Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Sharmishtha Mitra

Department of Mathematics and Statistics

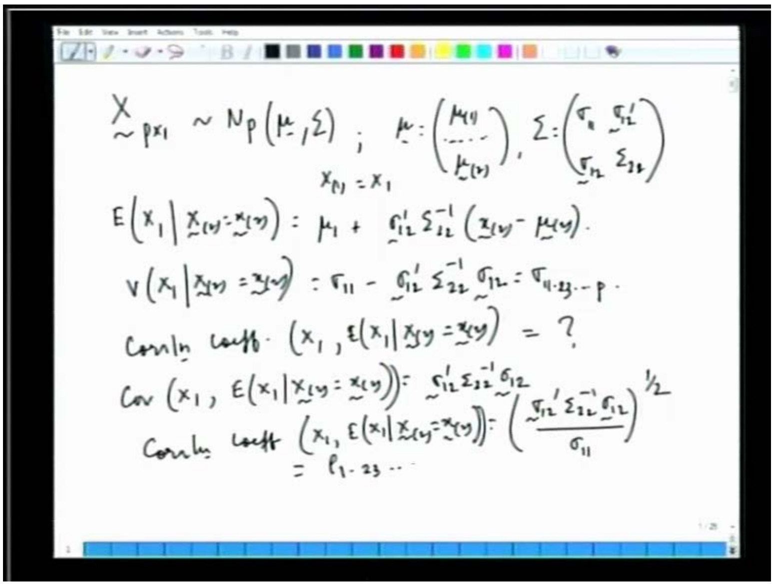
Indian Institute of Technology, Kanpur

Lecture No. # 22

Multiple Correlation Coefficient

After we had defined, what is multiple correlation coefficient recall that the setup, we had taken was totally distribution free and in a very general way, we had defined what is multiple correlation coefficient, we have talked about its bounds and we have given one or two interpretation of it. Now, after that we had talked about in the case of multivariate normal distribution what extra interpretation can be given to this multiple correlation coefficient.

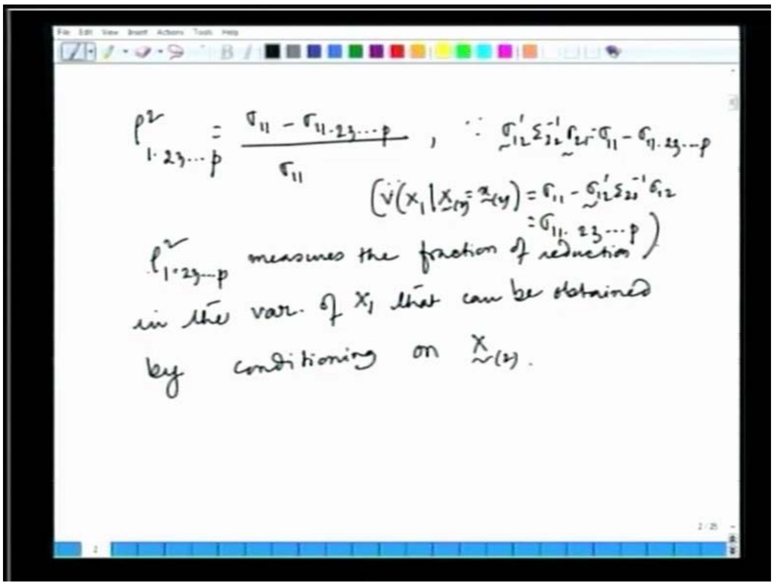
(Refer Slide Time: 00:51)



In this sense we had considered the p variate random vector which was following a multivariate normal distribution with mean mu and dispersion matrix sigma of course, as we would partition X we would have to partition mu and sigma. So, that mu was having a one component here and the rest of the p minus 1 elements coming in mu 2 and similarly we had a sigma matrix partitioned in the form of sigma 1 1 sigma 1 2 vector and then the p minus 1 ordered square matrix. We had said that if I consider the conditional expectation of X 1 let us, without loss of generality we take X 1 as the first random variable X 1. So, that I consider the conditional expectation of the first member of the partitioned vector and then conditioned on X 2 and, if you recall this was nothing but mu 1 plus sigma 1 2 this is from our earlier results and multivariate normal and this was then x 2 minus mu 2 and what was the conditional variance of X 1 given X 2 was nothing but sigma 1 1 minus sigma 1 2 sigma 2 2 inverse and sigma 1 2 and we are giving a special notation to this and we are calling it sigma 1 1 dot 2 3 to p and we had taken up the calculation of correlation coefficient between X 1 and this expectation of its conditional expectation conditioned on X 2 residual part of the random vector.

And then we had seen that, if I consider the numerator that is X 1 covariance between X 1 and conditional expectation of X 1 given X 2 this was coming to b we need to put a second bracket here and this was covariance was simply sigma 1 2 prime sigma 2 2 inverse and sigma 1 2. This fact lead us to the fact that correlation coefficient one that we were interested in this correlation coefficient between X 1 and the conditional expectation of X 1 conditioned on X 2 was nothing but probably we had come up to this as well, because the variance of this is nothing but this factor when we take the variance of this expectation again we get back this type of this same factors. So, with the root of it what we have is sigma 1 2 sigma 2 2 inverse sigma 1 2 with sigma 1 1 and raised to the square root. So, this is nothing but the usual multiple correlation coefficient that is rho 1 dot 2 3 to p.

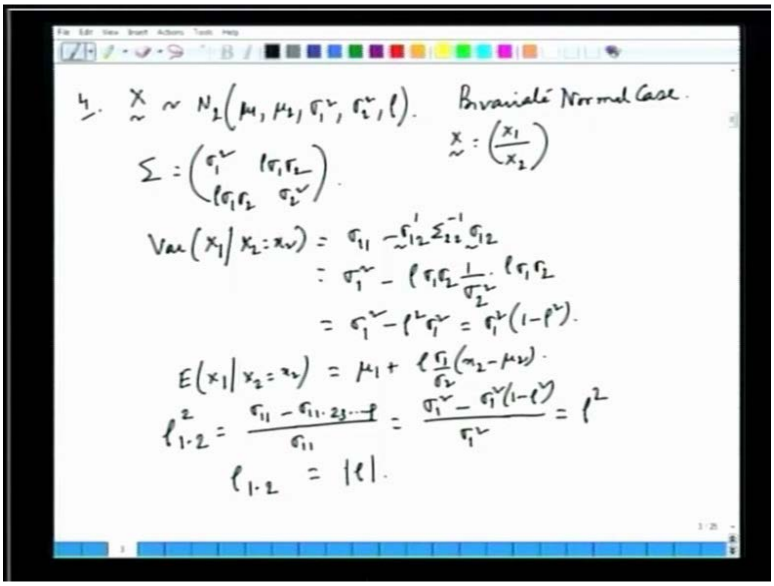
(Refer Slide Time: 04:35)



