Applied Multivariate Analysis

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

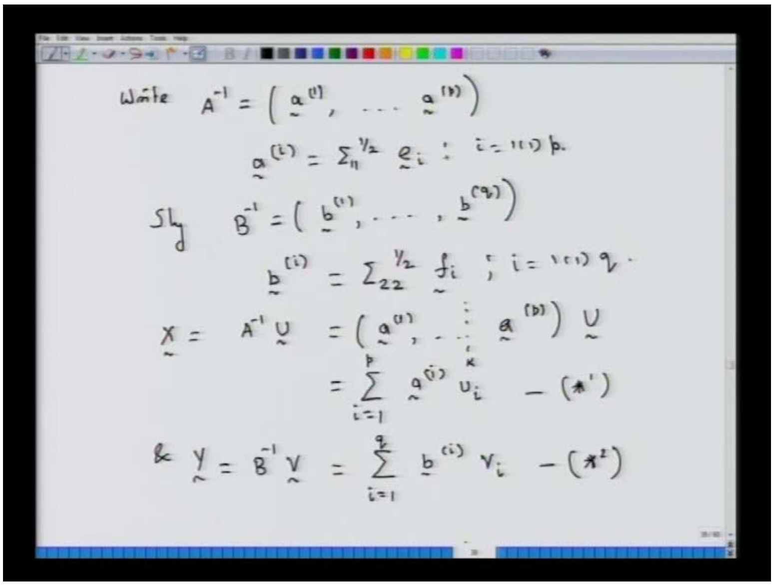
Indian Institute of Technology, Kanpur

Lecture No # 43

Cannonical Correlation Analysis

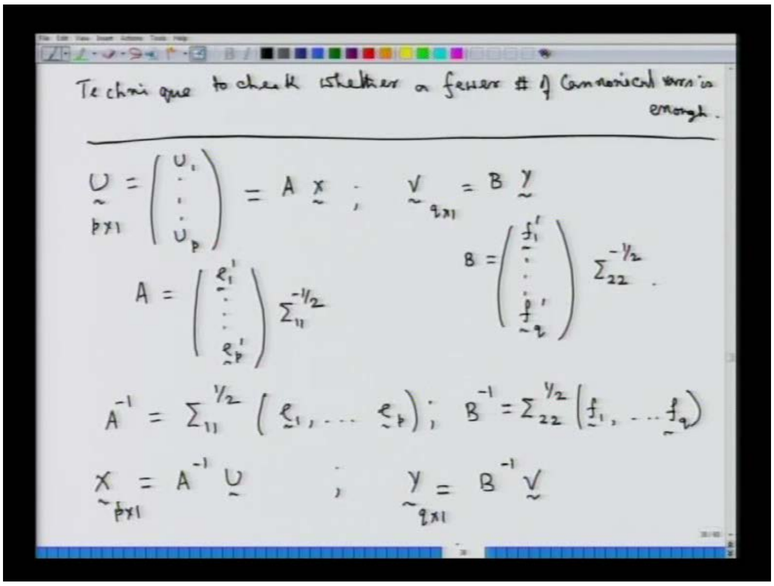
We continue our discussion on cannonical correlation analysis we were looking at when we are, have actually obtained the cannonical variables. Then, what sort of analysis can be done using those derived cannonical variables? In the last lecture, at the end, what we were looking at is, when we have a p dimensional original x random vector and we have got the corresponding cannonical variables. We were seeing, that if we are going to express the variance covariance structure of x matrix, x and y matrix jointly taken, then if we are concentrating on a fewer number of cannonical variables, then what sort of approximation to the variance covariance matrix, this k ordered cannonical variables gives?

(Refer Slide Time: 01:17)



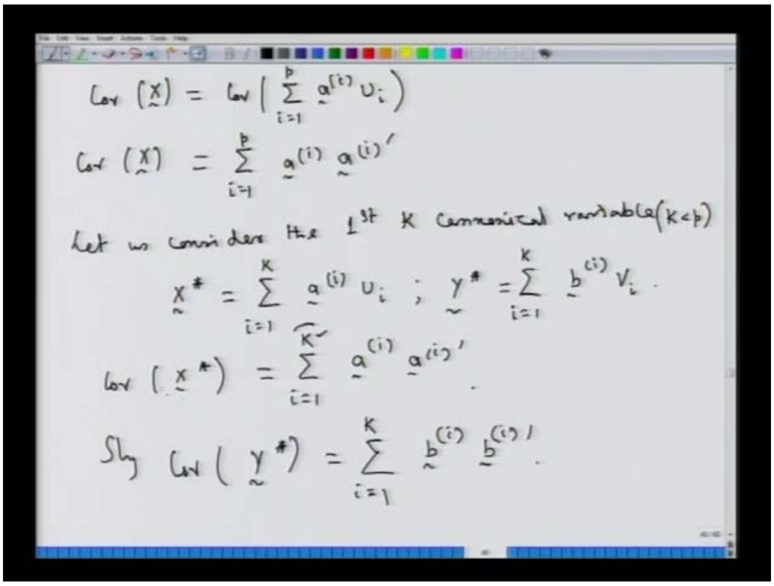
Specifically what we had seen in the last lecture was we had come up to this point, that we had seen, that this X vector, the original set of random variables, this X can be written in terms of this star 1, which is a i u i and similarly, v, Y can be written as summation i equal to 1 to q v i times v i, where a i, these are column vectors corresponding to the A inverse matrix. A inverse matrix was defined as the matrix, which was having the following structure.

(Refer Slide Time: 01:41)



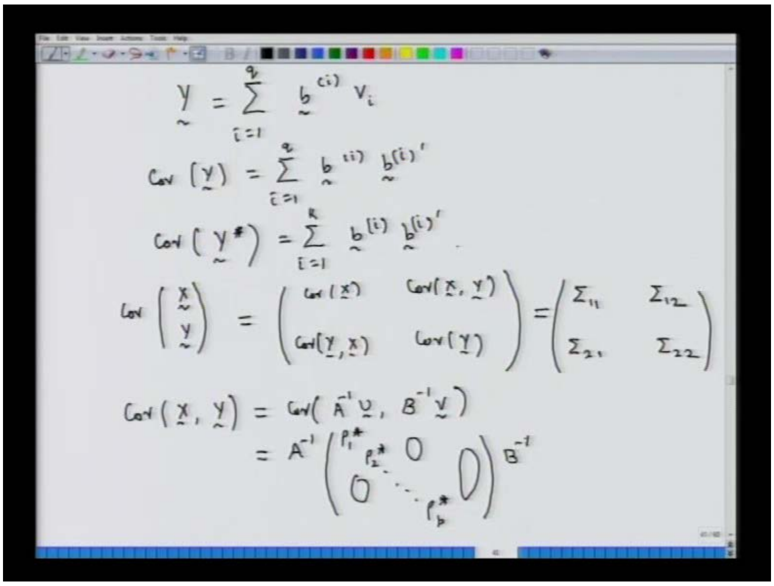
So, this A inverse matrix was this sigma 1 1 to the power half e 1 and. So, on. So, the i th column here of a inverse matrix was sigma 1 1 to the power half.

(Refer Slide Time: 01:58)



And hence, if we consider only first k cannonical variables, here k is say, less than p or less than or equal to p. If we are taking the entire set of variables in the cannonical variables, then x star can be written x star, which is x reconstructed using the first k cannonical variables. Then, that x star is i equal to 1 to k a i u i and similarly, Y star is equal to i equal to 1 to k v i times v i. And using those two representations, x star and y star, we see, that the covariance matrix of x star is represented as this, which is summation i equal to 1 to up to k only, where k is a number of cannonical variables chosen a i a i prime, and similarly, this y star is equal to this.

