Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Shramishtha Mitra

Department of Mathematics and Statistics

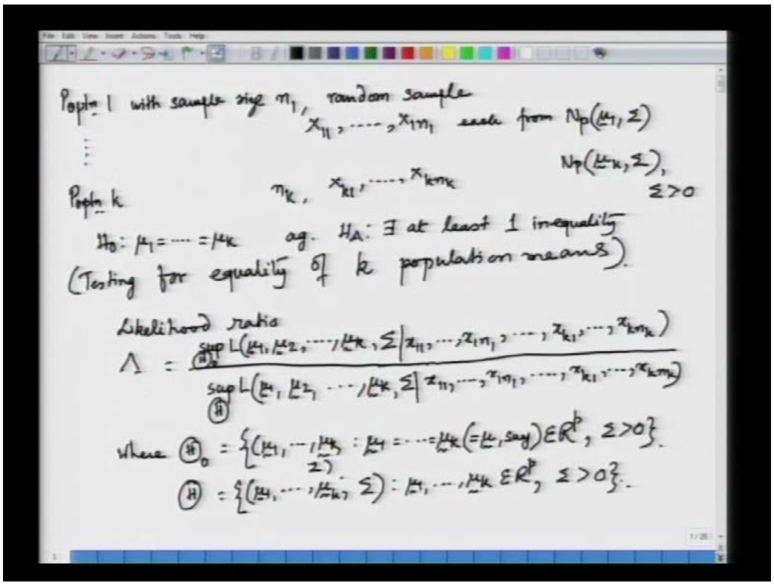
Indian Institute of Technology, Kanpur

Lecture No. #18

Manova – I

In the previous session, we have considered comparison of paired population means, that is comparison between population means from to distributions. The distributions are multivariate normal distributions. In this discussion, we are going to generalize this to the case, when we have more than two populations. We assume that, we have k, a finite number of populations, each of them multivariate normal. And we are interested in testing the equality of the k population means. Obliviously we can handle this by repeated paired comparison test, but if you can handle the whole thing in one go, this it is, obviously more preferable. So, what we have is essentially, we have now k populations.

(Refer Slide Time: 01:07)



So, that we have the first one say population one, the first population with sample size n 1, and the observations are random sample that we have. These are being the noted by x 11 to x 1 n 1. Note that, we have two subscripts now; the first one as you can see pertains to the first population that is y it is 1 throughout, and then the next the second subscript is for the number of observations within that population.

So, we have this random sample x 11 to x 1 n 1, each from the normal the multivariate normal, p variate normal mu 1 dispersion matrix sigma. Similarly this is, this setup is repeated to the k th situation, where we have population k, the k th population, sample size is n k n n k and the random sample is now k 1 to k n k, and we have them coming from the p variate, multivariate normal with mean mu k and sigma. Note that, to differentiate between the populations, we have different notations for the means. The i th mean mu i is pertaining to the i th population, which has sample size n 1 n i, but you we have to take care of the fact that, the covariance matrix is same throughout. It is the same sigma and we have only the usual assumption, that sigma is positive definite which we are denoting by the sigma greater than 0.

So the null hypothesis, we are interested in testing is H naught, the null hypothesis is mu 1 is equal to, up to mu k. So, essentially the problem is testing for equality of k population means. Obviously, when the populations are multivariate normal, the k populations, the i th population is multivariate normal p, dimension with mu i mean and variance matrixes covariance matrixes sigma.

So this is being tested, the null hypothesis is being tested against the alternative, that H naught is not true, which is nothing but, there exist at least one inequality. So, the equality relationship is violated at least once, that would violate the null hypothesis, that all the means are equal.

So we apply, the usual, we apply the likely hood ratio test principal for testing this null against this alternative. And if you recall the likely hood ratio test principal, it is centered around that likely hood ratio, which we denote by a lambda. So, we denote the likely hood ratio is lambda. And just recall the definition, we have the numerator is the supermom over the likely hood function.

Now, this is written as function of the, we are considering the parameters here. So, we denote all the parameters that are coming into the picture. So, they are k, such means mu

1 to mu k and obviously the dispersion matrix sigma. This supermom is being considered over the parameters space under the null hypothesis, that is the restricted parameters space, hence we have this theta, capital theta with the subscript 0 here. Just like for the hypothesis, we have H with a subscript 0, this is a same thing. So, this is supermom is considered is being considered over the parameter space under the null hypothesis, and this is over supermom of the likely hood function mu 1 mu 2 to mu k and sigma. And this supermom is over the unrestricted parameter space. If you want, you can also include the observations here.

