Multivariate Procedures with R

Prof. Shalabh

Department of Mathematics and Statistics

IIT Kanpur

Week – 12

Lecture – 55

Statistical Analysis of Canonical Variables

Hello friends, welcome to the course multivariate procedure with R. So, you can recall that in the last lecture we had introduced the concept of canonical correlation and canonical variables. I had just given you some brief idea that what are these things and how are we going to work on that and my objective was that because you already have learned the concept of principal component. So, extending a similar concept to a two sets of variable is not difficult. So, I wanted to make you comfortable in understanding the basic concept behind the canonical correlations and canonical variable. So, that we already have now done.

So, now in this lecture we are going to understand that how the statistical analysis related to canonical correlation is conducted. How the framework is created and what exactly do we do in order to execute those concepts which we had discussed in the last lecture. So, well I am going to take here a mathematical setup. I will not go into the mathematical details but I will try to give you the setup, the guidelines and what is the outcome.

You will see similar to the case of principal component if you try to recall we had done all the analysis but finally, it turned out that in order to solve the problem of principal component analysis essentially, we need to find out the eigenvectors and eigenvalues. The similar thing is going to happen in the case of canonical correlation also. But finally, the job will be to find out the eigenvectors and eigenvalues in a certain way of certain covariance matrices or which are transformed to correlation matrices. So, this will give you a hope that okay, that okay it may look a little bit mathematical but practically it is very simple to implement it. But in this lecture, I am going to take a very small example and I will try to manually show you that how these canonical variables and canonical correlations are obtained and in the next lecture I will try to show you how to implement it in the R software.

So, let us begin this lecture and try to understand it. Okay, so now in this chapter we are going to talk about the statistical analysis of canonical variable. So, our interest is that we want to measure the association between the two groups of variables. So, because there are two groups, so there are going to be two samples or two population and for that we have to define two random vectors. So, the first group of variables has suppose p variables which is represented by a $p \times 1$ random vector, so here x1 like this.

You can see here this superscript 1 that is indicating the population of the group 1 variable. So, that is how I write down here say x1 that is the first random variable corresponding to the group number 1. So, this 1 inside the parenthesis in the superscript is indicating the population number. So, similarly following the same notation I can say that the second group of variables has suppose q variables and which is represented by a $q \times 1$ random vector say x2. Right, so you have here variables x1, x2, xq and this 2 in the superscript inside the parenthesis is indicating that this is from the second group.

Now, between p and q one is going to be smaller, one is so what we try to assume that $p = q$. Actually, this is actually needed for the theoretical developments. So, we assume that x1 has less variable than x2. Well, in case if the opposite happen you simply have to interchange some notations, nothing more than or you rename the variables. So, now we have seen here, now we need to associate a mean vector and covariance matrix with the random vector in the first and second group.

So, we assume that here x1 that is the $p \times 1$ vector of the first group has mean vector mu1 which is indicated by here, subscript is given here as say, superscript is given here as say 1 inside the parenthesis and with covariance matrix here sigma11. Yeah, do not get confused that it is a sub matrix of sigma earlier we had partitioned the sigma like this here, but it is only here the covariance matrix of x1. Actually, the second group has q variables which are represented by a $q \times 1$ vector x2 and we assume that it has got a mean vector here mu2 and covariance matrix sigma22 and this covariance matrices are symmetric and positive definite. And the covariance between the x1 and x2 it is indicated by here is sigma12 which is the same as sigma21, this is of order $p \times q$. And we are assuming that x1 has less variable than x2.

So, this is our theoretical setup now. Now we can join this x1 and x2. So, we will have here say x = like as here like this we are trying to join this thing and we then the expected value of this here x will be become here mu which = here mu1 mu2 and the covariance matrix now this is sigma. So, remember the sigma is the covariance matrix of the entire sets of variables in x1 and x2 that is the p variable and q variable which is here given here by here sigma11, sigma12, sigma21 and sigma22. Since we are interested only in the variances and covariances.

Write $\underline{X} = \begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$, $E(\underline{X}) = \underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{bmatrix}$

$$E\left(\underline{X} - \underline{\mu}\right)\left(\underline{X} - \underline{\mu}\right)' = \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

So, without loss of any generality I can assume that expected value of x is 0 that is this mu is a null value vector. So, the covariance between the pairs of variables from the two sets are contained in this covariance matrix sigma. So, sigma is having p, q elements and these elements measure the association between the two sets. And now we understand that when p and q are large interpreting the elements of sigma12 collectively is extremely. In the case of physical component, I have shown you that if you try to increase the number of variables then how it affects the ah the number of elements in the variance covariance matrices right.

So, moreover it is often linear combination of variable that are that are interesting and useful for us for predictive and comparative performance they can be found right. Instead of those variables we are more interested in finding out the linear combinations of the variable and we believe that they will give us some interesting information for forecasting comparison etcetera. So, the main task of canonical call. So, the main task of canonical correlation analysis is to summarize the association between the variables in X1 and X2 sets in terms of a few carefully chosen covariances or equivalently the correlation instead of the entire p, q terms in the sigma12. And you have to ah recall that when you are talking of the covariance matrix if you try to make the observation scale then this becomes simply the correlation matrix also.

