**INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR**

**NPTEL
National Programme
on
Technology Enhanced Learning**

Applied Multivariate Statistical Modeling

**Prof. J. Maiti
Department of Industrial Engineering and Management
IIT Kharagpur**

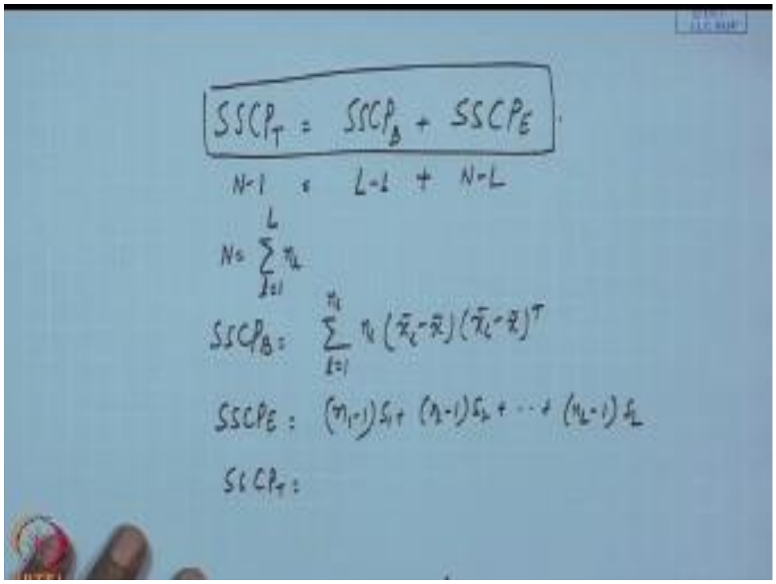
Lecture – 17

Topic

Multivariate Analysis of Variance (MANOVA) (Contd.)

Good morning. We will continue MANOVA Multivariate Analysis of Variance. What we have seen in last class.

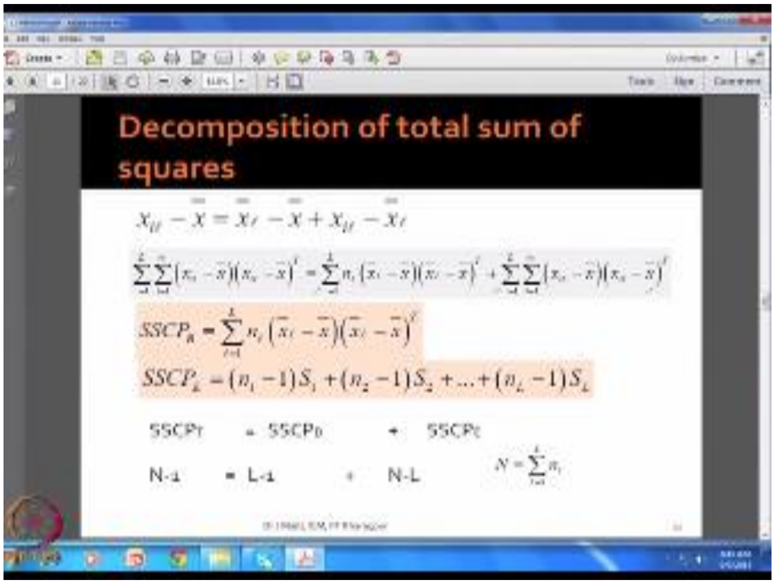
(Refer Slide Time: 00:28)



That we have decomposed SSCP total into two parts, SSCP between population plus SSCP error. We have also discussed that the degree of freedom for SSCP total is N -1, where N = L= 1 to capital L n_l, and there are L populations. So, the degrees of freedom for SSCPB is L - 1and rest the total of $SSCP_B$ and $SSCP_E$ degrees of freedom will be n - L, so the rest will be N - L. And as I told you in last class that $SSCP_B$ will be computed using this formulation l = 1 to n_l $\bar{x}_l$ - $\bar{x}$ and $\bar{x}_l$ - $\bar{x}$ transpose.

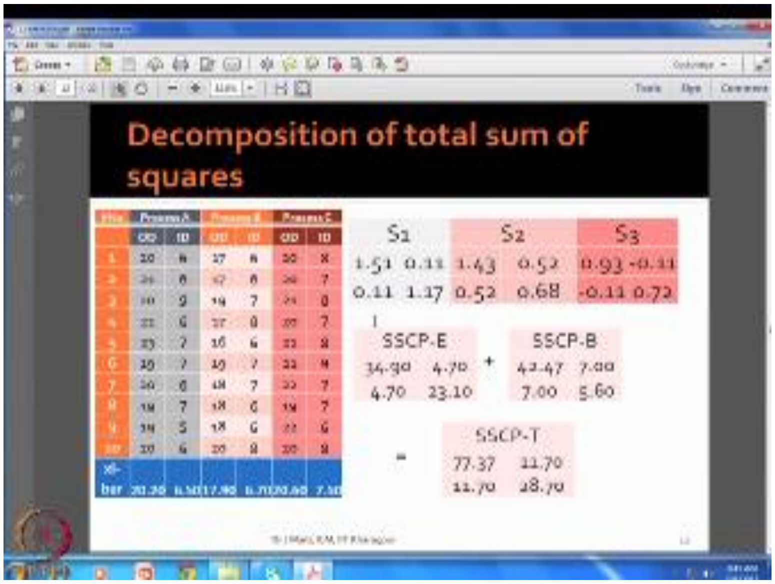
And $SSCP_E$ will be computed using this formula n_1 - 1 s_1 n_2 - 1 s_2 like up to n_l -1 s_l. Then SSCP T will be the sum of this two this equation.

(Refer Slide Time: 02:28)

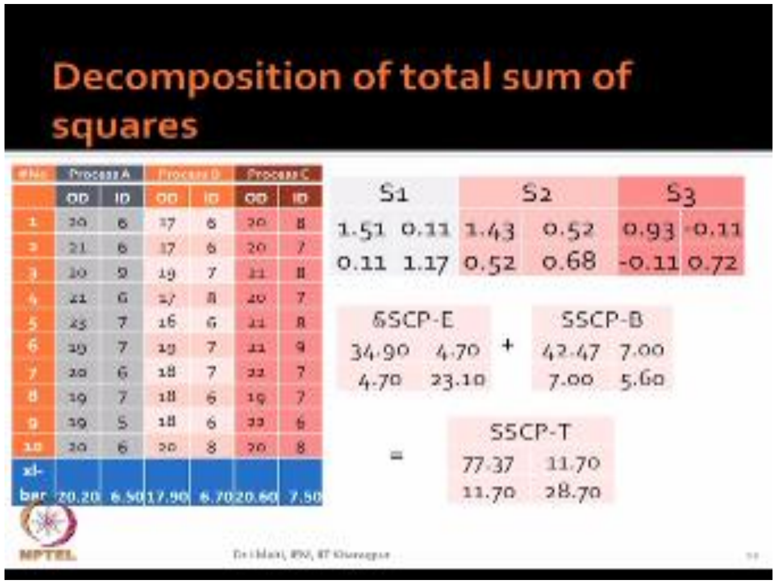


Now let us see one problem what we have discussed in last class.

(Refer Slide Time: 02:35)

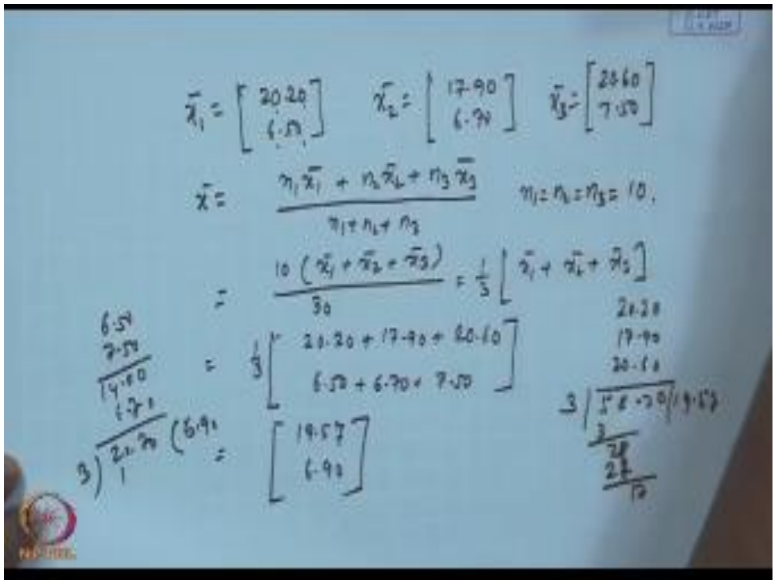


(Refer Slide Time: 02:38)



That there are three processes A, B and C, and we have sampled ten observations from all the processes and two variables outer diameter and inner diameter. You see the ∑ values for process A is 20.20 and 6.50.So if I write the ∑ values here.

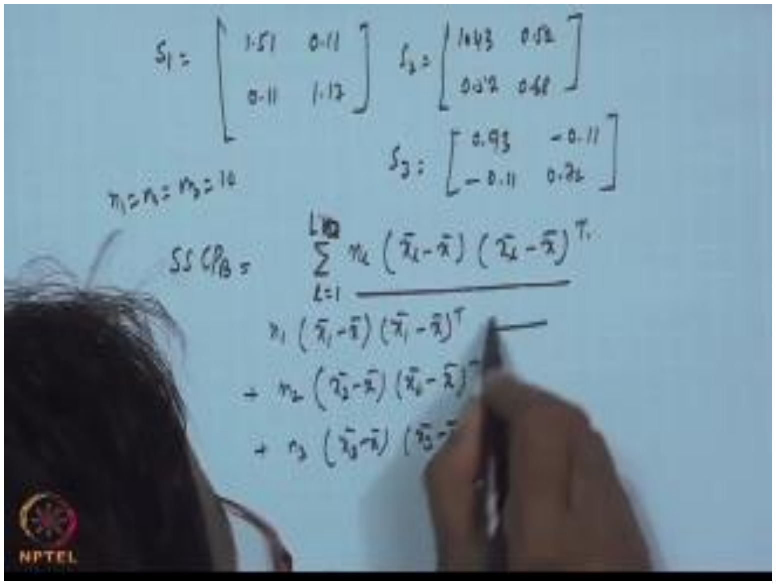
(Refer Slide Time: 03:05)



Then for process A, that ∑ value will be 20.20and 6.50 similarly for process B, it is17.90 and 6.70 and what about for process C? That is $\bar{x}_3$. This is your 20.60 and 7.50. Now, what will be your $\bar{x}$? Will be n_1 $\bar{x}_1$ n_2 $\bar{x}_2$ + n_3 $\bar{x}_3$ / n_1 + n_2 +n_3 now, we have $n_1 = n_2 = n_3 = 10$. So we can write this one as10 $\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ / 30, which is1 /3 into $\bar{x}_1$ + $\bar{x}_2$ + $\bar{x}_3$, so what will be this quantity? Then1 / 3 $\bar{x}_1$ is 20.20, 6.5, like this, your case will be20.20 + 17.90 + 20.60.

Second one will be 6.50, 6.70 + 7.50. If you add all those things 20.20, 17.90 and 20.60, quantity will be 17, 8, and 58.70 divided by 3; it will be 1, 9, 27, then17, then 5 almost 7 so this quantity will be 19.57 and second one will be 6.5 +7.5. This is14 + 6.70. This is 20.70 / 3 okay so this will be 6.90. So this is 6.90. Okay what more we want? We want to know what is s_1?