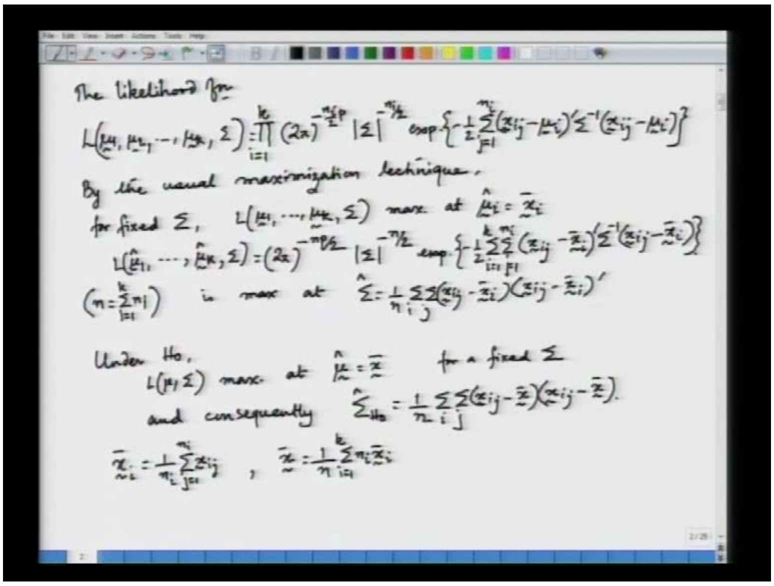
So we may as well write, the observations which are x 11 to x 1 n 1 and then for the k th one which is x k 1 to x k n k. So extend this, similarly we have the observations for the first population, and similarly for the k th population. So let us again go back to these parameters spaces, this notation that you have use the first one is… So, let us specifically just describe this for once. So what we have is, theta naught is the parameter space of mu 1 to mu k. Such that, basically the null hypothesis is all the populations means are equal. This is equal to say some common mu, and they are belonging to R p obviously, because we have the p variate, multivariate normal distribution and we have the variance covariance matrix of positive definite one.

So, this is the parameters space under a restriction, restricted by the null hypothesis, which we are going to consider in the numerator of the likely hood ratio of the likely hood ratio test principal. And the unrestricted parameters spaces, it is just rewriting the parameters space, because we do not have any assumption here. And we have its mu 1 we have not mentioned the sigma here. So obviously, we have to do this, let us raise a bracket over here, and then we have the sigma such that, we have mu 1, no question of equality here. We have each of these mu 1 to mu k, each distinct belonging to R p, and we have the positive definite variance covariance matrix.

Next, in question is the likely hood function, recall the likely hood function that you have considered exactly in the situation, where you had sampling distribution, of where you had considered multivariate normal distribution and sampling distribution from the multivariate normal distribution. When you had a single population, we handle the situation in a way like, the population size was n, or the sample size was n, the dimension was p, and we know how to handle, how to write the likely hood function.

Now, exactly the same thing will be done here, the differences that you have for the i th, you have k populations, that is the first difference. The i th population has sample size n i, it has the mean vector mu i. And since the populations are all independent, we just consider, the likely hood function for the ith population, and then we consider the product over I, over k such populations.

(Refer Slide Time: 09:31)



So, let us see how we handle it, the likely hood function, that we consider here is, we have L, and then this is nothing but, mu 1 to mu 2 to mu k, and then sigma, obviously conditioned on the observations, we are not rewriting it now. Now what we have is, usual for the usual case what we would write is, this is twice pi and instead of n by 2 we would write n i by 2, because the i th population has size n i, And then we would write the determinant of the variance covariance matrix, which is sigma. This raise to the power minus n by 2, no longer n by 2, I would write n i by 2 and then we have the exponent term, whatever what was that, so I have minus half and then I have two subscripts for the observations now.

Let us forget about the i th population for the time being. Let us consider the observations, when i is fixed. So this is sum over j from 1 to n i, and then I have the mean vector pertaining to this population, i being fixed is now. Obviously, these are vector observation. So, this is mu i and then we have the usual variance covariance matrix, its inverse, and then we have x i j minus mu i.

Now, what we have now is k such independent population. So at the final step, we consider the product of these over i going from 1 to k. So, that is the only difference. We are using two subscripts for the observations. we have a different mean vector in every populations. So it is mu i, and we consider the product over k such populations. So, we have to find firstly say, let us concentrate on the denominator part, where we do not have restriction.

So, we have got to maximize this likely hood function, to get the MLEs of this parameter. So the usual principal, recall what you had done, when you had sampling distribution from a single population. What we did was, initially we fix sigma and then try to obtain the MLE of mu. So, we do the same thing. So, by the usual technique. By the usual maximization technique, considering derivative. So, by the usual maximization technique, fix for fix sigma. We have L mu 1 to mu k sigma maximized at when mu i at is x i bar, and then this plugging this estimated values of mu i es. we have L mu 1 hat to L mu k hat, and then we have sigma.

And this is now, nothing but, 2 pi raise to the power, we will do something here, instead of this is, this is to the power, rewrite np by 2. Obviously, n is nothing but, summation n i, i from to k, and similarly for this determinant term also we can write this is minus n by 2. So, n is the total sample size considering all the populations. And then we have this is nothing, but minus half, and we have a double summation now. Here, first one over i for the groups of populations, next one is for observations within that group. So 1 to n I, and then we have x i j minus x i bar, because we are now putting the estimators of mu i es here, this is transpose, this sigma is for the variance covariance matrix, and then I have x i j minus x i bar.