So, now we are more interested in the linear combination of the variables in X1 and X2 and we want to concentrate on the ah the covariances or equivalently the correlations. So, these linear combinations provide simple summary measures of a set of variables. So, let us try to assume ah two linear combination from both the sets X1 and X2. So, we consider that u = a transpose X1, and v = b transpose X2 they are the linear combinations from this X1 and X2 sets of variables. And this a and b they are the coefficient vectors, a is the coefficient vector in u and b is the coefficient vector in b.

So, now since we already have assumed the mean and variance for this X, so we can find out the mean and variance of this linear combinations u and v also. So, the variance of u in this case comes out to be here ah because we are assuming the mean to be 0. So, the variance is the same as expected value of u square because expected value of u is 0. So, this is simply here the transpose covariance matrix of X1 into a which is by our assumption is a transpose sigma11 a. And similarly, variance of v is said simply expected value of v square which is b transpose covariance matrix of X2 into b which is b transpose sigma22 b.

$$\boldsymbol{Var(U) = E(U^2) = \underline{a}'Cov(\underline{X}^{(1)})\ \underline{a} = \underline{a}'\Sigma_{11}\ \underline{a}}$$

$$\boldsymbol{Var(V) = E(V^2) = \underline{b}'Cov(\underline{X}^{(2)})\ \underline{b} = \underline{b}'\Sigma_{22}\ \underline{b}}$$

$$\boldsymbol{Cov(U,V) = E(UV) = \underline{a}'Cov(\underline{X}^{(1)},\underline{X}^{(2)})\ \underline{b} = \underline{a}'\Sigma_{12}\ \underline{b}}$$

And the covariance between u and v it is ah expected value of u v which is the same here as a transpose covariance matrix between X1 and X2 into b which is here a transpose sigma12 b. So, now ah you can recall that if you have the information of variance and covariance then you can find out the correlation coefficient also. And yeah, you have to remember one thing all this a transpose sigma11 a b transpose sigma22 b they all are actually scalars they are not the ah matrix right. So, because u and v they are they are also scalars. So, now the correlation between u and v is written here as a like this a transpose sigma12 b divided by standard deviations of a transpose sigma11 a and b transpose sigma22 b.

$$\boldsymbol{Corr(U,V) = \frac{\underline{a}'\Sigma_{12}\ \underline{b}}{\sqrt{\underline{a}'\Sigma_{11}\ \underline{a}}\sqrt{\underline{b}'\Sigma_{22}\ \underline{b}}}}$$

The question here is now we do not know the values of here a and b. In case if I know the value of a and b then I note then I also know u and v also and the question is now how to find out here a and b. So, now we try to find out this a and b in such a way such that this correlation between u and v is as large as possible. So, essentially, we have to optimize this correlation u and v such that this value becomes maximum and whatsoever be the value of a and b for which this correlation is becoming maximum that will give us the solution in which we are interested. So, now based on this idea we try to define the correlation coefficient and canonical variables.

So, please try to understand how I am trying to do it and then I will try to extend it. So, the first pair of canonical variables is actually the pair of linear combination u and v such that they have got a unit variance, variance of u = 1 and variance of v = 1 and which maximizes the correlation between u and v and whatever that value comes out we try to indicate it by here u1 and v1 right. And this variance is taken to be here as a1 so that we can have a normalized linear combination. This is the same concept if you try to see we had used in the principle component also. So, now we have obtained the u1 and v1.

Now, in the second step we would like to find out the second pair of canonical variables which is indicated by here u2 and v2. So, they are also the linear combination having a unit variance that a variance of u2 = 1 and variance of v2 = 1 and they maximizing the correlation coefficient. So, this a and b are going to be chosen in such a way such that they are maximizing the correlation coefficient between u and v among all choices which are uncorrelated with the first pair of canonical variables u1 and v1. So, when you are trying to find out here u2 and v2 so this has to be find in such a way such then the correlation between u2 and v2 is maximum yes, I mean the first maximum has already been occupied by the first canonical first pair of canonical variables. So, now, what is going to be the second largest value corresponding to which what is going to be the value of here a and b they will give us u2 and v2, but now there is additional condition that this u2 and v2 has to be uncorrelated with u1 and v1.

So, this is a condition which is imposed while deriving the u2 and v2. Well, I am not going to demonstrate here the mathematical setup, but this is how we try to find it out. And this process is continued and at the kth step the kth pair of canonical variables is found in such a way which is having like here ak and bk here as the coefficient. So, they are obtained in such a way such that the variance of uk = 1 variance of vk = 1 and they maximize the correlation coefficient call between u and v among all choices which are uncorrelated with the previous k - 1 canonical variable pairs. So, the way I have told you that first you try to find out u1 v1 then u2 v2 and you ensure or you impose that u2 v2 is independent of u1 v1.