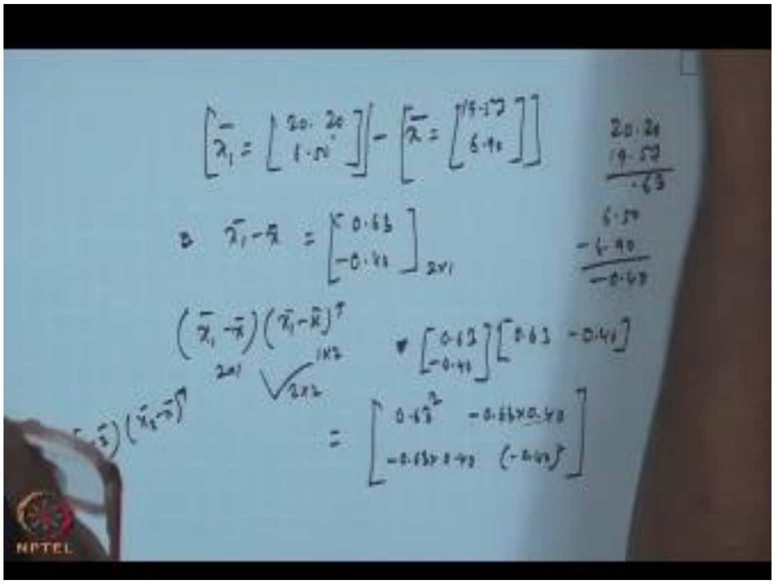
(Refer Slide Time: 05:58)



And in last class, we have computed s_1 is1.51, 0.11, 0.11, 1.17, your s_2 is1.43,0.52, 0.52, and 0.68 and s_3 is 0.93, -0.11, -0.11 and 0.72.Now we want to calculate SSCPB, first SSCPB, if you require to calculate $SSCP_B$, you require x l bar minus x̄ and x transpose to be multiplied. So, I am writing the formula first, SSC_B as $n_1 = n_2 = n_3 = 10$. So, I can write like this your l = 1 to n. n_l is basically n_1 to all those things n_l that is the formula $\bar{x}_1$ - x̄ $\bar{x}_1$ - x̄ transpose.

So we will first consider a thing, we have done a mistake l = 1 to L, capital L. This is the capital L. So, our L equal capital equal to 3. So, you will be getting 3 sets of values here. First one n_1 $\bar{x}_1$ - x̄ transpose + n_2 x $\bar{x}_2$ - x̄ + $\bar{x}_2$ - x̄ transpose + n_3 x $\bar{x}_3$ - x̄ and $\bar{x}_3$ - x̄ transpose. I will show you one calculation first one and then similarly, other also will be computed.

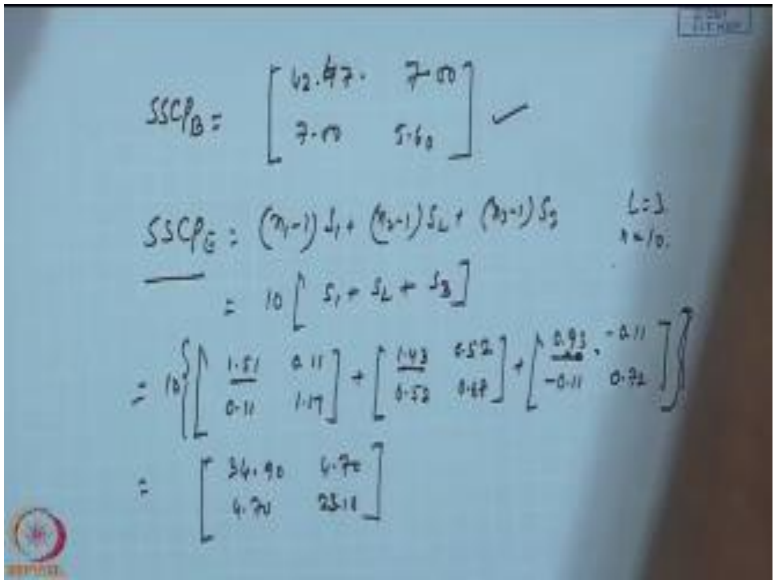
(Refer Slide Time: 08:27)



Now our $_1$ if you will see what our $\bar{x}_1$ here, you see $\bar{x}_1$ is 20.20 and 6.50, so 20.20, 6.50 minus, if we want to do $\bar{x}$, so I can see this minus $\bar{x}$ = our $\bar{x}$ value is this19.57. This is also a matrix of order 2 cross1. So, then this quantity, we can write $\bar{x}_1$ - $\bar{x}$ is 20.20 - 19.57, so this is 3,6, I think 0.63. Second one will be 6.50 -6.90, this will be -0.40, 0.40.So, what do you require? You require to calculate $\bar{x}_1$ - $\bar{x}$ x $\bar{x}_1$ - $\bar{x}$ transpose. This is 2 cross1. This is1cross2. Your resultant matrix will be 2 cross 2and that is what you want.

So, I can write like this now 0.63 -0.40 x 0.63 -0.40. The resultant matrix will be 0.63^2 -0.63 x 0.40 -0.63 x 0.49, then -0.40^2 okay free see that they are square matrix 2 cross 2 and symmetric matrix because this portion here this equal to this. In similar manner, you will be getting $\bar{x}_2$ - $\bar{x}$ $\bar{x}_2$ - $\bar{x}$ transpose and $\bar{x}_3$ - $\bar{x}$ this transpose. So, now what will be the resultant value? Resultant value is this n_1 = 10. So, this quantity will be 10 into we have already found out the value 0.63^2 -0.63-0.63 x 0.40 - 0.40^2 +10 into whatever matrix value you have get +10 x this.

(Refer Slide Time: 11:40)



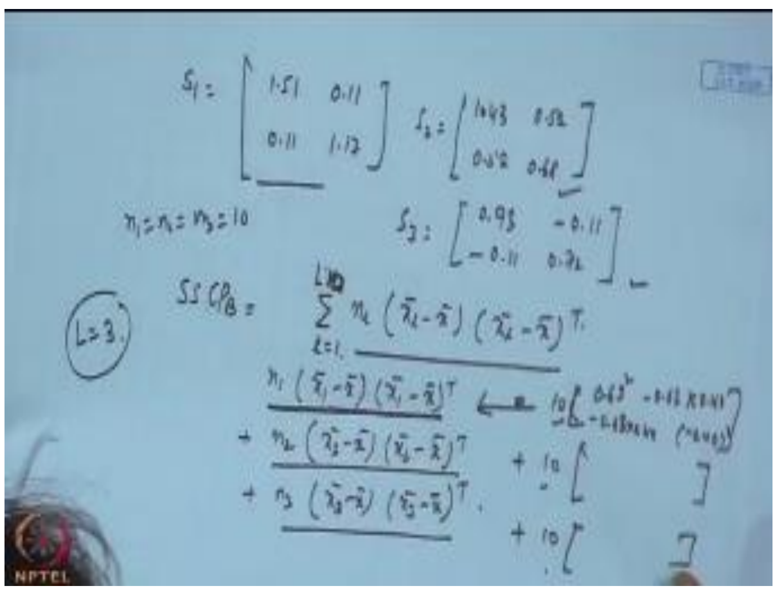
So resultant the ultimate thing that SSCP, SSCP what you will get here SSCP will be? $SSCP_B$ will be, we have computed that will be your 42.47 and 7.00, then 7.00 and 5.60. So, $SSCP_B$ is calculated like this. Now, we want to calculate $SSCP_E$. I told you this will be n_1 -1 s_1 + n_2 -1 s_2 + n_3 - 1 s_3. In this case, L = 3 capital L. So, we will go by like this n_1 x n_3, all equal to n = 10. This is nothing but 10 x s_1 + s_2 + s_3.

So, I will write down, now equal to 10. What is our s_1? You see our s_1 is this s_2 is this s_3 is this rather this 2 cross 2 matrices so you write down. So, s_1 is1.51, 0.11, 0.11, 1.17, so +1.43, 0.52, 0.52, 0.68+ 0.93 - 0.11 - 0.11, 0.72 this is your SSCPE. If you see that 1.51, 1.43and 0.93 this you, if you add what are the values you will be getting that 15, 14, the 29 + 9, almost 33 point something you will be getting. So this resultant quantity will be your 34.90, 4.70, 4.70, and 23.10.

So far what we have discussed if I recapitulate again, I think this is the computation part. We are considering that we said that in a MANOVA, in one way MANOVA, the SSCP total is decomposed into SSCP between population and SSCP Error. We also seen that the degrees of freedom is N -1 for SSCP total, L -1 for SSCP between and $SSCP_E$ is N - L. These are the formulation, computational formula for $SSCP_B$ and $SSCP_E$. Then, I have given you this, this one, this is the problem. What I said that let us compute $SSCP_B$ and SSCPE for the data sheet given for this. What we have done? You have to first find out what are the ∑ vectors and these ∑ vectors are like this. This ∑s vectors are $\bar{x}_1$ = 20.20, 6.50. This is a 2 cross1 vector. 17.70, 6.70, this is also 2 cross1 and x_3 is 20.60, 7.50.This is also 2 cross1. Using these 3 vectors, ∑ vectors, you have computed the grand ∑. We have already computed the grand ∑.

The grand ∑ is n1 $\bar{x}$ $_+$ n_2 $\bar{x}_2$ $_+$ n_3 $\bar{x}_3$ / n_1 + n 2 + n_3 and in our case it is equal sample in each case. So, finally, it goes to this one that $\bar{x}_1$ is 1, 1 / 3 times the sum of the 3 ∑ vectors and as resultant value s this. So, our grand ∑ is19.57 and 6.7, 6.90. Then, we computed SSCPB. SSCPB using these formulations, SSCPB is l = L to l capital L n_l into this and $SSCP_E$ is this value.

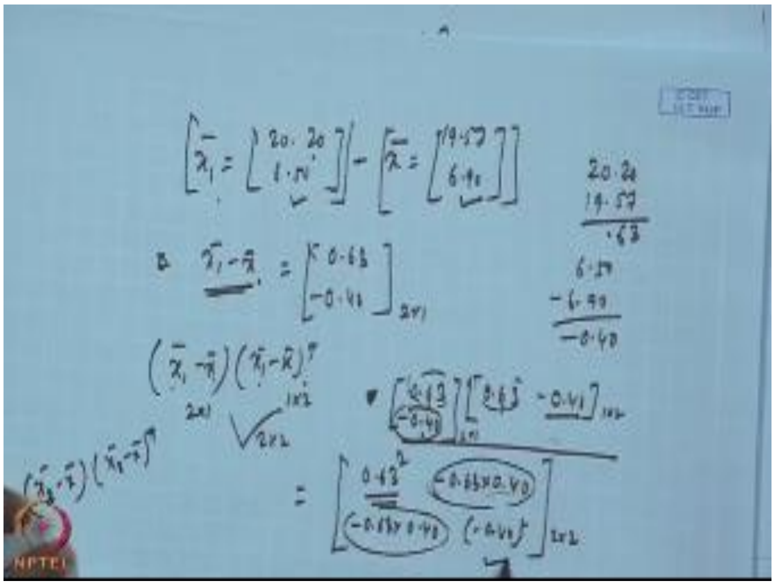
(Refer Slide Time: 16:34)



Now, from computation point of view what we have seen that I have shown you one computation because the $SSCP_B$ part if you see, it is basically sum of 3, 3 different elements. Those 3 different elements are like this. Three different element s that is n_1 into this because see this sum total of this l = 1 to capital L. So our capital L is 3. So, if I put l equal to 1, then this is the quantity if I put l = 2, then this will be the quantity if I put l = 3, then this will be the quantity.