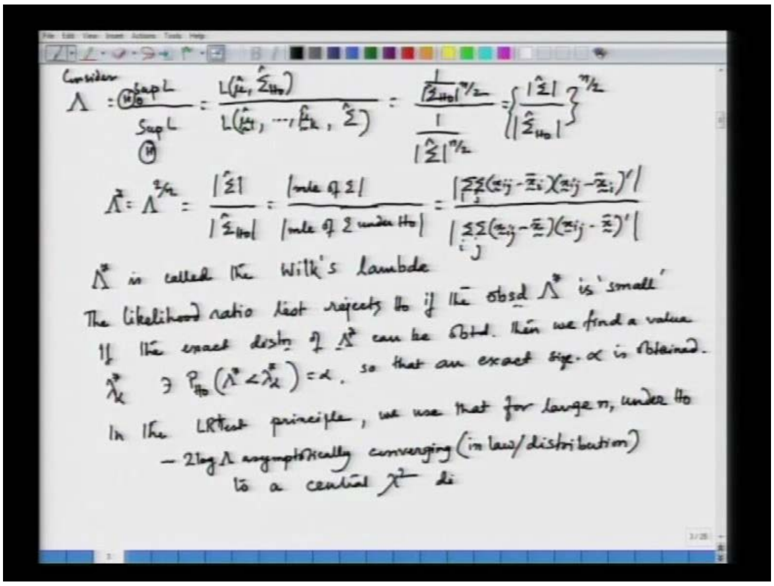
So this will give us, this is being maxed at sigma hat, that is the only parameter involve now, sigma. And this is nothing, but 1 by n, but easily seen that this is 1 by n, a double summation infect this term in the exponent. So, we have x i j minus x i bar x i j minus x i bar transpose, giving as a p by p matrix, square matrix. The first summation over a, the second over j. This is the situation, when we have the unrestricted parameters space for every mu I, we have mu i hat is x i bar. And the sigma hat is coming like a matrix like this.

Now if under H naught, we have simple situations like L mu sigma, because all the population means and now mu, and this is maxed at mu hat, this the overall mean for fixed sigma. And then, and consequently, the sigma hat under this restrictions. So, let us right the sigma hat with an H naught here, to distinguish from this estimator, when it is estimated on the unrestricted situation.

So, this is nothing, but we have 1 by n double summation. So, there is no question of x i bar. So now, we have a common x bar and this gives need a estimator. Let us though, we easily understand what was the, what are this x i bar and x bar r, still we just we may write it here, that x i bar is nothing, but when we have the i th population fixed and some is being considered over the observations in that fixed populations. So j is going from 1 to n i. This is obviously, element by element wise, because this is not a scalar, but a vector valued observation now, and similarly we have x bar is nothing, but the weighted means of the different population means.

So, this is going from 1 to k. So we say, that it is not really very different, or that all difficult to handle the likely hood function, when we have k population means. If there independent, we can simply consider the product of each of the likely hood function for i th population, and easily handle the maximum likely hood estimation case for the parameters involved in the unrestricted cases as well in the restricted case. Next we consider the prime factor of this likely hood ratio test principal, that the thing around around what the whole principal revolves, that is the likely hood ratio criterion, or the likely hood ratio test statistic.

(Refer Slide Time: 18:21)



So we go to the criterion, once again we consider lambda, this is well, we have, saying that this supermom of L under theta naught by supermom over L under theta. And we have actually seeing, what are the parameter estimate that are maximizing the likely hood function, and consequently we just put in this values here. So, we have now L mu hat sigma hat H naught according to our notation. And this is upon mu 1 hat up to mu k hat, and then our estimate of the variance covariance matrix sigma hat.

Now, if we do. So, we can easily see that, the whole thing can be simplified only to the this factor involving the sigma hat matrices. We will simply get 1 by determinant of sigma hat H naught, this raise to the power n by 2, and similarly 1 by determinant of sigma hat also raise to the power n by 2. So, this is what lambda is coming to, and this is simply sigma hat determinant of that sigma H naught hat determinant of that raise to the power n by 2.

So, what we do is we consider, something call something which we denote by lambda star, and this is nothing, but lambda raise to the power 2 by n. So, that we simply get this ratio here, which is determinant of sigma hat and this is determinant of sigma hat under H naught.

So, this is nothing, but well this is nothing, but determinant of MLE of sigma and this is by determinant of MLE of sigma under H naught. And we very well know what are they, because we have just seen that can be derived also very easily, that this is nothing, but 1 by n gets canceled, we have to p by p square matrices; the first one is x i j x i bar x i j minus x i bar transpose sum over i n j, and in the denominator we have the matrix x i j minus x bar x i j minus x bar transpose. So this lambda star, we would actually consider it is very easy now to see, that how to get the test statistic, if we consider the likely hood ratio test principle. Since we have the data with us, it is not a problem to calculate to get this matrices, the two matrices, and then to get the determinant also. This are now depending, this are depending on the observations, they depended on the estimated parameters also, but those again in turn depended on the observations.

