Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Sharmishtha Mitra

Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur

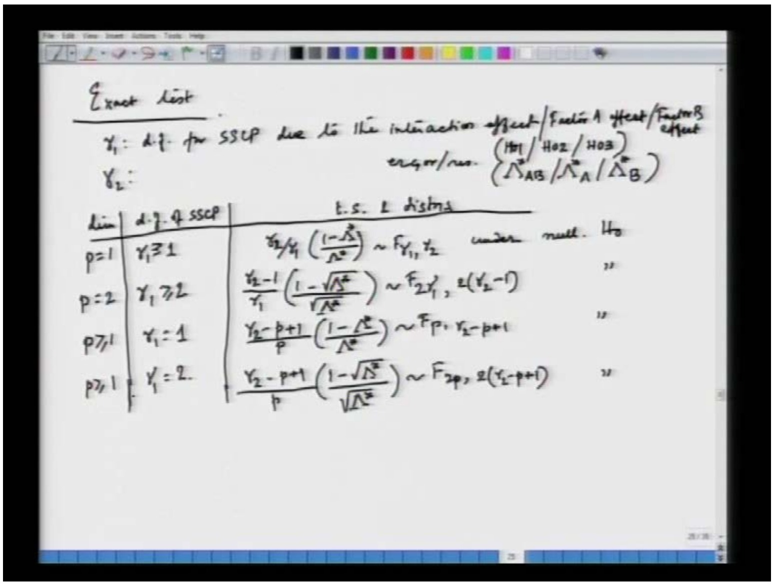
Module No. # 01

Lecture No. # 21

Manova and Multiple Correlation Coefficient

After the asymptotic tests of the two way manova technique we just take up a few exact tests few in the sense, that in the very few situations where we can apply the exact test. As in the case of one way manova we have seen that for some low dimensional cases where, there we had been given the number groups that is the value of k were given there, for which we could use the exact test technique. Here also in the similar manner, we will try to list down few situations, where we can use the exact test technique in the two way manova analysis.

(Refer Slide Time: 00:58)



So, what we have is essentially, we just denote by gamma 1 the degrees of freedom for the sum of squares cross product matrix, due to the interaction effect or the factor 1 effect; so that we can write that factor A effect due to factor A or factor B effect, whichever the hypothesis I am testing. So, essentially this means that this first degrees of freedom gamma 1 it is referring to the degrees of freedom for the case. Interaction effect essentially means, I am testing the null H naught 3 against the alternating way of H naught 3 or if I have the setup of testing the equality of factor a effects, I am testing the null H naught 2.

And similarly for factor B, I am testing the this is not H naught 3 therefore, first 1 is H naught 1; in fact, and H naught 2 for factor A and H naught 3 for factor B and the second subscript the second d F, all this refers to the degrees of the residuals or the error and then, the situations which we can pen down are these few situations, where I have dimension p equal to 1 which means that essentially I have an manova case and gamma 1 is greater than greater than or equal to 1. So, this gamma 1 refers to whichever hypothesis, I am testing if I am interested in testing the interaction effect, then I will have to look at the degrees of freedom of the interaction effect.

So, gamma 1 pertains to that in that case, and then that test statistic is given by gamma 2 by gamma 1, 1 minus again this lambda star you can very well understand the change with the hypothesis, so we have this is my table, I have 1 minus lambda star. Note that for H naught 1 hypothesis, I have lambda A B star for H naught 2 hypothesis, I have lambda a star this is how we have denoted the likelihood ratio criteria and for H naught 3 I have lambda B star. So, if I am testing the null H naught 1 in place of this lambda star, I am going to use this lambda A B star, and so on.

My gamma 1 will also change accordingly gamma 2 will remain fixed, so under null, so this is under null H naught, whichever one is coming. Similarly for the second situation, I will have just one increase in the dimensionality of the data for T equal to 2 and if I have this gamma 1, anything greater than or equal to 2, I can use, use an exact test given by the statistic is given by gamma 2 minus 1 by gamma 1 1 minus root of lambda star by root lambda star. And this is following an F distribution with twice gamma 1, twice gamma 2 minus 1 under the null hypothesis.

If the dimensionality generally will be considering data dimensionality of higher order say 3, 4 or even more, and in that case if this factor is reasonably small. In fact, of nu 1 or 2 only on those situations we can go for the exact test and in this situation the third situation, where we have p is greater than or equal to 1 gamma 1 equal to 1, I have the test statistic as gamma 2 minus p plus 1 by p, and then followed by 1 minus lambda star by lambda star this following an F distribution with p and gamma 2 minus p plus 1 under the null.

And the last case is when I have the dimensionality anything greater than equal to 1, and gamma 1 is equal to 2, I can use the test statistic gamma 2 minus p plus 1, factor in the denominator remains p and I have 1 minus root of lambda star by root lambda star following an F distribution again with 2 p and twice gamma 2 minus p plus 1, so then I stress upon the fact that when we are using the exact test we have the classifications according to the values of p which is always fixed that is the dimensionality of the data.