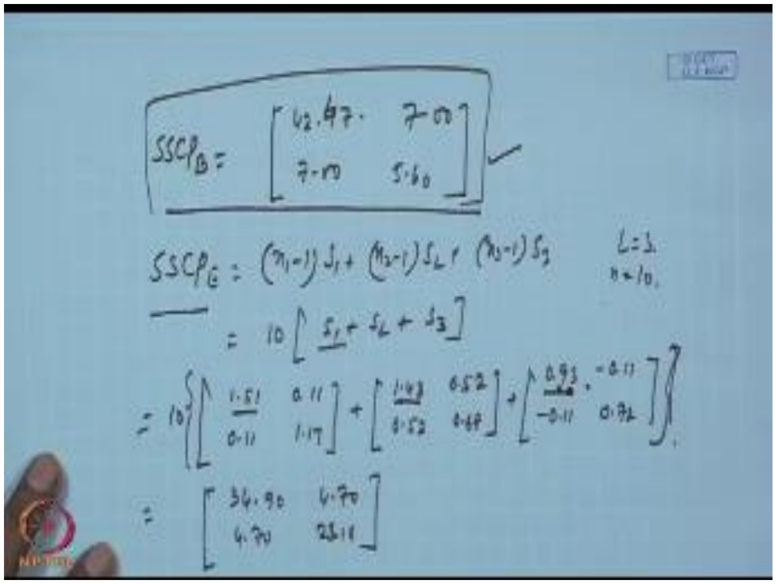
Our case is equal size case, so n_1, n_2, n_3;I am putting 10, 10, and 10. Then you are computing this one. How we are computing this $\bar{x}1$ minus $\bar{x}$ $\bar{x}1$ - $\bar{x}$ transpose? That one computation formulation I have shown you. This formulation is like this.

(Refer Slide Time: 17:34)



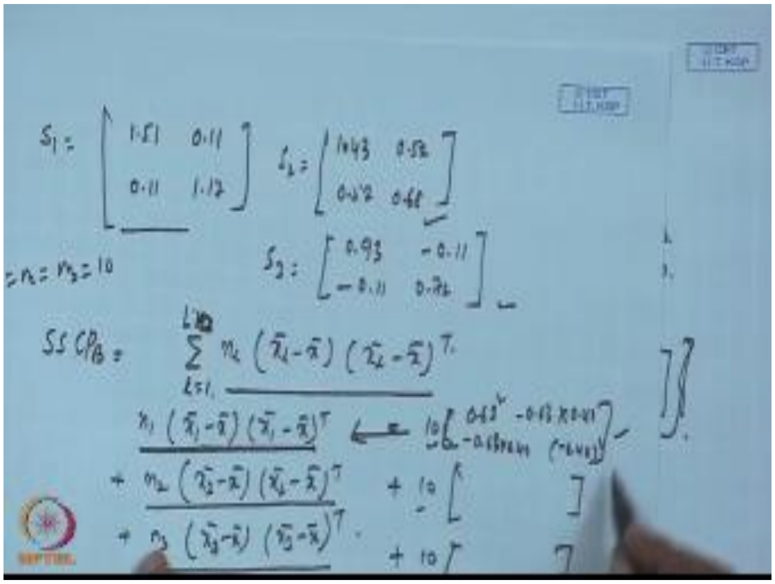
First your $\bar{x}1$ is this, x_2, $\bar{x}$ is this, and the difference will give you $\bar{x}1$ - $\bar{x}$, then $\bar{x}1$-$\bar{x}$ and its transpose. This is what you have written. So, it is a2 cross1 matrix this is1 cross 2 matrixes. You will be getting resultant 2 cross 2 matrixes, 0.63 into 0.63, 0.63^2, 0.63 into -0.40, this will be the value. Then -0.40 x 0.63, this is the value and - 0.40into minus 0.40, this is the square value. The resultant values are like this.

(Refer Slide Time: 18:14)



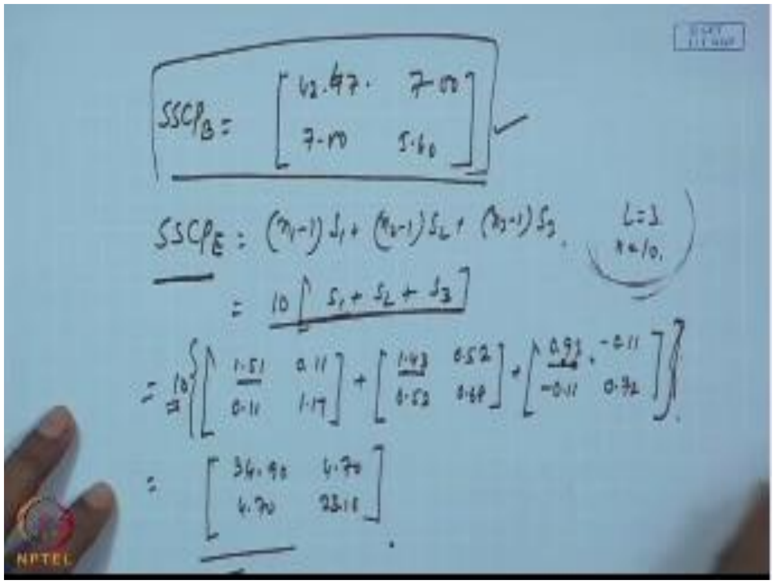
So for the first one, s_1, this SSCP resultant values is this where all three parts are added. All three parts are added.

(Refer Slide Time: 18:29)



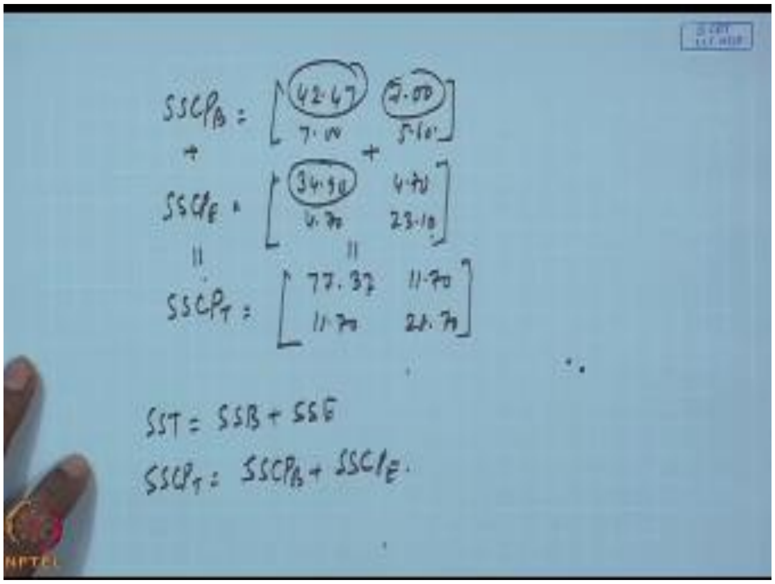
Parts related to this component, this component. This component when added, you will be getting like this.

(Refer Slide Time: 18:38)



Then what you require to know? We require knowing what will be the sum square cross product error matrix the formula is. It will be the sum total of n_1 - 1 s_1 n_2 -1 s_2 n_3 -1 s_3 because we have three population n = 10. So, our value is like this. We have started with this that our s_1 is this one, s_2 is this matrix, s_3 is this matrix. So, you are adding 1.51, 1.43 +0.93. This will be nothing but 3.49 and if you multiply by 10, it will be 34.90. In similar manner, you are getting $SSCP_E$.

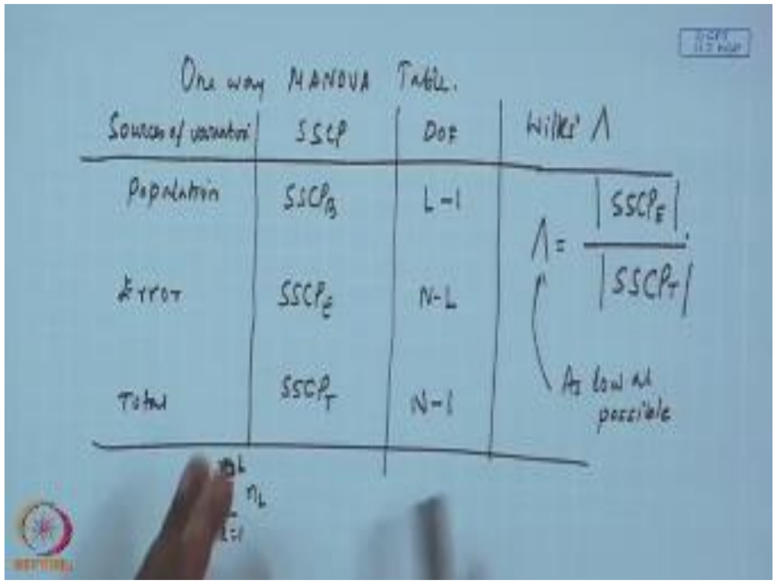
(Refer Slide Time: 19:24)



So what we have now with us? We have $SSCP_B$ matrix, which is 42.47, 7.00, 7.00, and 5.60. You have SSCP error matrix, which is 34.90, 4.70, 4.70, and 23.10. So, you know your SSCP total now, this plus this resultant will be 77.37; now 7 + 4.7, 11.70, and 11.70. Then 23.10 + 5.60, it will be 28.70.This is the decomposition of total matrix, this equal to this plus this or this equal to this plus this.

So in MANOVA, you have to decompose SST sum square total sum square between population and sum square error and in MANOVA, what you are doing, SSCP total is decomposed into SSCP between plus SSCP error. So, because of more than one variable, your computation will be in the matrix domain okay. Once you decompose the $SSCP_T$ into the parts, you are in a position to calculate the MANOVA table. Please remember this is one way MANOVA.

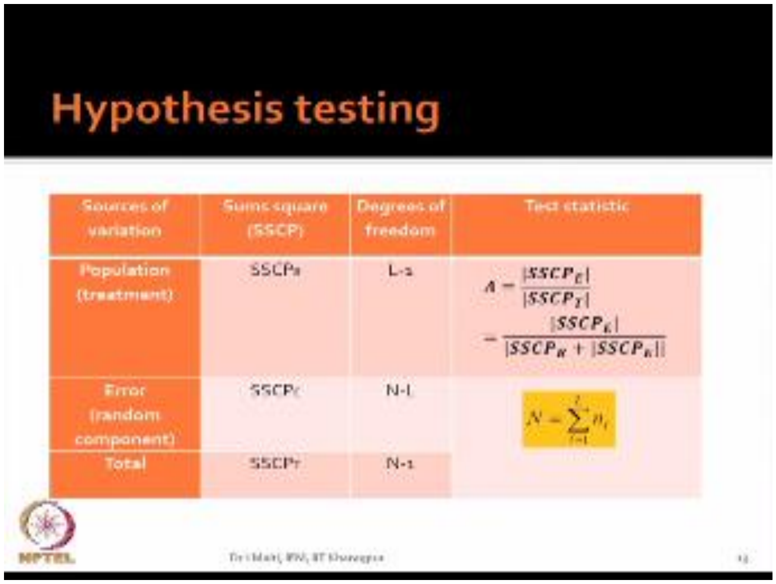
(Refer Slide Time: 21:17)



We were talking about one way MANOVA because we have taken only population of some kind one kind that population one to population and not some other kind of factors are considered in this table. You have to write down what is the source, similar manner sources of variation, so definitely your population is one source. Then error is another one, then total is coming, then you are writing SSCP matrix.