So obviously, this lambda, lambda star, whatever we talk of, is a test statistic. It is essentially got some numerical value, if we have the observations, if you have the data at hand. So this lambda star, that we have here is called the wilk's lambda, and this is used for the test criterion, and we say the likely hood ratio test rejects H naught, if the observed lambda star is small.

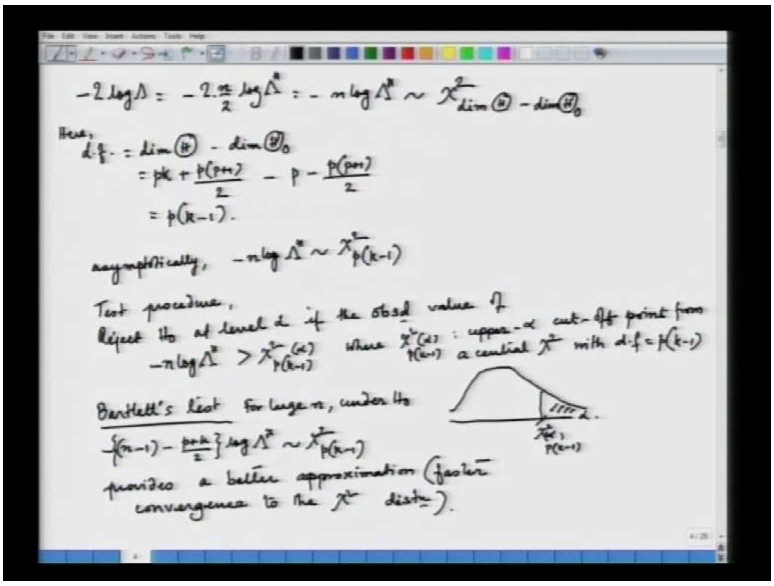
Now, why is this, if note what is the likely hood ratio, that we have considering it is supermom of L over the restricted parameter space, and then we take this over the supermom of L over the unrestricted parameter space. So, we are considering basically supermom over a sub set. So we can note one thing, at least that it has to be less than 1. This supermom in the numerator can never exceed the supermom in the denominator. So, this always has to lie within 0 and 1 essentially, because either we are considering joint p m a for joint p d f. So, positiveness there and this is also less than 1, and then when we see, that this is close to 1, what is happening is that, are assumption or the hypothesis that we have taken is very pretty much close to the actual situation, actual parameter. So, then in that situation, this ratio will be closer and closer to 1, and further and further we are from the actual situation, this will go closer and closer towards 0. So, that is why, the criterion is given in this way that, the rejection will be rejection of the null hypothesis over the alternative will be, when the test statistic lambda star is small.

Now, we can find the exact distribution, if the exact distribution of lambda star is obtainable, then what will happen? We will try to find some value say, if. So, if the exact distribution of lambda star can be obtained, then we might as well find a value, say small lambda star, such that what is happening? We are considering the rejection region, that is lambda star is less than this value lambda star. Let us put a second a subscript also alpha.

So, this probability of rejection under the null hypothesis, which means that this should be equal to alpha, if you have decided on the level of significant. So alpha. So, thus then we can find this value, such that probabilities is equal to alpha. So, that an exact size alpha test is obtained, but unfortunately, in most of the situations, it is very difficult to find the exact distribution of lambda star, and coupled it with the fact, that in the likely hood ratio test principle, we can use a very strong result that for very large n. For large n under the null hypothesis, we have a function of this lambda star is converging in low, or in distribution to the simple central chi square distribution.

So, we take help of the result. So, in the likely hood ratio test principle, let us use the acronym LR test principle, note that or we take, we use that for large n under H naught, we have minus. So, I said it is function of lambda, not exactly lambda. So, we have log of lambda is asymptotically converging.

(Refer Slide Time: 27:35)



So we are talking of convergence in distribution or law here, law or distribution, to a central chi square distribution. Consider, if we want have it in terms of lambda star, we will have minus twice of log lambda, which is nothing, but minus twice, then n by 2 log lambda star by its definition, and then we have this is simply minus n log lambda star. So, whether you calculate lambda or lambda star that estimators, and this will follow a chi square distribution. What about the degrees of freedom? Well, the degree of freedom is dimension of the parameters space theta, but we are losing certain degrees of freedom, and from where is it coming? It is coming from the fact, that we have put some restriction, which is basically our null hypothesis.

So, we have to take out that degrees of freedom from the whole, it is basically, the dimension of the unrestricted parameter space minus dimension of the restricted parameters space. Let us see what is the degrees of freedom in this situation. So, we have degrees of freedom is equal to dimension of theta, and this case here minus dimension of theta naught.

Now, recall how to this unrestricted parameters space looks like, we have k population means mu 1 to mu k, each of them is p dimensional vector. So, that for the mean vector part, for the mean part, we have p k number of parameters, and then we have the covariance matrix sigma. What is a number of parameters their? Recall that sigma, we have a covariance matrix is symmetric.