(Refer Slide Time: 02:50)



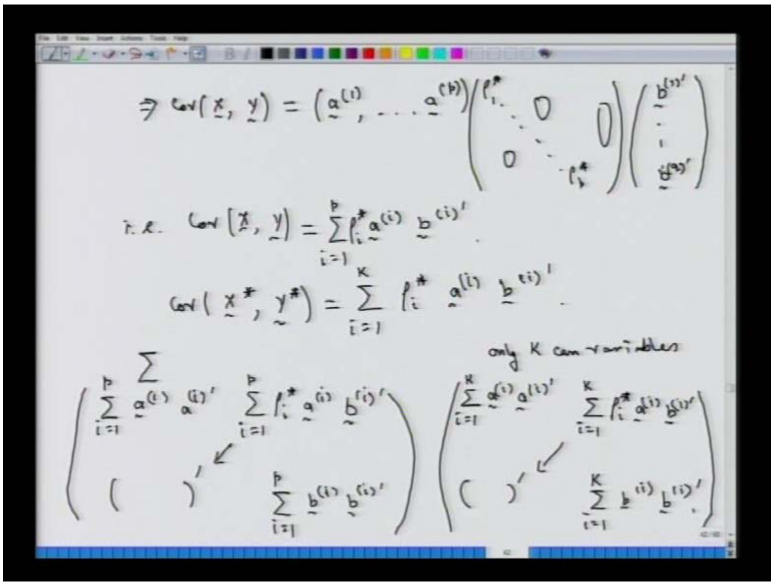
Otherwise this y corresponding to this form here, what we can write is summation i equal to 1 to up to q this v i vector, which is the original form of the y vector v i into v i. Now, from here this covariance matrix of y is straight forward to obtain, which is summation i equal to 1 to q v i vector and the transpose of the v i vector. Now, these are the constituent facts and the covariance of y star, which is this summation truncated at the point k, would be given by the corresponding sum as in i equal to 1 to k, which is v i v i transpose again.

And if we now concentrate on what is the covariance matrix between x and y, well the covariance matrix between x and y was covariance matrix of x, covariance matrix of y and this is the covariance matrix between x and y, and this is the transpose of this element, so this is covariance matrix of y with x. This, in terms of our original definition, is this sigma 1 1 matrix, sigma 1 2 matrix, sigma 2 2 matrix and sigma 2 1 matrix. Now, we have obtained this sigma 1 1 and sigma 2 2 in terms of this a i and v i vectors. Let us also see what this covariance between x and y is.

Covariance between x and y is in terms of the new notations, that is what we had introduced, that x is equal to A inverse U and Y is equal to B inverse V, X is equal to A inverse U vector and Y is equal to B inverse times this v vector. This is p by 1, this is q by 1. So, the covariance matrix of this is going to be given by A inverse times the covariance matrix between U and V, we have timed it. We have seen it time and again, that what that is equal to. This is rho 1 star, rho 2 star, rho p star; these are the p cannonical correlation coefficients that augmented with 0s.

So, this is the first p by p block wherein we are able to pair u i with v i and these are u i with v p plus 1 to up to v q. So, these are those columns here that multiplied by this b inverse. Now, in terms of this a i vectors and v i vectors, what we can write.

(Refer Slide Time: 06:04)



So, this will imply that the covariance matrix between x and y which is of course, sigma 1 2 that can be written as this is a 1 vector a 2 up to a p vector this multiplied by that matrix which we had taken this rho 1 star rho 2 star up to rho p star here all other entries are zeroes that augmented by this b 1 prime b q prime.

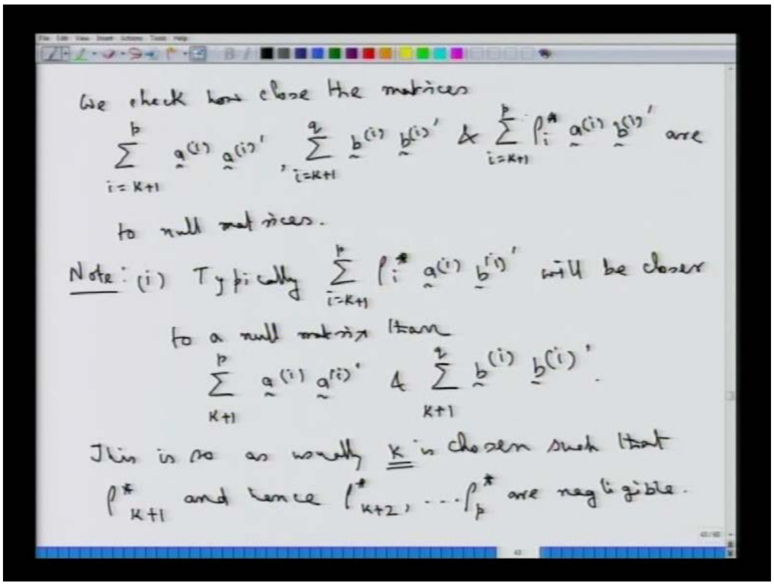
So, this is what it is coming this is b inverse is transpose because this inverse is coming on the other side. So, its transpose of that and this is what this is summation i equal to 1 to up to p a i vector that multiplied by well this rho i stars are all scalar quantities. So, i can write this rho i star in here that multiplied by the corresponding v i transpose that is this covariance matrix between x and y is finally, of this particular form now if we follow this same approach then the covariance matrix between x star the x vector approximated using the first k cannonical variables x star and y star

Note that nothing much changes except in here we will be truncating this up to the k th row here and we will be truncating this matrix up to the k th I am sorry this we are truncating here up to the k th column and here on the right hand side we are truncating up to the k th row here. So, that would lead us to summation i equal to 1 to up to k rho i star a i into v i transpose. So, what is that we are having the original sigma matrix now in terms of this a i v i vectors this element is summation i equal to 1 to p a i a i transpose the 2 2 th entry here is i equal to 1 to up to p v i v i transpose and this entry here is i equal to 1 to up to p rho i star times a i into b i transpose and this is just the transpose of this particular entry here

So, that is the covariance matrix in terms of this a i and v i vectors and when we are looking at only k cannonical variables then this matrix is what is now giving us the covariance structure. So, this is now i equal to 1 to up to k the same terms actually only the summations are having up to the k th terms. So, this is a i a i transpose this 1 is summation i equal to 1 to up to k rho i star a i v i transpose this is summation i equal to 1 to up to k v i v i transpose and this 1 is just the transpose of this particular entry here

So, when we have approximated or rather when we have just used k cannonical variables then we look at how good is that k number of cannonical variables by looking at the closeness of the respective entries here. So, we check how close is this sigma matrix to the new sigma matrix defined through the k cannonical variables and hence we check how close is this with this particular m k here 11 th block here how close is this 1 2 th block here with this which is using just k terms and how close is this term here the 2 two th block with this particular entry here. So, that is, what is, the technique to check for the closeness of approximation.

(Refer Slide Time: 10:35)



We check how close the matrices there are, three crucial matrices, I will write it as i equal to 1 to k plus 1 to up to p a i a i transpose summation i equal to k plus 1 to up to p b i. When we are looking at, this is up to q, this is b i b i transpose and the cross block here, which is summation i equal to k plus 1 to up to p rho i star a i b i transpose are to null matrices. If these three matrices are close to null matrices, then what will happen is, that well, this summation here is up to q because b i b i transpose as q entries.

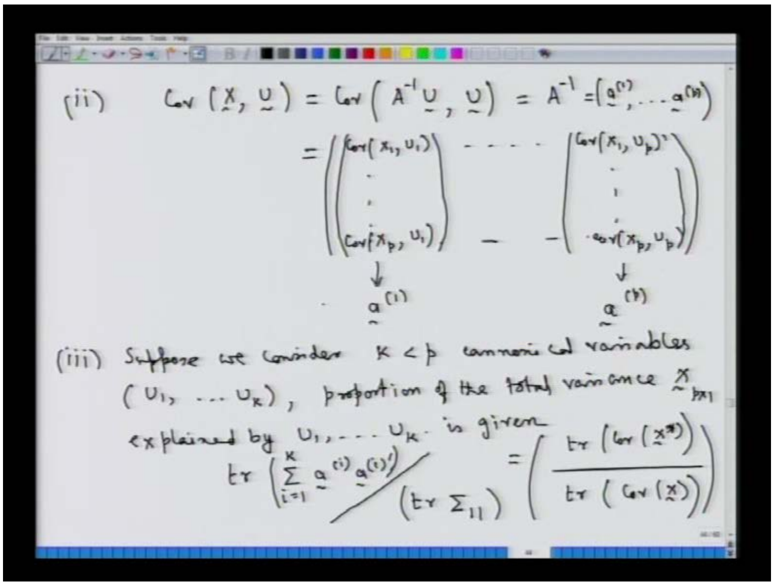
So, if we have these three matrices close to null matrices, then what we will be having is this matrix here, proves to this particular matrix because the term, that we are not considering, that is, i equal to k plus 1 to up to p here, i equal to k plus 1 up to q in this entry and i equal to k plus 1 to up to p in this entry here, will be close to the null matrices. And that this technique actually provides us a way to check, whether k number of cannonical variables is enough to explain the covariance structure of the original set of random variables.