So, that I can say that rho square of rho 1 2 3 2 p is nothing but sigma 1 1 minus the conditional variance with the notation sigma 1 1 dot 2 3 2 p and divided by the usual variance sigma 1 1 this is coming from the fact that sigma 1 2 sigma 2 2 inverse sigma 2 1 this was nothing but sigma 1 1 minus sigma 1 1 dot 2 3 2 p. Why if you recall that, we had also seen with the conditional expectation, we are also seen the conditional variance of X 1 given X 2 and this was nothing but this was my sigma 1 1 minus sigma 1 2 sigma 2 2 inverse sigma 1 2 with this notation. So, because of this fact what we have is the square of the multiple correlation coefficients is the unconditional variance of X 1 minus the conditional variance of X 1 conditioned on X 2 divided by the unconditional variance.

So, what we say is that the square of the correlation coefficient in the case of the multivariate normal distribution is giving me a interpretation like that it measures rho 1 to 3 dot p square of it measures the fraction of reduction in the variance that is the unconditional variance of X 1 that can be the reduction this reduction when can it be obtained when I condition this X 1 on the given information of X 2 that can be obtained by conditioning on X 2. So, that is the new interpretation that we can give when we have some distributional assumption namely the random vector follows a p variate multivariate normal distribution with some covariance matrix dispersion matrix sigma. So, this is one other interpretation and of course, if you recall the bivariate normal distribution.

(Refer Slide Time: 07:28)

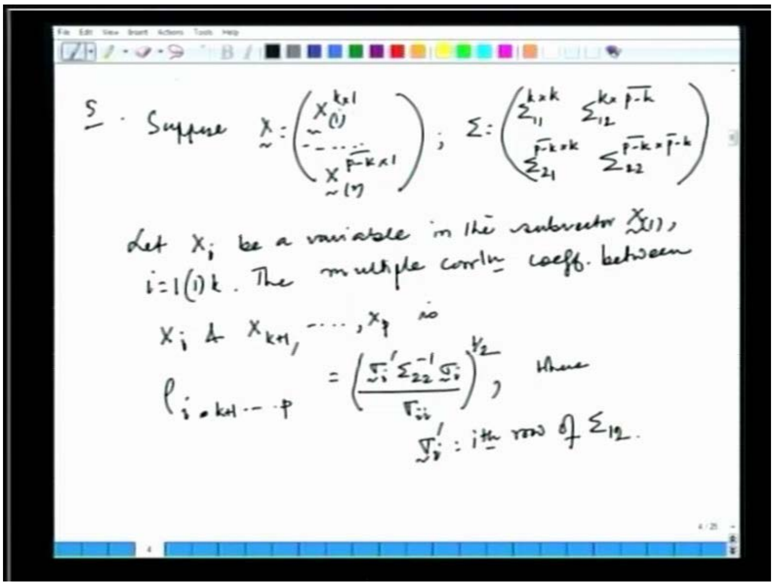


So, that we have X is now, a two- dimensional random vector. So, let this be our forth point to be noted and this is we have the bivariate normal case X following bivariate normal with the usual notation. Let us write mu 1 mu 2 sigma 1 square sigma 2 square and rho bivariate normal case. Now, here I will I have to consider a partition of X vector as X 1 and X 2 that is in the case I am considering multiple correlation coefficient of X 1 given X 2 now, it is basically a bivariate correlation coefficient type and in sigma matrix we have to consider a partitioning of the type, what is the usual sigma matrix for the sigma 1 1 I have only varying means of X 1. So, it is sigma 1 square according to my notation then it is covariance between X 1 X 2 that is rho sigma 1 sigma 2 and then here will the remaining part is again a one- dimensional case. So, I have sigma 2 square.

This is the situation what is variance of X 1 conditioned on X 2 well, it is if I just go back to my earlier formula it is nothing but sigma 1 1 sigma 1 2 sigma 2 2 inverse and sigma 1 2. Now, in this case what is the sigma 1 1 is sigma 1 square this is rho sigma 1 sigma 2 1 by sigma 2 square and then again I have rho sigma 1 sigma 2 this is giving me sigma 1 square and you have rho square sigma 1 square that is, what we are looking for this is sigma 1 square 1 minus rho square. If you recall in our bivariate normal well we have not done bivariate normal case separately, but then this is the conditional variance that we get in the case of bivariate normal and recall that the expectation X 1 given X 2 is nothing, but mu 1 plus rho sigma 1 by sigma 2 and X 2 minus mu 2.

So, that what is rho 1 dot 2 square well again what we had seen earlier that it is sigma 1 1 minus sigma 1 1 dot 2 3 to p by sigma 1 1. So, in this notation this is sigma 1 square minus this variance, the conditional variance which is sigma 1 square 1 minus rho square and then I have sigma 1 square this is leading mu to rho square which is the square of the correlation between X 1 and X 2. So, the multiple correlation coefficient here is actually the bivariate correlation coefficient, the usual correlation coefficient between X 1 and X 2. So, that this is I can simply write that the multiple correlation coefficient rho 1 dot 2 is nothing but the mod of the usual correlation coefficient, the square being equal to the square. So, I can have of course, this rho correlation coefficient the usual correlation coefficient can lie between minus 1 and 1, but the multiple correlation coefficient is always between 0 and one. So, this is the modulus of the usual correlation coefficient between X 1 and X 2.