Similarly, when you try to go for u3 v3 then u3 v3 are found in such a way such that they are independent of both and the first two canonical variable. And similarly if you come here say here at the end say uk vk then uk vk are found in such a way such that they are uncorrelated with all the first k - 1 pair of canonical variables right. The correlation between the kth pair of canonical variable uk and vk is called the kth canonical correlation right. So, whatsoever be the first canonical variable its correlation will be called as first canonical correlation for the second pair of canonical variable whatsoever be its correlation coefficient that will be called as second correlation coefficient and so on. So, you can see here it is not a very difficult.

So, now if I try to translate it in mathematical framework so suppose p <= q then what are we trying to find? We are trying to find out here the maximum correlation between u and v with respect to a and v and suppose this value comes out to be here rho1 star. So, this is attained by the linear combination u1 and v1 which are called as the first pair of canonical variates and it is like this u1 = a1 transpose x1 and v1 = b1 transpose x2 where this a1 and b1 they are the values of a and b such that this correlation u v is maximum that means in this case this correlation value is rho1 star. So, whatever are the values of a and b which are going to maximize this correlation coefficient between u and v they are indicated here as say rho1 star. Now similarly if you try to go for the second pair of canonical variates we can proceed exactly in the same way they are indicated by here u2 and v2. So, now you are trying to find out this u2 here is like a transpose x1 and say v2 will be here b transpose x2 where this a and b they are unknown.

Suppose $p \leq q$. Then

$$\max_{\underline{a},\underline{b}} \boldsymbol{Corr(U,V)} = \boldsymbol{\rho_1^*}$$

attained by the linear combinations U_1 and V_1– first pair of canonical variates – is $U_1 = \underline{a}_1'\underline{X}^{(1)}$ and $V_1 = \underline{b}_1'\underline{X}^{(2)}$ where $\underline{a}_1$ and $\underline{b}_1$ are the values of $\underline{a}$ and $\underline{b}$ for which the $Corr(U,V)$ is maximum.

The second pair of canonical variates is U_2 and V_2 such that

$$\max_{\underline{a},\underline{b}} \boldsymbol{Corr(U,V)} = \boldsymbol{\rho_2^*} \text{ with } {\rho_1^*}^2 \geq {\rho_2^*}^2,$$

$U_2 = \underline{a}_2'\underline{X}^{(1)}$ and $V_2 = \underline{b}_2'\underline{X}^{(2)}$ and (U_2, V_2) is uncorrelated with (U_1, V_1) where $\underline{a}_2$ and $\underline{b}_2$ are the values of $\underline{a}$ and $\underline{b}$ for which the $Corr(U,V)$ is maximum.

So, this a and b are found in such a way such that the correlation between u and v with respect to a b that is maximum from the remaining correlation coefficients. So, suppose this is equal to here rho2 star. So, rho2 star is something like second largest correlation coefficient and there is a condition that the first correlation coefficient rho1 is greater than the second correlation coefficient because rho can be in negative also. So, we are trying to use here the rho1 square is greater than or = rho2 square. So, now and then this whatever be the value of a and b suppose this value comes out to be here a2 and b2 that is corresponding to u2 and v2.

In general, $U_k = \underline{a}'_k \underline{X}^{(1)}$ and $V_k = \underline{b}'_k \underline{X}^{(2)}$, $k = 2, 3, \dots p$, is the k^{th} pair of canonical variables such that $Var(U_k) = E(U_k^2) = 1$ and $Var(V_k) = E(V_k^2) = 1$, i.e., having unit variances which maximises the correlation coefficient $Corr(U, V)$ among all choices which are uncorrelated with the previous (k - 1) canonical variable pairs, i.e.,

$$\max_{\underline{a},\underline{b}} Corr(U, V) = \rho_k^* \text{ with } {\rho_1^*}^2 \geq {\rho_2^*}^2 \geq \cdots \geq {\rho_k^*}^2 .$$

Here $U_k = \underline{a}'_k \underline{X}^{(1)}$ and $V_k = \underline{b}'_k \underline{X}^{(2)}$ and (U_k, V_k) is uncorrelated with $(U_1, V_1), (U_2, V_2), \dots, (U_{k-1}, V_{k-1})$, where $\underline{a}_k$ and $\underline{b}_k$ are the values of $\underline{a}$ and $\underline{b}$ for which the $Corr(U, V)$ is maximum.