Now, we make the following three important notes here. The first note is, what we expect is, that typically the 3rd matrix here, i equal to k plus 1 to up to p rho i star a i into b i transpose will be closer to null matrix, to a null matrix than the other two, the other two being a summation from k plus 1 to up to p a i a i transpose and the summation i equal to k plus 1 to up to q b i b i transpose. This is so because, this is so as usually, k is chosen as, to be chosen rather, k is to be chosen such that rho star k plus 1. And hence, rho star k plus 1 onwards, that is, rho star k plus 2 up to rho p star are negligible.

This is so because if we choose our k, the number of cannonical variables to be retained in such a way, that the Eigen values, the square root of the Eigen values of that sigma 1 1 to the power minus half sigma 1 2, sigma 2 2, inverse sigma 2 1, sigma 1 1 to the power minus half, that matrices Eigen values are rho i stars. So, if we choose rho star k plus 1 onwards, because it is an order to be negligible, then what will happen is, that this particular matrix, since rho k plus 1 star, rho k k plus 2 star, rho p star, all of these are negligible and close to 0. This third matrix here, which is the cross block matrix, this one will typically be close to a null matrix.

But that cannot actually be said about the first two blocks, the 1th block and 2 2th block because simply, because those does not involve this rho i star, which are negligible and that is basically the guideline of choosing the number of cannonical variables, that we are going to retain. So, typically, this particular matrix here, the cross block here, which actually looks at this covariance between x and y and that reconstructed through x star and y star, that particular block will be having a closer approximation to the original sigma 1 2 matrix.

(Refer Slide Time: 15:45)



Now, as a 2nd note here, we just, it is a small note, we have already obtained what is the covariance between x and u and also the correlation between x and u. Let us see, that this particular matrix here, the covariance matrix between x, the original set and the cannonical variables derived from the xs, both are p dimensional. So, this is in terms of the A matrix, we can write it straightaway as that equal to the A inverse matrix, why? Because x is equal to A inverse u according to our definition. So, this A inverse, which we had denoted by these column vectors a 1, a 2 up to a p, these p column vectors, these column vectors has got the interpretation, that this is covariance between x 1 and u 1, this is covariance between x p and u 1 and so on.

So, this A 1, which is the first column of the A inverse matrix, the first column of x and use covariance matrix, that holds all these entries here and similarly, all the other columns can be filled up. So, this is covariance between x 1 and u p and this last entries covariance between x p and u, u p. So, these are all the entries. So, this particular column is what we have as the a 1 column and the last column, the pth column here is, what is, that a p vector, this is in the other way definition, this is equal to, this one is equal to sigma 1 1 to the power minus half times u 1.

The 3rd note says, that suppose we consider k strictly less than p cannonical variables, as is a setup in the present situation, then these cannonical variables are just u 1, u 2, u p. Let, let me just u k, let me write, that u 1, u 2, up to u k, then the proportion of the total variance of x, this is a first set of variables. So, the proportion of the total variance of x explained by these k cannonical variables, u 1, u 2, u k, can be quantified by the following expression.

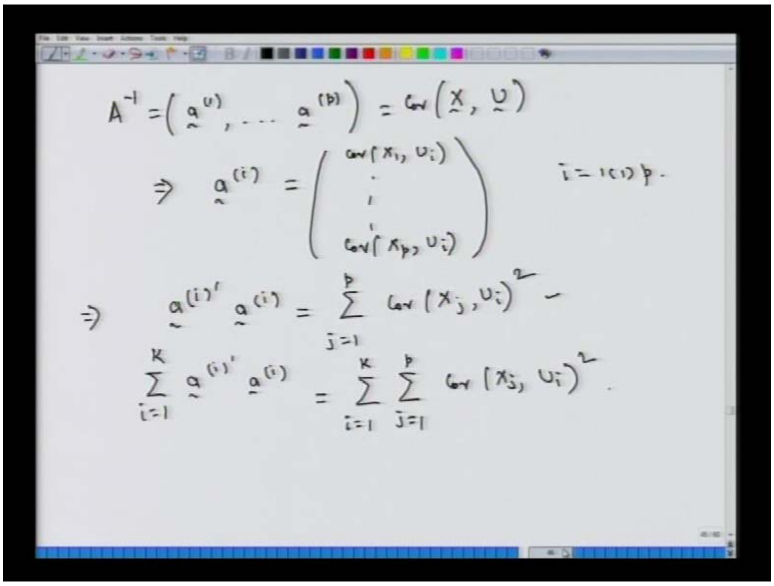
Now, how are we going to say this particular proportion? The logic behind quantifying this proportion of total variance, well, what is the total variance? The total variance is in x, is corresponding to the trace of the variance, covariance matrix of this x. So, that is trace of sigma 1 1. So, this is a total variance.

Now, this is going to be having a numerator if we are going to measure the proportion of total variance in x. That is explained by u 1, u 2, u k. This would be given by the reconstruction formula, so that this is equal to summation i equal to 1 to up to k a i a i transpose. So, this one is what? This one is the covariance matrix of x star; x star denotes, that reconstructed x using u 1, u 2, u k, the k cannonical variables. And hence, the total variation in x star can be measured by this quantity, which is trace of this matrix; this is nothing, but the covariance matrix.

So, just to recall, that this is nothing, but as we are denoting this by the trace of the covariance matrix, this is the trace of the covariance matrix of x star reconstructed using the k cannonical variables, that divided by the trace of the original covariance matrix, that is, I am sorry, this is the original covariance matrix, which is x. This trace of covariance matrix of x is nothing, but trace of sigma 1 1 and covariance matrix of x star was given by this if we are retaining up to k cannonical variables. So, that gives us the trace of this particular quantity. So, this is how we can quantify.

Now, we can get down with it further and see what is trace of a i a i prime matrix in terms of these covariances.

(Refer Slide Time: 20:40)

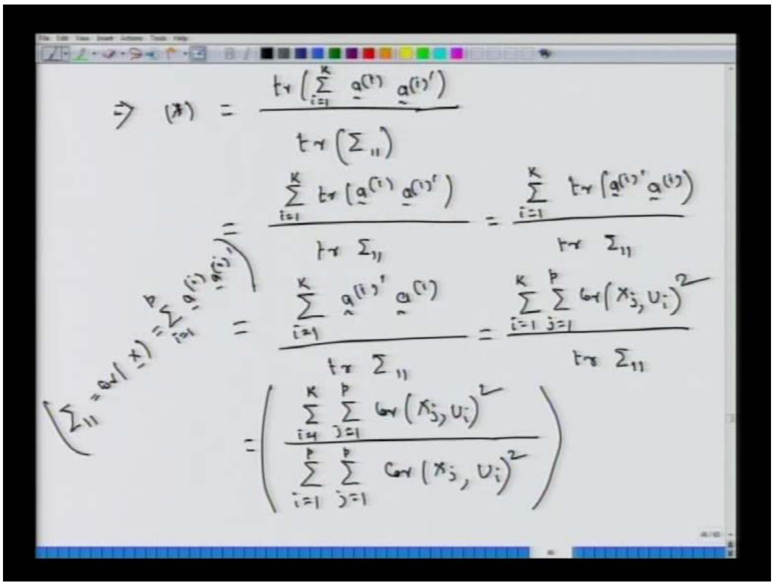


Now, as we can see, that this A inverse, which we have denoted in the previous slide as a 1, a 2, a p, which was that covariance matrix of x and u, so this would imply, that this a i vector is the ith column of this covariance matrix, the ith column here is the ith column here. So, that would be given by covariance between x 1 and u i up to covariance between x p and u i.

So, this is going to hold this element, which is covariance between x 1 and u i to up to covariance between x p and this ith cannonical variable. So, this is for i equal to 1 to up to p. So, what we are considering is this up to this term. So, if a i is equal to this, this will imply further, that a i transpose a i is just going to here hold these square entries. So, it is x j and u i covariance square of this and summation j equal to 1 to up to p. So, this is what we are going to have.

So, this will imply, that summation a i, which is going to play a role in that proportion a i transpose a i, i equal to 1 to up to k, the order of truncation. This would be given by summation i equal to 1 to up to k, summation j equal to 1 to up to p, this covariance between x j and u i this whole square.