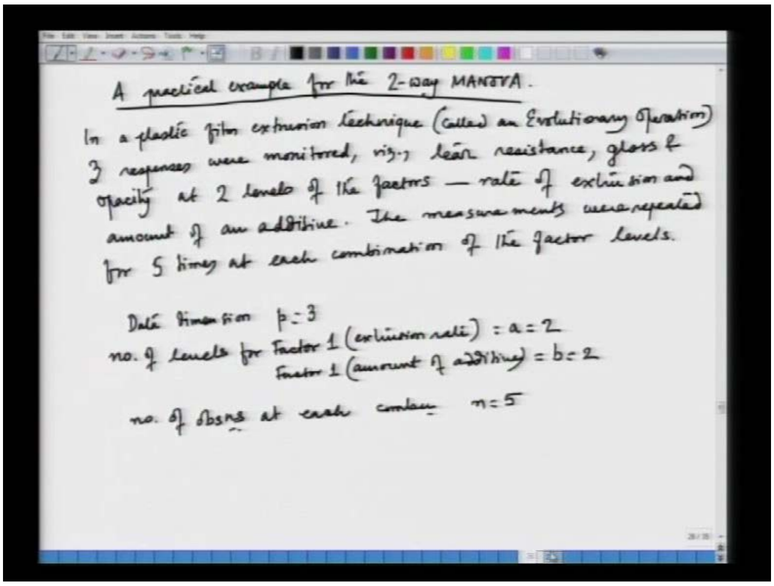
But the next classification is according to the value of gamma 1 which, is essentially the first degrees of freedom of the F statistic, so this keeps changing as we change our hypothesis from H naught 1, to H naught 2, to H naught 3 and along with this what changes is the likelihood ratio criterion walks lambda star, if it is the interaction effect essentially, we are testing H naught 1 then we have to look in to the degrees of freedom of the interaction effect put that value for gamma 1, and in place of lambda star we must put the value of lambda A B star and then get our F statistic.

If it is H naught 2 for gamma 1 we will have to use the degrees of freedom for factor A the first factor and in place of lambda star we have to use lambda a star and so on and when I am saying that the classification, the next classification is according to the first degrees of freedom of the F statistic it means, that the second subscript that is gamma 2 that is basically the degrees of freedom of the residuals that can be anything. So, this completes the exact test table of the two way manova we have these many situations they are listed for us.

And now let us go to a data Example, but here as in the case of one way manova we have discussed in details with the data here, we are not going to take up the data, I am just going to give a data setup, a typical data setup and try to give some interpretation and we are not will we would not be working with the actual data here, so here the example is of some chemical process which is called an evolutionary operation process, it is basically a process to execute plastic film and while doing. So, three valuables of interest are monitored and there are two factors of production that we are considered.

Note that now we are talking about two factors of production, because it is now two way manova and each of the factors having some levels two or three or more levels, and let us see how a typical data setup is in this situation, where we can apply the two way manova technique.

(Refer Slide Time: 08:47)



So, the setup is something like A practical example for the 2-way MANOVA what we are looking at is in a plastic film exclusion technique called an evolutionary pro operation, 3 responses, physically 3 variables are interest were monitored namely tear resistance, gloss and opacity.

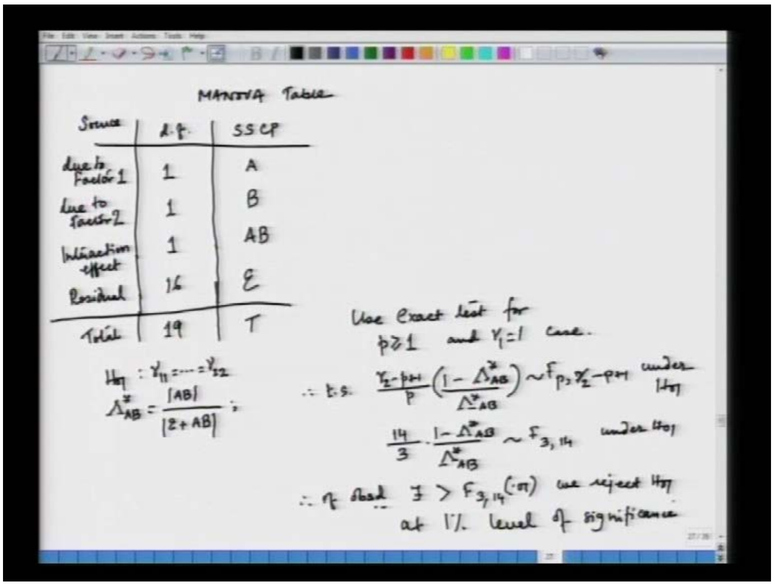
So, these are typical characteristics of a plastic film, how tear resistant they are, how glossy they are, how opaque or transparent they are, so the responses on these factors these were measured in some way in some units and further two factors of production were considered those were at two levels of the factors. First one is the rate of extrusion and the second one is amount of an additive used in the process essentially, so two levels of the factors, so we can see two levels for each of the two factors, these are the rate of extrusion and the amount of an additive.