So, I can create here u2 = a2 transpose x1 and v2 = b2 transpose x2 and u2 v2 is uncorrelated with u1 v1. So, this is what we have done we have found the values of a2 v2 which are the values of a and b such that the correlation between u and v is maximum and they are uncorrelated with the first pair of canonical variates u1 v1. So, in general if I come at the kth step what happened we are trying to find out the value of a and b like a suppose uk is a transpose x1 and vk here is b transpose x2 then we try to find out the value of a and b such that we have here uk = ak transpose x1 and vk = bk transpose x2 for all k goes from 2 to 3 such that the variance of uk is 1 variance of vk is 1 and a and b have been found in such a way which are maximizing the map correlation coefficient between u and v among all choices which are uncorrelated with the previous k-1 canonical variables pair. That is the value of the correlation coefficient at the kth stage is rho k star which is obtained by the values a = ak and b = bk and this correlation coefficient are such that that the rho1 square >= rho2 square is greater than up to = rho k square rho k star square. So, the idea is that the first correlation coefficient this is the maximum and then the last correlation coefficient will be the minimum.

$$Var(U_k) = 1\,, Var(V_k) = 1,$$

$$Cov(U_k, U_l) = Corr(U_k, U_l) = 0, k \neq l$$

$$Cov(V_k, V_l) = Corr(V_k, V_l) = 0, k \neq l$$

$$Cov(U_k, V_l) = Corr(U_k, V_l) = 0, k \neq l$$

So, we are trying to find out here uk and vk in such a way such that uk vk is uncorrelated with u1 v1 u2 v2 up to uk - 1 vk - 1 right and here ak and bk have been obtained such that the correlation between u and v is maximum right. So, now if you try to comprehend it what are the different conditions we have used here, we have used here or we have imposed here condition that variance of uk should be = 1 variance of vk should be = 1 covariance between uk ul covariance between and the correlation between uk ul will be = 0 right. Because once covariance = 0 then the correlation will also becomes 0. The covariance between vk and vl which is equivalently the correlation between vk and vl is 0 and the covariance between uk and vl which is similar to correlation between uk and vl = 0 for all k not = l and k and l goes from 1 to p right. So, these are the different condition which we have used, but now if you try to translate all these concepts into a statistical framework and try to understand the mathematics behind it then it goes like this that when we are trying to consider here uk and vk which are ak transpose x1 and bk transpose x2 for k goes from 1 to p the coefficients here ak and bk they are the solution of this characteristic equation.

In general, in $U_k = \underline{a}_k' \underline{X}^{(1)}$ and $V_k = \underline{b}_k' \underline{X}^{(2)}$, $k = 1, 2, 3, \ldots p$,

the coefficients $\underline{a}_k$ and $\underline{b}_k$, $k = 1, 2, \ldots p$, are the solutions of

$$\left(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^{*2}\Sigma_{11}\right)\underline{a} = 0$$

and

$$\left(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \rho^{*2}\Sigma_{22}\right)\underline{b} = 0$$

respectively and satisfy $|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^{*2}\Sigma_{11}| = 0$.

The quantities ρ^{*2} are the values $\rho_1^{*2}, \rho_2^{*2}, \ldots, \rho_p^{*2}$ which are the

eigen values obtained from $|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^{*2}\Sigma_{11}| = 0$.

Yes, you can see this and here this. These are two characteristic equation sigma12 sigma22 inverse sigma21 - rho star square sigma11 a = 0. So, a is obtained from this characteristic equation and b is obtained from this characteristic equation sigma21 sigma11 inverse sigma12 - rho star sigma22 b and we have to find that the determinant of sigma12 sigma22 inverse sigma21 - rho star square sigma11 = 0. So, if you try to see what you have done, you simply have taken this characteristic equation and this and then you have obtained the values of rho star square from this characteristic equation. So, you have taken this and because you have to taken here p <= q. So, from this equation you are going to get here the values rho1 square rho1 star square rho2 star square rho p star square.

And corresponding to these values of rho square if you try to obtain the eigenvectors a and b from these two equations 1 and here 2. So, if you try to see if I try to explain you very briefly the first step will be obtain eigenvalues from here. The second step will be use this eigenvectors to obtain the eigenvectors eigenvalues and obtain a and b from equation number 1 and 2 respectively that is all. And now what is ever be the value of ak and bk which you have obtained in uk = ak transpose x1 and vk = vk transpose x2 they are the eigenvectors. So, whatever a and b you have obtained here these are the eigenvectors and this is the eigenvectors.

And then based on that they are the values of ak and bk. So, the values of ak and bk are obtained as the values of the kth eigenvectors ak and bk and ak and bk are the eigenvectors corresponding to rho star square k that is the kth value. So, now I can say here that the coefficient a are the solution a1, a2, a p which satisfy this characteristic equation for rho star square = rho1 square rho2 square up to rho p square. So, this a1, a2, a p are the corresponding eigenvectors that is ak is the eigenvector corresponding to rho k star square for all k goes from 1 to p. So, similar a thing you can do for the second function also and the coefficient v are the solution v1, v2, v p which are obtained from the characteristic equation this one for the values of rho star square = rho1 square star square rho2 star square and rho p star square.

The coefficients $\underline{a}_k$ and $\underline{b}_k$ in $U_k = \underline{a}'_k\underline{X}^{(1)}$ and $V_k = \underline{b}'_k\underline{X}^{(2)}$, $k = 1, 2, 3, \ldots p$, are obtained as the eigen vectors.