So, we have p and then p minus 1, in this way to one, till we go down. So, it is basically p into p plus 1 by 2 unknown parameters in the covariance matrix. What is happening in the restricted situation? There we have said that, mu 1 to mu k are all equal and equal to common mean vector mu which is p dimensional. So, it has p unknown elements in it, and the sigma, the covariance matrix, since it remains the same, number of unknown parameters pertaining that also remaining same, and we have p times p plus 1 by 2, and this is giving me the value p times k minus 1.

So, we have asymptotically, to come to a decision, we are going to use that asymptotically, for large n that is under H naught minus n times log lambdas star follows a central chi square distribution with degrees a freedom p times k minus 1. It is recall n is the total sample size, that is sum of the sample size is over each populations, n is nothing summation of n i, p is the data dimension, and k is the number of groups or number of populations that we are handling.

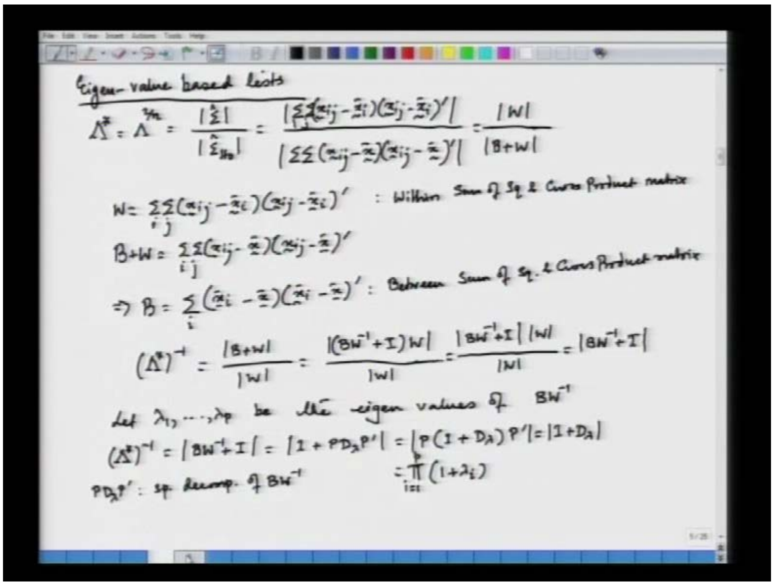
The test procedure, next is the decision, as we have already said, that it is we reject H naught, we reject H naught the null hypothesis, if the observe value is small. So, what is happening here? We are considering minus of n log lambda. So, the test procedure is obviously reject H naught in favor of H a at level alpha, if the observed value of minus n log lambda star or minus 2 log lambda, whatever you consider is greater than chi square p k minus 1, this is the cut of points. So, let us put some alpha here, where chi square alpha p k minus 1 is the upper alpha cut of point from a central chi square with degrees of freedom equal to p times k minus 1.

So roughly, will have a situation like, you have a upper alpha point here cut of point here chi square alpha p k minus 1 d f. So, this area is alpha. As an alternative to this, people sometimes use the bartlett's test. Bartlett's test is just a change in the constant terms for this log lambda star or log lambda, instead of minus n, we have little different expression here. It is minus n minus 1, and then we have a minus p plus k by 2, a little longish constant term attached with the main statistic, and then we have this log lambda star. And this is also following. So, we say for large n under the null hypothesis, this also asymptotically follows the central chi square with p k minus 1 degrees k the freedom.

And so, the test criterion is same as this one, only thing is the statistic is going to be this. So, this bartlett's test provides, a better approximation, in the sense faster convergence, the sense of faster convergence to the chi square distribution. Since, it is an approximation, if we know if we have something, that gives a better approximation of faster convergence, and since it is does not involve much of an extra labor from the usual likely hood ratio test, we can might as well use the bartlett's test.

Now, there are a few alternative tests to this, but pretty much based on the same principal. W e are going to discuss, the basis of such alternative test, and then going to just very briefly mention, what those alternative tests are. These are, as we see this will dependent on the eigen values of the matrices, that we have obtain that is sigma hat and sigma hat under H naught.

(Refer Slide Time: 34:58)



So now, we are considering the eigen value based tests, for which we once again recall what our wilk's lambda is, lambda star is nothing, but lambda raise to the power 2 by n, and which was nothing, but the estimate of the variance covariance matrix in the unrestricted situation by the estimate. If the restricted situation, and what we got was essentially, the determinant of two matrices x i j minus x i bar, and then we have x i j minus x i bar transpose. So, this is over i and j determinant of this matrix divided by the determinant of an our matrix of the type x i j minus x bar x i j minus x bar prime.

So, this is what our wilk's a lambda is, and now we give some special name to this, we call this matrix, which is coming in the numerator, that is the sigma hat matrix as w. And the denominator in the denominator we have the sigma hat under H naught; we give this the name B plus W.

