wishart matrix, so these two also are going to be independent. And further more, because we have the first set of statistic, coming from the first based on the first set of random samples, from the first population and the second set here, coming from the second population, we will have further independence of these statistic as well right.

Now, once we have this, we can say that this would imply further that x 1 bar minus x 2 bar, this would follow this is a random vector of the same dimension as that of the underline multivariate normal distribution this that is n. So, this will have the multivariate normal distribution m dimension, on mean vector as eta minus nu vector, and the covariance matrix sigma by n 1 this plus sigma by n 2.

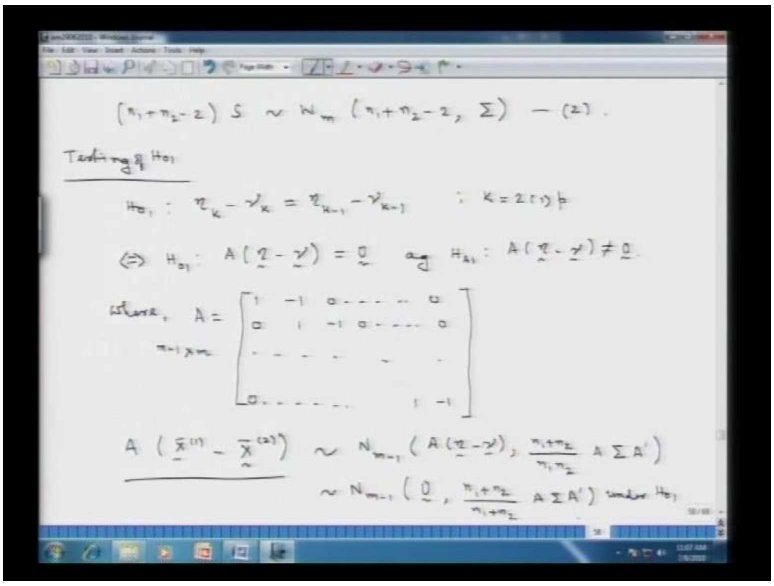
So, this reminds us of the two sample normal problem basically, so that this is multivariate normal eta minus mu vector and this is sigma is common, so we will have this as n 1 plus n 2 this divided by n 1 times n 2 (Refer Slide Time:07:52). Let me keep it up to this particular point, we will require that further also. Now, we will look at the pooled sample variance covariance matrix, why do we look at the pooled sample covariance matrix, because we have sigma the underline variance covariance matrix of the two populations to be exactly the same.

And hence we can look at the pooling of the two data, in order to obtain an estimate of the sample variance covariance matrix. So, pooled sample variance covariance matrix variance covariance matrix is what we have, that is n 1 plus n 2 minus 2 times S, say S is pooled sample variance covariance matrix, and that would be given by n 1 minus 1 S 1 this plus n 2 minus 1 S 2. So, that would be given by the two separate sum of squares and cross product matrices, so the first one would be this and the second one similarly, would follow.

So, that this is equal to let me make it complete, i equal to 1 to up to n 1 x i 1 this minus x bar 1 this multiplied by x i 1 this minus x bar 1 transpose this plus, the sum of squares and cross product matrix coming from the second population, this is x i 2 minus x bar 2 this multiplied by this x i 2 vectors minus x bar 2 vector its transpose (Refer Slide Time: 09:10). So, this is what we have as a pooled sample variance, covariance matrix.

It is easy to see that, since this has got a wishart distribution, wishart m n 1 minus 1 sigma, and this path has got a wishart distribution (( )) n n 2 minus 1 sigma, so we will have at the two wishart distributions are independent. So, we will have the distribution of this sum of two independent wishart distributions, with the same variance covariance matrix to be given by also a wishart distribution.

(Refer Slide Time: 10:17)



So, from the properties of wishart distribution, we will have the following that n 1 plus n 2 minus 2 times S, this would follow by the additive property of the wishart distribution, that it is wishart m n 1 plus n 2 minus 2 and with the same variance covariance matrix sigma. So, the two things that we are going to use in all the testing procedure is, this is one here, that we will have x 1 bar minus x 2 bar to have this multivariate normal distribution, and we will have this to follow a wishart distribution right.

Now, in the lines of this discussion, let us now move forward to testing of H naught 1 first, because that is the first hypothesis that one needs to test testing of this H naught 1 hypothesis. Now, what is H naught 1 hypothesis, H naught 1 hypothesis, if we remember its eta k minus mu k that is equal to eta k minus 1 minus mu k minus 1. So, this is going to test the parallelity of the two profiles in the population, so this is k is from 2 1 to up to p right.

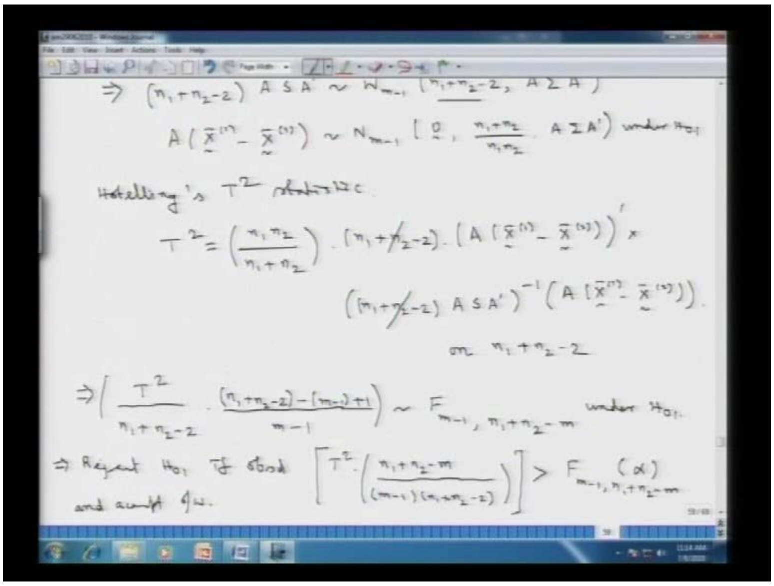
Now, this is equivalently written as through a matrix H naught H naught 1, which is a times eta vector minus this mu vector that equal to a null vector. Now, what is this A where, this A matrix is of the following structure, that it will have 1 minus 1 0, then the second row is 1 minus 1 0 and finally, the last row is what is going to be given by this minus 1 1. So, if we have this order as m by 1 this is going to be an m minus 1 cross m matrix, this is a matrix of constants. So, if we have A defined as 1 minus 1 1 minus 1 like this, along this block here, so what we are going to get here is that, from this first element remember that this eta and nu are the corresponding mean vectors of the multivariate normal distribution.

And hence, the first element would be eta 1 minus nu 1, and the second element would be eta 2 minus nu 2, so the first row here, when multiplied with that vector would lead us to having the first element that is eta 2 minus nu 2 is equal to eta 1 minus mu 1. So, we will have all these to be equal and thus, this basically is giving us the hypothesis of interest to be tested. This is to be tested against the alternate hypothesis H A 1, say this is to be tested against H A 1, which is A eta minus nu this is not equal to 0, so this is hypothesis to be tested.

Now, from the distribution theory, we had that this distribution that x 1 bar minus x 2 bar has this distribution, we will make this transformation that A times x 1 bar minus x 2 bar, what is the distribution of this, now this a matrix is m minus 1 cross m, so this is a matrix of constants. So, we will have this to have a multivariate normal on m minus 1 dimensions, and the mean vector would just be A times eta minus mu and what happens to the variance covariance matrix, we had the previous variance covariance matrix as this particular term (Refer Slide Time:14:03).

So, we will have A sigma A prime and hence this would be n 1 plus n 2 this divided by n 1 n 2 this is A sigma A prime where the matrix A is known to us. Now, it is important to note that, this would have a multivariate normal m minus 1 with a mean vector equal to a null vector and a covariance matrix as n 1 plus n 2 by n 1 plus n 2 A sigma A prime this is the distribution of this quantity on the left hand side under the null hypothesis H naught 1. Why, because this mean vector this mean vector of this random vector here is what is specified as a null vector under the null hypothesis, so we will have that, that is all right.

(Refer Slide Time: 15:03)



Then we will also look at similar transformation to this particular wishart distribution, this will also imply the following, that we will have n 1 plus n 2 minus 2 A S A prime, because S at n minus n 1 plus n 2 minus 2 S had a wishart distribution. So, if we pre and post multiply, pre multiply by A the matrix of constants post multiplied by A transpose, the transpose of that same matrix of constants, what we are going to have is also a wishart distribution, with dimension as n minus 1 degrees of freedom remaining the same n 1 plus n 2 minus 2. And the associated variance covariance matrix, now becomes a sigma a prime right.

So, we will use the two this distribution, under the null hypothesis, and this distribution under the alternate under I am sorry, not under the alternate hypothesis, this is corresponding to the wishart, so let me just copy it once again here, so that we have all the things before us. So, this x 1 bar minus x 2 bar this to follow a multivariate normal n minus 1 with a null vector, as its mean vector and the variance covariance matrix as n 1 plus n 2 by n 1 n 2 that into A sigma A prime as its variance covariance matrix; this under the null hypothesis H naught, and this is always a distribution of this.