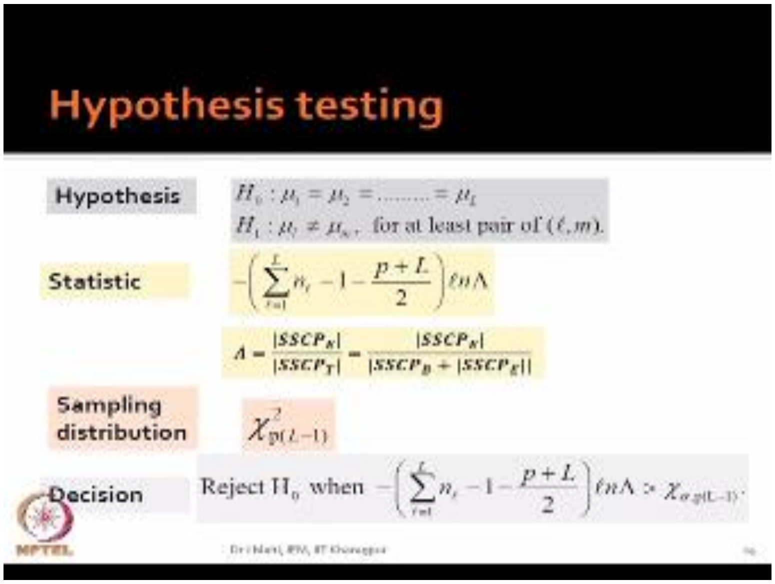
So, our case is SSCP between SSCP error and SSCP total. Then you will be writing degree of freedom and I told you that N is sum total of l = 1 to capital L n_l; this is the case. So, in that case, our degree of freedom for total will be N -1, for $SSCP_B$, L -1 and this will be N - L. Then in MANOVA, you find what the $\sum^2$ between $\sum^2$, error is, all those things here we will not use like this. We will create a matrix called Wilks λ. Wilks λ, this is basically a ratio; this wiki wills λ ratio will be computing where.

(Refer Slide Time: 23:08)



Wilks λ is determinant of SSCP error by determinant of SSCP total and what you want? We want that the SSCP error, this determinant must be a slow as possible compared to the total. So, this quantity will be as low as possible or otherwise we can say as small as possible, but there is certain distributions available using Wilks λ. If you go by multivariate books, you will be finding out that different combinations, different sampling distributions are available. We will show you a general one in this class.

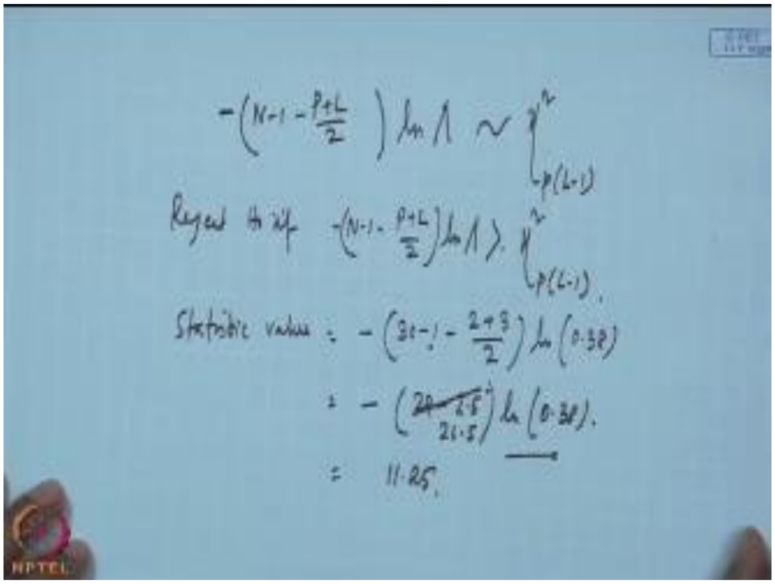
(Refer Slide Time: 24:21)



Okay now with using this, whatever we have developed so far, based on this, we want to test the hypothesis. What are the hypotheses? First hypothesis is null hypothesis and alternative hypothesis in our case. Null hypothesis is there is no difference in the ∑ vectors for all above lessons and alternative hypothesis is at least one pair of populations is different in terms of their ∑ vectors.

We will use a statistics. This statistics is minus within bracket l = 1 to L n_l; this is capital N -1 - p plus L by 2 log λ, λ you know that how to compute λ. Then this statistics follows chi^2 distribution with p x L - 1degrees of freedom. Okay then what is our statistics?

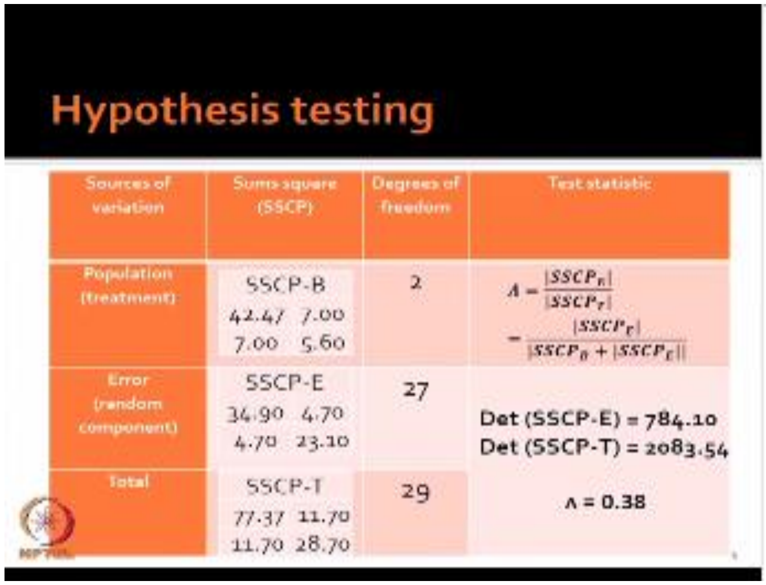
(Refer Slide Time: 25:32)



Our statistics is that N -1 - p + L / 2 log λ lagging λ. What we are saying this follows chi^2 p x L - 1. What will be your decision? The decision will be if I think this is -1 minus is there, because Wilks λ this is a low value keeping minus here making it positive. So, our hypothesis is reject H_0that may no difference among the population ∑ vectors if - N -1 - p + L / 2 log λ, this value greater than chi^2 p L -1.

We will reject null hypothesis under this condition. Now, we will see that for the problem given, what is the status?

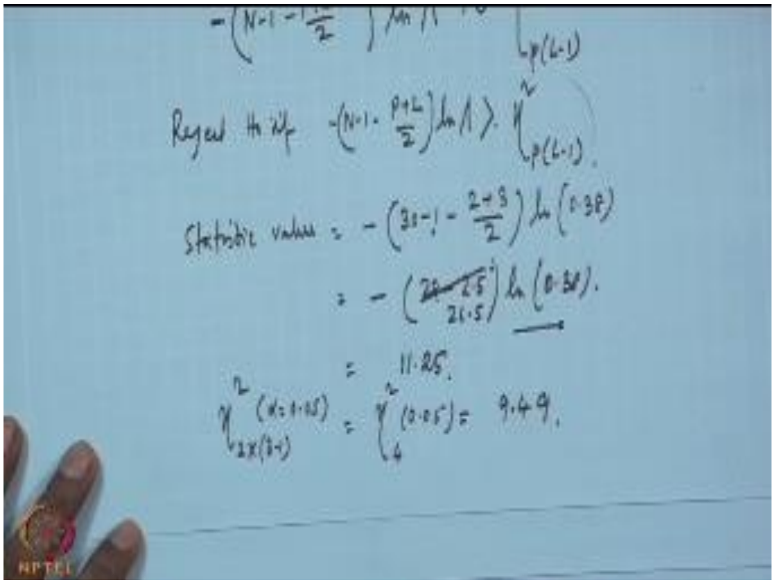
(Refer Slide Time: 26:54)



You see the slide. So, $SSCP_B$, you have already computed 42.47, 7 like this. $SSCP_E$ is also computed. SSCP total is also computed. Our capital N is 30. So, these degrees of freedom, I hope there will be no problem, 29, 27 and 2, and then test statistics is lambda. If you would, if you compute determinant $SSCP_E$, this one 34 into23.10 minus this square, you will be getting784.10. Total case, total case will be getting this one, 77.37 x 28, this will be getting this one.

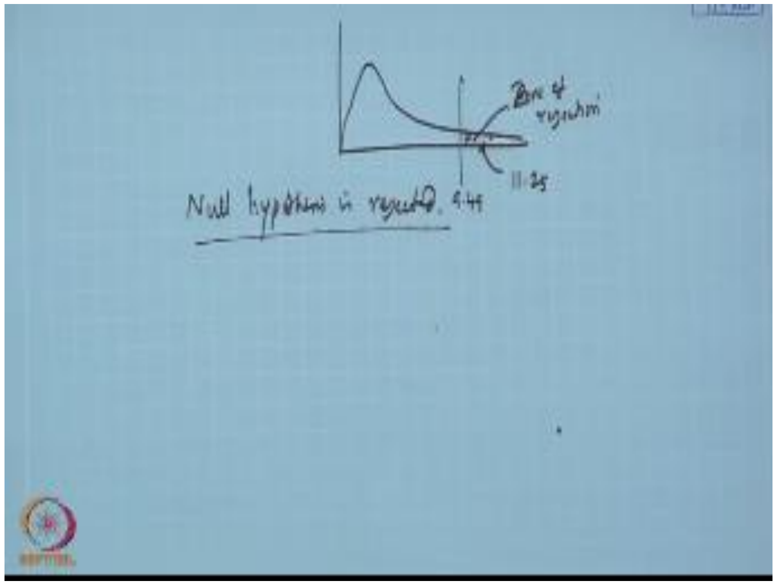
So, your determinant of $SSCP_E$ is784.10, determinant of SSCP total is 2.83.54.Now, if you take the ratio, you will be getting λ= 0.38. Now, what will be the log λ? Our statistics value will be - N is 30-1 - p is 2 + L is 3/ 2 log of 0.38, this one will be minus. This is 29, 29 - 2.5. So, that means the resultant will be, this resultant will be 26.5 into log 0.38. Find out this value. I think the resultant quantity, this will be 11.25.