(Refer Slide Time: 11:39)



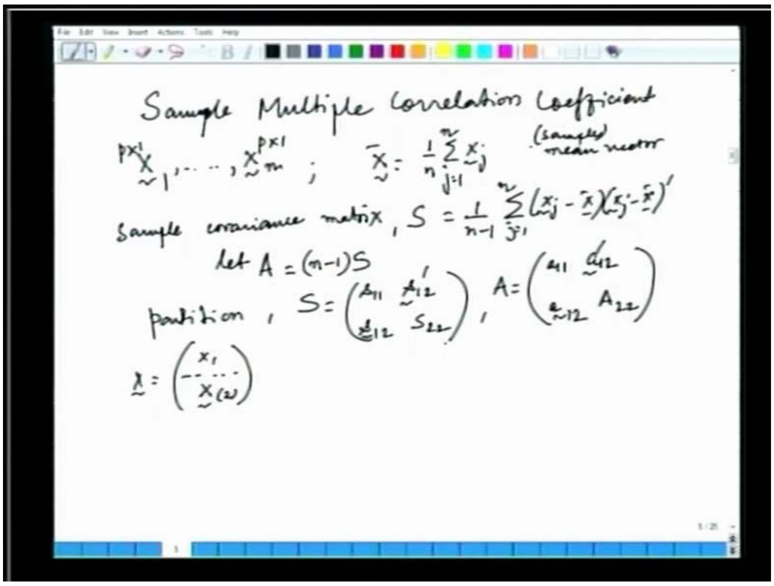
So, this is one way I can think about the bivariate normal case and after this a very small point that is if you have. Suppose till now, we were considering a partition of the random vector where there was only one element in the first sub vector. Now, if we partition the random vector in a way. So, that I have a sub vector initially which is of dimension k and the remaining how many p minus k coming in the next sub vector. So, that I have to talk about similarly partitioning the variance covariance matrix. So, this is now, going to be partitioned in this form. So, at the first element is no longer a scalar it is again a matrix.

So, let us call it sigma 1 1 it has to be of dimension k, because this is talking about the variance covariance matrix of all the k variables that are included in the first sub vector X 1 and then I have sigma 1 2 which is talking about the covariance between the elements of this X 1 and X 2. So, that this has to have dimension k cross p minus k the transpose of this matrix here and this is p minus k cross k and lastly the variance covariance matrix of the p minus k random variables coming in X 2. So, this is a square matrix of dimension p minus k. So, this is how I am partitioning the variance covariance matrix when the random vector is partitioned in this way and then I can think of the multiple correlation coefficient between any one element of this X 1 conditioned on the elements of the X 2. Let me say that let X i be a variable in the sub vector X 1. So, obviously, I can be any one of the k elements that are members of X 1. So, i goes from 1 to k and then I say the multiple correlation coefficient between X i and the elements of X 2.

Let them be without loss of generality let them be the last p minus k variable this. So, that is k plus 1 upto p we should write it X i with X k plus 1 to X p. So, let us upto X p is the notation is rho i dot k plus 1 upto p let me put a bigger dot for this and this is nothing but sigma i transpose sigma 2 2 inverse sigma i by sigma i i. So, obviously, we are talking about here the variance of X i and this vector sigma i is nothing but the we are talking about a particular k rho of this sigma 1 2 matrix. So, this is nothing but, where the rho, where covariance between X i and the other variables that is k plus 1 to p these are considered. So, where X i say if its transpose it is nothing but the i th rho of since we are saying it is i th rho. So, let us say it is of sigma 1 2.

This is how I can simply give the multiple correlation coefficient when not necessarily the first sub vector is a scalar it can also be a random vector it is just like here we have partitioned it in this case. So, that there are k variables in the first vector and I consider any one of the elements from here and consider it is multiple correlation coefficient with the members of the other sub vector of the remaining variables. So, it is just a case of properly handling the covariance matrix and then considering which row we should take up from this covariance part that is sigma 1 2 rest of it remains same. We will have to consider sigma 2 to inverse and here also we will have to consider the scalar which is the variance of that i th the variable which we are considering.

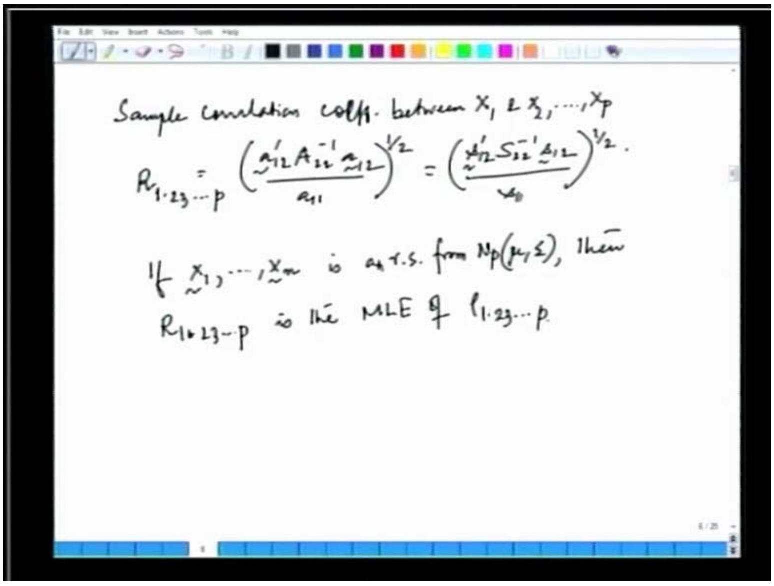
(Refer Slide Time: 16:41)



Let us now, talk about the sample correlation coefficient, if you have a random sample multivariate data. So, that sample multiple correlation coefficient how do we calculate it. So, that is and now, we have what we have is a random sample of size n say of multivariate data. So, each of them is the same p variate random variables and we have a Sample of size n X 1 X 2 to X n and we are going to calculate the multiple correlation coefficient from the given data. So, what we do is, we simply consider again the mean vector X bar which is nothing but 1 by n we consider the mean over the observations of over the n observation and for each of this X 1 and in this way we get the mean. So, to complete the mean vector the sample mean vector and similarly we also have the (( )) this is the mean vector and let us put the word sample in bracket and sample covariance matrix, we have the notation S for this and let our this is actually 1 by n minus 1 with divisor n minus 1.

So, that I have X j minus X bar X j minus X bar k j from 1 to n and then we define some a matrix let A be this 1 with this deviser coming with S. So, that is n minus 1 S. So, that I simply consider this type of matrix sum of X j minus X bar X j minus X bar transpose and then I will partition well I consider the partitioning of S. So, that I write this is S 1 1 S 1 2 this like the way we handled sigma S 1 2 and S 2 2 similarly, I can have a partitioning of A also not a problem. So, I have here A 1 1 say the vector A 1 2 prime A 1 2 and the matrix A 2 2. So, it is obvious that I am partitioning X into the form that I have one element here say X 1 again without loss of generality and the remaining p minus 1 here vector X 2. So, that is the S and A matrices have been partitioned in this way.