Now, we can frame the Hotelling's T square remember, we had defined Hotelling's T square with the two sets of random variables, one which was having a multivariate normal distribution, and the other which was having a wishart distribution. So, we have exactly the same setup out here, so we will have Hotelling's T square statistic to be given by this T square, which is going to be C. Now, C C inverse is this particular term, so we will have n 1 n 2 by n 1 plus n 2 this serving the purpose of the C which we had defined for the Hotelling's T square, this into n, n was the degrees of freedom. So, the degree of freedom here is n 1 plus n 2 minus 2. And then we have the transpose of this, so the transpose of this is A x 1 bar minus x 2 bar transpose of that, now that multiplied by the inverse of the wishart here.

So, that this is, that n 1 plus n 2 minus 2 this entire term, now is having that wishart distribution, so we will have inverse of that, and that multiplied by this A x 1 bar minus x 2 bar this term. So, this is the Hotelling's T square on what degrees of freedom, the degrees of freedom are the associated degrees of freedom of the wishart distribution. So, this is the Hotelling's T square on n 1 plus n 2 minus 2 degrees of freedom, it is easy to see that this term cancels out with this one.

And what remains is what we have as a wishart distribution note that, we cannot do anything further on the inverse of this A S A A transpose matrix, because A is a rectangular matrix. So, we will have this on n 1 plus n 2 minus n 1 plus n 2 minus 2 degrees of freedom, and then what we will be doing further is that, this would imply that this T square divided by T square by n.

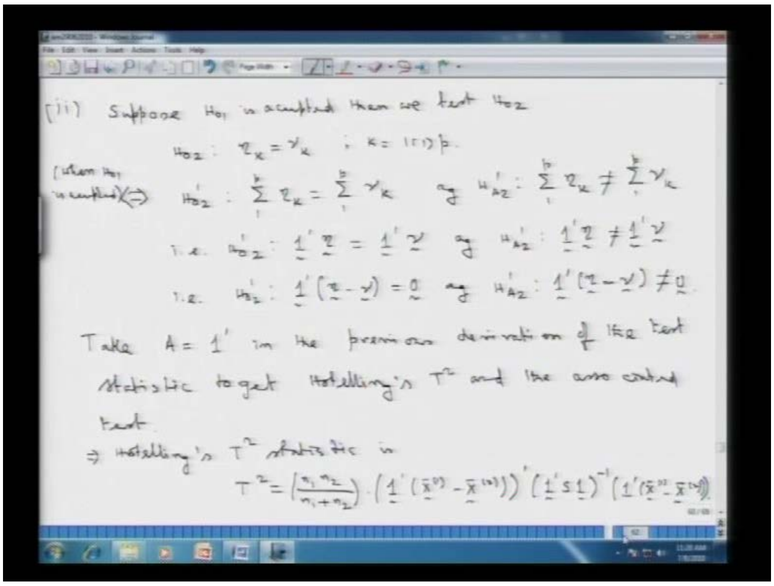
Now, n is the degrees of freedom that is n 1 plus n 2 minus 2 that multiplied by N minus m plus 1. So, n is the degrees of freedom that is n 1 plus n 2 minus 2 this minus m, now is what, m was the dimension the dimension here is m minus 1, this plus 1 that divided by the dimension which is m minus 1. So, this is T square by n m minus m plus 1 by m, now this will follow a central F distribution under the null hypothesis with the degrees of freedom as m minus 1, and whatever this comes, so this is going to be n 1 plus n 2.

So, this two cancels out with this plus 1 here, and the another plus 1 out here, so we will have n 1 plus n 2 minus m this under the null hypothesis H naught right. So, since we have the null distribution of T square multiplied by this particular constant to follow a central F distribution on m minus 1 n 1 plus n 2 minus m degrees of freedom, the testing of under H naught 1 actually, here also H naught 1, the hypothesis to be tested here is H naught 1.

And hence, we will reject the null hypothesis this implies, we will reject H naught 1, if observed value of this T square that multiplied by this term is n 1 plus n 2 minus m that divided by m minus 1 into m 1 plus m 2 m 1 plus m 2 minus 2, if the observed value of this quantity is greater than F m minus 1 n 1 plus n 2 minus m alpha, where this particular point here is a given point, which is the upper alpha percent cut off point of a central F distribution on m minus 1 n 1 plus n 2 minus m degrees of freedom. So, that is how we are going to test this H naught 1, we will reject H naught at of course, level alpha and with the associated level at alpha.

If the observed value of this is greater than this and except otherwise, so we have been able to obtain the testing for the H naught 1 hypothesis, the hypothesis which tests similarity or parallelity of the profiles from this particular Hotelling's T square statistic. So, that is it for the first set of testing problem.

(Refer Slide Time: 21:44)



Now, the second hypothesis of interest was to test H naught 2, now remember that we will only test H naught 2, if the first hypothesis is accepted. Suppose, H naught 1 is accepted, then we test H naught 2, if H naught 1 is rejected, then we do not test H naught 2 we stop at that particular point.

Now, this H naught 2 as it was given earlier was, equality of the profiles and that is what we have out here, that this is the hypothesis of interest for k equal to 1 to up to p. Now, this is equivalent to once we have H naught 1 to be accepted, when H naught 1 is accepted. Then this H naught 2 is equivalent to the hypothesis H naught prime which is the summation eta k equal to summation of this nu k values k equal to 1 to up to p k equal to 1 to up to p, now this is to be tested against the alternate hypothesis H A 2 prime which is summation eta k is not equal to this summation nu k under the condition that H naught 1 already is accepted. So, this thus is going to test whether the two profiles of the group, groups are actually equal given that they are similar or they are parallel right.

Now, this hypothesis is H naught 2 prime this is 1 prime eta vector equal to 1 prime nu vector this is to be tested against H A 2 prime, which is 1 prime eta vector not equal to 1 prime nu vector this or in other words, what you can write is that this is 1 prime eta minus nu vector, that is equal to a null vector. So, this is H naught 2 prime which is this is to be tested against H A 2 prime which is 1 prime eta minus nu vector this is not equal to 0.

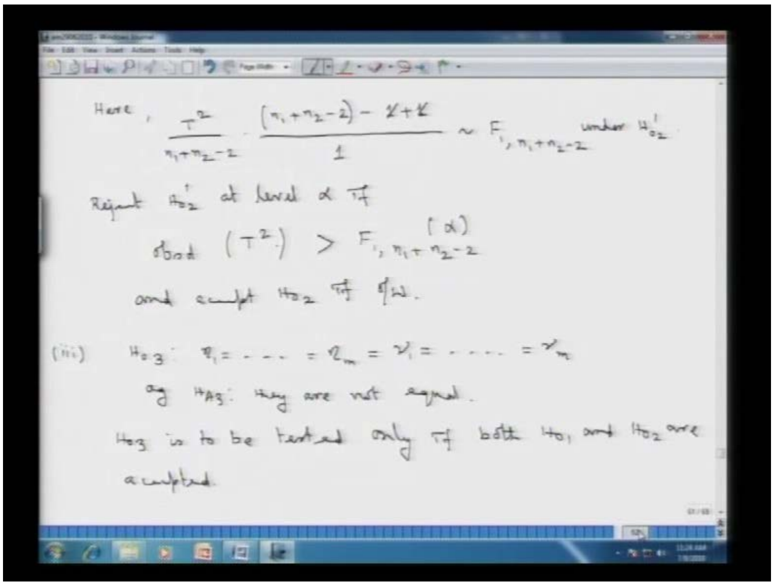
So, the testing for this second hypothesis H naught 2, that is what is testing the equality of the profiles given that the two profiles are similar or parallel, reduces to this particular problem. Now, the point in reducing this particular H naught 2 in this form is totally this particular problem of testing, this hypothesis with the previous problem, that we had tackled and discussed in detail, that we had the H naught 1 hypothesis given in terms of this sub matrix A of constants times eta minus nu, that to be equal to a null vector to be tested against that, this is not equal to a null vector (Refer Slide Time: 24:38).

Now, as we see the H naught 2 hypothesis, we have clearly reduced it in to a form which is exactly same form as that of the previous one, with the previous a being replaced now by 1 prime. So, that testing follows exactly in the same way, so we take A equal to 1 transpose in the previous formulation, in the previous derivation of the test statistic statistic to get Hotelling's T square, and the associated test right.

And hence, we can straight straight away use this particular test statistic, what we had got there with just A being replaced by 1 transpose. So, the Hotelling's T square statistic would exactly looks like the same with A being replaced by 1 transpose. So, this would imply that first up the Hotelling's T square statistic is now, this T square which is now going to be given by n 1 n 2 that divided by n 1 plus n 2, which was previously there and what we will be having here is 1 transpose x 1 bar minus x 2 bar whole transpose and then we have that A S A transpose.