(Refer Slide Time: 22:27)



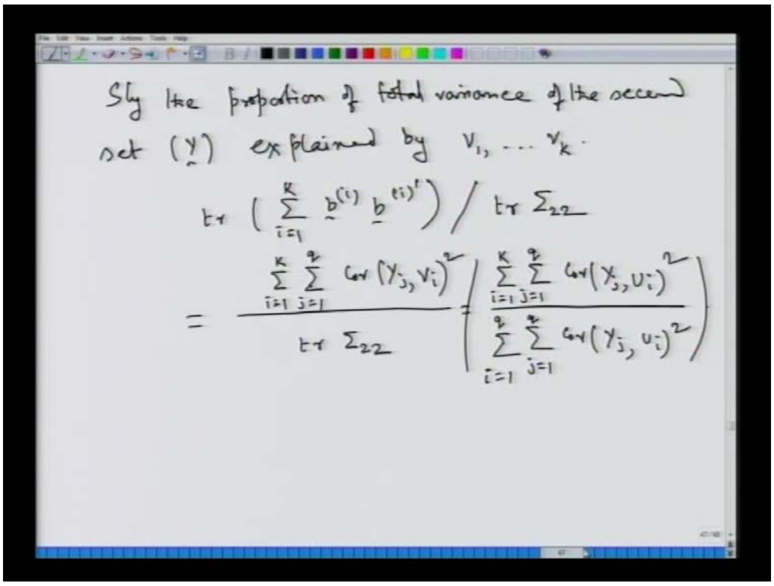
So, that this will imply, that this term, that we were considering, let us write this as star term. So, this star term, which was equal to the trace of the summation i equal to 1 to up to k a i vector into the transpose of that a i vector, that divided by trace of sigma 1 1 matrix, this is going to be given by, now trace, you can take trace of sum equal to some of the respective traces.

So, what this is going to be? This is going to be i equal to 1 to up to k trace of a i a i transpose, that divided by trace of sigma 1 1. You can further write trace of a b equal to trace of b a, so that this is equal to i equal to 1 to k, trace of a i transpose into a i, that divided by trace of sigma 1 1. Now, a i transpose a i is the scalar quantity and hence, this is just equal to summation i equal to 1 to up to k a i transpose into a i, which we have just now computed divided by trace of sigma 1 1, so that this term is equal to summation i equal to 1 to k, summation j equal to 1 to up to p, covariance between x j and u i square, this divided by trace of sigma 1 1. Well, this trace of sigma 1 1 can also be written in terms of such covariances.

Recall, that this sigma 1 1 matrix in terms of a i vectors is just equal to a i. So, we recall, that this sigma 1 1, which is the covariance matrix of x is nothing, but in terms of these definitions, that is, a i, a i transpose i equal to 1 to up to p. So, this is what we have and hence this one further can, one can write as summation i equal to 1 to up to k, summation j equal to 1 to up to p. This is covariance x j and u i square, that divided by once again the double summation, wherein the first summation i will be equal to 1 to up to p j equal to 1 to up to p, that divided by the same term covariance x j and u i square.

So, this is a crucial thing, which actually tries to explain the proportion of total variance in the first set of variables, that is, x in terms of only the k k less than p cannonical variables.

(Refer Slide Time: 25:25)

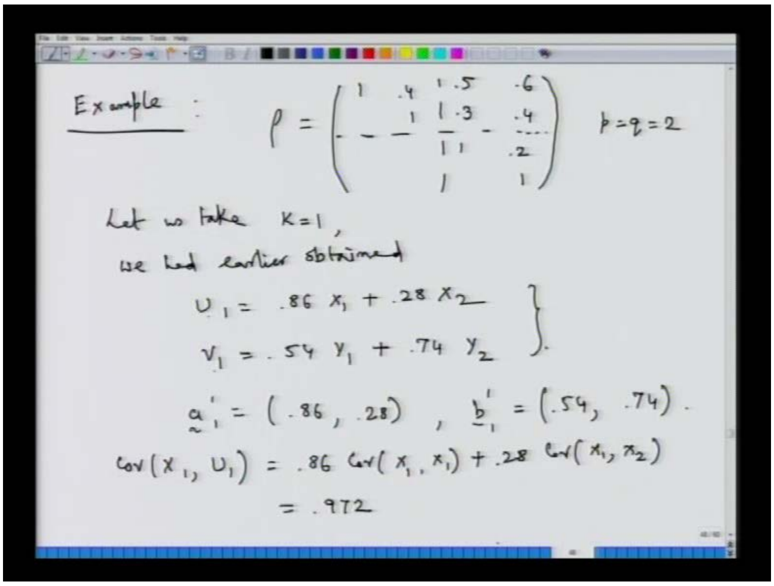


So, if this is the expression, we can similarly obtain the expression for the second set as, similarly the proportion of total variance in the second set, total variance of the second set, that is, y set of vectors explained by the retained 2nd set of cannonical variables, which are v 1, v 2, v k would be given by trace of that summation i equal to 1 to k once again of b i b i transpose entries, that divided by the trace of the covariance matrix of y, which is trace of sigma 2 2.

Now, as in the previous example, one can reduce it in to this particular form. So, this is going to be given by summation i equal to 1 to up to k, let me write this as summation i, summation j equal to 1 to up to q. And then, this is going to be given by covariance between y j and v i, this whole square, that divided by trace of sigma 2 2 or once again this trace of sigma 2 2 can be written in terms of this covariance square terms and this can thus be written as summation, the numerators does not change. Of course, i equal to 1 to k and summation j equal to 1 to q because q is the order of y random vectors. So, this is y j and u i square, that divided by summation i equal to 1 to q.

Now, because we are looking at the entire x and its covariance matrix as sigma 2 2, this is summation j equal to 1 to q and this is covariance between y j and u i square once again. So, these two quantities are crucial in looking at how much of the total variation in the respective sets, actually xs and ys are explained, if we are retaining only k cannonical variables in the process.

(Refer Slide Time: 27:43)



Now, let us look at an example, a numerical example to see how this actually works, the concept that we try to learn today. So, let us take that previous rho matrix with which we had obtained the cannonical variables. So, that had got this structure, that it was the correlation or the covariance matrix of that 2 by 2 standardized variables, which had values as 11.4 as the 1st block here. The 2nd block was given by 0.5, 0.6, 0.3 and 0.4 and this block here is, see rho 2 2 block is 1, 0.2, 1, it is a symmetric matrix. So, this entry is a transpose of this and this entry is of course, 0.4 and this one is also 0.2. So, we have p equal to q equal to 2.

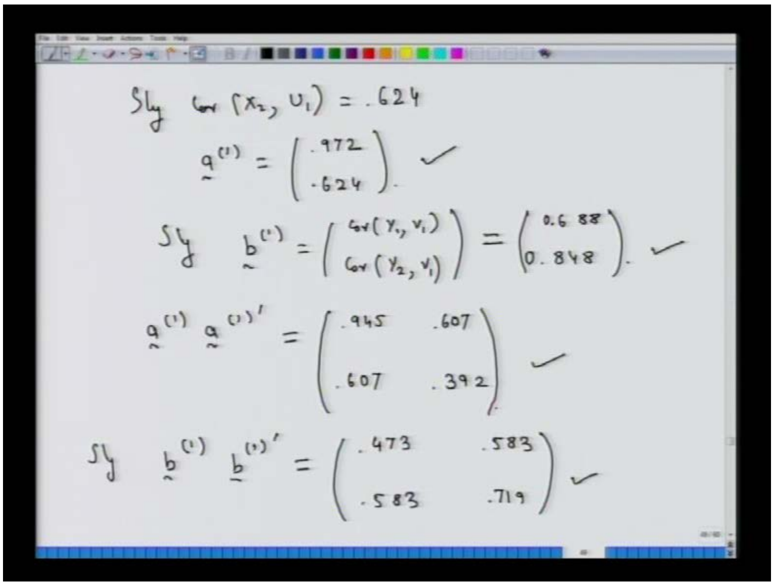
Now, let us take, as in the previous example, as we had seen in other previous lecture, that this, this matrix has got the corresponding cannonical correlation coefficients, the first one significant and the second one was negligible actually, so that rho 2 star was equal to 0.03, if you recall. So, we are going to neglect that particular term and take, let us take this k equal to 1 and we will say that since the 2nd cannonical correlation coefficient was negligible, we are going to retain only one cannonical variable.

We had earlier obtained the two cannonical variables associated with k equal to 1 as the following. I will, just trying those cannonical variables, which was 0.86 times x 1, the first original variable, this plus 0.28 times x 2. And v 2 was computed as 0.54 times y 1, this plus 0.74 times y 2. So, this is what we had derived earlier.

Now, using this, of course, it uses this a 1 vector transpose, that to be equal to 0.86, which is the coefficient vector here, 0.86, 0.28 and this b 1 vector was 0.54 and 0.74, these are the two terms.