(Refer Slide Time: 29:10)



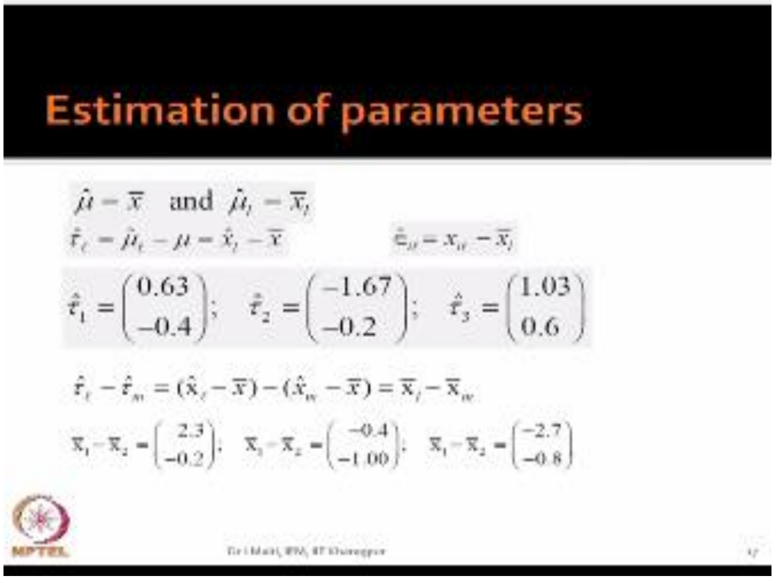
Now, we require finding out, what we require to find out chi^2 p L -1. p is 2, L-1 is also 3 -1 that is 2 and you have to take a α value. Let α = 0.05 and this one is chi^2 4, 0.05,this value is 9.49.The statistics value is more than the tabulated value.

(Refer Slide Time: 29:40)



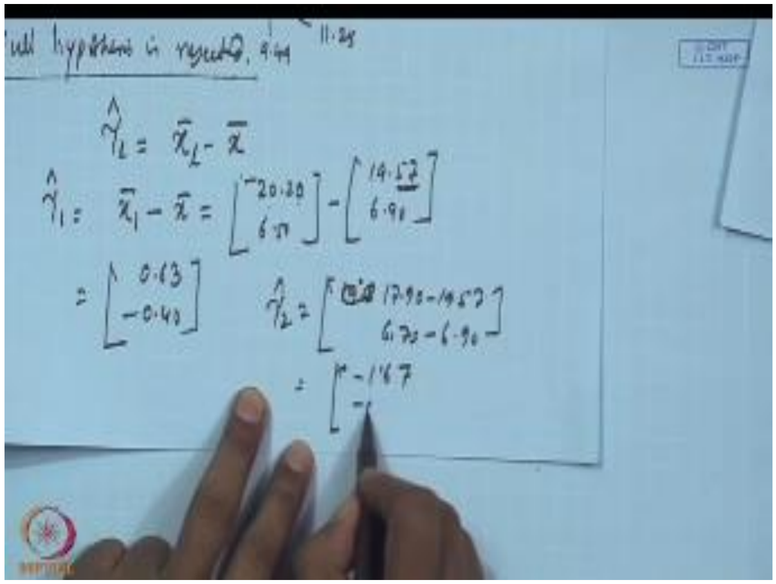
The case is coming like this. This is my tabulated value, 9.49 and my computed value is here, 11.25. This is the zone of rejection. Okay so null hypothesis is rejected or some space. Now this null hypothesis is rejected. There is alternative hypothesis that there is some difference among the population means.

(Refer Slide Time: 30:29)



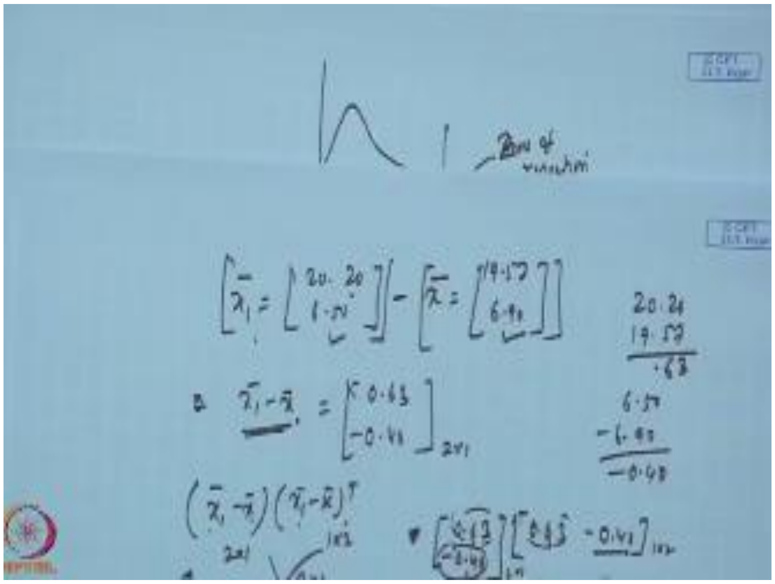
So as null hypothesis is rejected, so we can say that population effects are there. Now, we want to know the population effect. First, you see the slide. I think if you can recall my last class, then you will find out that we have computed population effect in terms of τ. τ_1 is the population first population effect, τ_2 second population effect, τ_3 third population, so we can calculate τ how we calculate τl? τl is $\bar{x}_l$ - $\bar{x}$.

(Refer Slide Time: 31:16)



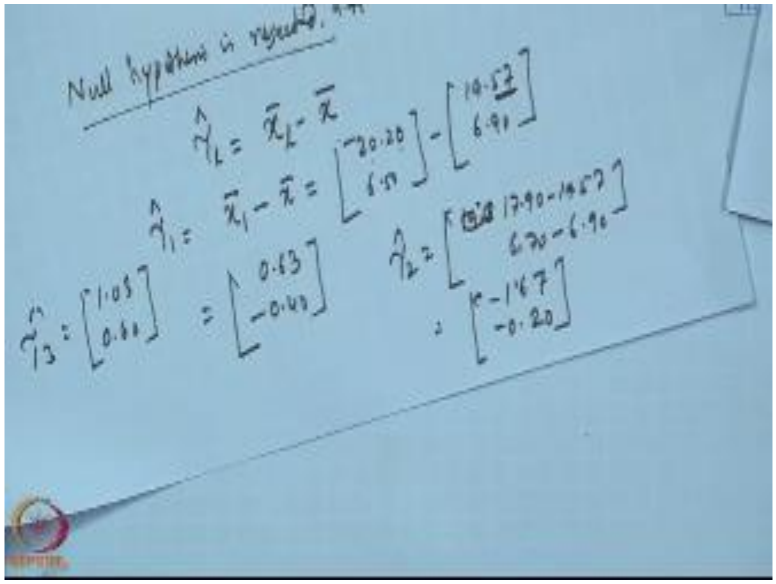
So if you want to compute τ_1, if I say this is the estimated value, then τ_1, one then you have to write down what is $\bar{x}_1$ - $\bar{x}$. I think we have already seen the $\bar{x}_1$ and values are there with you, $\bar{x}_1$ and $\bar{x}$ values.

(Refer Slide Time: 31:46)



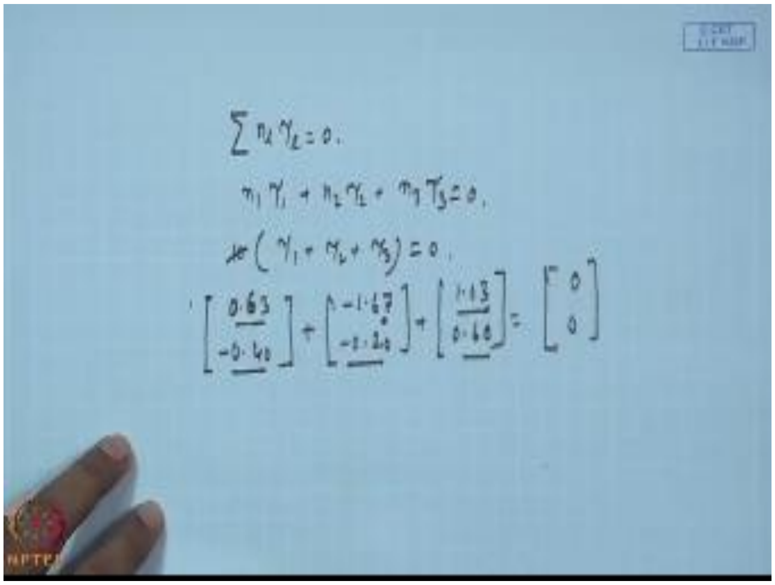
If $\bar{x}_1$ is 20.20, so your value is 20.20, 6.50 - $\bar{x}$ is, what we got $\bar{x}$, 19.something, I think 19.57 and 6.90. What will be this value? Then, this value 20.20 will be 0.63- 0.40. So, similarly, t τ_2 value, you will be getting $\tau_2$value also find out in the same manner that 19. τ_2 value will be 17.90 -19.57 and 6.70 -6.90, this value will be, and τ_2 value will be -1.67 - 0.20.

(Refer Slide Time: 33:07)



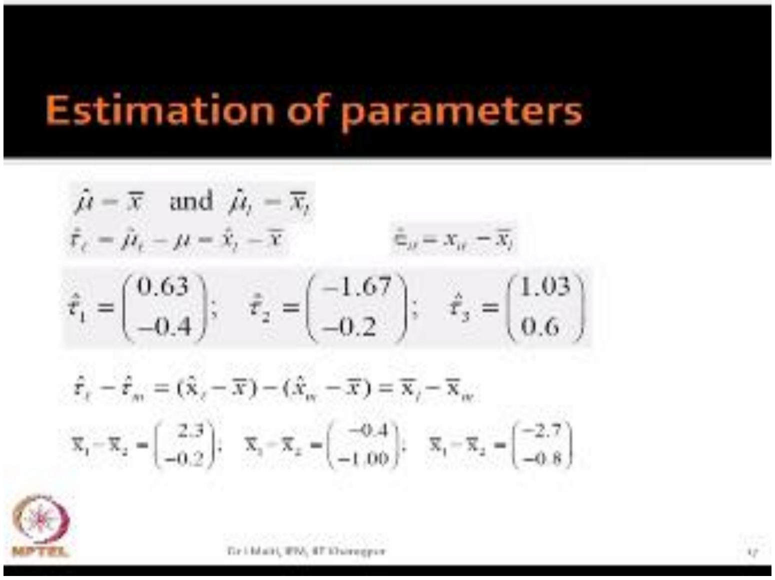
Similarly, your τ_3 value will be 1.03, 0.60. Okay I told you in beginning of ANOVA as well as in MANOVA, we say that the parameters, populated parameters are basically that tau l basically are having a certain property.