So, we will have 1 transpose S 1 this is that vector out there, so you can, if you want you can just put this vector sign here. So, we will have this and then inverse of that that multiplied by the transpose of this vector, and this actually is a scalar quantity 1 transpose x bar 1 minus x 2 bar right.

(Refer Slide Time: 27:19)



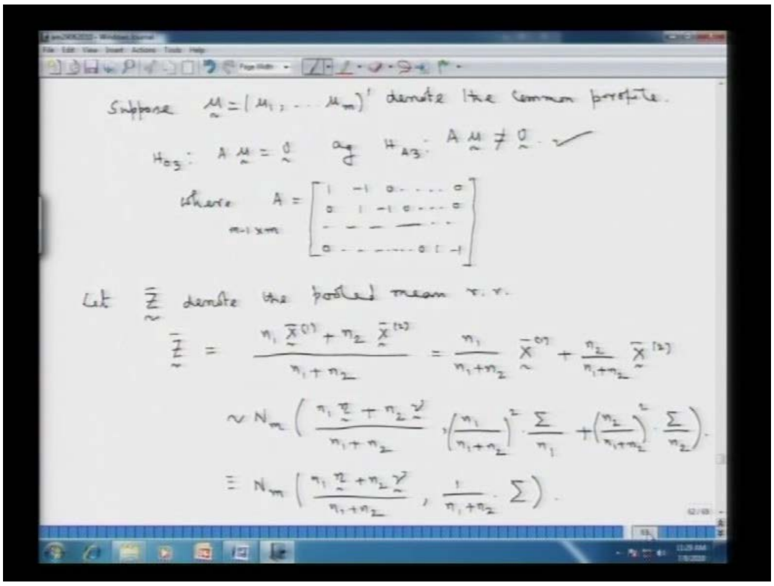
So, that is the Hotelling's T square statistic in this testing problem, H naught 2 prime and then, we can say that what is the test statistic; now here what will be having is T square divided by n 1 plus n 2 minus 2, that what we have was n 1 plus n 2 minus this term, if we look back here this we have as m 1 plus m 2 minus m (Refer Slide Time: 27:34). So, we will have the same term here, now what is m here, m is going to be equal to that term there.

So, that this is n 1 plus n 2 minus 2 this minus m minus 1, m minus 1 is nothing but, 1 here this plus 1 this divided by 1, because that is n minus 1 is equal to 1 here, this would follow a central F distribution on 1, n 1 plus n 2 minus 2 degrees of freedom under the null hypothesis H naught 2 prime (Refer Slide Time: 28:03). And we will reject H naught 2 prime at level alpha, if observed value of this T square, so this cancels out. So, what will be having is just this n 1 plus n 2 minus 2 also will cancel out.

So, if the observed value of T square only is greater than F 1 n 1 plus n 2 minus 2 alpha, where this point is the upper alpha percent cut off point of a central F distribution on n 1 plus n 2 minus 2 degrees of freedom. And accepted if otherwise, accept H naught 2, if otherwise, so using the similar type of derivation as to what we are used for the first H naught 1 hypothesis, we have obtained the test statistic and the testing procedure for testing, the second set of hypothesis testing for the equality of the two profile. Then we move on to the third set of hypothesis of interest, what we had there was testing for H naught 3; now testing for H naught 3 was eta 1 equal to eta 2 equal to eta, what was the dimension m, that equal to nu 1 equal to nu 2 equal to nu m against, that they are not equal. So, this is to be tested against H A 3, that they are not equal, all of them are not equal right, now remember that H naught 3 is to be tested only if both H naught 1 and H naught 2, in that order H naught 1 and H naught 2 are accepted.

So, if after H naught 1 being accepted, we have moved on to H naught 2, and then we have observed that H naught 2 is rejected, then we will not proceed for testing of H naught 3, and we will stop at that particular point, if H naught 2 also is accepted. Then we will move on to testing this H naught 3 hypothesis which is going to test, whether the common profile that is what we have is a level profile, that is all of the components are same. Now, let us see how we can we are going to test this particular hypothesis, now this H naught 3, we can write it as following.

(Refer Slide Time: 31:07)



Now, let me first denote give this notation suppose, now since we are only testing H naught 3, when H naught 1 and H naught 2 are accepted equality of the two profiles would imply, that there is a common profile. Suppose, this mu vector which is denoted by mu 1, mu 2, mu m denote the common profile of the two groups, so if we have this to denote the common profile then H naught 3 is what we are trying to test as A mu equal to A null vector, this is to be tested against H A 3 that A mu is not equal to a null vector.

So, we are testing this at A mu is not equal to a null vector, where A is as what we have used, if previously, so this can be now, remember that there is no uniqueness in defining this particular matrix of constants A, A can alternatively be defined in order to translate, the hypothesis that all the components are equal also. So, that there is no uniqueness in representing this particular A matrix, this is a minus 1 1 previously and (( )) all zeroes before that, so this is an m minus 1 cross m matrix of constants.

So, this is now, the hypothesis of interest, what we are going to test is that the common profile, what we have which is being denoted common profile between the two groups. Now, that is we are going to test that all the components are equal, and through this matrix a we are having mu 1 equal to mu 2, mu 2 equal to mu 3, mu 3 equal to mu 4, m minus 1 equal to mu m and the thus we have all the components to be equal.

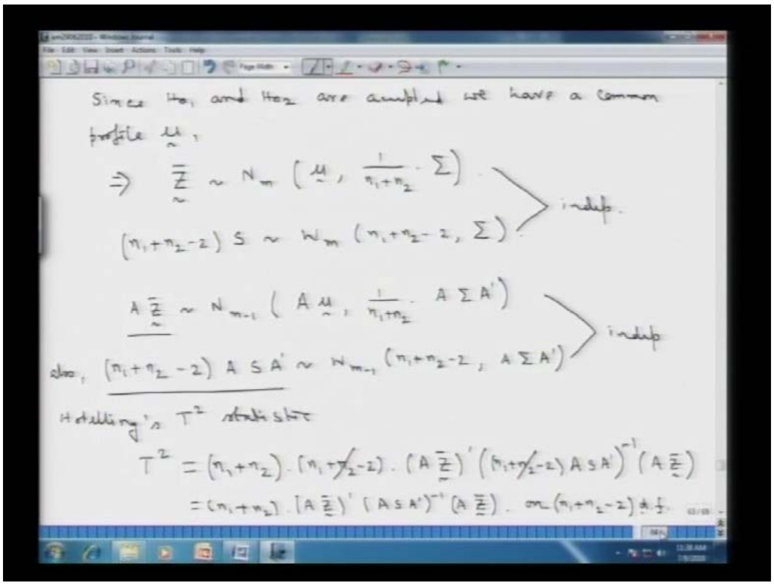
Now, let z bar denote the pooled mean denote the pooled mean random vector, where this z bar is given by, so since it is a pooled mean, we will have that to be equal to n 1 times x 1 bar plus n 2 times x 2 bar. So, this is going to give us the sum of all the observations, that divided by n 1 plus n 2 right. Now, when we have this being defined then note that, this is just n 1 by n 1 plus n 2 times x 1 bar which is having a multivariate normal distribution, that plus n 2 by n 1 plus n 2 times x 2 bar, which is also having a multivariate normal distribution the two are independent.

So, this would imply, that this z bar the pooled sample mean random vector will have a multivariate normal distribution, with mean vector as n 1 times, the mean vector of this which is eta this plus n 2 times this nu vector divided by n 1 plus n 2, so that is what is the mean vector corresponding to this (Refer Slide Time: 34:11). And then the covariance matrix of this element would be n 1 by n 1 plus n 2 whole square this is scalar constant, that multiplied by the variance covariance matrix of x 1 bar, which is sigma divided by n 1, this plus n 2 by n 1 plus n 2 whole square that multiplied by the variance covariance matrix of x 2 bar, which is sigma by n 2 right.

Now, we have this particular term some simplification can be done, that is this is a multivariate normal distribution n 1 eta this n 2 nu vector this divided by n 1 plus n 2. So, what we see here is that 1 n 1 gets cancelled out with this one, and 1 n 2 gets cancelled out with this one. So, if we take sigma outside, we will have an n 1 plus n 2 in the numerator and n 1 plus n 2 whole squares in the denominator. So, what this will lead us to is that this is one upon n 1 plus n 2 one of them cancels out, that multiplied by sigma.

Now, since we have the two hypothesis sequentially being accepted, H naught 1 and H naught 2 being accepted, we have these 2 eta vector which is leading us to the profile or the first group, and this eta eta and nu both of them are accepted to be having a common profile; and hence we can replace them by mu, and hence what we will be having is the following.

(Refer Slide Time: 35:58)



Since, H naught 1 and H naught 2 are accepted are accepted we have a common profile we have a common profile mu which would imply that our z vector, the pooled mean vector has got a multivariate normal distribution with a mean vector what, if we look back here, if both of them are replaced by mu, then the mean vector of this pooled sample mean would just be equal to mu. And the variance covariance matrix to be equal to 1 upon n 1 plus n 2 times sigma.