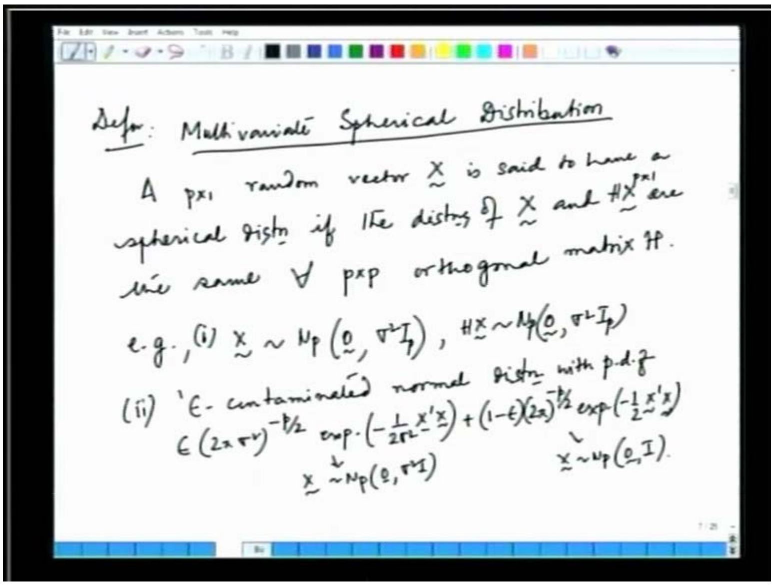
(Refer Slide Time: 20:01)



And then we can define sample correlation coefficient between X 1 and X 2 to X p note that these are again symbolic this X 1 can be any X i and the only thing that will result is the absence of this X i from this p minus 1 remaining variables. So, this I will write as sample correlation coefficient between X 1 and X 2 to X p is nothing but our usual R 1 2 now, the sample notations. So, the (( )) been replaced by the letter R and we have this as A 1 to transpose A 2 2 inverse A 1 2 by A 1 1 root. Note that, if I write with the elements of the S matrix then this does not change and this is exactly, because of the relationship between A and S. So, this is S 1 2 S 2 2 inverse S 1 2 and then I have S 1 1 leads to the power half. One thing can be noted from here that if X 1 to now, we are always interested in knowing what is happening if these random variables are coming from the normal population. So, if this is R random sample from p variate normal distribution normal p mu sigma then this R 1 dot 2 3 p is the MLE maximum likelihood estimator of the multiple the population multiple correlation coefficient rho 1 dot 2 3 p.

Now, that we are talking of the sample correlation coefficient, we can actually think of devising tests on the population correlation coefficient rho 1 dot 2 3 p. So, for this we have to know something about the sampling distribution of this sample correlation coefficient or some function of this sample correlation coefficient, if you recall the usual bivariate case you can think that the you may recall that the testing procedure is devised on a function of the correlation coefficient R and it is of the form R square by root of 1 minus R square with some constant and this following a t distribution. So, let us see what is happening here and, if this generalization leads to the case of the test procedure on the simple correlation coefficient. Now, before we talk about the sampling distribution of this R or some function thereof we have to talk about 2 very special type of distributions the multivariate spherical distribution and the elliptical distribution.

(Refer Slide Time: 23:31)

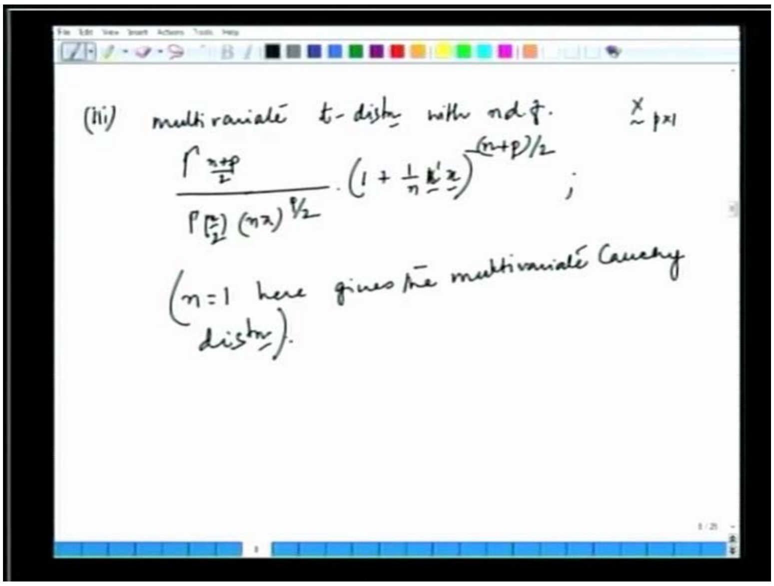


So, we first define, what is the multivariate spherical distribution the definition is of multivariate spherical distribution, we say that a p-dimensional random vector X is said to have a spherical distribution obviously, in this case the multivariate spherical distribution, if the distribution of X and H X the distributions of X and H X are the same for all orthogonal for all p by p obviously, we are considering a square matrix this being p plus 1.So, we have for all p cross p orthogonal matrices for all such type of orthogonal matrix H. Example well you can very easily think of one example that is the multivariate normal distribution X following normal p with 0 see sigma square i. So, that the we have no problem of correlation if the only the variance are present and that to the homoscedastic case. So, that the variances are each of X 1 X 2 to X p is sigma square. So, that I have the sigma matrix reduced to a diagonal matrix sigma square i and I say that X is following a normal with mean null vector and variance covariance matrix sigma square i.

Now, if H is an orthogonal matrix the distribution of H X from our earlier results on multivariate we know that, this also has to be p variate normal with mean a mu, if you recall what was the distribution of a X when X is following multivariate normal this is coming from that. So, this mean will be H mu which is null and sigma square H transpose H as the variance covariance matrix now, H being an orthogonal matrix is actually giving back to sigma square and this is i p cases obviously. So, we have this X particularly following this type of multivariate normal distribution this is a case of a multivariate spherical distribution secondly, if you think of. So, this is the first one there is distribution called the special type of distribution E contaminated normal distribution with p. d. f type E times 2 pi sigma square to the power minus p by 2 and then I have the exponent term that is 1 minus 1 by 2 sigma square X transpose X and then comes a term like 1 minus E times 2 pi to the power minus p by 2 and the exponent term minus half X transpose X.