The coefficients $\underline{a}$ are the solutions $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_p$ which satisfy

$(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^{*2}\Sigma_{11})\underline{a} = 0$ for $\rho^{*2} = {\rho_1^*}^2, {\rho_2^*}^2, \ldots, {\rho_p^*}^2$

where $\underline{a} = (\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_p)'$.

So $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_p$ are the corresponding eigen vectors, i.e., $\underline{a}_k$ is the eigen vector corresponding to ${\rho_k^*}^2, k = 1, 2, \ldots, p$.

So, you can see here this v1, v2, vp are the corresponding eigenvector that is vk is the eigenvector corresponding to rho k star square for k goes from 1 to p. So, if you see here you started with here uk and vk which were depending on a and here b and those values we are obtained here as ak and bk, but now this ak and bk they are simply the eigenvectors corresponding to say rho star square k that is all. So, what you have to do? First you have to write down the characteristic equation from there you try to find out the values of rho star square and then using each of the value of rho star square try to find out the corresponding eigenvector. And now you can see what are what you have done? You have obtained here the p values of rho stars mean the collision coefficient. Now, some values will be less, some values will be more and one of them will be the maximum and say another will be minimum.

The coefficients $\underline{b}$ are the solutions $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_p$ which satisfy

$$\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} - \rho^{*2}\boldsymbol{\Sigma}_{22}\right)\underline{b} = 0 \text{ for } \rho^{*2} = {\rho_1^*}^2, {\rho_2^*}^2, \ldots, {\rho_p^*}^2$$

where $\underline{b} = (\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_p)'$.

So $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_p$ are the corresponding eigen vectors, i.e., $\underline{b}_k$ is the eigen vector corresponding to ${\rho_k^*}^2, k = 1, 2, \ldots, p$.

So, now you can order them and whatsoever be the maximum value out of this rho1 square, rho2 star square etcetera that will become the correlation coefficient of the first canonical variate or in simple word the maximum value out of this will be the first canonical correlation and corresponding to which whatever eigenvector you are going to obtain that will give you the kth pair of canonical variate. So, if you try to see the whole process is very simple. Now, if I try to show you that what really I mean suppose first group has 5 variables say x1, x2, x3, x4, x5 and second group has 7 variables x1, x2, x3 up to x7. So, the number of canonical variables that can be obtained from here is the minimum of p and q which is here 5. So, I try to find out I try to use here this characteristic equation and from this equation I try to find out here the values of here say rho1 square, rho2 square, rho5 square which I am indicating here as a lambda1, lambda2, lambda3, lambda4, lambda5 right.

Now, this lambda1, lambda2, lambda3, lambda4, lambda5 they have been arranged in such a way so that lambda1 is greater than lambda2, greater than lambda3, greater than lambda4 and greater than lambda5. So, I can say in some word here that this lambda1 is the maximum value among rho1 star square up to here rho star square here p and or here in this case 5 and lambda5 is here the minimum value between rho1 star square up to here rho5 star square. Now, you have here 5 values of rho square up to here rho square also equivalently lambdas. So, corresponding to each eigenvalue find out the eigenvectors say a1, a2, a3, a4, a5 and b1, b2, b3, b4, b5 and based on that try to create here the uk as ak transpose x1 and vk = bk transpose x2 right. And then here first pair of u1, v1 that will be based on the eigenvalues lambda1 which is the maximum correlation.

Suppose first group has p = 5 variables $\underline{X}^{(1)} = (X_1^{(1)}, X_2^{(1)}, \ldots, X_5^{(1)})'$

Second group has q = 7 variables $\underline{X}^{(2)} = (X_1^{(2)}, X_2^{(2)}, \ldots, X_7^{(2)})'$

Number of canonical variables = $min(p, q)$ = 5

Suppose the eigen values are found to be

$$\lambda_1 = 0.55, \lambda_2 = 0.20, \lambda_3 = 0.12, \lambda_4 = 0.08, \lambda_5 = 0.05$$

$$\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \lambda_5$$

Corresponding to each eigen value, find the eigen vectors $\underline{a}_1$, $\underline{a}_2$, $\underline{a}_3$, $\underline{a}_4$, $\underline{a}_5$ and $\underline{b}_1$, $\underline{b}_2$, $\underline{b}_3$, $\underline{b}_4$, $\underline{b}_5$.

Then find $U_k = \underline{a}_k'\underline{X}^{(1)}$ and $V_k = \underline{b}_k'\underline{X}^{(2)}$, $k = 1, 2, 3, 4, 5$.

So, this u1, v1 they will be called as first pair of canonical variates and so on right. So, now if you try to see here how much is the relative contribution of this lambda1, lambda2, lambda3, lambda4, lambda5. So, if you try to choose the first two canonical pairs of variables then they take care here this 0.55 + 0.20. And if you try to divide it by the 2 to this lambda1 + lambda2 + lambda3 + lambda4 + lambda5 then this is 0.75 which is equivalent to 75 percent of the correlation. Similarly, if you try to choose here first three pairs of canonical variation then you can see here this is here coming out to be like this and this is here the lambda1 + lambda2 + lambda3 divided by the sum of here, sum of here lambda i. This is here close to 87 percent of the total correlation and similarly if you try to choose all the value lambda1, lambda2, lambda3, lambda4 and lambda5 this will come out to be 1 because that is taking care of the maximum value. So, now and so you can see here that first two canonical variables can take care of 75 percent of the correlation, first three canonical variables can take care of the 87 percent of the correlation and first four canonical variables can take care of the 95 percent of the correlation.