So, we have the ratio of determinant of W by determinant of B plus W, and then obviously my W is, let me write it again, it is the matrix x i j minus x i bar x i j minus x i bar prime ,and B is in our B, but B plus W is whatever is given there. And from here, I can see that, if I consider B, it is nothing, but. So, let me write this once again. So, this is x i j minus x bar x i j minus x bar prime, giving me B matrix, as what we have to do is a actually, take a plus and minus of x i bar in this expression B plus W, and obtain by B matrix as x i j minus x i bar x i j minus x i bar transpose, sorry this little correction here, this is x i bar. So, we have x i bar minus x bar, that is why, we had the single summation, that is x i bar minus x i bar prime over i.

Note this, W matrix is nothing, but the if you think of situation, where you have the p equal to one case, that is you do not have vector valued, but scalar valued random variable, then what is this? This is simply nothing is it this W is a summation x i j minus x i bar, and we have whole square over it, because p is equal to 1, and we have a scalar value for that. And in that case, it is the within sum of square, if we recall. So, this in the multivariate analog, this is call the sum of squares, within some of squares and cross product, this is the extra term we use in the multivariate case cross product matrix and B is nothing, but the between some of squares and cross product matrix.

So, what is then, lambda star, the wilk's lambda, the reciprocal of it. Well it is nothing, but determinant of B plus W by determinant of W, and let us, to us little bit of manipulation here, that we consider, we try to take out this determinant W common from the numerator. So, what we do is, we post multiplies, say this by W inverse and we have the identity here.

So that, we have W matrix be taken common in the numerator, and we have this here. So, since determinant of a B is determinant a times determinant B, we simply have this as B W inverse plus i, and we have determinant of W by determinant of W, which is now giving me the determinant of B W inverse plus i. I have a matrix here, no longer ratio, I have a simple determinant now. W W inverse exists, because W after all gives the MLE of sigma. So, that is that is not the determinant of which is naught 0, and inverse exist as a result of it. So, we have this is coming out, and then let us consider the spectral decomposition of this matrix. So, that we have p D lambda p prime, or usual spectral decomposition method. So, let us write the eigen values, and now coming into the picture as soon as talk about spectral decomposition.

So, face all being p dimensional square matrixes, lets lambda 1 to lambda p be the eigen values of the matrix B W inverse plus the identity matrix. So, if we have this the eigen values, then we can a spectral decomposition of this matrix, and we can write that, this is nothing, but the wilk's lambda criterion, reciprocal of it, which is nothing, but determinant of B W inverse plus I, this is giving me determinant of i plus p D lambda p prime. So, this p D lambda p prime is spectral decomposition of the matrix p omega inverse. So, that D lambda contains the eigen values. So, if that is the spectral decomposition of the W inverse, then this are the eigen values of B W inverse, and not of the B W inverse plus I, because we are keeping this i separate, and using p D p prime for this part only B W inverse.

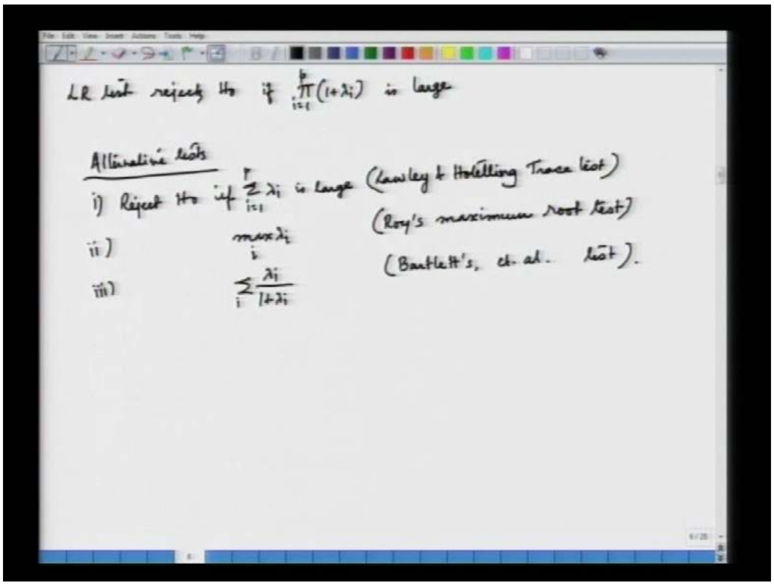
So, it is spectral decomposition of this matrix, and as a result this are eigen values of B W inverse, and p is the orthogonal matrix, where columns are the corresponding orthonormal eigen vectors. So, this is again we can handle this by, we consider determinant of i plus p D lambda p transpose, because we have p matrix as orthogonal, we can very easily take p common out of here, and we try it D lambda here, and p transposes getting outside from here. And then we have this as nothing, but the determinant of 1 plus D lambda, and then it is determinant of p p transpose, which is the identity matrix, which is equal to 1.

So, this is basically. Now, we have a matrix i plus D lambda. Both are diagonal matrices, till now we had a diagonal matrix, with lambda 1 to lambda p as the diagonal elements, now we have lambda i plus 1 as the diagonal elements, and we are considering its determinant, a diagonal matrix. So, this is nothing, but product of 1 plus lambda i, i from 1 to p.