And what we also have is this quantity which is we have already seen it, that n 1 plus n 2 minus 2 the pooled sample variance covariance matrix, this has a wishart distribution, wishart m n minus I am sorry, it is n 1 plus n 2 minus 2 as its degrees of freedom and sigma as its associated variance covariance matrix.

So, we have the two distributions z bar being a multivariate normal and n 1 plus n 2 minus 2 times S, where S is the pooled variance covariance matrix to have a wishart distribution this. Now, since we have this S to be independent of x 1 bar and x 2 bar this z bar which is the derived random vector from x 1 bar and x 2 bar that is going to be independent of the pooled sample variance covariance matrix; and hence, these two distributions, we can say that this these two are independently distributed.

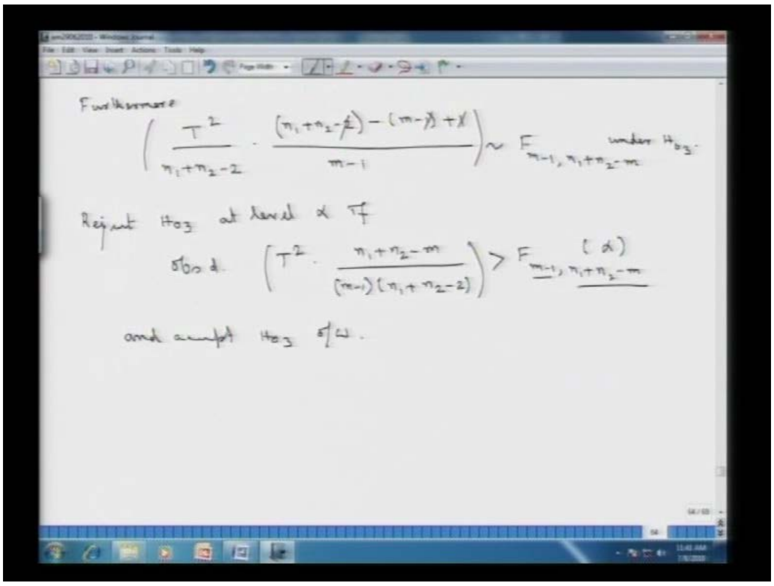
Now, let us go back to what we have to test, the null hypothesis we have translated as A mu equal to a null vector, against A mu is not equal to a null vector, so we need to introduce this A somewhere here. So, from these two what will be having is that A times z bar vector, now A is a matrix of constants which is m minus 1 cross m dimension. So, this random vector is going to be, a random vector of dimension n minus 1 it is going to have a multivariate normal distribution with a mean vector as A mu and a covariance matrix as 1 upon n 1 plus n 2 times A sigma A prime.

And we have a similar, we make a similar transformation to this wishart matrix here, so we will have also n 1 plus n 2, we will also have n 1 plus n 2 minus 2 that into A S A transpose, this will have a wishart distribution on the dimension m minus 1, with the degrees of freedom as the previous degrees of freedom of that n 1 plus n 2 minus S wishart matrix. And with the variance covariance matrix now, being given by A sigma A prime and since, the previous two distributions where mutually independent, we will also be having the distribution of A z bar and this quantity here to be independent.

So, we have once again the constituents constituent parts actually to frame the Hotelling's T square statistic, because in the Hotelling's T square statistic we have 1 A wishart distribution and the other having a multivariate normal distribution, the independence of the two will actually lead us to forming the Hotelling's T square statistic. So, using these two, the Hotelling's T square statistic would turn out to be the following, Hotelling's T square statistic is going to be given by T square which is going to be equal to C times N, now C is what, C is the inverse of this particular quantity (Refer Slide Time: 39:51).

So, we will have C as n 1 plus n 2, then comes the degrees of freedom, degrees of freedom is associated with the wishart, so that this is n 1 plus n 2 minus 2 and then we will have the transpose of this. So, we have A z bar transpose, and then the inverse of the wishart matrix which is for the given case n 1 plus n 2 minus 2 times A S A transpose and whole inverse of that that multiplied by this vector which has got the multivariate normal distribution, this is A z bar. So, what is that we get, we can see that this cancels out and this is just equal to n 1 plus n 2 times A z bar transpose A S A transpose whole inverse A is that rectangular matrix, which we have in H naught 3, so this is multiplied by A z bar. So, this is the Hotelling's T square on what is degrees of freedom, degrees of freedom is the same as that associated with the associated wishart distribution, on this degrees of freedom.

(Refer Slide Time: 41:11)

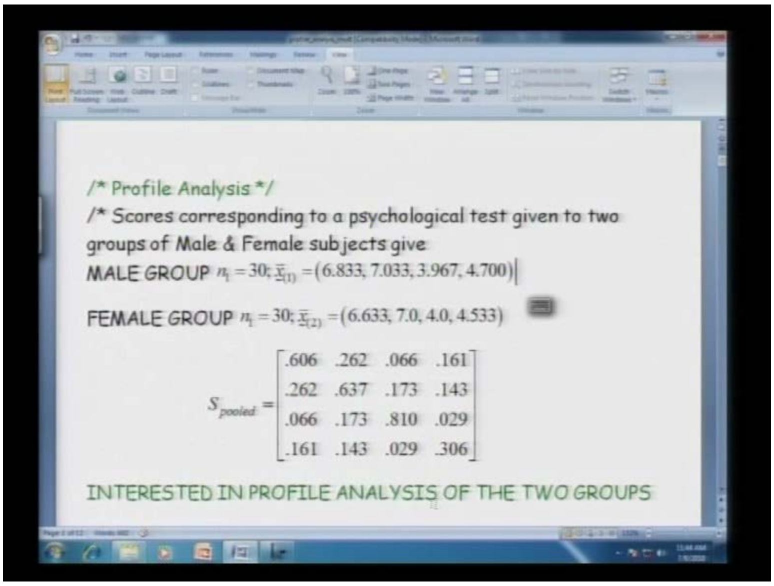


So, if we have that, we can also say that furthermore this T square divided by its degrees of freedom n 1 plus n 2 minus 2 multiplied by degrees of freedom n 1 plus n 2 minus 2 then minus m m here is m minus 1, so that is m minus 1 plus 1 this divided by m minus 1. So, this is going to have, this is going to follow an F distribution central under the null hypothesis H naught 3 only, so we will have that to have an F distribution m minus 1 as the first degrees of freedom.

Let us see, what is the second degrees of freedom, this is plus 1 plus 1 minus 2, so that cancels out and what we will be having here is n 1 plus n 2 minus m under the null hypothesis H naught 3. Now, under the alternate hypothesis actually, this statistic will still have an F distribution what, but it will be a non central F distribution. So, since we have the distribution of this is statistic, the null distribution do have an F distribution, we have the testing procedure that, we will reject H naught, so this term here 1 minus 1 and minus 2, these terms cancel out. Reject H naught 3 at level alpha, if observed value of T square this multiplied by n 1 plus n 2 minus m that divided by m minus 1 into n 1 pus n 2 minus 2, this is if the observed value of this quantity for the given sample exceeds the upper alpha percent cut off point of a central F distribution, on n m minus 1 and n 1 plus n 2 minus m degrees of freedom. So, this denoting the upper alpha percent point of a central F distribution of m minus 1 n 1 plus n 2 minus m degrees of freedom, and accept H naught 3, if it is otherwise right.

So, this is how the three testing procedures, H naught 1, H naught 2, H naught 3 are going to be tested for the profile analysis, questions the questions of interest to be tested, and they are to be tested in a sequential manner, first H naught 1 if accepted, H naught 2 if accepted, H naught 3 if at any point H naught 1 or H naught 2 are any one of them is rejected. Then we do not proceed for testing the next level of hypothesis, and if at the end of H naught 3, all the previous H naught 1, H naught 2, H naught 3 all are accepted, then we will say that we have the profile of the groups to be not only similar or parallel, not only equal they are also level profiles.

(Refer Slide Time: 44:14)



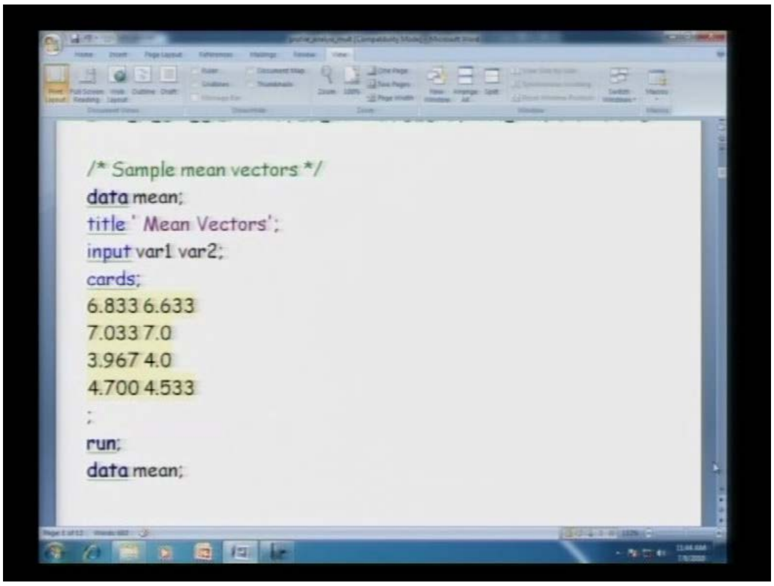
Now, what we are going to see next is some actual data analysis based on some real life data, where profile analysis is going to be carried out. So, this is what we have here, this is some given data, so this we have some practical data, where the scores corresponding to a psychological test given to two groups of male and female subjects. So, there we have two groups here, the first group is a male group and the second group is a female group, and then the same set of psychological tests, four tests actually are given to both the groups. And then we are interested in type of questions that we have just now answered theoretically, in profile analysis.