The rate of extrusion having two levels it can be anything like high or low rate and the amount of additive used again two levels can be moderate, or excessive or it could be the permissible level or the non beyond the permissible level something like that, so we have essentially two Factors each having two levels of each having two levels, and then the measurements, if you recall we have something more in the theoretical setup we said that there are n observations for each of the A B combinations, so the measurements here also were repeated for five times at each combination of the factor levels.

So, what do you know now, what can we write here, without knowing the actual data, I have an idea about something at least, I have data dimension p is equal to 3, because we have 3 responses measurements on 3 variables tear resistance, gloss and opacity. We have number of levels for factor one, factor one is the extrusion rate for factor one say the extrusion rate here, each factor has the same levels on a problem extrusion rate this is number of levels, so this is equal to A and this is equal to 2 here and similarly number of levels for factor 2 that is amount of additive, this we are using a notation b, so that is also equal to 2.

Number of observations, number of independent observations at each combination is n which is 5 here.

(Refer Slide Time: 13:32)



So, we can sort of write a dummy MANOVA Table. So, MANOVA Table if we have known the data, or at least the group means, or this and the sample variance covariance matrices, we could have completed the entire MANOVA Table, so this is a sort of a dumb MANOVA Table, where we have the sources of variation is due to factor one, the degrees of freedom is a minus 1, so this is one, I have some sum of squares cross product matrix for this which are denoting by A.

And then I have due to factor 2, the degrees of freedom is B minus 1, 1 again this matrix 3 by 3 matrices, I am denoting this by A this one by B and the interaction of 2 factors, interaction effect due to interaction of the 2 factors that is a minus 1 times B minus 1 that is also 1 and this is A B, what is the total number of observations well we have how many 4 combinations and we have 5 observations each. So, I have twenty minus 1 that is 19 is the total degrees of freedom.

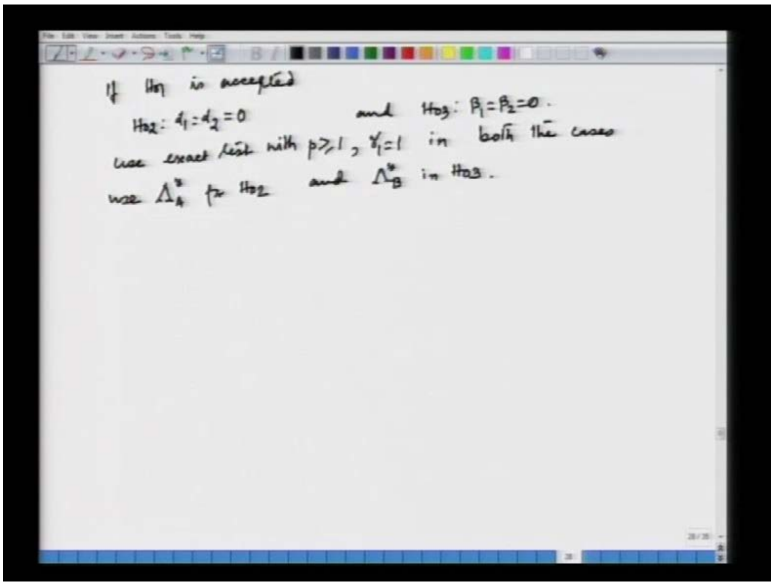
And hence I am going to have 16 d F for the residual, which can be otherwise checked also because I know the degrees of freedom for residual is A B times n minus 1 here A is 2B is 2 and n minus 1 is we have n is 5. So, n minus 1 is 4, so that is that gives me 16 and this matrix is denoted by E, and this is the total variability matrix and now if I am interested in testing the interaction effect what I am checking is gamma 1 1 is equal to up to I have gamma 2 2.

So, that is gamma 1 2 and gamma 2 1 coming in between against the null hypothesis, against the alternative hypothesis, I will use the criterion AB lambda AB star which is given by determinant of the A B matrix by determinant of E plus AB matrix and this is going to some slight change in this with some constant, I can use the asymptotic test, but what is suggested here is the use exact test, which exact test can be used, used the exact test for the p greater than equal to 1 and gamma 1 equal to 1 case.

Because I have the degrees of freedom for the interaction effect is equal to 1, and then therefore, the test statistic is nothing, but gamma 2 minus p plus 1 by p 1 minus lambda AB star by lambda AB star and this is following the F distribution with p gamma 2 minus 2 p plus sorry minus p plus 1 gamma 2 minus p plus 1 under H naught, well you can partially calculate this, because we do not have the value of this gamma of lambda A B star this is equal to gamma 2 is the residual degrees of freedom and p is the dimensionality of the data and 1 is, so this is giving me this is 16 minus 3, so that is 13 plus 1 and I am getting 4, 14 and p is 3, so 1 minus lambda AB star by lambda star this is following F with 3 and 14 under this is H naught 1, under H naught 1.

So, if therefore, if observed value that is this value if observed F exceeds F 3 14 at 0 1 we reject H naught 1 at 1 percent level of significance.