So, you see very easily, once we have these matrices obtain from the data, which is the matrix B and the matrix W, then we obtain the W inverse matrix, what we do is, then consider a product of B and W inverse, and then calculate the eigen values of that matrix. So, we get p eigen values lambda 1 to lambda p, add 1 with each of them, consider the product, and that immediately gives me the value of the test statistic. That is the wilk's lambda.

So, this is now my value of the lambda stack of the wilk's lambda, and I can form my test criteria based on this value. So, that is why it is call the eigen value based test, because all the or the whole thing falls down to eigen values of the matrix B W inverse.

(Refer Slide Time: 44:54)



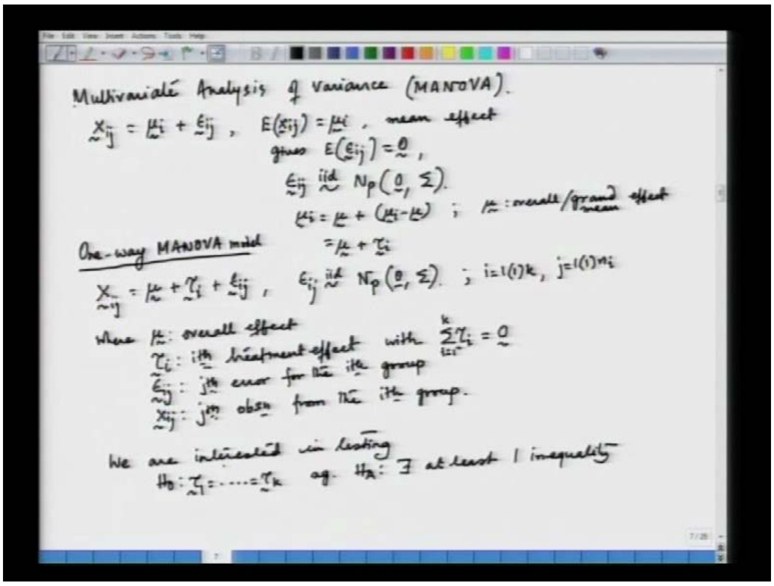
So, let us list down the few test, which consider these, the statistics in this form. So, our first one is, obviously, if we talk about the likely hood ratio test, the test criteria is rejects H naught in favor of the alternative. Consider what happened in the beginning, we have talked of lambda star, if we lambda star is small. Now, we are considering reciprocal of lambda star. So, the test is going to reject H naught, when this reciprocal is large, now we have the reciprocal in terms of the eigen values. So, we say, we rejects H naught, if the test statistics reciprocal of lambda star or the product of 1 plus lambda i i from 1 to p is large, because this is equal to not the statistic, but its reciprocal. That likely hood criterion, but it is reciprocal.

So, the alternative tests are, this is the basic one and then there a some other tests also using eigen values, but in slightly different from, the criteria basically slightly different form. So, this the first one says, that reject H naught, if some of lambda i which is nothing, but the trace of the B W inverse matrix, i from 1 to k. Sorry its naught i from 1 to k, but i from 1 to p, that is dimension of the data is large, because this is the trace, and this is due to lawley and hoteling, and this is called the lawley and hoteling trace test. The next one is reject H naught, if maximum of this is large, make sense also going by this principle, and this is called roy’s maximum root test. Root means here the characteristics root, it is now the name for the eigen value. Something more, we have reject H naught, if summation lambda i over 1 plus lambda i, this is large, and this is by bartlett’s test at all.(No audio from 47:26 to 47:35)

So, these are the few test, where we can use the eigen values of the estimated variance covariance matrix, the unrestricted situation giving us the matrix W and the estimated variance covariance matrix in the restricted situation giving us symmetric B, again what we calculate is the product come of B W inverse, we get the eigen values of this of this matrix. And just considered different forms, different functions of this eigen values to reach a two different testing criterion.

Now, we extend this discussion on the equality of k population means, when the observations are coming from multivariate normal distribution for the purpose of moving to our next topic, which is manova, that is the multivariate analog of anova or analysis of variance. If you recall, what we do in analysis of variance, we have a data, and we try to look at the data and try to try to assign try to look into the variability of the data, and to assign them to different sources of variation. Here also same thing is being done, but the only difference is, instead of the random variable, we have a vector valued random vectors, now the data that we have at hand, they are multidimensional data and we try to generalize the anova to the multivariate situation.

(Refer Slide Time: 49:16)



So now the next topic is multivariate analysis of variance,(No audio from 49:17 to 49:32) in short manova, and we consider as in the usual anova context, what we do here is in the simpler situation, we consider a model. So, we have x i j and then we have its mean effects, mu i and error term that is p i j, the difference here is, these are all vectors, multidimensional data are being considered. So, these are all vectors and we have expectation of x i j, well mu i is the mean effect mu I, this is the mean vector. So, giving us, this gives expectation of the error term, this is obviously equal to 0. Now additionally, we assume (( )) we assume the equality of variances, we assume uncorelatedness, and we also assume, that the data the variables are coming from multivariate normal population.