Based on $\lambda_1 = 0.55, \lambda_2 = 0.20, \lambda_3 = 0.12, \lambda_4 = 0.08, \lambda_5 = 0.05$

- **If we choose, first two canonical variables, then they take care of**

$$\frac{0.55+0.20}{0.55+0.20+0.12+0.08+0.5} = 0.75 \equiv 75\% \text{ of the correlation.}$$

- **If we choose, first three canonical variables, then they take care of**

$$\frac{0.55+0.20+0.12}{0.55+0.20+0.12+0.08+0.5} = 0.87 \equiv 87\% \text{ of the correlation.}$$

- **If we choose, first four canonical variables, then they take care of**

$$\frac{0.55+0.20+0.12+0.08}{0.55+0.20+0.12+0.08+0.5} = 0.95 \equiv 95\% \text{ of the correlation.}$$

So, now it is up to you that how do you want to choose it, it is similar to choosing the principal component in the number of principal component in the principal component analysis, but I will try to show you some technique by which you can do it in the R software. Now, the next question comes here how are you going to estimate it on the basis of given sample of data. So, this is again very simple just like as in the case of principal component analysis, right. So, we consider here the maximum likelihood estimation of this canonical correlation coefficient and its variable. So, suppose x1, x2, x n be a random sample from normal mu sigma and we try to partition x here as x1, x2 mean here as mu as mu1, mu2 and random observation x alpha as here x alpha1 and x alpha2 and the covariance matrix here is sigma which is sigma11, sigma12, sigma21, sigma22.

So, now if I you try to see here the maximum likelihood estimator of mu that is straightforward that will be here simply the sample mu, right like this one and the maximum likelihood estimator of sigma will be like as here, right that we already have discussed couple of time. So, this is the covariance matrix based on the observation x1 and this is here the covariance matrix which is based on the observation base 2 and these are the covariance matrices which are based on the sample observation from x1 and x2, right. Now, we have to first find out the maximum likelihood estimators of rho1 star square, rho2 star square etcetera. So, the maximum likelihood estimators of rho1 star square, rho2 star square, rho p star square are the roots of this equation, right and ak and bk are obtained to satisfy these conditions. So, if you try to see here I have replace here whatsoever I am obtaining the value of here rho star square from here this is substituted here you can see here and all this sigma11, sigma22, sigma12 they are replaced by their estimators, sample based estimators the maximum likelihood estimators.

Let $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ $(n > p)$ be a random sample from $N\left(\underline{\mu}, \Sigma\right)$.

Write $\underline{X} = \begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$, $E(\underline{X}) = \underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{bmatrix}$, $\underline{x}_\alpha = \begin{bmatrix} \underline{x}_\alpha^{(1)} \\ \underline{x}_\alpha^{(2)} \end{bmatrix}$, $\alpha = 1, 2, \dots, n$

$$E\left(\underline{X} - \underline{\mu}\right)\left(\underline{X} - \underline{\mu}\right)' = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, p < q.$$

The maximum likelihood estimators of $\underline{\mu}$ and Σ are $\underline{\bar{x}} = \begin{bmatrix} \underline{\bar{x}}^{(1)} \\ \underline{\bar{x}}^{(2)} \end{bmatrix}$ **and**

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{\alpha=1}^{n}\left(\underline{x}_\alpha^{(1)} - \underline{\bar{x}}^{(1)}\right)\left(\underline{x}_\alpha^{(1)} - \underline{\bar{x}}^{(1)}\right)' & \frac{1}{n}\sum_{\alpha=1}^{n}\left(\underline{x}_\alpha^{(1)} - \underline{\bar{x}}^{(1)}\right)\left(\underline{x}_\alpha^{(2)} - \underline{\bar{x}}^{(2)}\right)' \\ \frac{1}{n}\sum_{\alpha=1}^{n}\left(\underline{x}_\alpha^{(2)} - \underline{\bar{x}}^{(2)}\right)\left(\underline{x}_\alpha^{(1)} - \underline{\bar{x}}^{(1)}\right)' & \frac{1}{n}\sum_{\alpha=1}^{n}\left(\underline{x}_\alpha^{(2)} - \underline{\bar{x}}^{(2)}\right)\left(\underline{x}_\alpha^{(2)} - \underline{\bar{x}}^{(2)}\right)' \end{bmatrix}$$

respectively.