(Refer Slide Time: 33:31)

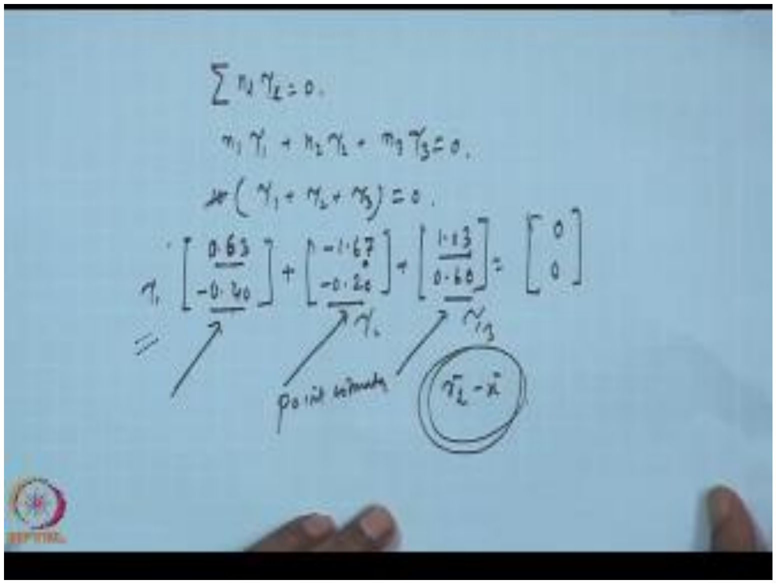


That property is $\sum n_l \tau_1$, this will be equal to 0. So, that means in this case, $n_l \tau_{1\,+} n_2 \tau_2 + n_3 \tau_3$ this will be 0. Our n_1 n_2 n_3 equal that is10 plus, so τ_{1+} τ_1 τ_2+ $\tau_{31\,=}$ 0. Now see what is given here. What is τ_1,1.03 this, not τ_1, this is τ_1, 0.63 is τ_1 - 0.40, so I can say cancel 10 + 2 is - 1.67 -0.20 plus τ_3 is1.03,0.60, what the value is coming? You see 0.63+1.03, this is1.66 -1.67, it is0 basically that because of rounding error, this space comes, so this is 0; similarly,0.40 - 0.20 that - 0.60 + 0.60.

(Refer Slide Time: 35:05)

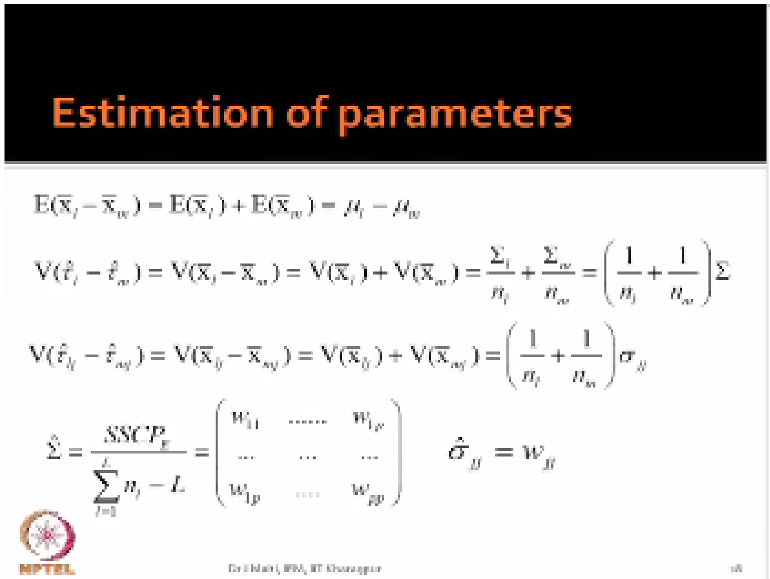


(Refer Slide Time: 35:06)



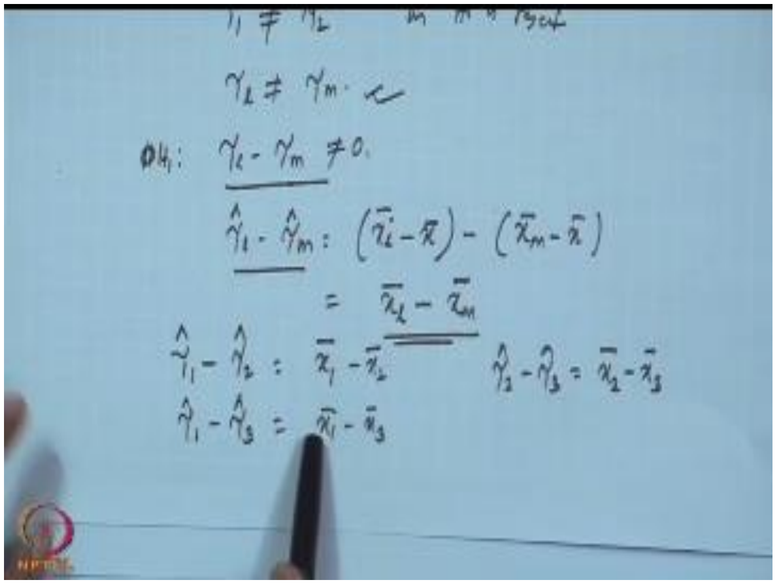
Okay so we found out now the effects of each of the population and well also found out collectively there is a difference. Now we want to find out which variables are making the difference. That is our next objective.

(Refer Slide Time: 35:30)



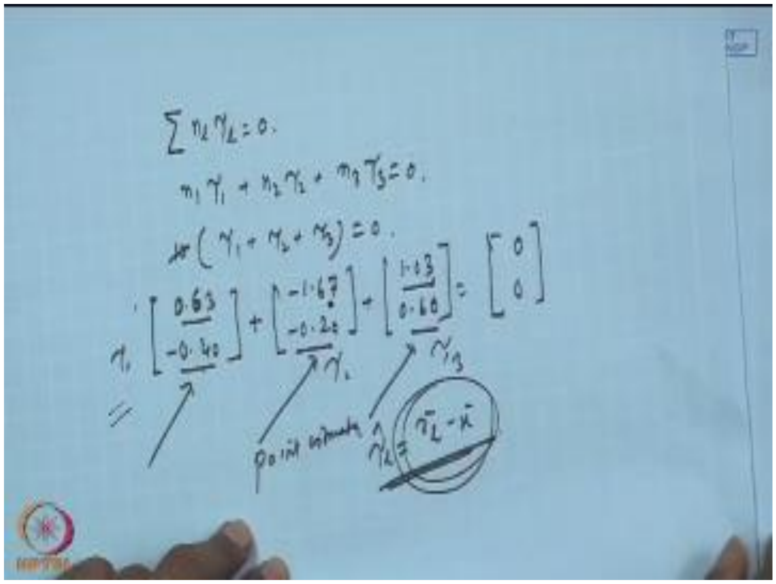
So, in order to do so I will first show you the slide; here I am showing you something more. What is that? You have already found out that this is your τ_1, this one is τ_2. This one is τ_3. Other one is τ_3. So, there are all known as point estimates. Now we want to know the interval estimate of all those things. So, if you require, you want to find out the interval estimate, then you will definitely require finding out that $\bar{x}_1$-$\bar{x}$, what will be the estimate, all those things, so we have seen earlier.

(Refer Slide Time: 36:26)



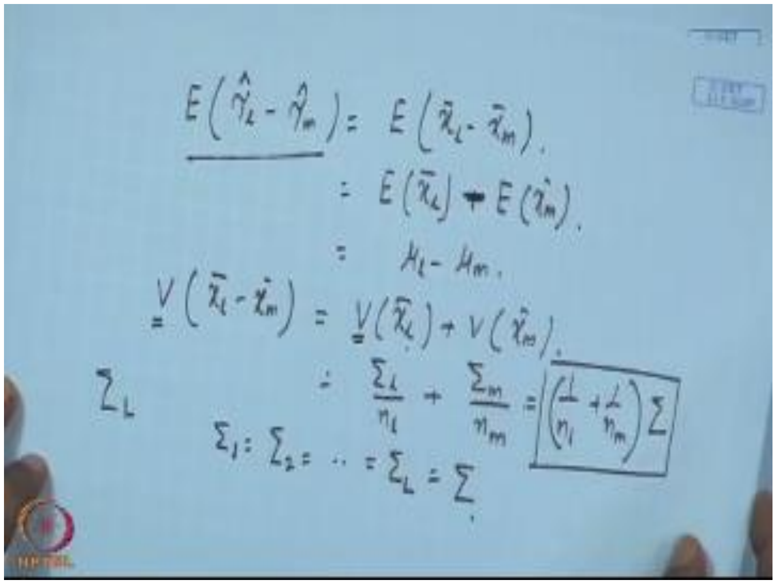
In this class, I will show you that as we are saying that τ_1 may be different from τ_2 as collectively H_0 is rejected. So, I want to create a situation that τ_1 not equal to τ_m, then $\tau_{1-}\tau_m$ that not equal to 0 that is your alternate hypothesis. So, what I said earlier that for individual τ_1, you can go for interval estimation also, we are here going for the interval estimation of difference because that is what we want; H_1 is accepted. Now, τ_1- τ_m cannot be written like this that $\bar{x}_l$ I am taking the estimate, $\bar{x}_l$ - $\bar{x}$ - $\bar{x}_{m\,-}$ $\bar{x}$. Let us see here.

(Refer Slide Time: 37:43)



What I said that τ_1 is this every population $\sum$ minus the grand $\sum$.So, then this one is τ $\bar{x}_l$ - $\bar{x}_m$, $\bar{x}$ is cancelling out. So if I want know the point estimate of these, this is nothing but this one. So, you have already seen what this $\bar{x}_1$ is. Suppose I want see the value $\tau_1 - \tau_2$, then it will be $\bar{x}_1$ - $\bar{x}_2$, yes or no? So, similarly, $\tau_1 - \tau_3 = \bar{x}_1 - \bar{x}_3$ and $\tau_2 - \tau_3$ $\bar{x}_l - \bar{x}_3$ τ_2, - τ_3 will be $\bar{x}_2 - \bar{x}_3$ and using the matrix values, you can find out what are the difference value, which is nota big problem. You can easily find out. So, once you get these values, it will be point estimate. Now, I want to find out the interval estimate of the same.

(Refer Slide Time: 39:15)



So, what you require to do for interval estimation? You require to know what is the expected value of suppose $\tau_{\hat{l}}$ estimate - τ_m estimate; this is nothing but expected value of $\bar{x}_l$ - $\bar{x}_m$ what you have seen earlier. We also know that this nothing but expected value of $\bar{x}_l$ - $\bar{x}_m$ what you seen earlier. We also know that this is nothing but both expected value of $\bar{x}_l$ expected value of $\bar{x}_m$ because the two populations are independent. μ So, this will be μ_l - μ_m. Similarly, you require computing variance component of $\bar{x}_l$ - $\bar{x}_m$, which will be variance of $\tau_{\hat{l}}$ - $\tau_{\hat{m}}$.

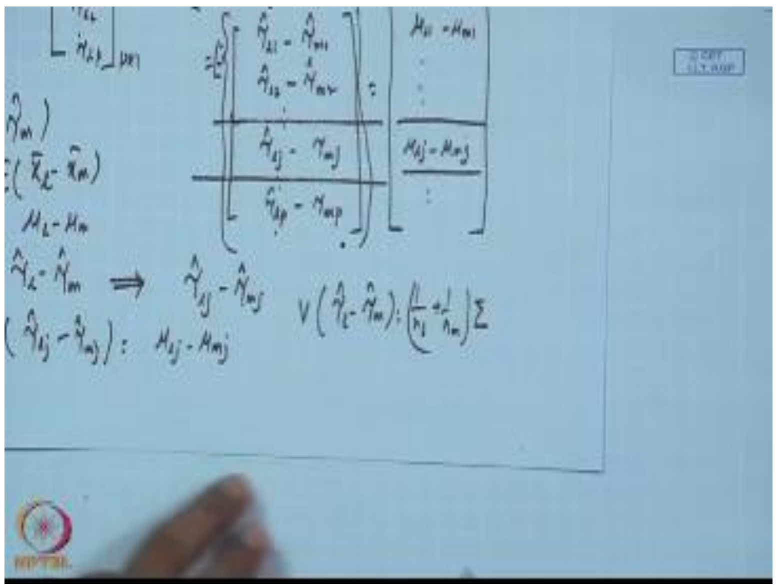
Now, variance of this is equal to variance of $\bar{x}_l$ plus variance $\bar{x}_m$. We have seen earlier this one. So, this is what the variance of $\bar{x}_l$ is vector quantity. So, it is coming from the population l. Now, population l covariance matrix is capital σl and it is the $\sum$covariance of the $\sum$ value. So, it will be σ_l / n_l +$\sum$m/n_m. Now, you see that I have, although I have written here variance, but you can write also co variance.