Now, this, from this expression here, u, u 1 and v 1, what we can compute is what is the covariance between x 1 and u 1, that is straight forward. So, what this will be? This will be 0.8 times covariance between x 1 and x 1. Now, these xs were standardized actually. So, this covariance will be equal to 1, this plus 0.28 times the covariance between x 1 and x 2, whatever. So, this covariance between x 1 and x 2 is going to be this 0.4 covariance between x 1 and x 1, that is, the variance is equal to 1. So, this turns out, the numerical value of this covariance between x 1 and u 1 turns out to be, it is 0.972.

(Refer Slide Time: 31:11)



Now, similarly, one can compute the other term, this covariance between x 2 and u 1 will turn out to be it having a numerical value 0.624. Now, why are we computing these two terms from the cannonical variables covariance x 1 and u 1 and covariance between x 2 and u 1? Because we want to construct, that a upper bound vector, if you come back to this particular point here, that this a 1 is what? It is the 1st vector associated with covariance of x 1 and u 1, x p and u p.

So, that is the 1st column of this A inverse matrix, and why is that required? That is required essentially, in order to look at these quantities, which are going to tell us about the proportion of total variance in x and in y. Later on, that is being explained by the first cannonical variable u 1, cannonical variable pair u 1 and v 1. So, once we have these two, we have a 1 vector, upper one, that is, the first column of the A inverse matrix. So, this has got the entries that we have computed; these are the two quantities.

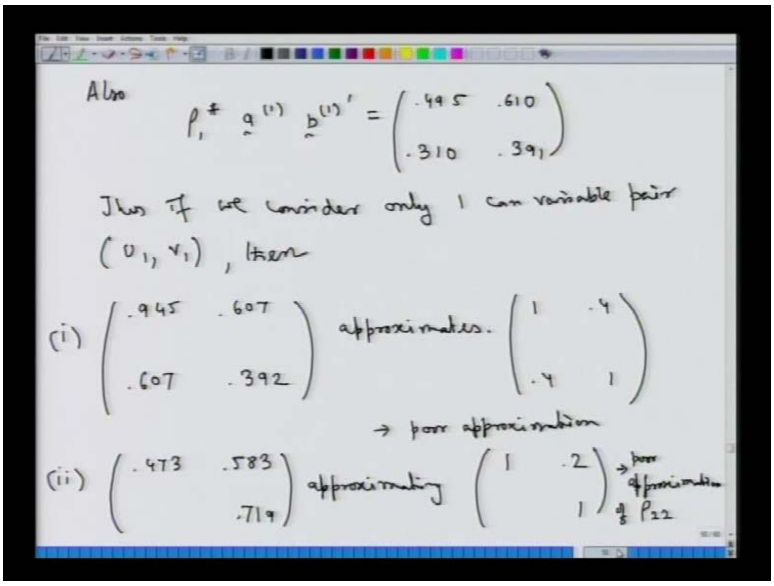
Now, similarly, this b 1 vector can be computed. Now, this b 1 vector will hold the first entry here, will be covariance between y 1 and v 1, and the second entry will be covariance between y 2 and v 2. So, given the form of v 1, this one can find out what is the covariance between v 1 and y 1, v 1 and y 1 and v 1 and y 2. This is covariance between y 2 and v 1, this can be computed once again, looking at the form. So, the numerical value of this turns out, that this is 0.688 and this is equal to, point, 0.848.

Now, using this one and v 1 vector, one can compute what is a 1 a 1 transpose, this particular matrix. So, this matrix turns out, that it has the following entries, which is 0.945, 0.607, this is 0.607, this is 0.392. Now, similarly, from this b 1 vector we can obtain this b 1 b 1 transpose matrix, which turns out to have the following numerical values, which is 0.473, 0.583. This is, of course, a symmetric matrix, no need as such to write these of diagonal entries.

So, this b 1 b 1 transpose and b 1 b 1 transpose and a 1 a 1 transpose are the two matrices, which are going to approximate. What these, this a 1 a 1 transpose is that proportion in summation, a i a i transpose summation, a i a i transpose summation, i equal to 1 to up to p is actually giving us the covariance matrix of the original set of random variables. And if we are looking at only one cannonical variable, that is, if we are choosing k equal to 1, then a 1 a 1 transpose, this is, that sum, summation i equal to 1 to k, we are choosing as 1.

So, it is that portion, which is actually going to approximate the original sigma 1 1 matrix and this b 1 b 1 transpose is that portion of the covariance matrix of y, which is just going to explain the covariance structure of the 2nd original set of random variables, which is the vector y having the covariance matrix as sigma 2 2. Here, we have rho 2 and we also need to compute another term, which is the off diagonal term.

(Refer Slide Time: 35:28)



Let me also write that off diagonal term first. Also, this rho 1 star a 1 b 1 transpose, why are we looking at that? Because if we recall, that in the off diagonal block we have summation i equal to 1 to p in the original set of random variables. So, the covariance matrix between x and y is in the off diagonal block, which is summation i equal to 1 to p rho i star a i b i transpose.

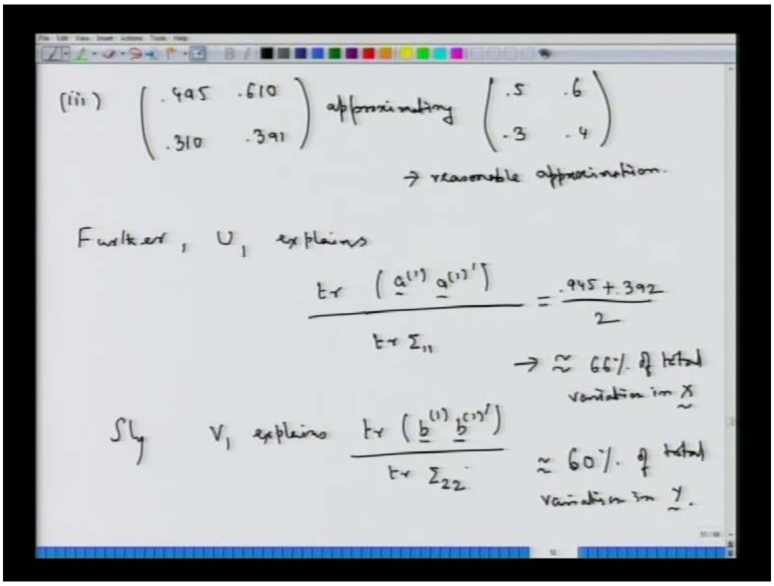
Now, we are truncating the numbers or number of variables at k equal to 1, we are choosing only one cannonical variable and hence, this is one, this is the matrix, which is having the numerical values as 0.495, 0.61, 0 .310, .391. This is that portion, which is going to approximate the rho 1 2 block there.

So, what we are going to have? Thus, if we consider only one cannonical variable pair, as in the present situation, u 1 v 1, then this matrix here, what we have obtained, I will have to write these numerical values once again. So, these numerical values here, 0.945, 0.607, 0.607, 0.392 approximates. This is that portion, which is explained through the first two canonical, first pair of cannonical variables, this as approximating this block here, which is the rho 1 1 block.

So, that is one, which is prime to approximate 1, 0.4, 0.41, that is number one, this is one, which approximates these. Now, what you, what we can see is that this matrix, while we are considering one pair of cannonical variables, as such is a poor approximation of the original matrix. So, this is a poor approximation. Now, the second 2 2th block, if we look at the 2 2th block, which is b 1 b 1 transpose, that is, particular matrix has numerical values as this 0.719, this as approximating the rho 2 2 block. Now, the rho 2 2 block has entries 1, 0.2, 0.21.

So, once again we see, that this b 1 b 1 transpose matrix as such is a poor approximation of this rho 2 2, so this is a poor approximation. However, if we are now looking at the off diagonal block, the off diagonal block has entries.

(Refer Slide Time: 38:54)



So, after these two what we see is the third approximating matrix. The 3rd approximating matrix is what we have obtained in here, which is rho 1 star, a 1 a 1, a 1 b 1 transpose. So, that approximation is what we have the numerical values as 0.495, 0.610, 0.310, 0.391. This, as approximating, which quantity the rho 1 2 matrix, this as approximating the rho 1 2 entry, which is 0.5, 0.6, 0.3, 0.4.

Now, in contrast to the two previous approximations, which we had seen, that this was trying to approximate rho 1 1, this matrix b 1 b 1 transpose trying to approximate is rho 2 2 matrix and this matrix here trying to approximate the rho 1 2 matrix while the two previous approximations were quite poor.