Now, for the male group, we have taken n 1 observation which we have n 1 equal to 30, so from the male group we have 30 observations, leading us to this sample mean vector which is a four-dimensional vector, which has these as the corresponding constituents. Similarly, for the female group we have got 30 observations with a mean vector computed from a two observations, this actually should be n 2 not n 1. So, this is from the second group we should write it as n 2 well they are same, so does not matter much, but technically one should write this as n 2.

This is the mean vector corresponding to the second group, and we in the profile analysis we have been assuming all along that the variance covariance matrix, associated with the different groups are same. And hence, if we do not have that, then the testing procedure for the Hotelling's T square actually gets spoiled, because at the point of looking at the common sigma, then pooling the two different wishart distributions will have problem there. And the additive property of the wishart cannot be used in such a situation.

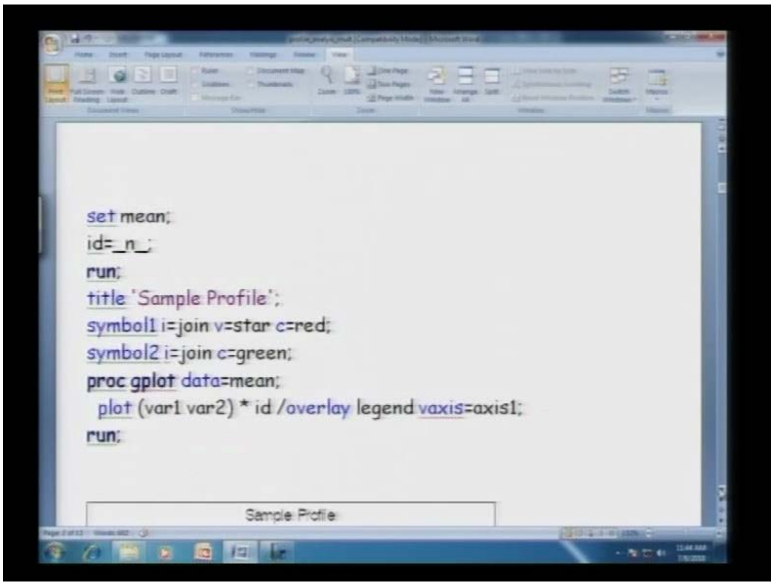
So, in order to actually have the testing, carried out through the Hotelling's T square we would require the common sigma, assumption to be inserted there, and for a given data here n 1 n 2 as 30 30. So, from the 60 observations, we have obtained this pooled sample variance covariance matrix from the data; now we are interested in profile analysis of the two groups. How do we proceed, now first we would like to see, how the profiles of the groups actually are behaving so we would require the sample mean vectors in order to lean us to those profiles.

(Refer Slide Time: 46:28)



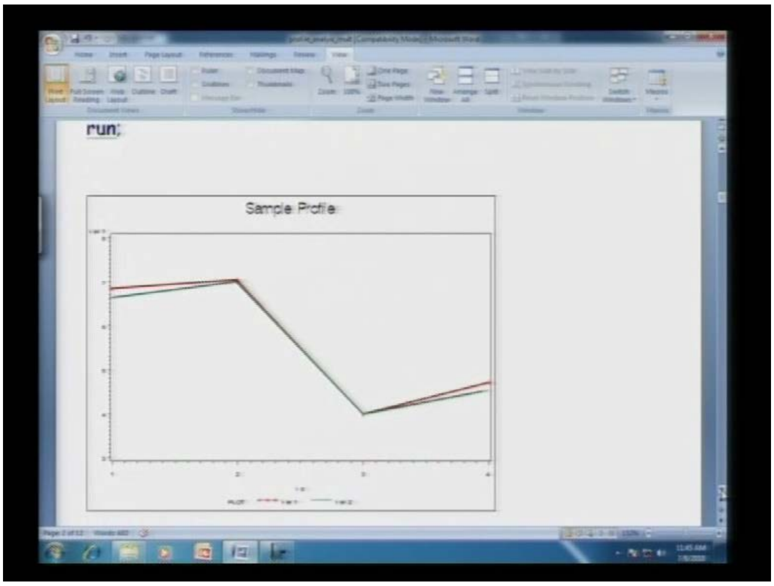
So, this is implemented using a SAS routine, so we are looking at this basically are the mean vectors in inserted in SAS, and then what we have is the following.

(Refer Slide Time: 46:52)



So, we are looking at construction of the profile, these are the SAS statements in order to get to that profile of the two groups.

(Refer Slide Time: 47:02)

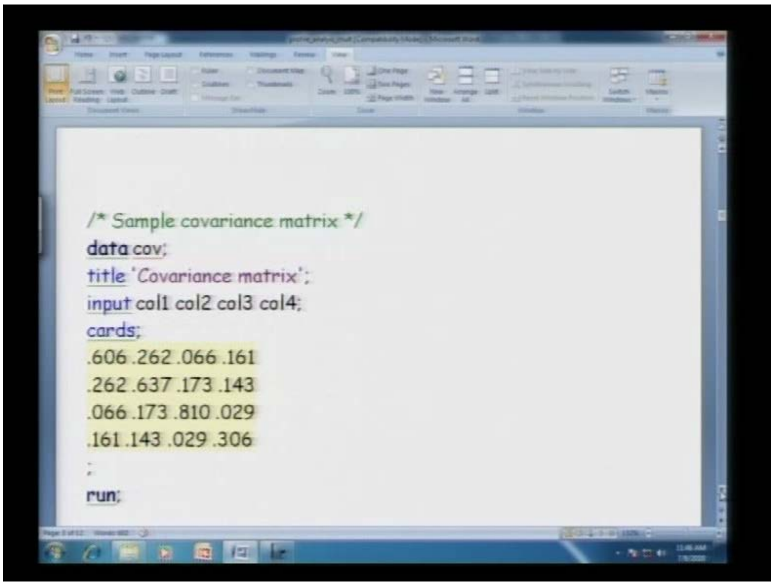


So, the profiles this is going to be the sample profile, sample profile is what is obtained, when we join the points which are associated with the sample mean vectors of the two groups. So, this is the profile for the first group in red colour here, and the green colour denoting the profile of the second group. So, this point is basically corresponding to the first group, we will have that to be that the mean of the first component, in the in the second group there. So, if we look back at the data it would be clear what these points actually are.

So, we have for the male group the first mean to be 6.8, so in the profile, sample profile of the first group, we will join this point with the next point here, to get the profile of the first group, first direction. So, this all this points are joints joint consecutively to get to the first sample profile, and similarly this is going to be, so we see that this is higher than this particular term here, these two are almost the same, these two are almost the same there is a minor difference between these two (Refer Slide Time: 48:01).

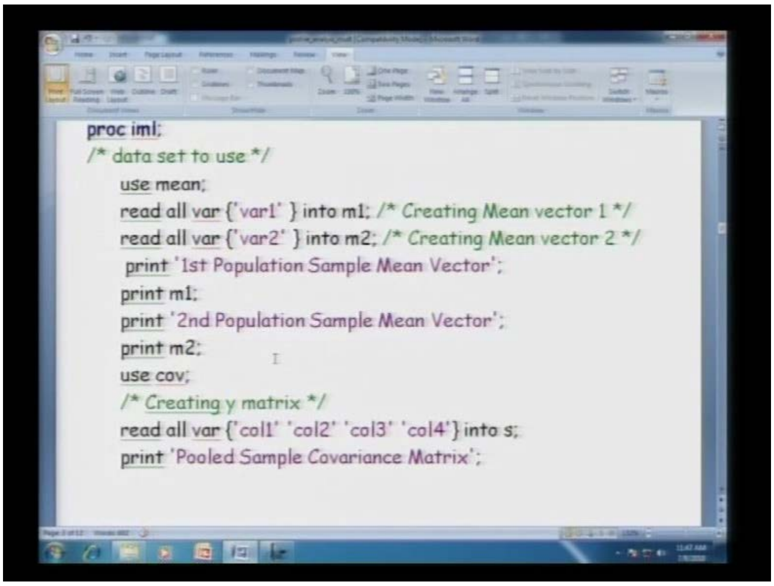
So, that we have got this as the sample profile, so we have this a bit higher than this, the second one almost the same, the third one almost the same there is a minor deviation between the last point of the two groups right. So, this is the male profile, this is the female profile; now we are going to test the three types of hypothesis, that we say we are interested in.