And then we are trying to impose a similar condition that we had condition ak transpose sigma11 Ak = 1, so that my normal, so that my linear combination has normalized. So, I am replacing by their estimators. So, we have two conditions that ak hat transpose sigma11 hat ak and bk hat transpose sigma22 hat bk both are = 1 and bk hat transpose k sigma22 hat Al hat transpose hat this value 1 and 0. So, for all the values p = l, what is now p + 1 to q, right. Because you can recall that we have p elements in one vector and q elements in another vector.

The maximum likelihood estimators of $\rho_1^{*2}, \rho_2^{*2}, \ldots, \rho_p^{*2}$ are the roots of $\begin{vmatrix} -\rho^{*2}\widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{21} & -\rho^{*2}\widehat{\Sigma}_{22} \end{vmatrix} = 0$, and $\underline{a}_k$ and $\underline{b}_k$ are obtained to satisfy

$$\begin{pmatrix} -\widehat{\rho_k^{*2}}\widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{21} & -\widehat{\rho_k^{*2}}\widehat{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \widehat{\underline{a}}_k \\ \widehat{\underline{b}}_k \end{pmatrix} = 0, \qquad k = 1, 2, \ldots, p$$

$$\widehat{\underline{a}}_k' \widehat{\Sigma}_{11} \widehat{\underline{a}}_k = 1,$$

$$\widehat{\underline{b}}_k' \widehat{\Sigma}_{22} \widehat{\underline{b}}_k = 1,$$

$$\widehat{\underline{b}}_k' \widehat{\Sigma}_{22} \widehat{\underline{a}}_l = \begin{cases} 1 & k = l = p+1, p+2, \ldots, q \\ 0 & k \neq l = p+1, p+2, \ldots, q. \end{cases}$$

So, we have obtained the canonical variance only for the p. So, what we have to do for q. So, this is trying to indicate that, right. So, as you can see here finding out the maximum likelihood estimator is not difficult, you simply try to take the data try to and then try to find out the maximum likelihood estimators of mu and sigma. Substitute them in the characteristic equation and characteristic and then find out the characteristic roots and then whatsoever be the root corresponding to them try to find out the characteristic vectors and whatsoever value you get that will be the maximum likelihood estimator of canonical correlation and canonical variables.

Now, we try to show you this through. So, in this example actually the head length and head breadth of two children are collected in a sample of 25 families, right. So, for example, if you try to see here that we have here four variables x1, x2, x3, x4 and then there is x1 and x2 they are indicating the head length and head breadth of the first child in alpha-th family. And similarly, x3 and x4 they are indicating the head length and head breadth of the second child in the alpha-th family and alpha goes from 1 to 25. And here the aim is to investigate the relation between the measurement for the first child and for the second child. So, we try to create here two data vectors here x alpha1 and x alpha2 here, x alpha1 has x1 alpha and x2 alpha observation and x alpha2 has x3 alpha and x4 alpha, right.

The head length and head breadth of two children are collected in a sample of 25 families. Let for $\alpha = 1, 2, \dots, 25$,

$x_{1\alpha}$: Head length of the first child in α^{th} family,

$x_{2\alpha}$: Head breadth of the first child in α^{th} family,

$x_{3\alpha}$: Head length of the second child in α^{th} family, and

$x_{4\alpha}$: Head breadth of the second child in α^{th} family.

The aim is to investigate the relations between the measurements for the first child and for the second child.

Thus $\underline{x}_\alpha^{(1)} = (x_{1\alpha}, x_{2\alpha})'$ and $\underline{x}_\alpha^{(2)} = (x_{3\alpha}, x_{4\alpha})'$.

Now, the sample has been obtained and based on that the summarized data is here like this that sample mean vector is here like this and the sample covariance matrix has been obtained here like this. So, you can see here this is like this here for this x1 and x2. So, this is your here say a11, this is a12, this is a21 and this is here a22, right. And based on that the correlation matrix has been obtained. Well, in R software if you simply use the scale format of the data then this covariance matrix can be converted into the correlation matrix.

So, this correlation matrix R here is obtained. So, it is something like this on the diagonal element this is the correlation coefficient between x1 and x1. So, this is 1 and the off-diagonal elements and for example, this is here the correlation between x1 and say here x2, right. So, this is and this this matrix here has been indicated by here say R11, R12, R21, R22. So, this is R11, this is here R22, this is here R12 and this is here R21, right.

The sample data is summarized as sample mean vector and sample covariance matrix as

$$\underline{\bar{x}} = \begin{bmatrix} 185.72 \\ 151.12 \\ 183.84 \\ 149.24 \end{bmatrix}, \quad S = \frac{1}{24}A = \begin{bmatrix} 95.29 & 52.86 & 69.66 & 46.11 \\ 52.86 & 54.36 & 51.31 & 35.05 \\ 69.66 & 51.31 & 100.81 & 56.54 \\ 46.11 & 35.05 & 56.54 & 45.02 \end{bmatrix}$$

The correlation matrix (scaled data of covariance matrix) is

$$R = \begin{bmatrix} 1.00 & 0.73 & 0.71 & 0.70 \\ 0.73 & 1.00 & 0.69 & 0.71 \\ 0.71 & 0.69 & 1.00 & 0.84 \\ 0.70 & 0.71 & 0.84 & 1.00 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

Now, we try to find out the characteristic roots. They are obtained here as say rho1 star square = this and rho2 square star = this. So, that I can say here rho1 hat is like this, rho2 hat is like this. Well, one thing I am not saying at all that you can find out the estimator of rho1 star just by taking the square root of the estimator of rho1 star square, but this is only one possible way out to consider the feasible version otherwise you have to obtain it separately. So, now we have obtained here the two values of here correlation coefficient rho1 and rho2 on the basis of given set of data and you can see here that this rho1 star hat is much much larger than rho2 hat star. But anyway, first we try to find out the eigenvectors corresponding to rho1 star and rho2 star and they are here obtained here like this.