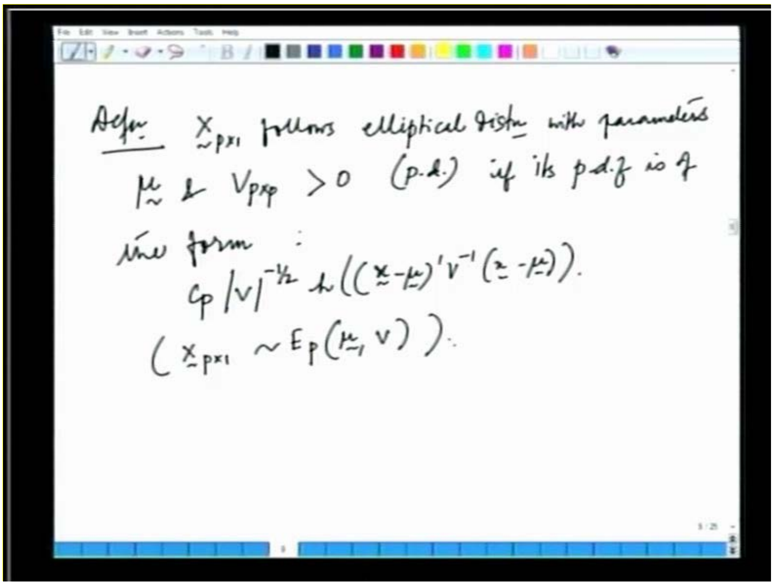
Now, note that this part is actually coming from X following multivariate normal the one, we have talked about just now with mean 0 and dispersion matrix sigma i here this is even simpler I have a case of multivariate normal from 0 i and a combination of these with convex combination of this type that is E with and 1 minus E multiplied with this p d f is giving me the p d f of the E contaminated normal distribution and this is also an example of spherical distribution.

(Refer Slide Time: 28:15)



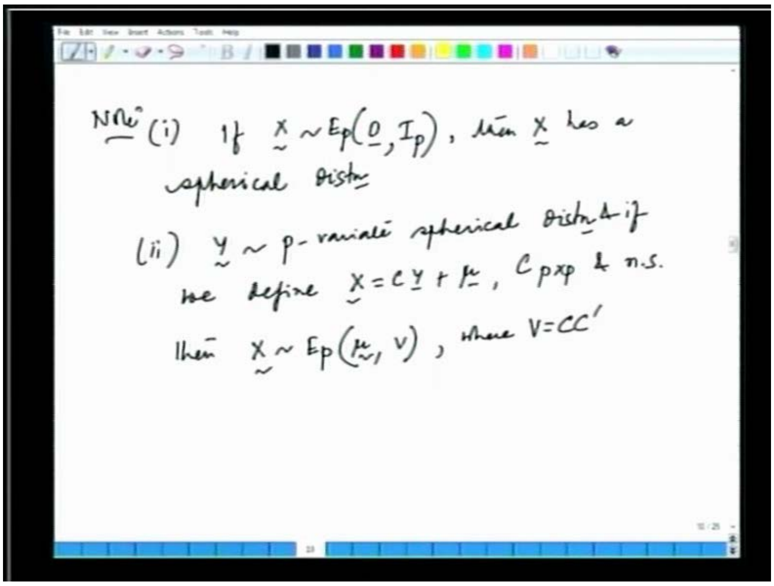
Otherwise, we can also talked about one more this is the multivariate t- distribution with n. d. f and it is having a p. d. f like gamma n plus p by 2 by gamma n by 2 and then we have n pi to the power p by 2 with 1 plus 1 by n X transpose X actually (( )) X transpose X with n plus p by 2 and p comes from the dimension of X is p cross 1 and if you have n equal to one here gives the multivariate Cauchy distribution. So, these are the three examples that I am giving of multivariate spherical distribution. Next we talk about the elliptical distribution and see the connection between the elliptical distribution and the spherical distribution and we must not forget about the main connection that is the connection of all these with the distribution of the sample multiple correlation coefficient.

(Refer Slide Time: 30:00)



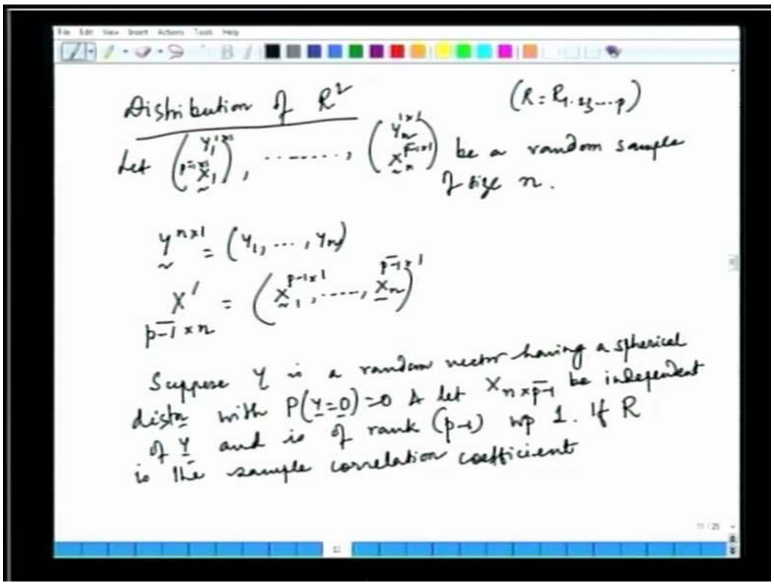
So, the definition of elliptical distribution is we will say that X a random vector p variate random vector X is following follows elliptical distribution with parameters mu and say a V matrix this is like the covariance matrix V square matrix of dimensional p, a positive definite, the notation we are using for positive definiteness. So, we say the random vector follows an elliptical distribution with parameters mu and V if its p. d. f is of the form C p. Let us write it in the next line C p determinant of V raise to the power minus half and then we have a function of this type X minus mu transpose V inverse X minus mu and the notation the symbol that we say is X is following E p with parameters mu and V.

(Refer Slide Time: 31:36)



And if note that if X follows E p with mu replaced by the null vector and I p V replaced by the identity matrix sort of standard elliptical distribution mu being replaced by null vector and sigma replaced V replaced by I p then in this case, X has a spherical distribution. Second point is that if Y is a coming from a p variate spherical distributions. So, if Y satisfies the condition of spherical distribution. So, it follows p variates spherical distribution and, if we define linear transformation like X which is C times Y plus mu where C is p by p and non -singular. So, this also remains the p-dimensional vector X and in that case X will follow an elliptical distribution of dimension p C being p cross p with parameters mu and v, but where V is it has to be in terms of C it is actually C transpose. So, we have define multivariate spherical distribution and the elliptical distribution and shown the connection between the elliptical distribution and the spherical distribution. Now, we go back to our sample correlation coefficient and look into the distribution of not R, but R square.