Whenever there are more than none variable, we say variance structure $\sum$ variance plus co variance structure. Now, in MANOVA, the assumption is all the population variable, it is equal this equal to capital $\sum$. So, then we can write this one as n_l +1 / n_m x $\sum$, because $\sum_l$ = $\sum_m$ =$\sum$

because all the population have equal covariance. So, your variability part is taken care of by this covariance part.

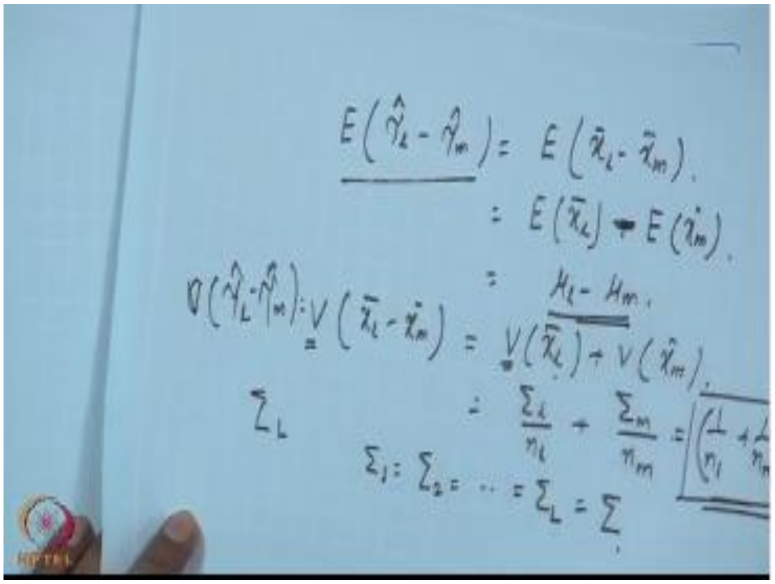
I think in vertleing t square time you have also seen this one. Suppose we are interested to this is from the overall that population f x point of view, you are finding out the difference. Now, there is j = 1 to p variables.

(Refer Slide Time: 42:31)



As I told you that I want to know what are the variables making the effect, so that means τ_1 is nothing but your τ_{11}, τ_{12} like this τ_{1p}, p cross1. When I say that τ_{l}- $\tau_{\hat{m}}$ that means I say here this is τ this l $_1$ - τ $_{m1}$ τ_2 - τ_m 2. So, like this, I can get one point where $\tau_{lj} - \tau_{mj}$, then slowly up to the last variable $\tau_{lp} - \tau_{mp}$. If I say that $\tau_l - \tau_m$ not equal to 0 that means I am saying that this vector τ_{l1} to τ_{lp} is not equal to $\tau_{m1}$to τ_{mp}. So, we have seen the variance covariance structure of this one.

(Refer Slide Time: 43:57)

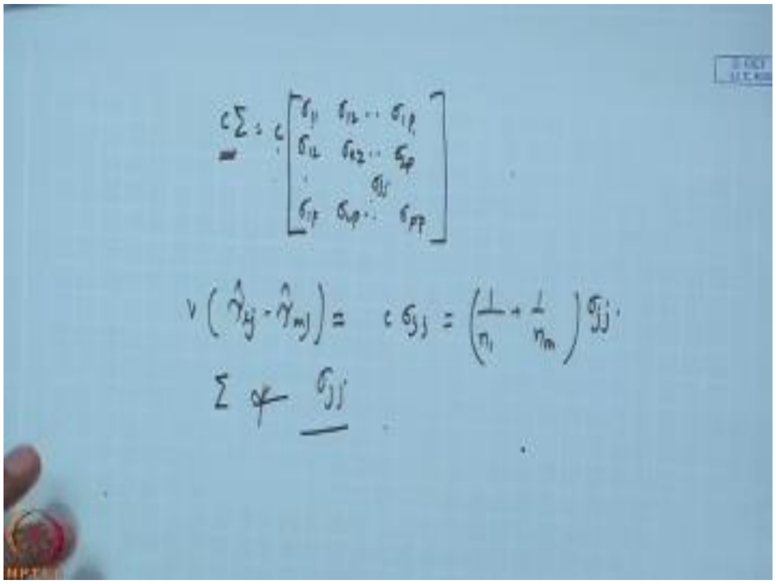


What I can say that the variance structure of $\tau_{\hat{l}}$ - $\tau_{\hat{m}}$ that you have seen this is the covariance structure expected value also you have seen, so I want to know. Also, suppose if I take a particular variable here τ_{lj} - τ_{mj}, what will be the covariance variability and $\sum$ value for this? So, when I expected value of $\tau_{\hat{l}}$ - $\tau_{\hat{m}}$, it is nothing but expected value of $\bar{x}_l$ - $\bar{x}_m$, which we say it is nothing but μ_l - μ_m. Then, this one is nothing but if we write down the expected value of this, expected value of this or expected value of this totality.

Then what you will get here you will get here, μ_l - μ_m 1, like this here, you will be getting mu l j minus μ_{lj} - μ_{mj} Yes or no? So, for a particular variable case, the difference if I see and then the expected value if I want to find out that $\sum$ what I am interested. Now, instead of $\tau_{\hat{l}}$ - $\tau_{\hat{m}}$, I am interested now with a variable that is τ_{lj} - τ_{mj}. So, expected value of τ_{lj} - τ_{mj}, this is nothing but μ_{lj} - μ_{mj}; this is third quantity.

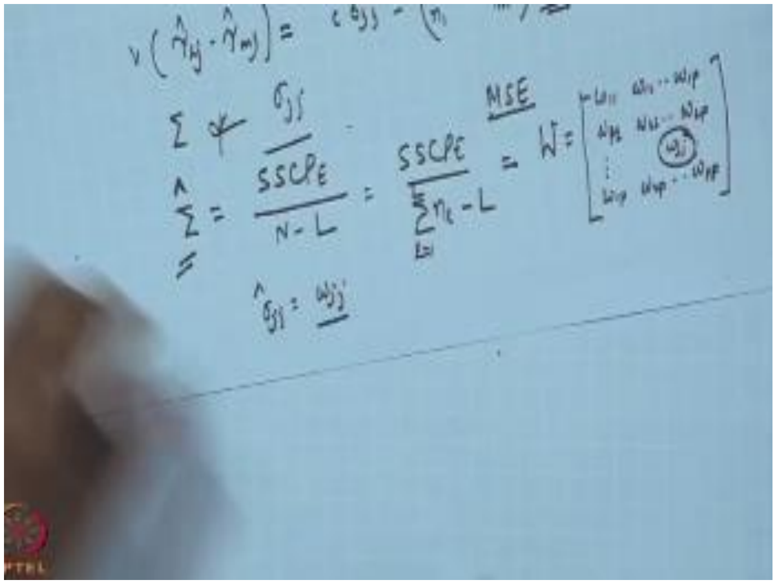
Then, what will be the variability part here? Variability of μ_l - μ_m that one you have seen as1 / n_l +1 /n_m into $\sum$. So, we will find out the value. So, if I write down that this one, this one, this quantity as c, so it is basically c $\sum$,c $\sum$ then, what is happening?

(Refer Slide Time: 46:28)



Then you are getting this co variance; c ∑ is something like this c x ∑ 11, ∑ 12 that ∑ 1p, 12, ∑ 22, ∑ 11, ∑ 12, ∑ 2p, where we have started that c ∑ is c into ∑ 11, ∑ 2p. Then ∑ 2p is coming like this ∑ 1p, ∑ 2p, then ∑ pp, somewhere in between there will be ∑ jj, we have taken this, ∑ jj, somewhere in between there will ∑ jj. So, for this variable, what we have taken our case is like this;

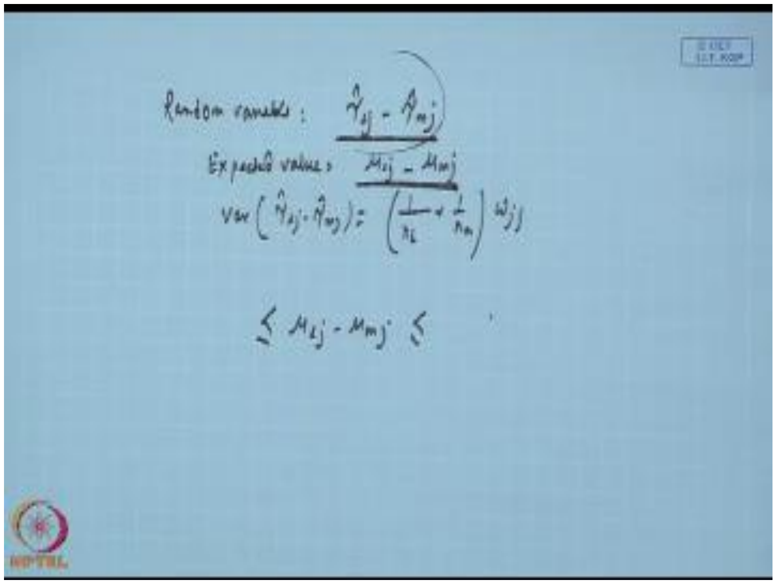
(Refer Slide Time: 47:17)



For this $\tau_{\hat{l}}$ - $\tau_{l\hat{m}}$ that is μ_{lj} - μ_{mj}, this one the variability part τ_{lj} - τ_{mj}, this will be definitely c ∑jj. There is absolutely no problem for you. That ∑s this is1 / nl n m into ∑jj, but please remember, we do not know ∑. What is the capital ∑? We do not know. That ∑ we do not know, we do not know ∑jj also. So, what will be the estimate of ∑ ? In ANOVA, we have seen we said that I think you can remember that MSE we talked about that MSE is the estimate of $\sum^2$. So, here also you have found out $SSCP_E$, if you divide it by degrees of freedom, this is the estimate of∑.

So, $SSCP_E$ divided by sum total l = l to capital L n l - L, this is your estimate. Suppose if we write down like this as W, which is which is nothing but w 11, w 12, then w1p, w 12, w 22, w 2p, so like this, you will be getting w 1p, w 2p, w pp in between somewhere w_{jj}. So, this w_{jj} as I told you that ∑ is estimated like this, now you are getting w_{jj} also; this w jj will be estimate of $\sum_{jj.}$ So, you can say $\sum_{jj}$ is equal to w jj.