(Refer Slide Time: 18:37)



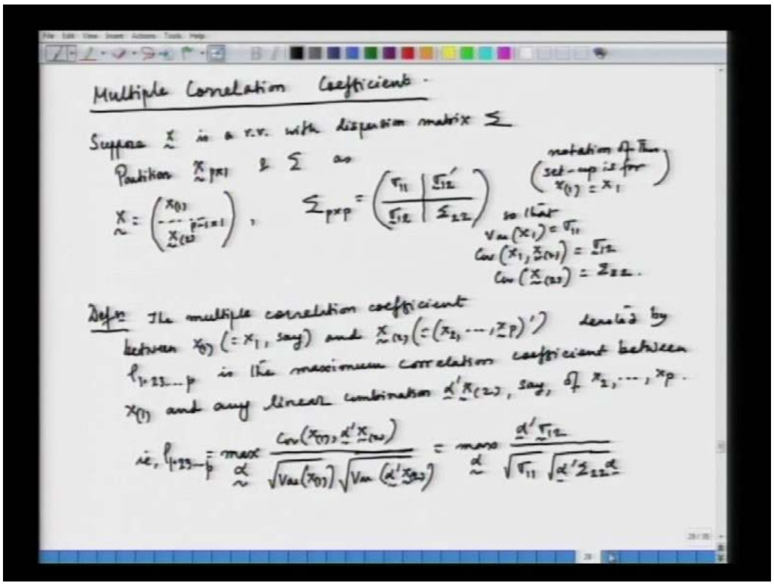
Now, if H naught is accepted, if H naught is accepted which means there is actually no interaction effect by the of the between the 2 factors, that is the rate of extrusion and the amount of additive use that is whether the rate is high or low that there is no connection if the fact that I have to use the moderate level or high level of additive, if that is true there is no interaction effect, I go for the next 2 hypothesis that is I test H naught 2 which says which test that alpha 1 equal to alpha 2 and only 2 levels of the first factor that is equal to 0.

And I also test separately H naught 3 which says that beta 1 is equal to beta 2 equal to 0, so for H naught 2 I use again if you have to use exact test, so use exact tests for this p greater than with p greater than equal to 1 and gamma 1 equal to 1 in both the cases why because I have the degrees of freedom for both the factors equal to 1 because both have 2 2 different levels. So, that is 1 for both of them in both the cases and the test statistic also remains the same both the cases same, but the value of the lambda criterion that is changing, the lambda star in the first case is lambda a star while in the case of H naught 3 its lambda B star ok.

So, use lambda star for H naught 2, and lambda B star in H naught 3, so this is the sort of the data setup at a bigger setup, where we can apply the 2 way manova technique and bring some bring out some significant conclusion about the about this chemical process. So, this sort and we have not taken in to account the actual data, it is going to be little cumbersome though the data dimensionality is 3, but still, so this can be actually done in any computer software the what is important is to know the theoretical foundation and how to give the interpretation.

So, this with this we close our discussion on manova and our next topic is going to be on multiple correlation coefficient.

(Refer Slide Time: 21:25)



So, that is our next topic Multiple Correlation Coefficient, it talks about the degree of association among all p variables, now here essentially we are start we are dealing with the set of p variables the data dimensionality is p, and we are talking about a degree of association among all these p variables, now if you if you think about the usual correlation coefficient it is giving you a pair wise association it is always correlation coefficient between X and y.

So, that is a pair wise association talking about 2 variables only, so it is association between two variables, and if we do the same thing here we will be actually getting number of such correlation coefficients in fact, we will get about p choose two correlation coefficients and still that would not give us the complete picture about the association among all the p variables, so we have to modify our usual definition, I mean you have to take in to account something new now, again if we if you recall the usual definition of correlation coefficient, you see that the data dimensionality there is two we are talking about X and Y, it is Bivariate, now here how do we handle that.

So, in multiple correlation coefficient we say that we are measuring the association between X and say X 1 of them say X 1 say, and the rest of them X 2 to X p that is 1 in 1 hand and p minus 1 on the other, but how do we handle this p minus 1 dimensionality in the other group, well we consider a linear combination of this p minus 1 variables and, so reduce it to the data dimensionality from p minus 1 to 1 and then, we talk about the usual correlation coefficient between X 1 and this linear combination of the p minus 1 variables say some alpha prime X this X contains X 2 to X p.

So, now if this is the situation or 1 may say that well from p choose 2 number of correlation coefficients, we are now talking about an infinite number of correlation coefficients, because there can be an infinite such choices of linear combinations, so what is the multiple correlation coefficient between this X 1 and the other variables, so what we do is, we maximize this simple correlation coefficient maximizing over the choice of alpha. So, now, we have handled the problem of data dimensionality we are essentially talking of two dimensional case, now X 1 and a linear combination of any number of variables.

So, what we have is two variables, but that linear combination factor is also taken care of by considering the maximum correlation coefficient of these two variables, so let us formally define multiple correlation coefficient we say that; before that let us just give the introduction to the setup. So, suppose X is a random variable now with dispersion matrix the mean vector does not have much role to play here, so we have dispersion matrix sigma now partition X the dimensionality is p and sigma as I am saying that X has suppose it has X 1 and X 2.