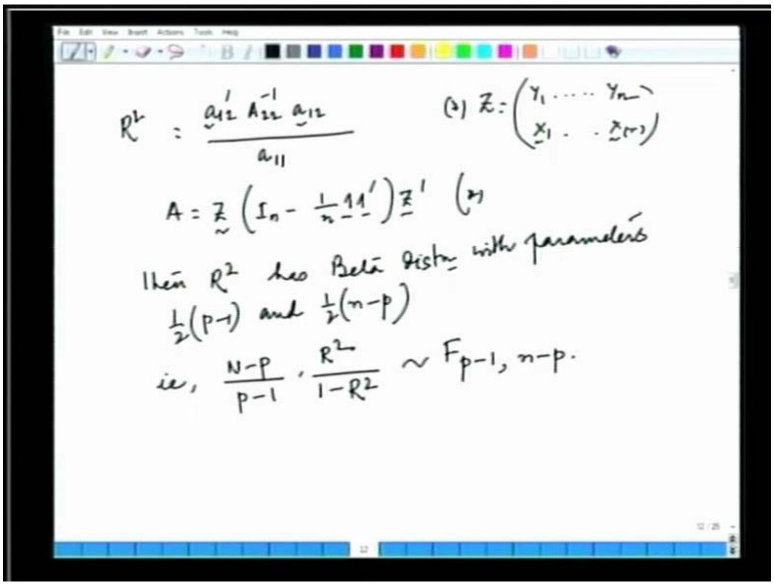
(Refer Slide Time: 33:48)



So, we are now, talking about after having defined this we talk about the distribution of R square. Let us write R for R 1 to the multiple the sample multiple correlation coefficient. So, I am simply writing R for this again this symbolic this is this can be sample multiple correlation coefficient between any member of X 1 X 2 to X p and the remaining. So, distribution of R square is we talk of let Y augmented with X 1 then we have the sample. So, that is Y 1 to X 1 and we have Y n augmented with X n be a random Sample of size n. Let us denote the n-dimensional vector Y n as Y 1 Y 2 to Y n the n observations of the random variable Y and let X the observations on X be arranged in a matrix form. So, let this X transpose be a p minus 1 cross n-dimensional matrix which means basically these are all one dimensional 1 cross 1 and these are all p minus 1 cross 1 dimensional vectors.

So, these elements can be arranged in the way X 1 to X n these are the columns of these matrix each of these having p minus 1 dimension and then we will see that if Y has some distribution. Then it will lead to some distribution of the square of the sample correlation coefficient and in the case of multivariate normal distribution of Y what can be the distribution of the sample correlation coefficient. Now, we are considering the distribution of R square with our background of the definition of spherical distribution and elliptical distribution. Now, we said that from a random sample of this type of size n. We are considering arranging the observation and we have the vector Y and the matrix X transpose, where we have incorporated our full data and then we say that suppose Y is a random vector having a spherical distribution with the probability of this is equal to the null vector being 0 and let X be independent of Y and is of rank. Now, if I say that if R is the sample correlation coefficient, it is between Y and X 1 X 2 to X n.

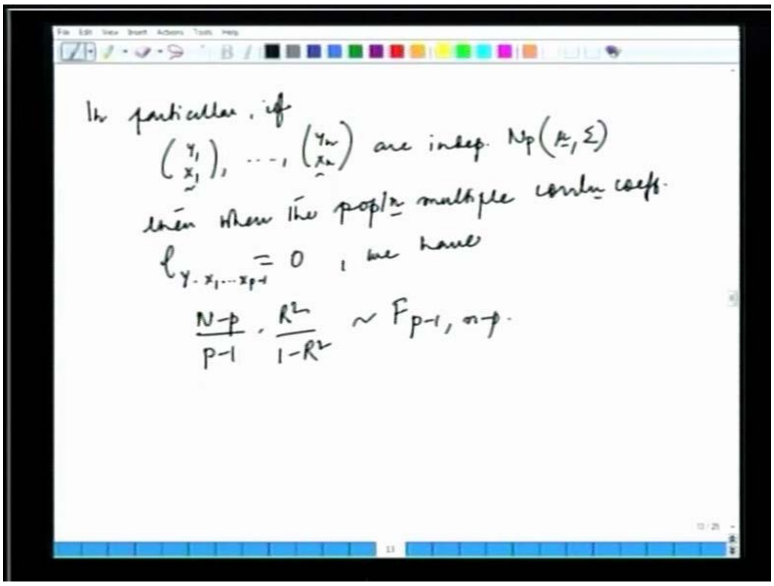
(Refer Slide Time: 38:07)



So, this is now denoted by I have R square as a 1 2 transpose with A 2 2 inverse then I have a 1 2 a 1 1 as we have noted earlier with A matrix how do I form this A matrix. Now, that I have actually the random vector is say it is of sum form Z and how does this Z look like well I have Y 1 to Y n. Let this be the first part, the one variable which has been separated out and then I have X 1 to X n these are p minus 1 dimensional. So, that together we have a p-dimensional case, but now we are using different where notational variables Y for the one dimensional case coming in the first part and then the p minus 1 coming in the lower part which have been denoted by X 1 to X 1 X 2 etcetera. So, this A matrix is the sample variance covariance matrix along with the divisor, how can we express it in terms of Y and X well it is nothing but I have with the Z with I n minus 1 by n 1 1 transpose these are vectors of 1 and this is Z transpose, where Z is defined in this way.

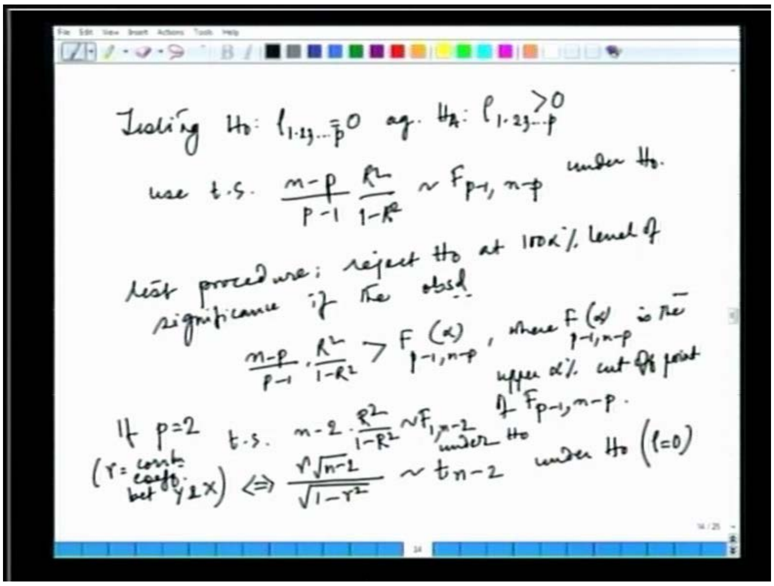
So, this is our Z and then this is the sample correlation coefficient I have and this R square has beta distribution with parameters half of p minus 1 and half of n minus p that is the data dimensionality including Y and this p minus 1 dimensional vectors X and half n minus p n being the sample size or the number of observations on the random variable that is, if this is so by our well known result, we have n minus p by p minus 1 times R square by 1 minus R square thus following and F distribution with degrees of freedom p minus 1 and n minus p. Now, when we talk about the distribution of the sample correlation coefficient or square of eta or some function of this. We have to have some distributional assumption in the background and we saw that we are forming a case like Y the partitioning of the random vector is in this form where we have the random variable Y and the random variable X 1 to X p minus 1 giving in total a p variate random vector and if we have some assumption on the distribution of Y, we say that it follows a spherical distribution actually the Y vector which includes the all the observations on Y 1 to Y n that follows a spherical distribution. Then only we can have a distribution of this type, where R is coming into the picture in this from R square by 1 minus R square times n minus p by p minus 1 is the very well known F distribution with degrees of freedom p minus 1 and N minus p.