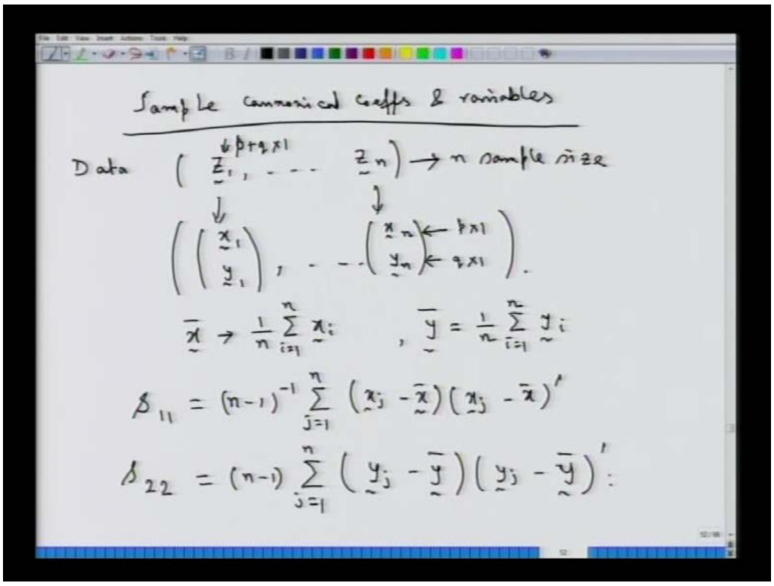
This is a reasonable approximation, this actually highlights, that the note number one, that we had written here, if we go back to that particular note here, we had said, that typically what will happen is that this matrix, the residual matrix will be closer to the null matrix, that is, summation i equal to 1 to k will be a closer approximation to the corresponding matrix, which is sigma 1 2 matrix. Then, the other two matrices, which are the rho 1 1 or sigma 1 1 and rho 2 2 or sigma 2 2 matrix, this is what it is, basically highlighting, that although we had poor approximation of the two previous matrices, this is a reasonable approximation.

Now, the 2nd, the 3rd note, what we had done today is the proportion we explained by the 1st cannonical variables, 1st pair of cannonical variable. This u 1 would explain the trace of a 1 a 1 transpose here because k is equal to 1 here. So, this is case of this matrix, this divided by trace of sigma 1 1 or rho 1 1, whatever it is, the starting matrix. So, for this, for the given problem this trace turns out, that this is equal to, I will just write it the values here, that is, 0.945 plus 0.392, that divided by 2. So, this approximately is explaining 66 percent of total variation in the original set of random variables x 1 and x 2. This is a reasonable amount, which the first cannonical variable is explaining.

And similarly, this v 1 explains this trace of b 1 b 1 transpose matrix. This divided by the trace of sigma 2 2 matrix, this further can be obtained or rather, the numerical value turns out, that it is 60 percent of the total variation in y vector. Now, in the, in this example, this is the covariance matrix is corresponding to the standardized variables and hence, this sigma 1 1 and sigma 2 2 are actually rho 1 1 and rho 2 2.

So, this is how for a given problem, one actually can see how close will be a particular order of cannonical variable pairs chosen, that is, the k term in this approximation and how close will the corresponding approximations to the variance covariance structure of x b if we are considering up to k cannonical variables. And also, by considering those many cannonical variables, what is the proportion of the total variance in the first set, was the proportion in the variance in the second set of variables are getting explained in the, through the retained number of cannonical variables.

(Refer Slide Time: 43:18)



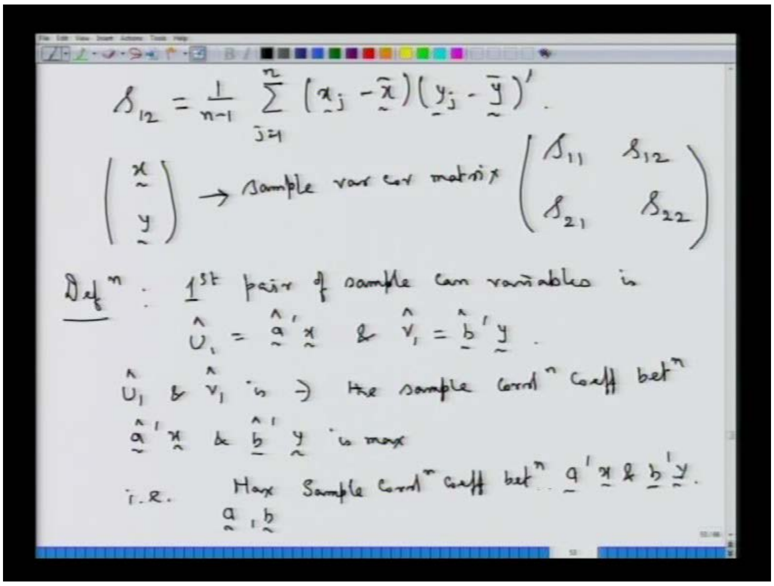
Now to end this concept and to end this lecture series what we are going to just look at is, in this cannonical correlation analysis how we are going to have the sample cannonical variables and the cannonical correlation coefficients because that is what we are going to have in practice.

So, in practice what we are going to have is a data, which might look like the following, that it is z 1, z 2, z n. So, we have n data vectors, each of these data vectors has got two components, x and a y component. So, this is a first observation corresponding to the x set of random variables and this one is the nth set here, which we denote by x n and this is y n. Now, we assume, that this x here is p by 1 set here and this is q by 1 set here. So, this is the original set of random variables, the first set, the second set p by 1, q by 1.

So, each of these z i's are actually, p plus q by 1 random vectors. Now, given these two sets of random variables observations and m replications of the, that is, the sample size n is the sample size, how we are going to calculate the sample cannonical correlation coefficients and the corresponding sample cannonical variables. Now, given this set, data set here, what we can compute is the x bar vector, the sample mean random vector corresponding to the first set of random variables. So, that is equal to this summation i equal to 1 to n of x i's. And similarly, this y bar can be computed, this is 1 upon n summation i equal to 1 to n of these q dimensional observations.

Now, once we have these two, we can also calculate the sample covariance matrices respectively, for the respective components. Now, S 1 1 is 1, that is going to be computed from x 1 vector, x 2 vector, x n vector and that, let us consider, that we are considering this 1 upon n minus 1 as the divisor, which is going to correspond as we have seen in the theory, that this is the unbiased estimator. So, this is summation j equal to 1 to n x j minus this mean vector x j minus this x bar transpose S 2 2 matrix, which is a sample covariance matrix variance covariance corresponding to the observations y 1, y 2 , y n random vectors observations So, summation j equal to 1 to n once again of this y j vector minus the corresponding y bar vector, y j vector minus this y bar vector transpose.

(Refer Slide Time: 46:28)

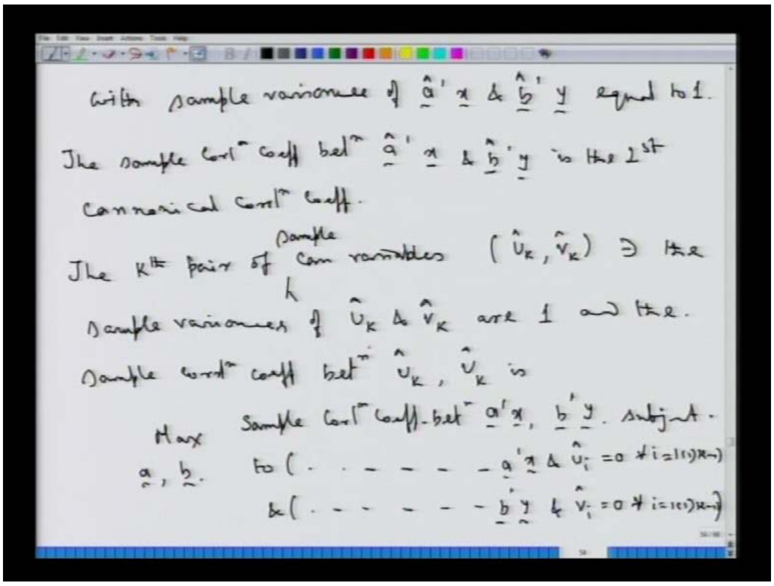


And also the covariance matrix, the sample covariance matrix, this is once again with a divisor 1 upon n minus 1, this is summation j equal to 1 to n x j minus x bar into y j minus y bar transpose. So, we will have, corresponding to this random variables x and y, will have the sample variance covariance matrix, sample variance covariance matrix is now given by this S 1 1, S 2 2, S 1 2, S 2 1. So, this is now the sample variance covariance matrix.