Now, this is a scalar while this is a p minus 1 dimensional vector, now in this X 1 I can have anything, I can have X 1 or X 2 anyone of the p variables right. So, instead of writing just X 1 I prefer to write 1 with within a bracket meaning that this X 1 may contain anyone of the p variables, and consequently the other part the other sub vector p minus 1 dimensional sub vector will contain the remaining. So, if this is X 1 the first variable X 2 will contain X 2 to X p if this is X 2 then this will contain X 1 X 3 to X p and, so on.

Now, then as we have partitioned X similarly we must partition the dispersion matrix, now this is a p p dimensional square matrix and this is going to have the variance say sigma 1 1 strictly speaking, I should write in this way, but for simplicity sake let me write just sigma 1 1 for this. So, it essentially is the variance of the variable which is coming here, if this is X 1 it is sigma 1 1, if this is X 2 it is actually sigma 2 2 and so on.

But instead of writing that, I am just writing sigma 1 1 for this and then, I have sigma 1 2 this is a row vector sigma 1 2, a column vector and a p minus 1 dimensional square matrix sigma 2 2. So, this is setup is for or let us write notation of thus of this setup is for X 1 is equal to X 1 see if I have the first variable as the first member here then, I will write this sigma matrix a sigma 1 1 variance of X 1 sigma 1, 2 this is actually the covariance vector the covariance of X 1 with X 2, X 1 with X 3, X 1 with X p and so on and this is sigma 2 2 is the dispersion matrix for all the remaining p minus 1 variables from X 2 to X p right.

So, that let us write here variance of X 1 is sigma 1 1 covariance of X 1 and X 2 that is sigma 1 2 and the dispersion matrix of X 2 is sigma 2 2 and after that, I write the between again same in a general way, I write between X 1 and X 2 this is a scalar while this is a vector is denoted by or if we do not want to create confusion, let us write X 1 sorry this is a scalar and X 2 well X 2 is again let us just write that this is let us put it in a bracket X 1 say it could be any X I and X 2, so if this is X 1 this is nothing, but X 2 to X p between X 1 and X 2 is denoted by rho 1 dot 2 3 to p.

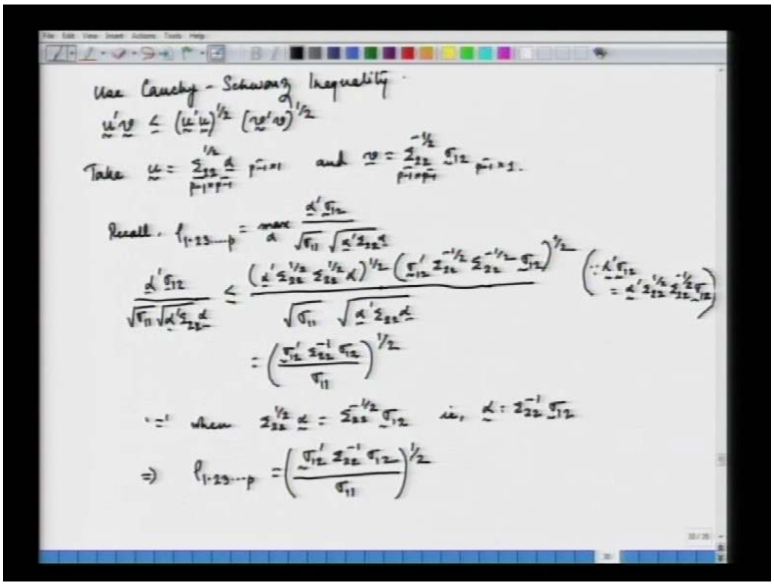
So, this is actually a dot coming after this break up 1 dot and 2 3 up to p, so this are coming a little bit below this point, this is the maximum correlation coefficient between X 1 and any linear combination of a prime X 2 say of X 2 to X p, just what I have said that we consider the correlation coefficient between any one of them and the linear combination of the remaining p minus 1 variables, we consider the maximum over this choice of the scalars, which we are using for the linear combination alpha prime X 2 basically for particular choice of alpha the maximum over the alpha and that gives me the multiple correlation coefficients.

So, our notation is that, is we are using this by rho 1 and 2 3 to p is given by covariance, then what comes is the usual definition of correlation coefficients, so we have covariance between X 1 and the next scalar valued variable that is alpha prime X 2 not a problem, and then what we have is standard deviation root of variance of X 1 times root variance of alpha prime X 2, this also comes in a bracket, so we have a typical Bivariate correlation here we have reduced the dimensionality.

And then, what we have to consider we are not left any space for this, so this is rho 1 dot 2 3 to p this equals maximum of this simple correlation coefficient the maximum being taken over alpha, alpha is having the same dimensionality as X 2 bracket that is p minus 1 vector can take the maximum over any choice of such vector for this gives the multiple correlation coefficient. What is the covariance between X 1 and alpha prime X 2 according to our notation? So, basically we are going to consider maximum over alpha the covariance between X 1 and alpha prime X 2.