(Refer Slide Time: 42:10)



And if we have the additional assumption that in particular if the combine random vector that is if Y 1 X 1 to Y n X n each of them they are independent, if Y these are independent normal p variate normal, because this is one dimensional this is p minus 1 dimensional these are independent normal p mu sigma then when the population correlation coefficient. when the population multiple correlation coefficient rho actually, we can write here in the form rho Y X 1 to X p minus 1 or simple the population coefficient rho 1 2 rho p are usual notation. So, that these are now 2 to p instead of 1. So, this is equal to 0 when this is true, we have N minus p by p minus 1 R square 1 minus R square follows a F distribution with this degrees of freedom. So, this fact can be very well used for the purpose of testing of this multiple population correlation coefficient being equal to zero. When the observations are assumed to be coming from the multivariate normal distribution what we will do is, we will simply say that, we have a test statistic of this form following the F distribution under H naught. So, this can be very well used in the testing of inference for the multiple correlation coefficient.

(Refer Slide Time: 44:22)

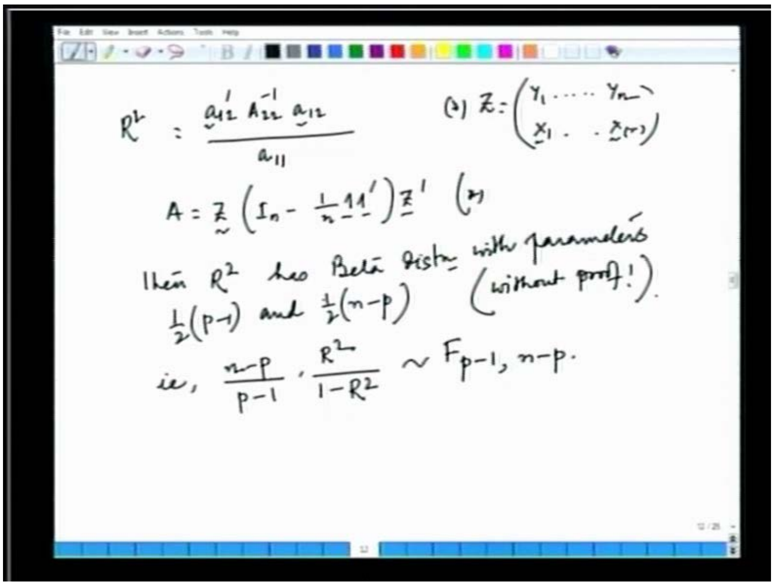


So, when we are saying that this is known to us, then for testing H naught that rho 1 all the notation that we are using now, Y X 1 to X p minus 1 is equal to 0 against the alternative H A that this is greater than zero, because it is lying between 0 and one. So, this is greater than zero. So, we can use test statistic n minus p I have written capital well it is no need to write the N here, it is the n the sample size what about earlier case similar mistake made here. So, this is simply n and we have this following. So, we use the test statistic n minus P P minus 1 R square 1 minus R square following F p minus 1 n minus p under H naught.

So, the test procedure is simply reject H naught at 100 alpha percent level of significance if the observed we need to calculate the sample multiple correlation coefficient and then this factor if the observed test statistic that is in minus P minus 1 R square 1 minus R square is greater than F of p minus 1 n minus p at alpha, where F alpha p minus 1 n minus p is the upper alpha percent cut off point of F distribution with degrees of freedom p minus 1 and n minus p and recall that, if you have if p is equal to two, we have the usual bivariate case and what is this statistic, in that case the test statistic reduces to n minus 2 and this is R square well let us, write R no problem we can write R square for this also R square by 1 minus R square this is going to follow an F distribution with F 1 and n minus 2 under H naught and this is equivalent to saying that root of this that is R root n minus 2. Let us write this R is equal to correlation coefficient between Y and X now, there is no question of X 1 to X 2 to X p minus 1. So, this the two- dimensional case and we have R root n minus 2 by 1 minus root 1 minus R square following a t with n minus 2 degrees of freedom under H naught which is the population correlation coefficient is equal to 0 again this is the usual bivariate correlation coefficient between X and Y Y and X here.

So, we get back the case, where when we consider p equal to two in this general form for p is any number. So that this the generalization works nicely here, and we get back the usual two-dimensional case for p equal to 2. So, after this we have only one thing to be considered, but note one factor, we have not actually derived the distribution of this thing R square by 1 minus R square etcetera or the beta distribution that we are talking about that is what we have is R square has the beta distribution with parameters p minus 1 by 2 and n minus p by 2. This derivation that coming from this assumption of spherical distribution of Y is extremely cumbersome. So, we avoid that derivation what we have done is, we have only defined what is spherical distribution and what is elliptical distribution the connection with the sample distribution of sample correlation coefficient, but this is without prove to be noted.