So, we might as well say, that actually we have this error terms E ij, they are iid, uncorrelated and normality giving them independence, and we have them coming from the multivariate normal distribution N p with mean 0, and variance covariance matrix sigma. So, let us write the, we do a we also do a further manipulate not manipulation, we would say to obtain upon the treatment effect, what we do is, we break this mean effect, and we try to get, try to obtain a grand mean effect out of it.

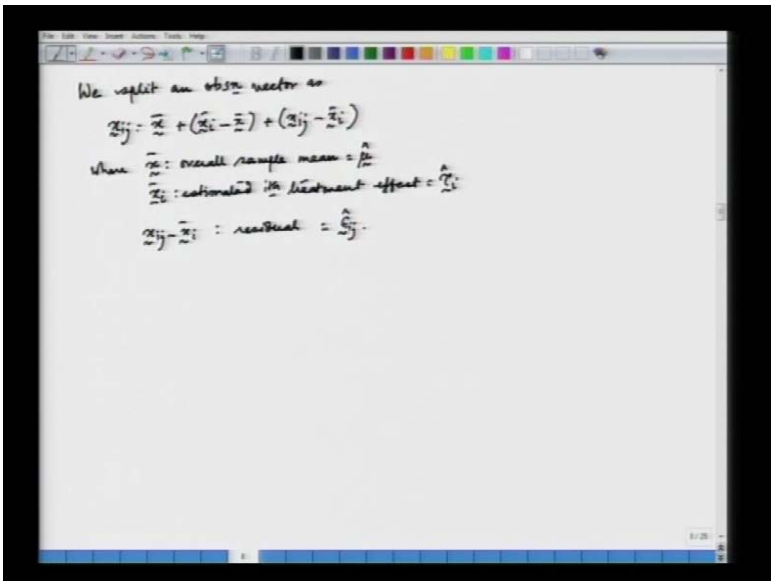
So, what we do is we say, that mu i is the mean effect is equal to the grand mean effect or the overall effect, and obviously, then we have to write the rest of it in this way, all of this are vectors here, that this mu is the overall or the grand mean effect. So, for this second part, we use a notation for the treatment effect, that is what we do in anova, that is this being the treatment effect now I, now a vector again here. And so, the one way anova, let us now write the one way manova model, explicitly you have the data x i j, the j th observation from the i th group, that is equal to the overall effect mu plus the treatment effect tau i, and the error E ij, where E ij this is important are assumption, these are iid normal 0 sigma, let us also give brief definition of brief introduction of this also.

So, we have mu is the overall or the grand mean effect, overall effect tau i is the i th treatment effect, (No audio from 53:18 to 53:27) with note that what is tau, it is nothing, but mu i minus mu. So, we have a restriction here, we have a constraints, that some of tau i i from 1 to k, this is equal to 0. Here we said, that we have k group. So, i goes from 1 to k, same thing like it has k populations. So, i from 1 to k and number of observations in each group, that need not be constant.

So, we have j going from 1 to n I, and E ij is the error, j th error for the i th group. This is like we have x i j, that is the j th observation from the i th group. Now, we are trying to look at the variability of the observations, and trying to assign them to the different sources of variation. So for this purpose, the hypothesis that we are testing is nothing, but the hypothesis, we just looked at now. And we are interested in anova or in manova, we are interested in testing that, same thing that is happening. We have instead of writing the null hypothesis in terms of mu I, it is equivalent to write in terms of the treatment effects. So, now, that we write, this is all the tau i is are equal tau 1 to tau k against the alternative, where exists at least one inequality.

So, this is the setup of the one way anova model, where we see that we are going to use the testing procedure whatever we have learned just now, how to test, how to compare between the k population means, when the random samples are coming from the multivariate normal distribution. Here we have this assumption in place, and the observations are also coming from multivariate normal distribution, and then we are using this equality of the k treatment effects, which is equivalent to saying that we are basically testing the equality of the mean effects mu 1 to mu k.

(Refer Slide Time: 56:14)



We simply split, a data vector, we split and observation vector haves, be have x i j looking at it, we bring down the overall mean, sample mean, and then the group means x i bar minus x bar, and then obviously the rest of it, which is x i j minus x i bar.

Now, this is giving me the overall sample mean, let us write these. So, where the overall sample mean is very comfortably giving me an estimate of the grand mean, overall sample mean, and this is nothing but mu hat, and we have x i bar, the group mean x i bar, this is estimated treatment effect. We are calling this as not the mean effect now, because we have separated out an overall effect, and we call this as strictly the treatment effect, the i th treatment effect, estimated i th treatment effect. And there is tau i hat, and the rest of it is nothing but the residual.

So, that has got the notation residual, which was E ij hat. So, we are going to look at, how we obtain the treatment sum of squares, and the residual sum of squares, and the total sum of squares, and we follow it up with the one way Manova table, and also look at the test statistic for the hypothesis, that we have stated.