So, we have to look at the sigma matrix, so what is this according to this notation again we come back to this notation and this is nothing, but we have alpha prime, and then covariance between X 1 and X 2 which is sigma 1 2 variance of X 1 is root of sigma 1 1 and variance of alpha prime X 2 is nothing, but alpha prime sigma 2 2 alpha. So, we have to essentially maximize this correlation coefficient this measure of correlation coefficient with respect to this alpha which is coming here as well as its coming here.

(Refer Slide Time: 34:11)



So, for this purpose we are going to use a very common inequality, used several times since statistical analysis mainly the Cauchy Schwarz inequality, use Cauchy Schwarz inequality given by in the vector form. You have, if you have two vectors u and v, I have the scalar product is less than or equal to product of the norms, and we have u transpose u raise to the power half similarly, v transpose v raise to the power half. So, this is the one we are going to use and what we do is take for typical particular choice of u and v. So, it is almost apparent from the expression of the correlation coefficient what we have written.

So, we take u as sigma 2 2 half and then we have alpha both it is not a problem of conformability, because we have a p minus 1 dimensional matrix here and a p minus 1 dimensional vector here, so no problem and v as sigma 2 2 minus half with sigma 1 2. So, again no problem this is p minus 1 dimensional square matrix, this is also p minus 1 dimensional vector. Note that these square root and the inverse square root matrices are defined, since sigma 2 2 is a dispersion matrix and we can use the spectral decomposition of sigma 2 2 to obtain what is sigma 2 2 the square root matrix of sigma 2 2 and the inverse of square root of sigma 2 2.

Because, we have non zero in fact, we have positive Eigen values for sigma 2 2, now with this choice of u and v we directly apply the Cauchy Schwarz inequality to get we have them from the expression I have what was it? So, it is rho 1 dot 2 3 up to p and this was max of over alpha, alpha transpose sigma 1 2 and then we have root of sigma 1 1, alpha transpose sigma 2 2 alpha, so just recall this, so by definition we have got this the multiple correlation coefficient, so what we consider here is alpha prime sigma 1 2 and try to give an upper bound to that using the Cauchy Schwarz inequality.

So, we have root sigma 1 this is less than or equal to because, what we have here essentially with this choice of, this choices of u and v what we have here is essentially, u transpose v and then this is less than equal to u transpose u, which gives me alpha transpose sigma 2 2 square root and then root of that and we also have v transpose v and root of that, so we have here sigma 2 2 square root inverse of that and sigma 1 2, so this is u transpose u, we have 2 write this again with its better to write the full form right with the half notation.

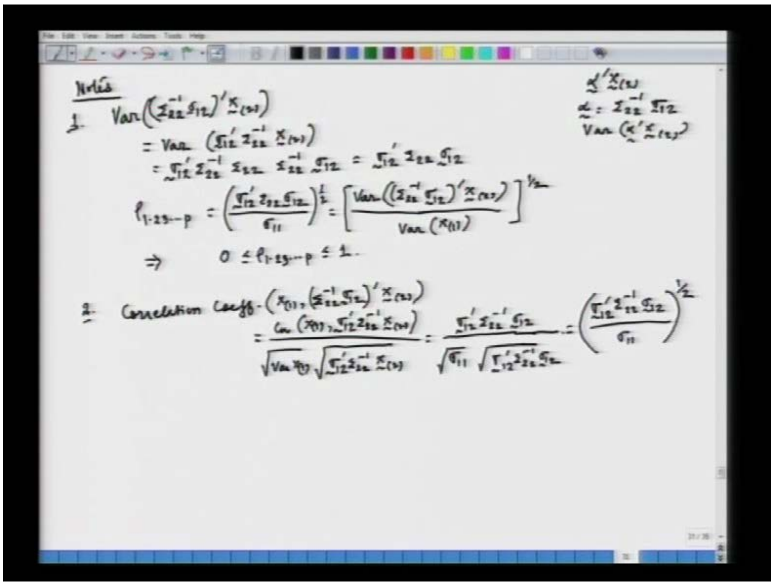
So, that we have alpha transpose sigma 2 2 half u transpose u sigma 2 2 half and alpha this with the half, and then we have v transpose v, so that is sigma 1 2 half sigma 2 2 minus half these are symmetric matrices, so I am just repeating the matrices without the transpose sigma 1 2 this also have a root and this is as it is alpha transpose sigma 2 2 alpha, so we are going to reduce this to the form to get the upper bound of the multiple correlation coefficient. Now, note that this is due to the fact, because we have since alpha prime sigma 1 2 is nothing, but what we are doing is bringing in this sigma 2 2 half and sigma 2 2 minus half matrices in to the picture to use the Cauchy Schwarz inequality in this form.

So, this has been replaced by this first, and then this the inequality has been used now what we have at this stage after simplification is nothing, but note that this alpha this variance term actually cancels out with this and what remains here is sigma 1 2 transpose with sigma 2 2 inverse defines, so not a problem and sigma 1 2 and I have a sigma 1 1 this goes away cancels away with this whole thing raise to the power half, so equality when, there is equality in the Cauchy Schwarz inequality situation.