Now, from here how are we going to define the sample cannonical correlation coefficients? We will see how that is given, so we will first write the definition. So, the first pair of sample cannonical variables are the first pair of cannonical variables is given by, this is u 1 hat, which is given by A hat times, A hat transpose times x and v 1 hat, we are using the same notation still and putting a hat in order to indicate, that they are some sort of estimated components. This is v hat prime, y this u 1 hat and v 1 hat is such that the sample correlation coefficient between A hat prime x and this v hat prime y is maximum. That is, we are looking at maximum over a and b, the two linear combining vectors of the sample correlation coefficient between this a prime x and b prime y.

So, this one is going to be the optimum value, optimum linear combining vector, here u 1 hat and v 1 hat will be with the optimum coefficients here. So, A 1 hat or rather A hat is that a vector and b hat is that b vector, which is going to maximize the sample correlation between these two linear combination comprising of xs and ys with, of course, the restriction, that with the sample variance of A hat prime x and b hat prime y equal to 1.

(Refer Slide Time: 49:28)



So, that is basically the similar type of situation that we had seen for the population cannonical correlation coefficients computed from the sigma matrix. When we had sigma matrix, then we had also considered a prime x, b prime y and then, we tried to find out that linear combining vector, A hat, a and b, such that the population correlation coefficient between a prime x and b prime y is going to be the maximum among all such linear combinations with the restriction, that the variances of u 1 and v 1 will be equal to 1.

Now, when we are looking at the sample counterpart, we are also having the same philosophy, that you look at all such linear combinations with the, with the condition, that the sample correlation coefficient between u 1 hat and v 1 hat would be the maximum subject to the condition, that the sample variances of a 1 prime x and b prime y will be equal to unity. Now, we will have the similar definition for the second, the non-words cannonical variable pairs, which we are going to give.

Well, with that sample we should also say, that the sample correlation coefficient between these two linear combining vectors, a hat prime x and b hat prime y is the first cannonical correlation coefficient, that can be derived from the data and in general, the kth pair of cannonical variables, say, given by u k hat and v k hat. So, this is the kth pair of sample cannonical correlation variables, I should say kth pair of sample cannonical variables. This is such, that the sample variance of this u k hat and v k hat, sample variances of a k hat and v k hat is equal to 1.

And the sample correlation coefficient between this u k hat and v k hat is given by the maximum over a and b of all possible sample correlation coefficients, between a prime x and b prime y subject to the condition, to the two conditions. Number one, that the sample correlation coefficient between this a i prime, a i hat prime x, let me just write this as a prime only, a prime x and u i hat will be equal to 0 for every i equal to 1 to up to k minus 1. And number two, the sample correlation coefficient between b prime y and v i hat will also be equal to 0 for every i equal to 1 to up to a minus 1.

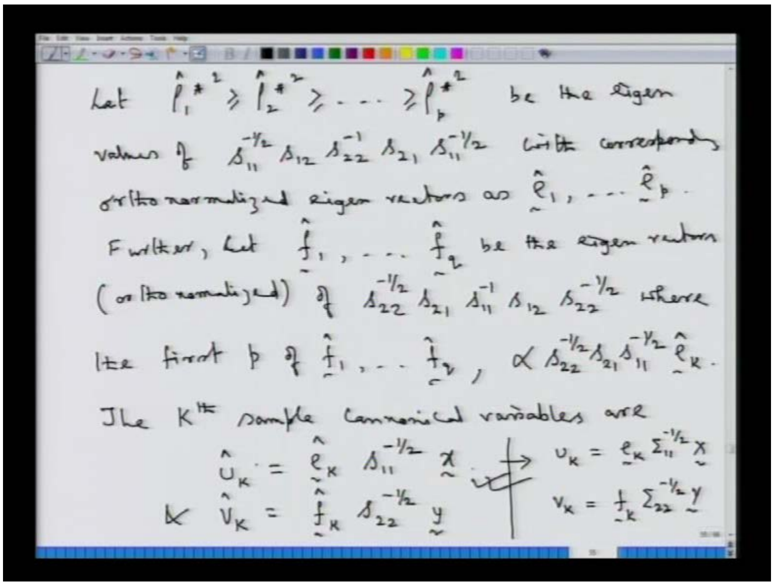
So, this basically translates, that concept of the population correlation coefficient, cannonical correlation coefficient in terms of the sample cannonical correlation coefficients because as you can see here, that the sample correlation coefficient, that we are going to define for the kth pair of the cannonical variables is, that this u k hat and v k hat are going to be such, that this is the kth pair of sample cannonical variables are going to be such, that the sample variance of each of these quantities will be equal to 1.

And the sample correlation coefficient between u k hat and v k hat is of course, such that it is going to maximize over all possible choices of a and b, such that the sample correlation coefficient between the linear combining terms a prime x and b prime y is going to be the maximum subject to the condition, that we will be having. The sample correlation coefficient between this a prime x at the kth step, uncorrelated with the all the previous k minus 1. So, this u i hat are the previous k minus 1 cannonical, sample cannonical variables. Now, these are going to be uncorrelated with the previous k minus 1 terms and we are also going to have this being satisfied, that b prime y, all such bs, such that those b prime ys will have the sample correlation coefficient with v i hat for every i equal to 1 to k minus 1, all previous k minus 1 cannonical variables obtained through the linear combining y will also be equal to 0.

Now, if we have defined this sample cannonical correlation coefficients in this particular way, the last thing, that we will be doing is, we will just look at how these are going to be done in relationship with the S matrix, that we have as a sample variance covariance matrix this and it is, it turns out, so that it is exactly the similar type of expressions, that we are going to get, we had actually got when we had considered the population correlation population variance covariance matrix as sigma.

So, with this definition of sample cannonical variables and the sample cannonical correlation coefficients, we will just look at how to compute the sample cannonical correlation coefficients and the sample canonical, sample cannonical variables.

(Refer Slide Time: 56:01)



Now, let this rho 1 hat star square greater than or equal to rho 2 star hat square up to rho p hat star square be the Eigen values of the matrix, which is going to play the same role as the corresponding matrix in the population setup. So, it is S 1 1 to the power minus half S 1 2 S 2 2 to the power minus 1 S 2 1 S 1 1 to the power minus half.

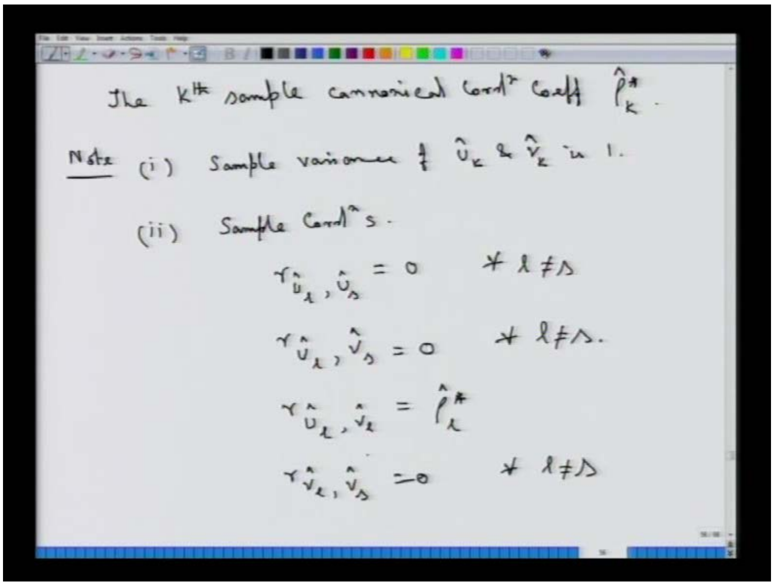
So, suppose these are the ordered Eigen values of this particular matrix with corresponding orthonormalized, with corresponding orthonormalized Eigen vectors as e 1 hat, e 2 hat and e p hat. The reason why we are putting hats here is that we are looking at the estimator of this quantity from the sample variance covariance matrix. The Eigen values are denoted by hats, the Eigen vectors are also denoted by the corresponding hat vectors.