(Refer Slide Time: 48:28)



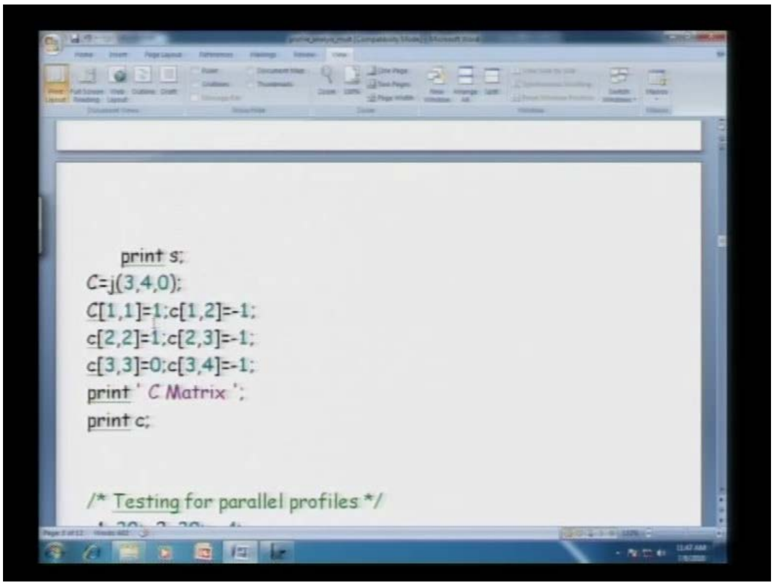
So, we have to test the profile analysis hypothesis one after the other, so that we have this as the sample variance covariance matrix now; this is we performed this profile analysis using iml procedure of the SAS, and then we will be implementing what theory we have learnt.

(Refer Slide Time: 48:41)



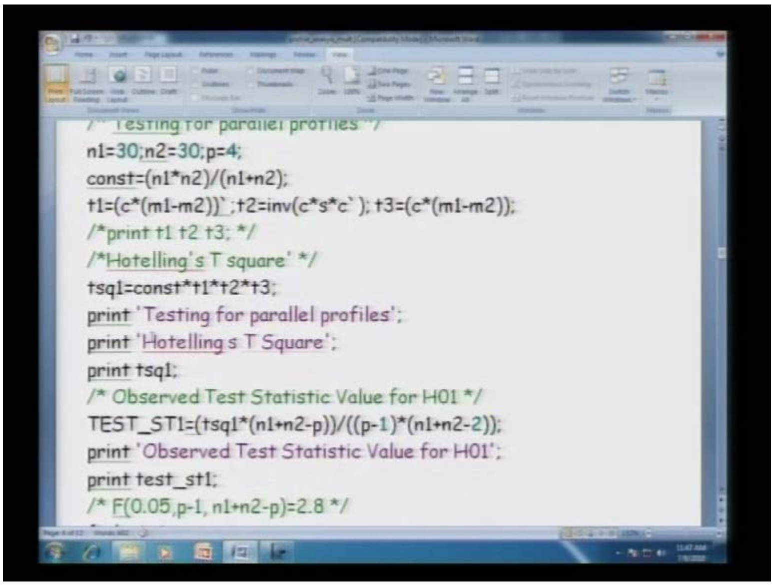
So, this is what is the procedure iml doing, so we are reading all the variable ones and two this is creating the mean vector for the first group, creating the mean vector for the second group, so first population mean second population mean etcetera.

(Refer Slide Time: 49:11)



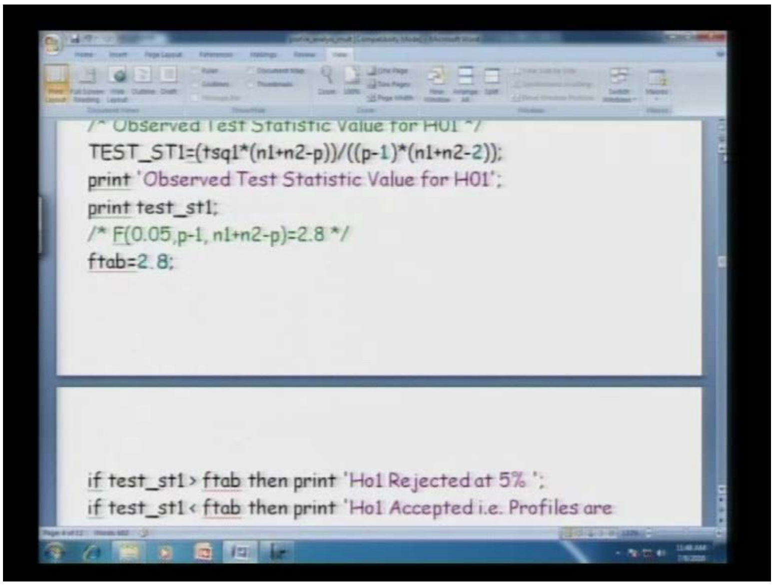
So, we will also be requiring to we will also require to have the pooled sample variance covariance matrix, because that is what is going to be used, then we need to construct that A matrix what we had set, when we had said that we are going to test this H naught 1, using an A matrix 1 minus 1. Remember, so this C matrix here is basically having that 1 minus 1 type of structure along the main diagonal block, so that we will have that H naught 1 to be tested.

(Refer Slide Time: 49:34)



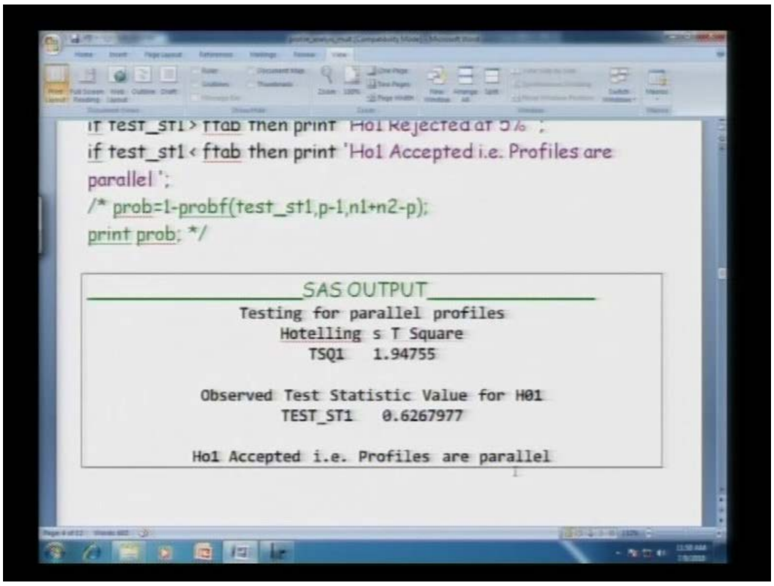
And then, once all these are in place, we are testing for the parallel profiles, we are when we are testing for this parallel profiles here, we have this n 1 equal to this n to equal to this p the dimension is 4 and then we will have to compute the Hotelling's T square statistic. We are looking at construction of computing that Hotelling's T square statistic, and then we will be looking at the tabulated value of the F distribution, will compare that with the value which is the tabulated value of the F distribution. And comparing that, we will have the Hotelling's T square statistic being either accepted or rejected.

(Refer Slide Time: 50:12)



So, we have for the parallelity of the profiles, we will require this observed test statistic value for H naught 1, now the testing procedure as what is given in this particular piece of program in it is basically, going to have that sequential mode of testing. That first, we will test H naught 1, if this is accepted we will proceed to test for H naught 2; if it is not, then we will exit from that particular program and that particular point of time. So, these are the program statements what is more important is to look at the output of these particular program.

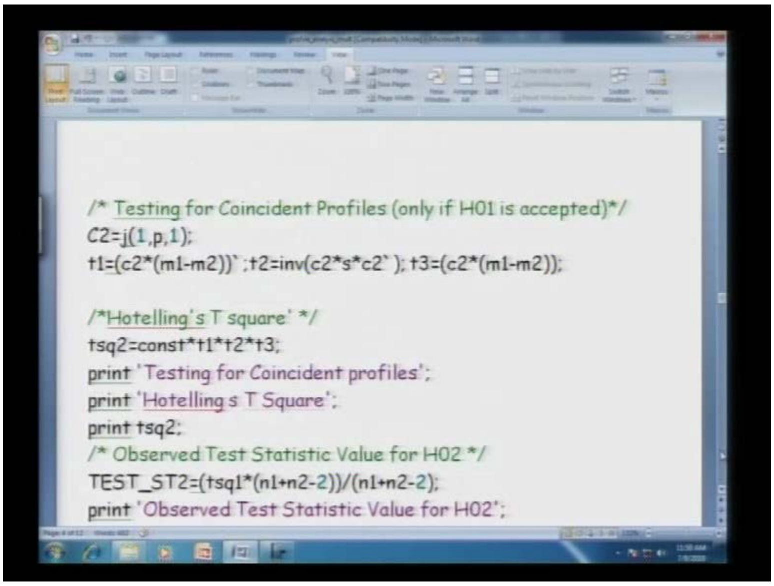
(Refer Slide Time: 50:35)



Now, when we have that profile being given, it appears as if that the two are almost the same, they are of course, parallel with minor deviations. So, the point of interest is to see whether that minor deviation, actually is significant or not that will make the two profiles deviating from parallelity and equality of the two profiles also. Now, when we look at the result for testing of the parallel profile, the Hotelling's T square statistic turns out to be 1.947 and the observed T statistic, which is a constant multiplier of this Hotelling's T square statistic turns out to be it is 0.626.