So, when this when u is equal to v, so v have then sigma 2 2 half and alpha is equal to sigma 2 2 minus half and sigma 1 2 right that is, I have a choice of alpha this alpha is nothing, but sigma 2 2 inverse sigma 1 2. So, this particular choice of alpha gives me the maximum correlation coefficient, and the value of the maximum correlation coefficient is given finally, by this expression which you see is free of alpha, so not a the problem this is the value of the correlation giving me.

So, implying rho 1 2 3 to p this equal to because that is the maximum value of this, so that is sigma 1 2 sigma 2 2 inverse sigma 1 2 sigma 1 1 right, now we know that the usual correlation coefficient between X and y it always lies between minus and plus 1 let us see what happens to the multiple correlation coefficient does it have the same bounds or is it something different.

(Refer Slide Time: 41:53)



So, for that we Note some things, so we are few important motes after the definition the 1st one, note that if you consider the values of the variables scalar variable, now what we have what we are started with was a linear combination of the remaining variables.

So, we have started with some alpha prime X 2 and the choice of alpha that we get that we got for maximizing the correlation coefficient was sigma 2 2 inverse X 2 right. So, if I consider variance of alpha prime X 2 for this choice of alpha what do I get? So, essentially I am considering variance of sigma 2 2 inverse sigma sorry that is not X 2, but sigma 1 2 right, so that is sigma 1 2 transpose, so this is the particular choice of alpha transpose X 2, now this variance is nothing but I have to consider the dispersion of this.

So, this is basically variance of sigma 1 2 transpose sigma 2 2 inverse transpose that is transpose inverse of basically, we are getting inverse only and X 2, so that is giving me sigma 1 2 transpose sigma 2 2 inverse dispersion matrix for X 2 that is sigma 2 2 and then again transpose of this whole thing, so that is sigma 2 2 inverse and sigma 1 2 which is sigma 1 2 prime sigma 2 2 sigma 1 2. So, what I see is the correlation coefficient, the multiple correlation coefficient 1 dot 2 3 to p is nothing, but values of which is again recall which is this 1 1 sigma 1 1 root half is actually the root over of the ratio of 2 variances the first variances in the numerator, it is the variance of sigma 2 2 inverse sigma 1 2 transpose X 2 and the denominator is variance of the separated out variable X 1 raise to the power half.

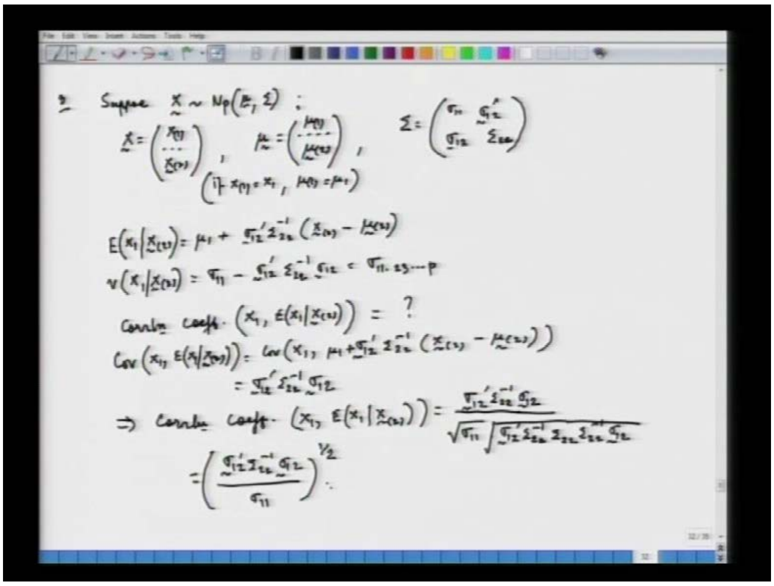
So, this implies that although the usual correlation coefficient lies between minus 1 and plus 1, now although we have used more or less the same techniques, the same thought processes to obtain the multiple correlation coefficients still the multiple correlation coefficient lies between 0 and 1 it cannot be negative, and it is lying between the other part is we are getting another sharper bound for this, so giving me the value of rho 1, 2, 3 to p. This part the other part of the inequality is quite obvious again it is it comes from the Cauchy Schwarz inequality whichever, applies for the simple correlation coefficient the same logic will apply here to get this part of the bound.

But for this part we are using this the reasoning which we have just stated, because it is the ratio of 2 standard deviations essentially after the bound, let us talk about some interpretation of the some other interpretation of the multiple correlation coefficient suppose you have let us consider the correlation coefficient between 2 scalar random variables, so that is X 1 and now we are basically considering the linear combination that this that is my second variable which I have used, so that is first 1 is X 1 the scalar value random variable and the next 1 is alpha prime X 2, but with the typical choice of alpha of this particular choice of alpha, so that is sigma 2 2 inverse sigma 1 2 transpose X 2.