Further, let this f 1 hat, f 2 hat and f q hat be the Eigen vectors orthonormalized, once again of course, orthonormalized Eigen vectors of S 2 2 to the power minus half S 2 1 S 1 1 to the power minus 1 S 1 2 S 2 2 to the power minus half matrix, where the first p of f 1 hat, f 2 hat, f q hat will be proportional to, similar to the population setup, will be proportional to S 2 2 to the power minus half S 2 1 S 1 1 to the power minus half times e k hat, because the first p non-zero Eigen values of this matrix, which is S 2 to the S 2 2 to the power minus half S 2 1 S 1 1 inverse S, 1 1, 1 2 S 2 2 to the power minus half will be same as the corresponding p non-zero Eigen values, the largest Eigen values of this particular matrix.

And the corresponding orthonormalized Eigen vectors, first p of them, the corresponding to the first p largest Eigen values of this matrix will be proportional to this particular element here. The kth sample cannonical variables are going to be given by u k hat, which is going to be given exactly in the same way as to what we had done for the population term. So, it is going to be e k hat times S 1 1 to the power minus half times this x vector and v k, the kth estimated sample cannonical variable is going to be given by f k hat vector S 2 2 to the power minus half times this y vector. So, this actually gives us a way to compute the sample cannonical variables in pairs.

Now, it has a striking resemblance, of course, with what we had for the population setup. For the population setup u k, the kth, the first component of the kth cannonical variable was given by, if you recall, e k times sigma 1 1 to the power minus 1 times this x vector and v k was given by this f k vector times sigma 2 2 to the power minus half times y vector. So, what we are doing here is just using the corresponding estimates and getting all the k cannonical variables.

(Refer Slide Time: 1:00:51)

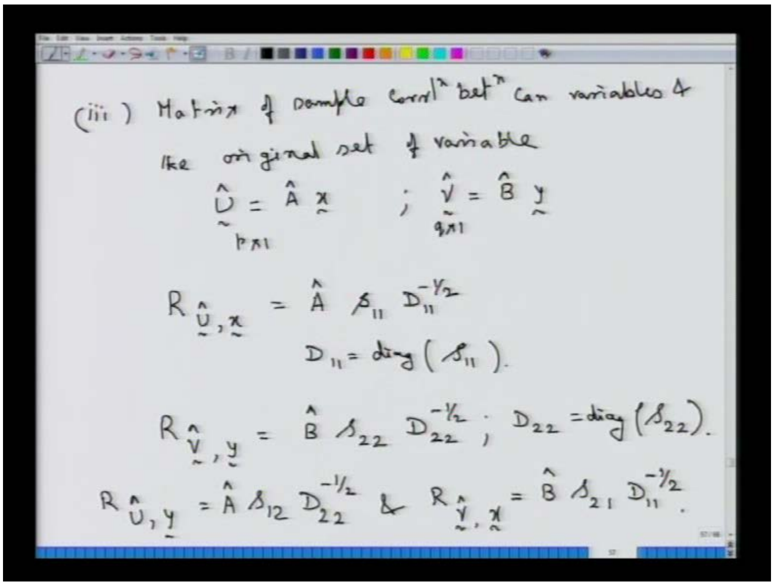


Now, all the results actually, that is what we had done for the population canonical correlation coefficient and the corresponding cannonical variables, they are applicable for the sample setup in particular. The important things, of course, what we are going to write in here now, after this of course, one just will recall, that when we had rho 1 star square, rho 2 star square, rho p star square as the population cannonical correlation coefficients, the cannonical correlation coefficients. Here, the kth cannonical correlation coefficient will just be equal to rho k star hat. So, before we proceed to the note we will just say that the kth sample cannonical correlation coefficient, the sample cannonical correlation coefficient is given by rho k star hat.

Now, we go on to this last few notes here, that once this sample, sample cannonical variables are constructed, the sample variance of u k hat and v k hat can be easily shown to be equal to 1, they are equal to 1. This follows from the orthogonality of e k hat vectors and f, f k hat vectors. Then, sample correlations between these constructed terms, say r u l hat and u s hat, this is going to be equal to 0 for every l not equal to s. Sample correlation between u l hat and v s hat, say, will be all so equal to 0 for every s not equal to l.

This sample correlation coefficient between u l hat and the corresponding v l hat, this will be equal to rho l hat star, and the similar terms can also be filled up. This is, say, between v l hat and v s hat, this will be equal to 0 for every l not equal to s and so on. So, this is, these similar type of results we had also got for the population quantity as well.

(Refer Slide Time: 1:03:32)



Now, the 3rd note is what we will say, that the matrix of sample correlations between the original variables and the cannonical variables, between the cannonical variables and the original set of variables, remember we had similar expression for the correlation matrix between the cannonical variables and the original set of variables, that is, rho u and x are rho v and y and the other combinations in the population setup of the cannonical variables.

Now, in case of the sample cannonical correlation, sample cannonical variables the sample correlation matrix between these cannonical variables and the original variables can similarly be obtained. So, they have the following form with the entire set of sample cannonical variables, say u hat, which would be given by A hat times this x vector and v hat would be given by this B hat times this y vector, this is going to be our p by 1 vector, this is going to be our q by 1 vector.

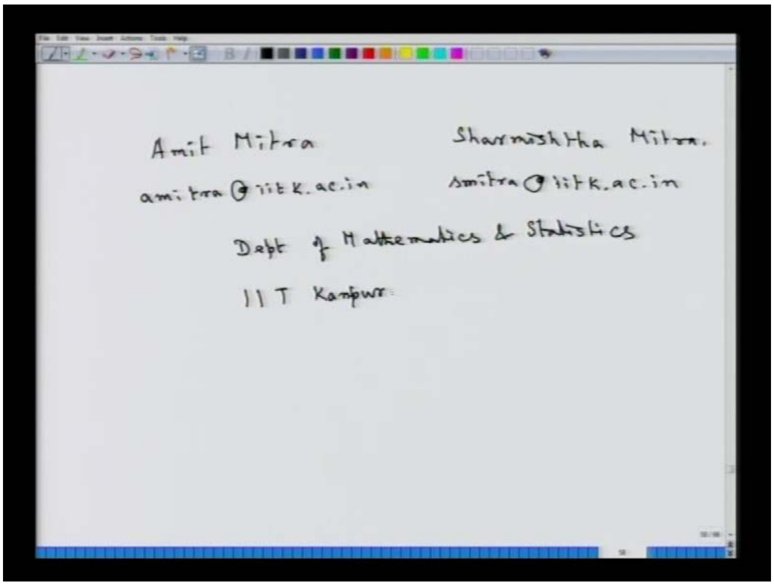
So, the sample correlation matrix between u hat vector and this x vector would be given in a similar expression to that of the population quantities with S 1 1 taking the place of sigma 1 1 and d 1 1 matrix taking the place of v 1 1 matrix, where this d 1 1 is the diagonal matrix, which is computed by choosing the diagonal elements of the s 1 1 matrix. And we will also have the correlation matrix between the other quantities, say, the correlation matrix of v hat vector with y vector. So, this y vector also, is that original set here in analogy with the population terms, this will just be given by S 2 2 matrix and this will be given by the d 2 2 matrix to the power minus half, wherein this d 2 2 matrix is the diagonal matrix, which holds the diagonal entries of the S 2 2 matrix.

We also have the cross terms, which is the correlation matrix between u hat and y vector, that would be given by A hat times S 1 2, which is taking the place of sigma 1 2 matrix, this is multiplied by d 2 2 to the power minus half. And the last thing is, that the correlation, the sample correlation matrix between this v hat vector and the x vector, that would be given by this B hat matrix, that multiplied by the S 2 1 matrix, that post multiplied by D 1 1 to the power minus half.

So, all the next set of other results and how to do these computations, therefore, exactly in the same way as to what we had done for the population cannonical correlation variables, which we had started using, either sigma matrix or the rho matrix. Now, here we have the sample, the estimated sample variance covariance matrix S and as we see in these notes, all the things, that we had done for the population cannonical variables they are exactly, they follow exactly in this same way for the sample cannonical variables, pairs and the sample cannonical correlation coefficients.

So, all these results actually just reminds us, that we basically had the similar type of expression for the population where S 1 1 was, in place of S 1 1 we had sigma 1 1 and so on. So, this thus concludes actually this particular course on applied multivariate analysis.

(Refer Slide Time: 1:07:42)



So, for further queries one can contact the instructors of this particular course. The instructors, I am Amit Mitra and my co-instructor in this course was Sharmishtha Mitra, one can contact us on our email ids for any queries, any clarification regarding any lecture in this course. This is email id of my co-instructor, at iitk dot ac dot in. One can also contact us on our department address, department of mathematics and statistics at IIT Kanpur. So, this concludes this course, on this video course on multivariate, applied multivariate analysis.

Thank you.