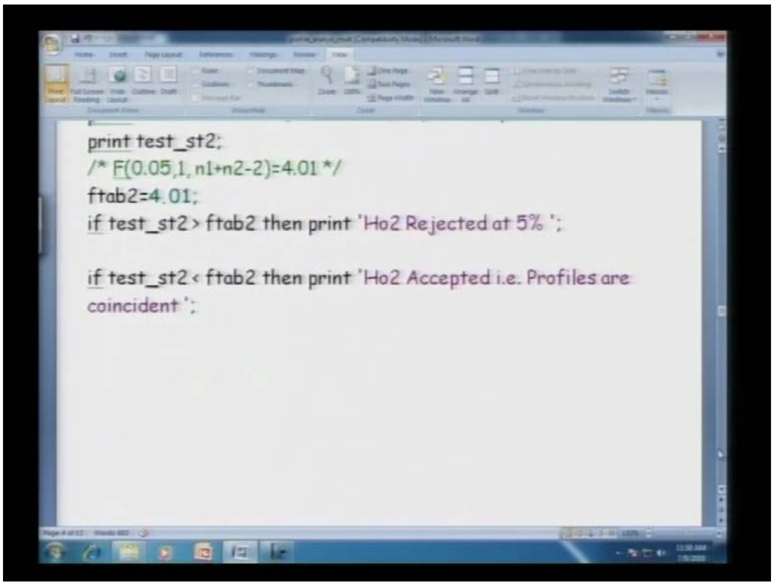
And hence, the test statistic value is lower than the tabulated value of the F distribution at the respective, degrees of freedom. And hence we accept this H naught 1, that is the profile say parallel whatever, deviation we have it is minor deviation basically.

(Refer Slide Time: 51:38)



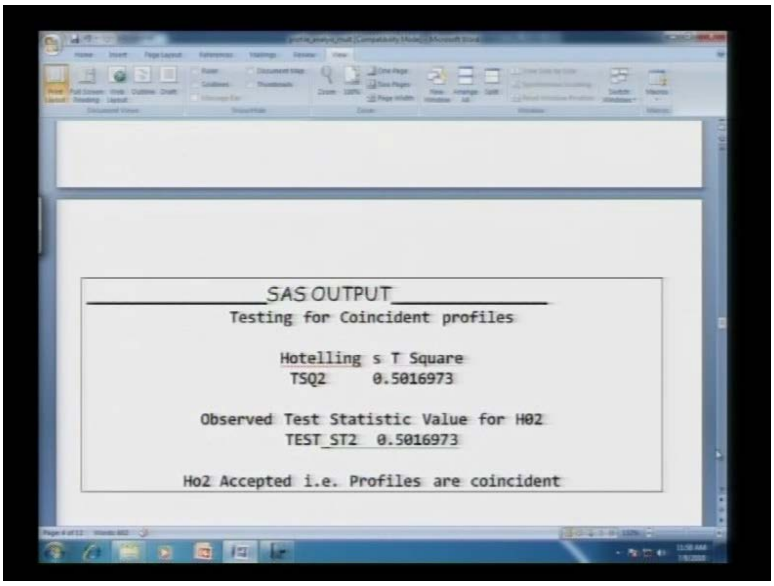
Now, once we have this H naught 1 being accepted, then we proceed for testing of H naught 2 hypothesis. So, testing for the coincident profile, only if H naught 1 is accepted that is what we have, then we once again have the similar type of programming, in order to compute the Hotelling's T square statistic and hence to look at the observed value of the test statistic for H naught 2 hypothesis which is given by this.

(Refer Slide Time: 51:57)



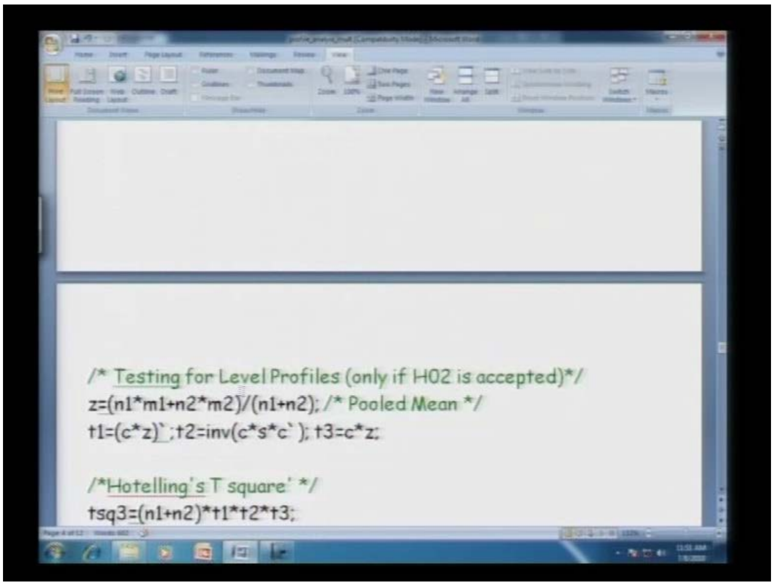
The output tells us, that we will be testing for coincident profiles, this is the Hotelling's T square, this is the value of the test statistic, what we have.

(Refer Slide Time: 52:00)



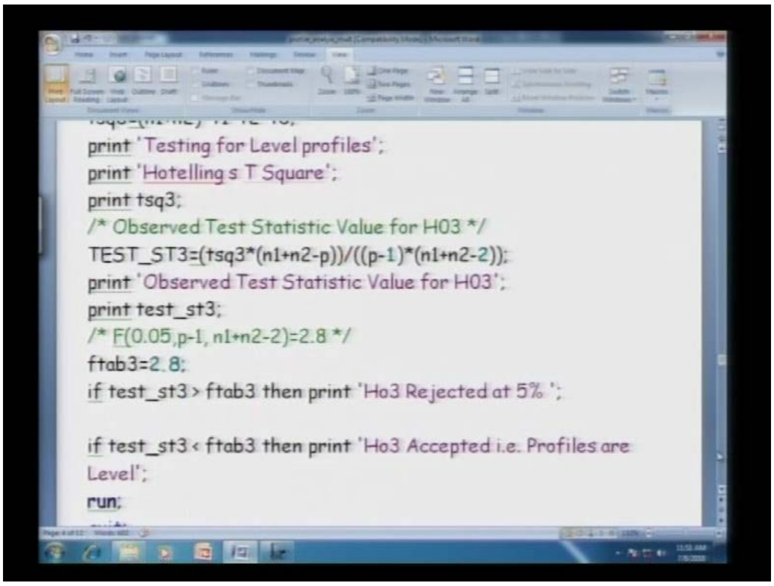
And once again H naught 2 is accepted that the profiles are coincident, now if H naught 2 is accepted as what is done here, we will have proceed to H naught 3, which would test, whether the two profiles are actually level.

(Refer Slide Time: 52:23)



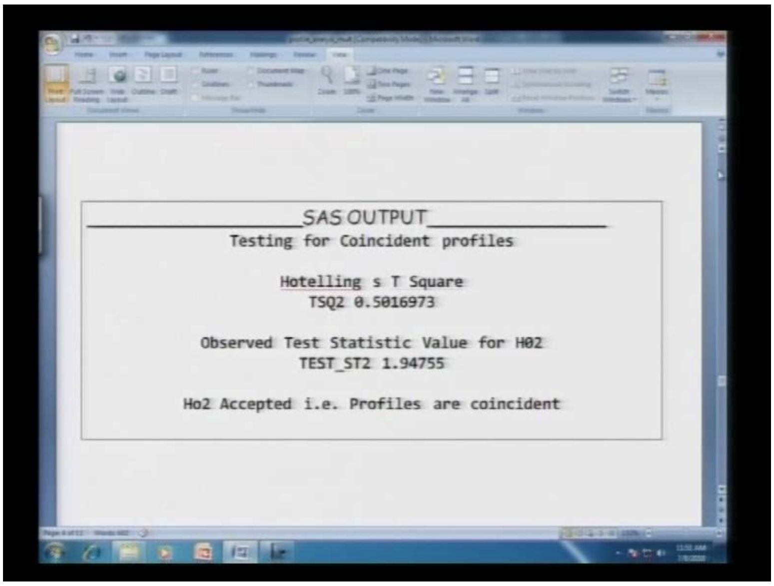
Therefore, the common profile is level that is testing for level profiles.