So, I am considering correlation coefficient between these 2 variables and this is giving me essentially then, what is this I will have covariance between X 1, and covariance between X 1 and sigma 1 2 transpose sigma 2 2 inverse X 2 and here I have variance of X 1 with variance of sigma 1 2 transpose sigma 2 2 inverse X 2, I have already obtained this expressions, so this was sigma 1 1 and I saw that this variance is nothing, but the root of root of this variance it is sigma 1 2 transpose sigma 2 2 inverse sigma 1 2 and this is nothing, but sigma 1 2 transpose sigma 2 2 inverse and covariance between X 1 and X 2 that sigma 1 2.

And I am getting the multiple correlation coefficient between X 1, and the other remaining p minus 1 variables, so that sigma 1 2 and sigma 1 2 here and sigma 1 1 here root of this. So, basically the multiple correlation coefficient that I have obtained is nothing, but the simple correlation coefficient between X 1, and this linear combination of the other p minus 1 variables namely it is the simple correlation coefficient between X 1 and sigma 1 2 transpose sigma 2 2 inverse X 2 where this matrices this parameters have the usual interpretation, this will vary accordingly as I vary my basket of X 2 variables, so that I get a different X I consider a different X I each time the other basket will change and similarly consequently the composition of the dispersion matrix the original dispersion matrix sigma will change, so I will have sigma I either first element and in this way, but this can be handled very easily.

(Refer Slide Time: 49:21)



The third point that we consider is suppose, I have X is following the p variate normal distribution mean mu and dispersion matrix sigma, I have mu also partitioned now as mu my partition of the X vector the data random vector remains the same as before I consider a single 1 here X 1 and X 2, so accordingly nu is nu 1 and the rest of it and sigma is sigma 1 1, sigma 1 2 and sigma 2 2.

So obviously, if X 1 is nothing but the first random variable X 1, then mu 1 is also equal to mu 1 and this holds for any I and accordingly this also is changing. So, if this is the situation what is the additional here is this distributional assumption of multivariate normality and we have already seen that the conditional expectation of X 1 given X 2. Now, let us specifically consider this the first random variable as X 1, so I am writing it without the bracket, so this X 2 contains the usual X 2 2 X p, so this conditional expectation is nothing, but mu 1 plus sigma 1 2 transpose sigma 2 2 inverse than X 2 minus mu 2.

If you recall our earlier discussions on multivariate normal distribution, this will come immediately and variance the conditional variance of X 1 given X 2 is nothing, but sigma 1 1 minus sigma 1 2 sigma 2 2 inverse sigma 1 2, and let us use a special notation for this conditional variance. Let us write this as sigma 1 1 dot 2 3 up to p, and then if I consider the correlation coefficient between X 1 and the second random variable then I consider here is a particular choice of X 2, but that is not the usual choice of X 2 that we have been considering all along till now, but I the second variable is in fact, here that I consider is the conditional expectation of X 1 given X 2.

Note that this conditional expectation is nothing, but a linear combination of X 2, so let me consider the correlation coefficient between X 1 and this second variable which is expectation of X 1 given X 2 that is another linear combination of X 2, now what is this, this is giving me put a question mark here, I am trying to find what is this equal to, so for this first I have to look in to the covariance of between these 2 variables, so that is X 1 expectation X 1 given X 2, let me write what is this conditional expectation specifically.

So, that is mu 1 plus sigma 1 2 transpose sigma 2 2 inverse X 2 minus mu 2 right and this gives me for the first term X 1 and mu 1 that 0, because this mu 1 being a constant, so what will I have is for the second term that is X 1 with this term sigma 1 2 transpose sigma 2 2 inverse and, so on. So, that gives me sigma 1 2 transpose sigma 2 2 inverse note that I am considering the covariance of X 1 with this expression again if, I whether I consider it with a whole thing or just this thing its immaterial because even if I separate out this mu 2 factor again, I am going to get a zero covariance because this part will consist of only constant terms.

So, I have the dispersion matrix of X 2 only and that is sigma 2 2 and I have sigma 2 2 inverse and sigma 1 2 here sorry this is not in this way, but what I am is sigma 1 2, I am considering the covariance between X 1 and sigma 1 2 transpose with sigma 2 2 inverse X 2, so I have sigma 1 2 inverse with sigma 2 sigma 1 2 transpose sigma 2 2 inverse and this covariance. So, that sigma 1 2 basically the same thing is coming, but more easily.

And what I have next is implying that correlation coefficient in that case is equal to correlation coefficient between X 1 and the conditional expectation of X 1 given X 2 and that is equal to the covariance is sigma 1 2 transpose sigma 2 2 inverse sigma 1 2 and the variance is sigma 1 1 and the dispersion of this matrix which is nothing, but we are considering the dispersion of this the test variable for mu 1, I have no contribution at all and then this part is giving me sigma 1 2 transpose sigma 2 2 inverse sigma 2 2 and sigma 2 2 inverse with sigma 1 2.

So, that is being equal to nothing, but the it is the usual correlation coefficient between 1 and 2 2 p variables my usual correlation coefficient, so you are going to continue we have seen for the multivariate normal case the multiple correlation coefficient has an extra interpretation that it is the correlation coefficient between X 1 and the conditional expectation of X 1 given X 2 with this, we are also we are arriving at the same multiple correlation coefficient between X 1 and X 2 to X p, so we should continue from this.