The characteristics roots are obtained as $\widehat{\rho_1^*}^2 = 0.622$,

$\widehat{\rho_2^*}^2 = 0.622$, so that $\widehat{\rho_1^*} = 0.789$, $\widehat{\rho_2^*} = 0.054$.

Then for

$$\widehat{\rho_1^*} = 0.789, \quad \underline{\hat{a}}_1 = \begin{pmatrix} 0.057 \\ 0.071 \end{pmatrix}, \quad \underline{\hat{b}}_1 = \begin{pmatrix} 0.050 \\ 0.080 \end{pmatrix};$$

$$\widehat{\rho_2^*} = 0.054, \quad \underline{\hat{a}}_2 = \begin{pmatrix} 0.140 \\ -0.187 \end{pmatrix}, \quad \underline{\hat{b}}_2 = \begin{pmatrix} 0.176 \\ -0.262 \end{pmatrix}.$$

Since $\widehat{\rho_1^*} = 0.789 > \widehat{\rho_2^*} = 0.054$. So we can confine our attention to the first canonical variates to study the relation between two head dimensions of first child and second child. The second canonical variates are weakly correlated.

Corresponding to rho1 star these are the eigenvectors corresponding to which we get here the values of a1 hat and b1 hat here like this. Similarly, for the second value of rho2 star hat we have here the eigenvectors and based on that we obtain here the a2 hat and b2 hat here like this. So, now I can show you here that because rho1 star hat is much much greater than the rho2 star hat. So, we can confine our attention to the first canonical variates to study the relation between two hat dimension of first side and second side right. The second canonical variates whereas they are weakly correlated right and based on that if you try to see I can write down here the first and second pair of canonical variates using here this a1 hat and b1 hat I can write down here u1 = 0.057 times x1 + 0.071 times x2 and if p 1 will be your here 0.50 x1 + 0.80 times x2. And similarly using this here a2 we can write down here u2 and using here b2 hat we can write down here v2 right. So, we are going to consider only here the first pair of canonical variate which is this is the u1 and v1 right. So, now we come to an end to this lecture and you can see here that we have taken a very small example to explain you that what is really going to happen.

$\widehat{\rho_1^*} = 0.789, \quad \underline{\hat{a}}_1 = \begin{pmatrix} 0.057 \\ 0.071 \end{pmatrix}, \quad \underline{\hat{b}}_1 = \begin{pmatrix} 0.050 \\ 0.080 \end{pmatrix};$

$\widehat{\rho_2^*} = 0.054, \quad \underline{\hat{a}}_2 = \begin{pmatrix} 0.140 \\ -0.187 \end{pmatrix}, \quad \underline{\hat{b}}_2 = \begin{pmatrix} 0.176 \\ -0.262 \end{pmatrix}.$

The canonical variables are as follows:

$U_1 = 0.057X_1 + 0.071X_2, \quad V_1 = 0.050X_1 + 0.080X_2$

$U_2 = 0.140X_1 - 0.187X_2, \quad V_2 = 0.176X_1 - 0.262X_2$

The first set of canonical variable s are as follows:

$U_1 = 0.057X_1 + 0.071X_2, \quad V_1 = 0.050X_1 + 0.080X_2$

And I have tried my best to very briefly explain you the statistical and mathematical framework behind this canonical correlation analysis. And I have given you the concept and finally, the we have learned that it is equivalent to finding out the canonical correlation and canonical variables by finding out the eigenvalues and eigenvectors of certain covariance matrices. And based on that you can take this call. Well in this example I have taken only these 4 variables and then the data is summarized. So, my idea was to give you an idea that how you can obtain from the eigenvectors and eigenvalues you can obtain the canonical vectors and canonical correlations.

But definitely it has to be done on the R software only and in R software you will have different types of options by looking into the graphical way, analytical way and in R software it is very easy to implement. But the main challenge comes that what is that outcome indicating. So, now once you have understood the basic fundamental behind the canonical correlation analysis, I am confident that it will not be difficult for you to understand its implementation in the R software. Basically, you need only a command that how you can do this canonical correlation analysis is not it? Yes. So, you try to have a look into this lecture try to understand the basic concepts and try to understand them.

They will help you in understanding the concept in the next lecture. So, you try to practice it and I will see you in the next lecture till then good bye.