(Refer Slide Time: 52:28)



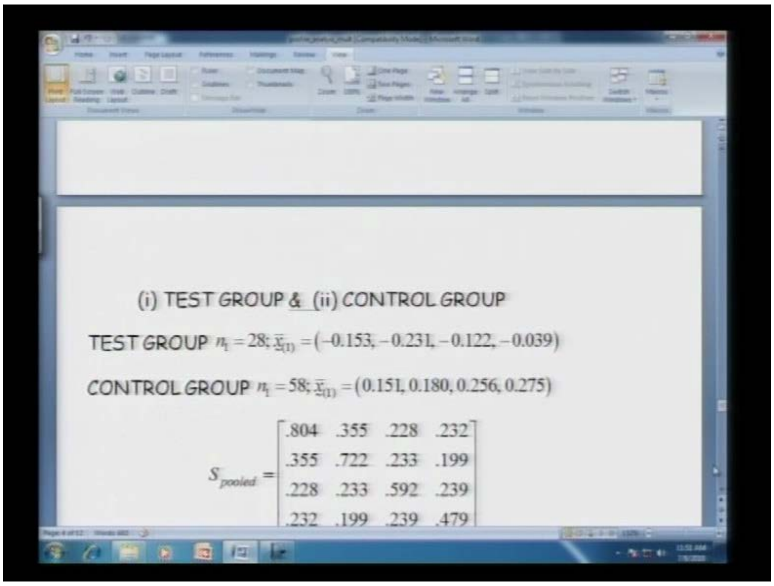
Level profiles will test that all the components are same; they do not differ significantly in terms of statistical hypothesis testing.

(Refer Slide Time: 52:40)



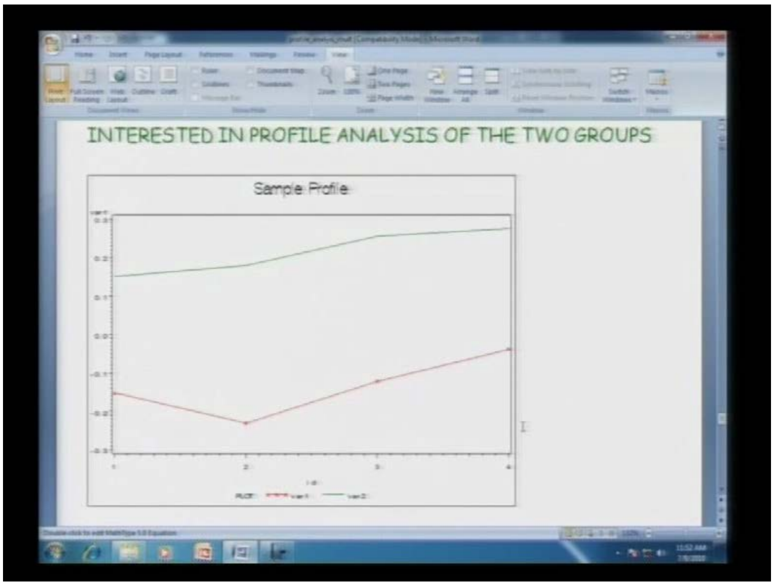
So, one can similarly, implement this particular problem here, as what we have seen sample profiles does not actually lead us to believing, that we have got the third type of hypothesis, it cannot be actually accepted that the profiles are level.

(Refer Slide Time: 52:44)



So, we will have this, as the coincident profiles the result corresponding to this is the second result actually, testing for the level profiles that level profile hypothesis actually is rejected, we can H naught 3 is rejected at 5 percent level of significance. We have another example, this is relating to groups of observations which is one is a test group, and the other one is a control group.

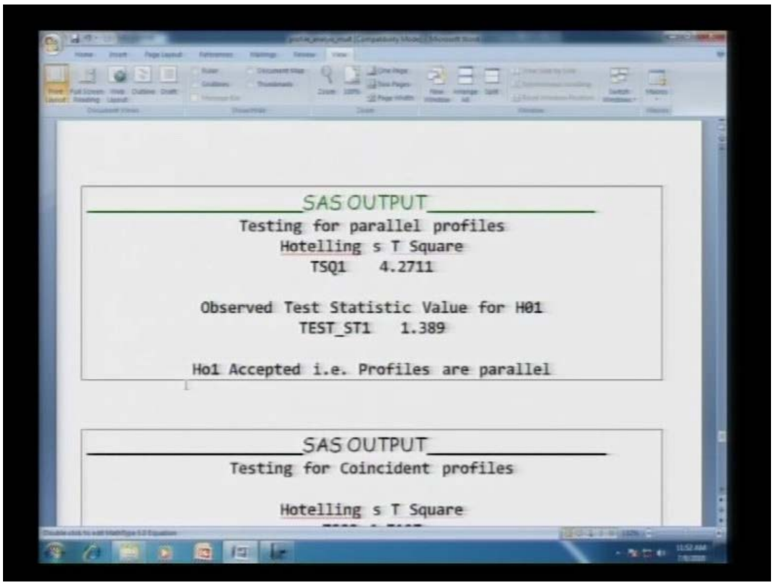
(Refer Slide Time: 53:33)



We have once again four-dimensional data here n 1 is 28, n 2 is 50 58 and this second mean vector this is x 2 bar is this, where the pooled sample covariance matrix as this.

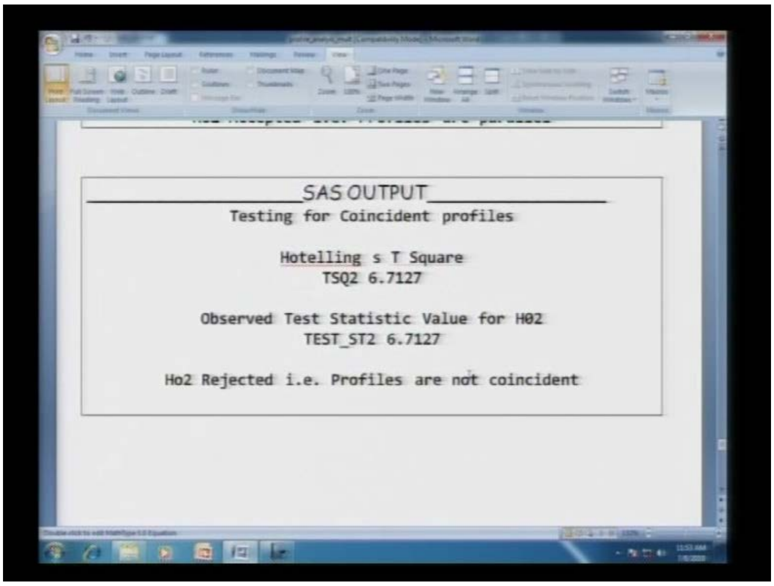
Now, once again we are interested in profile analysis of the two groups; we see that the sample profile here is of this particular nature, this one corresponding to the control group, and one corresponding to the test group. If we once again have to perform profile analysis will have to go sequentially, in order to test this hypothesis, the SAS output is what is just given here.

(Refer Slide Time: 53:53)



So, for the testing of parallelity of the profiles of this test and the control groups, we have H naught 1 to be accepted, that is the two profiles are parallel. Then once that is accepted, we will have to test for the hypothesis that they are coincident profiles that is they have a common profile.

(Refer Slide Time: 54:16)



So, this basically is a output corresponding to that, now as we had seen for the from the sample profile of the two groups, that they do not appear to the naked eye actually to be coincident profile. We have the test statistic value, the Hotelling's T square first to be 6.7, the observed value of the T square statistic to be 6.7, because as we had seen further second hypothesis for the given problem is that only. And H naught 2 is rejected for this particular set of data of coincident profile of the test group, and the control group.

And hence, we say it we can say that the profiles are not coincident, in this particular setup, and hence we do not have the coincident profile nature from this two as it is quite evident from this that. We have clearly two different profiles actually, this is profile for one group, and this is profile for the other group, although they appear to be statistically parallel. But they are not coincident profiles, and hence the level profile hypothesis does not come in to picture here, and we will have to be we will have to stop at that particular point, when we say that we have well parallel profile. But we do not have coincident profile that is the profiles are not equal, they are different.

So, this is how profile analysis actually is carried out in practice, there are other applications also of Hotelling's T square statistic like paired comparison, then repeated measure designs and all. So, we are not going in to detail of those, so they are similar problems which can be tackled under the similar type of approach, what we have adapted for the profile analysis, it is basically Hotelling's T square is what is going to be used there as well. So, from the next lecture, what we are going to start is, we are going to look at multivariate analysis of variance comparison of more than two groups of means, is what we are first going to see. And then we are going to look at extension univariate theory of analysis of variance, to the multivariate data, thank you.