(Refer Slide Time: 49:33)

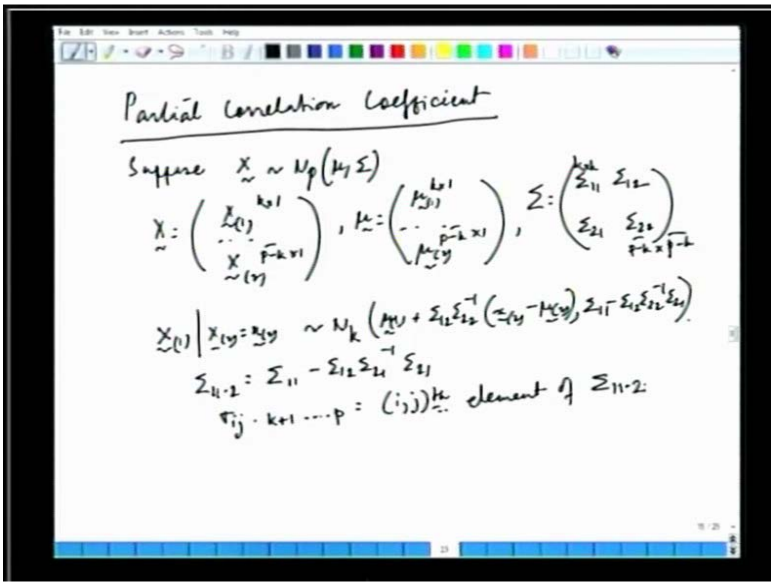


So, this is without proof. So, this derivation has not actually be done we have only mentioning it here just this and after this comes quite easily, we have from simple knowledge of sampling distribution, if we have beta distribution then we can have this type of thing as an F distribution and the fact that helps us in the testing of hypothesis for of the population correlation coefficient is the situation, where we use the distributional assumption of a multivariate normal of Y and X, the joint distribution of this as the p variate multivariate normal. So, as we were saying that the distribution that of R square is a better distribution with this parameters p minus 1 by 2 and n minus p by 2, this we have accepted without actually deriving. So, this is as gone without proof and after that this is simple knowledge of sampling distribution that, if R square follows beta distribution with these parameters we will have this type of function of R square following an F distribution.

However for the purpose of testing of the hypothesis on the multiple correlation coefficient, we are making that distributional assumption that the combined distribution of y X 1 to y n X n, this is a p variant normal distribution with mean mu and variance covariance matrix sigma and then, if we have the null hypothesis that this is in fact, equal to 0 against we are testing this hypothesis against that this is greater than zero, we can very easily use this test statistic which will follow an every distribution under H naught. So, that the test procedure is simple procedure based on the upper alpha percent cut off point of F distribution of the relevant F distribution.

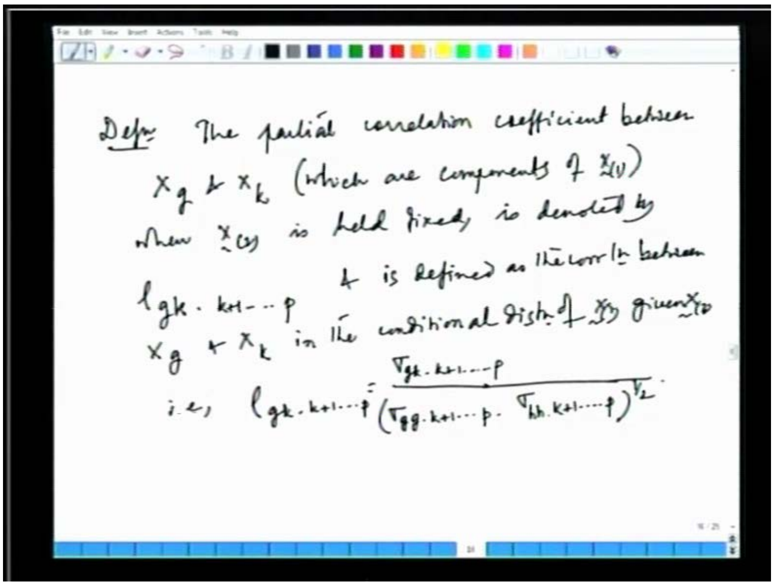
Here after putting the value of p equal to two, we can also easily establish the case of the bivariate case, where we have the sample correlation coefficient R is a simple correlation coefficient between Y and X and we see that if we put p equal to two here we get back our t statistic, because n minus 2 R square by 1 minus R square this is an F distribution with degrees of freedom 1 and n minus 2 which is equivalent to saying that the root of this the square root of this term here is following a t distribution with degrees of freedom n minus 2.

(Refer Slide Time: 52:20)



After this we have just 1 small point to be considered which is the partial correlation coefficient we go very briefly with this and we say that suppose we have X following the usual p variate normal distribution with mean mu and variance covariance matrix sigma. We partitioned X in the form I have X 1 having k elements. So, this is k- dimensional and X 2 which is p minus k- dimensional. So, you have partitioned X in this way consequently I have mu also partitioned in similar manner mu 1 and mu 2 p minus k cross 1 sigma variance covariance matrix has been partitioned in the way sigma 1 1 sigma 1 2 sigma 2 1 and sigma 2 2 this is k cross k this is p minus k cross p minus k and incident coincidently this will be of the relevant appropriate dimensions k by p minus k p minus k by k, and then we have X 1 the conditional distribution of X 1 conditioned on X 2. So, distribution of X 1 given X 2 is nothing but a k- dimensional normal distribution, if you recall with mean mu 1 plus sigma 1 2 sigma 2 2 inverse X 2 minus mu 2 and the variance covariance matrix is sigma 1 1 minus sigma 1 2 sigma 2 2 inverse sigma 2 1. So, this we are denoting by sigma 1 1 dot 2. So, let sigma 1 1 dot 2 is sigma 1 1 minus sigma 1 2 sigma 2 1 inverse sigma 2 1 and we say that let, sigma i j dot k plus 1 to p be the i j th element of sigma 1 1 dot 2.

(Refer Slide Time: 55:09)



Then we can define the partial correlation coefficient between X g and X k suppose, X g and X k which are components of the first sub vector, which are components of X 1, the partial correlation coefficient between this which are components of X 1, when X 2 is held fixed is denoted by rho g k, then I have the elements of X 2 that is k plus 1 to p and is defined as the correlation between X g and X k in the conditional distribution of X 1 given X 2 that is by the simple notation I can write rho of g k dot k plus 1 upto p by my notation of sigma i j or whatever we had written that sigma i j the i j th element of sigma 1 1 dot 2 this is sigma g K plus 1 to p, and we have in the denominator sigma gg k plus 1 upto p with sigma hh k plus 1 upto p this raised to the power half, because these are the variances.

So, this is the partial the definition of partial correlation coefficient which has the interpretation that it is the correlation coefficient the simple correlation coefficient between X g and X k when the distribution that we are considered is the conditional distribution of X 2 given X k note that X g and X k are both members of this sub vector X 1. So, we conclude our discussion on multiple correlation coefficient with a very little introduction about the partial correlation coefficient with this our next topic. Hence, both is going to be the principle component analysis.