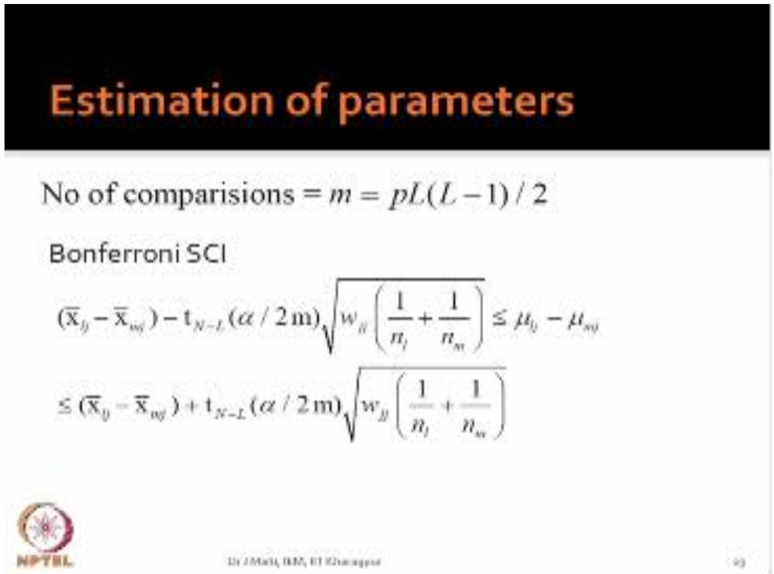
(Refer Slide Time: 49:59)



So, essentially what you got? Now, you got very interesting things; one is your random variable that is $\tau_{\hat{l}j} - \tau_{l\hat{m}}$ and its expected value also you got, $\mu_{lj} - \mu_{mj}$, that is the ∑ value. You also got the variance of $\tau_{\hat{l}j} - \tau_{l\hat{m}}$ that one is 1 / n1 + l n l +1 / n m x w jj. So, you have everything now. Now, what you want to know? We want to know the interval estimate of this that is what we have started. So, interval estimates of, again please remember, this is a random variable, the ∑ interval estimate for the∑ we want to compute not that interval estimate of this.

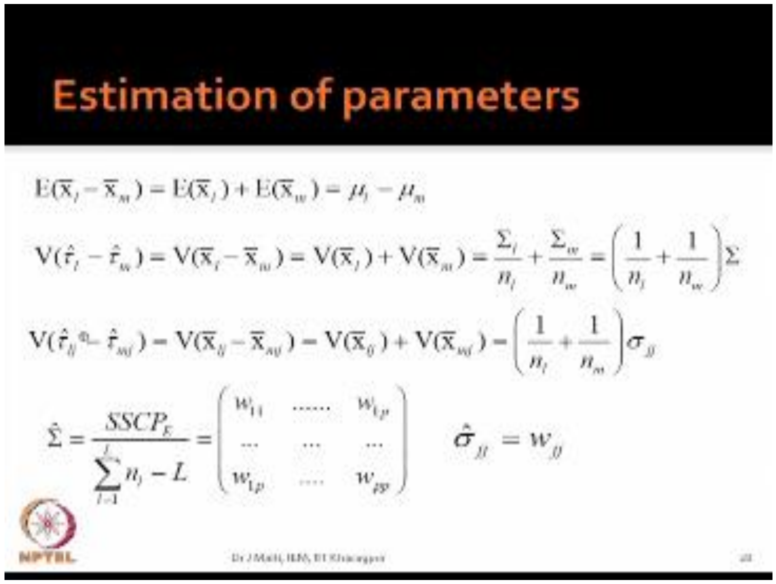
With the help of this, we want to compute the interval estimate of its∑ value. So, what you want $\mu_{lj} - \mu_{mj}$, you want to create an interval.

(Refer Slide Time: 51:38)



So, what will be that interval? You see that we will approach will go by Bonferroni approach. Why we are interested in Bonferroni approach? Here, you see this slide, see the slide here.

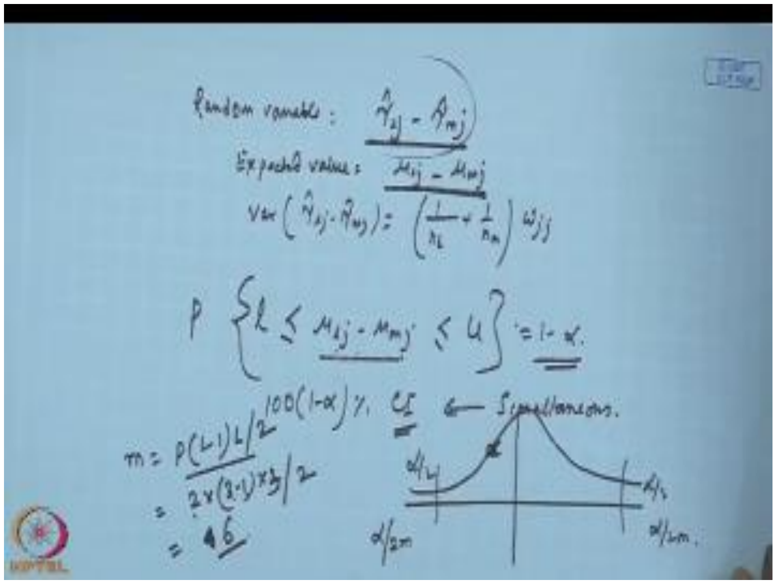
(Refer Slide Time: 51:56)



Now, how many comparisons are possible? How many comparisons are possible from the variable point of view? You see there are L populations. You are comparing two at a time. Then, L x L -1 / 2 will be the number of comparisons because if there are 3 populations, 3 x 3 -1 / 2 that 3 comparisons if there are 4 populations, then 4 x 4-1 / 2 that is 6 comparisons possible, 6 comparisons possible will 6 ∑ vectors comparisons.

Now, we are comparing again the variables. So, the p values here coming, how many variables are there? P variables are there. So, in total, you have m comparisons, but what you want? You want an interval here, l and u in such a manner.

(Refer Slide Time: 52:51)

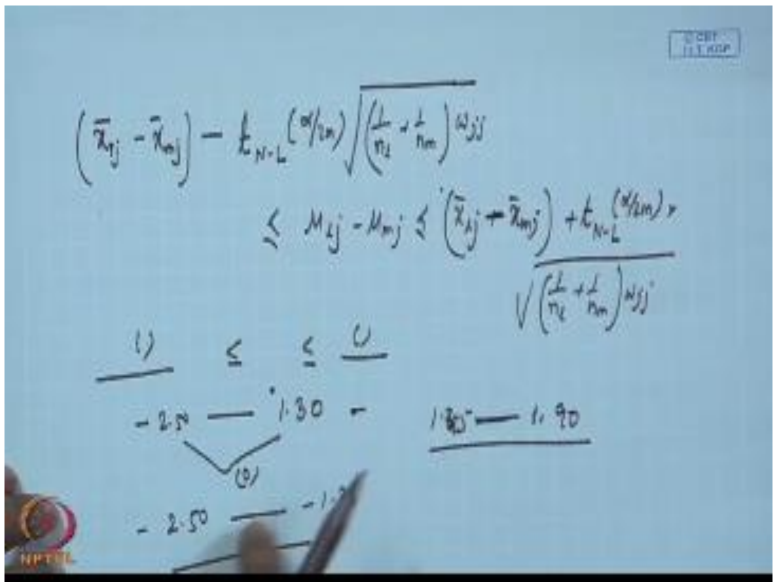


That this will probability of this less than l less than is equal to, this equal to 1 minus alpha in such a manner that you will achieve 100 - 1 per cent confidence interval that should be simultaneous. That should be simultaneous. Now, we have seen earlier. Now, there are two variables, two variables, two population cases that there are two approach maximization lama, your Bonferroni approach and Bonferroni approach is easier. So, we are considering Bonferroni approach.

In Bonferroni approach, what happened odd? It is basically first saying that how many comparisons are there which ∑s how many μ_{lj} - μ_{mj} that is m? m = p L -1. L in our case, it is p is equal to 2, L -1 will be or3 -1 x 2 / 2 divided by 2.So, it is 4 p is 2, 3-1, this 2, so 4 comparisons are possible. 3 comparisons, p is, p is how much? P is 2, L is three. So, this is 6 comparisons. L is 3, so 6 comparisons. Simultaneously, we are to make sure our α it is a two tail case t distribution.

We will be using two tails, this side α / 2, this side α / 2, but what happened? We have 6 comparisons that Bonferroni says you divide it either equally or with certain vertex, you are dividing actually. So, that mean α/2 m and this side also α/2m, so then this is the case. So, if this is the case, then what you are required to know? Now, that ∑s we are creating a t statistic here and t statistic; the degree of freedom will be N - L.

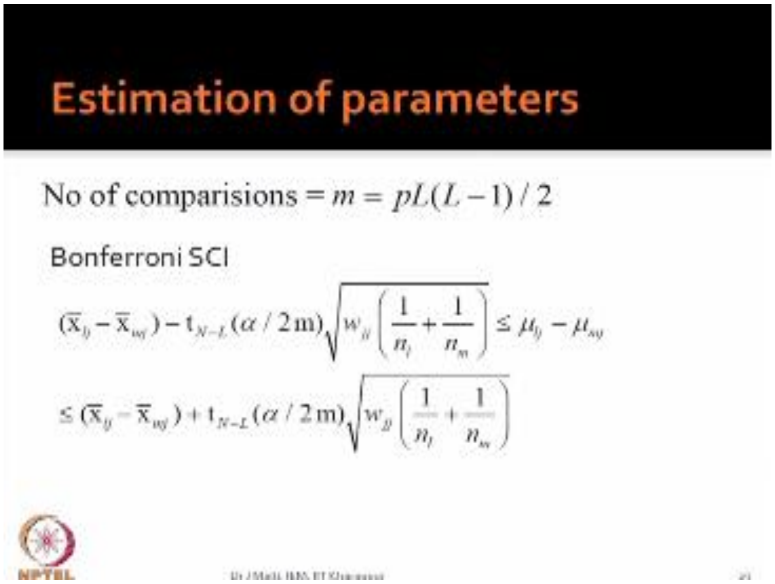
(Refer Slide Time: 55:22)



You are considering α/2 m k because m, m comparison is there and you are also multiplying this with what the variable be component, what is this variable be component 1 / n_l +1 / n_m x w jj. This one will be subtracted. That portion would be subtracted from $\bar{x}_{lj}$ - $\bar{x}_{mj}$, this minus this. This is same manner the way we have done earlier. Also, we know the distribution, and then the critical value and we multiplied with the variability. This will be less than is equal to μ_{lj} - μ_{mj} less than is equal to $\bar{x}_{lj}$ - $\bar{x}_{mj}$ + t N – L α/2 m into the variance part.

So, in this manner, you have to find out the difference. That is what I said the confidence interval for the differences.

(Refer Slide Time: 56:56)



(Refer Slide Time: 57:02)

Then you find out who is difference content of μ that means under null hypothesis, there is no difference; so μ_{lj} - μ_{mj} will be 0.If you find out any interval, any interval this side or that side, this value plus this value, it contains 0, then that variable is not differentiate. Suppose the difference will come for a particular variable, let it be -2.50 to 1.3, then see within in between, there is 0. So, this variable is not creating the difference.

Suppose some value is like this 2.50 -1.30, this is a differentiating variable or 1.45 to 1.90, that is correct, that is also different; 0 is not there. This means null hypothesis is not satisfied with this interval because we have created t distribution t values ∑ that 0 will be there in between. I think this is what is one way MANOVA and you have seen in totality like this that you are comparing several population means then fine of coming to the conclusion that population means then you are trying to find out that which of the variables are causing this difference.

You are going for the interval estimation of each of the variable for different pair of comparisons, and then find out which of the variables are creating the difference. Okay next class, I will show you one case study. Thank you.

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians