Data Analysis and Decision Making - I
Prof. Raghu Nandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 35
MLE Estimates

A warm welcome to my dear students and friends a very good morning, good afternoon, good evening to all of you and welcome to this DADM which is Data Analysis and Decision Making course under NPTEL MOOC series, then we are in the 35th lecture. So, as you know this course is a total duration is 12 weeks. So, this 35th would basically signify we are ending the 7th week. And the total number of lectures is 60 and total hours content is 30 hours and each week we have 5 lectures and each lecture being of half an hour 30 minutes and I am Raghu Nandan Sengupta from the IME department IIT Kanpur.

So, if you remember we were discussing about the very briefly we discussed about the multivariate extreme value distribution. What I do is was not possible to give the example for two dimensional in a higher dimensional case I took a very simple one dimensional multivariate extreme value distribution for the four indices in the stock markets which is Nikkei from Japan, Footsie from United Kingdom.

And then KOSPI from Korea and NIFTY from India and I just do or explain that how you can take the blocks for each some duration of time, the blocks can be overlapping also. You take the maximum on the minimum depending when they are going to for the positive returns on them negative returns. Take the maximum for each blocks and then concatenate it or do bootstrapping to increase the sample size make it almost equal to the population and do find out the average values for the sample means or the variances whatever it is.

And then find out in the long run the expected value of the parameter of the random very distribution which is basically a univariate case of the extreme value distribution for the maximum. And then from that you can find out the alpha beta gamma, which are base I am using the general symbol for alpha beta gamma which is location, scale and shape.

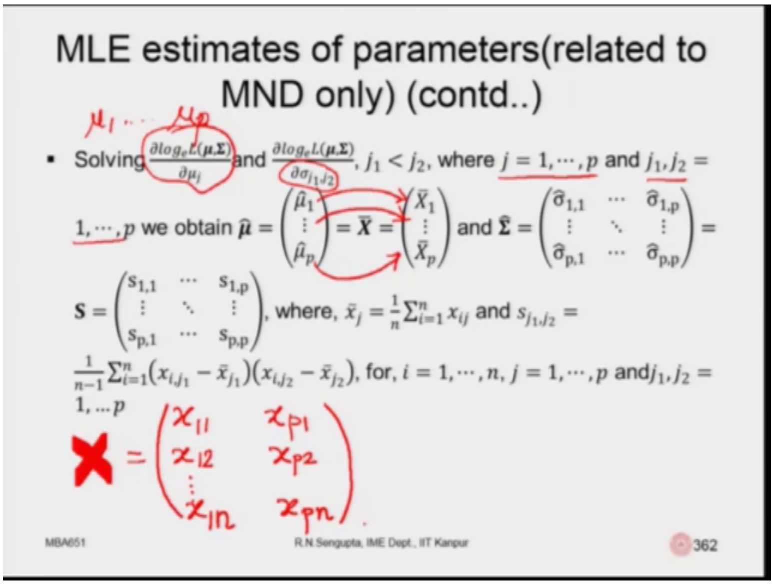
And then you can basically find out the how the distribution looks like, similarly you can do it for the minimum also and do your analysis accordingly. Then in the last, in the last part the very briefly I mentioned, that given the multivariate normal distribution case you want to find other parameters. So, I said that and generally this is norm for the maximum likelihood estimation problem, you find out the corresponding probability.

Supposing that capital X equals to x 1 capital X 2 is equal to x 2 capital X 3 is equal to x 3 find out the corresponding probability take the log and differentiate. So, the difference if partial differentiation, when I am using the word differentiation here, it is partial differentiation partial differentiation with respect to the parameters put it them to 0 and solve it.

So; obviously, they are closed form solution no problem you will get exact solutions which are basically alpha hat beta hat, gamma hat or the location shape and scale estimated value. And then you check the unbiasness consistency and then basically proceed further. In the method of moments, you basically take a sample find out the moments with respect to the parameter values, then equate it to the values which you have found, if you have the I have only dealt with only two slides there. So, equate it and find out the estimated values and proceed accordingly. So, this concept will be utilizing here also.

So, once you have the logline crude function. So, logline crude functions consists of something to do with the fixed values n and p, with p is the number of random variables we are taking n is the sample size, second term would be something to the variance covariance and the third term would be the difference of x minus mu whole square transpose and all these things. So, basically the mean values and the variance covariance matrix. So, if you different these parameters are mean which is mu 1 mu 2 mu 3 till mu p similarly sigma 1 and sigma 2 will be the standard deviation. So, you take them accordingly.

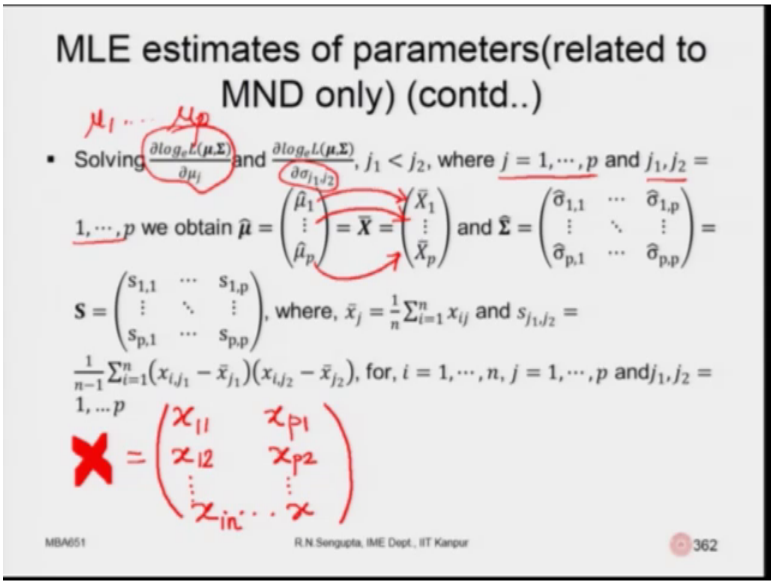
(Refer Slide Time: 05:10)



So, solving so, you basically partially differentiate with respect the log likelihood with respect to all the mean values. So, these are to be found out these are the parameters. Similarly you will differentiate with respect to the standard deviation of the ith and jth 1 and take j is equal to 1 to p j 1 j 2 is equal to 1 to p find it out. So, once you find it out you will find the hat values. So, these are the hat values so, the mu hat would be equal to the sample mean for the first set of observation. Similarly mu 2 hat will be the sample mean for the second random variable set of observation, similarly for the last one which is mu hat suffix p which is the sample mean for the pth observation.

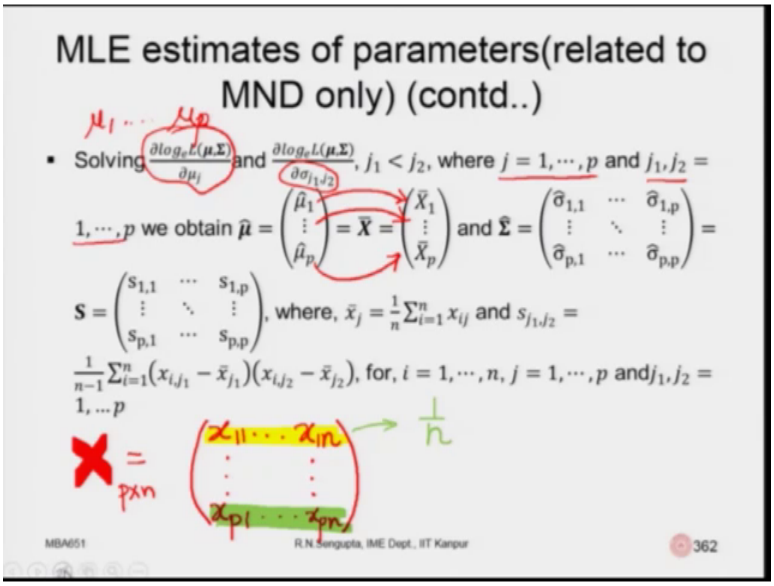
So, what you have is you have this observations x 1 1 let me do it for them, I am going to consider this whether it will be easier for me to discuss hope this basically signifies the bold. First readings you should go to the pth reading 1, for reading first random variable 2 to pth reading 2, first reading n pth reading n. So, this is of values would be number of rows would be 1, this I have thing I have not written it correctly wait please.

(Refer Slide Time: 07:20)



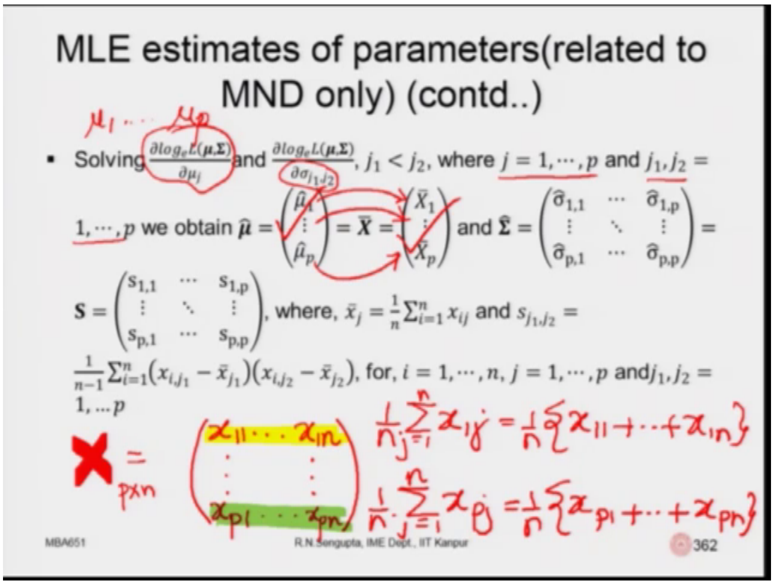
So, this is the first training first prostrating second 1 2, I think I have think I should redo it sorry just going to this.

(Refer Slide Time: 07:55)



So, it will be p rows and n columns. So, it is so, this one so, this one. This is basically the first random variable for n number of readings. And similarly this is the pth random variable n number of readings. So, I consider from here the average so, this is my colour scheme is different sorry this is it.

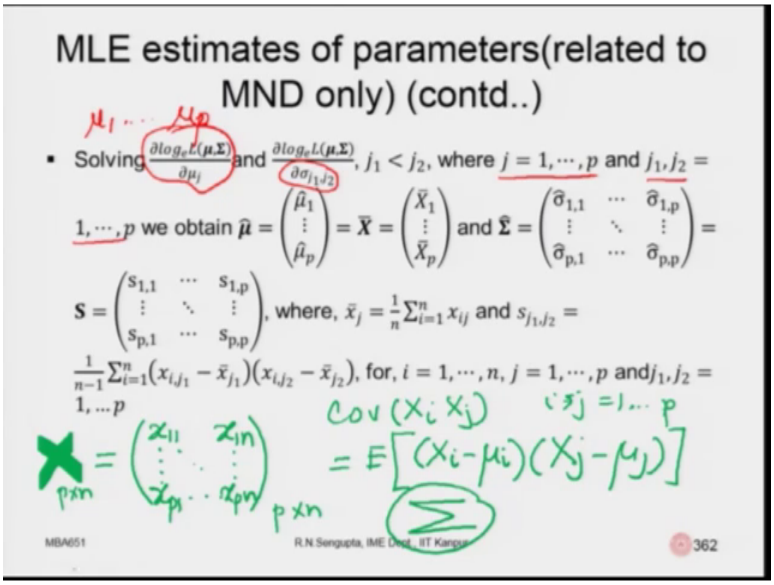
(Refer Slide Time: 09:03)



So, if you take the average it is 1 by n, summation of x 1 j j is equal to 1 to n. Similarly for the last one so, I am going a little bit slow please bear with me, just gives me the averages. So, this average is done average is done. Now, I want to find out so, this is for the averages or for the sample remember that. So, this would basically come here, this I am using the red colour. So, let me highlight it using the highlighter red. So, this would come here ok.

Now, I want to go to the first part with the expected value or the population this is for the mean, mean for the sample. So, I erase it. So, this I will highlight. So, these are the mean values for the sample. So, erase it is it again, I will again draw it, but using a different colour. So, it is not confusing for all and ok, this I need to arise also this is done. So, now, I used the green colour. So, consider coming to the variance covariance matrix.

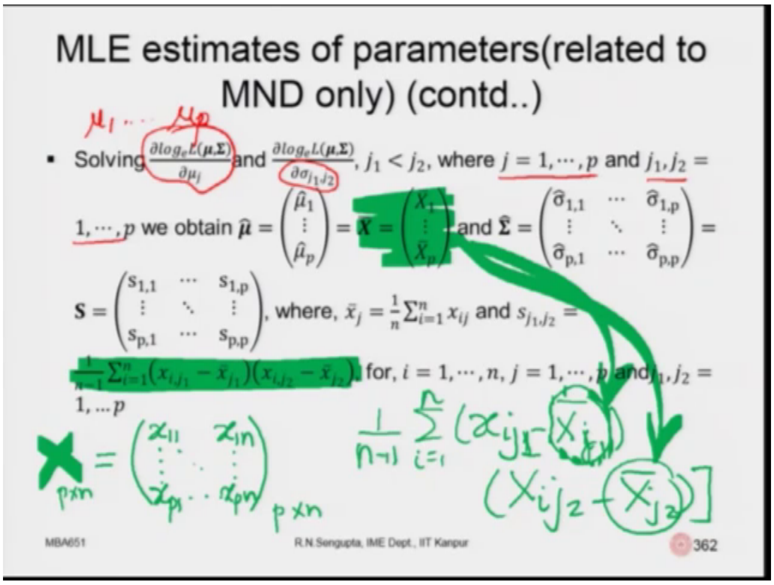
(Refer Slide Time: 11:30)



So, that you have this is the matrix, which they have same thing, but I am trying to basically give a different colour. So, now so say for example, I want to find the covariance of X i X j, so, i and j are basically changing from 1 to p. So, both values can be same or different so, that is equal to the expected value X i minus mu i into X j minus mu j. And, I have discussed this that how you can find out in the joint distribution in the discrete case on the calculus case, I just discussed the formula, if you remember those bullet points were there 1 2 3 4, they were about 9 bullet points in one of the lectures.

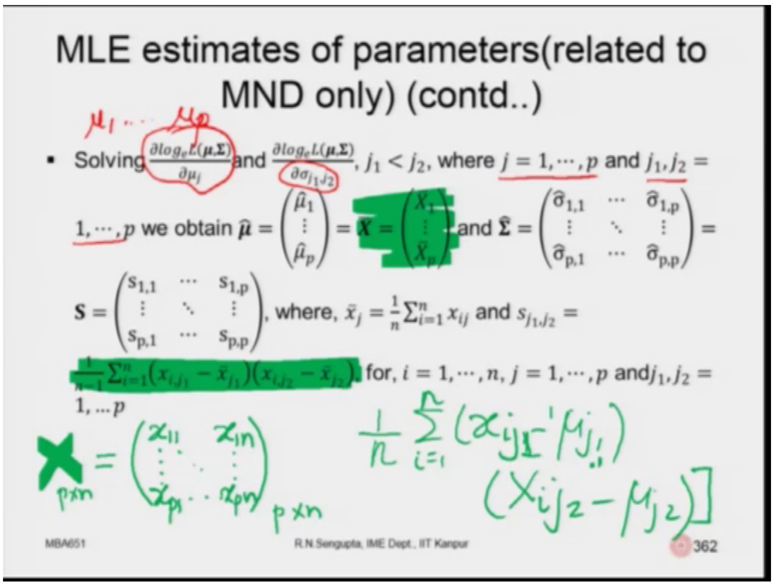
Now, in case if i is equal to j you will have the variances and in case that the not i is equal to T we will have the covariances; obviously, we will have the covariance variance matrix which is this one. So, this is the covariance matrix, but now in the case when you are taking the standard error square so; obviously, you would have today would be given by the formula here, it will be 1 n minus 1.

(Refer Slide Time: 13:54)



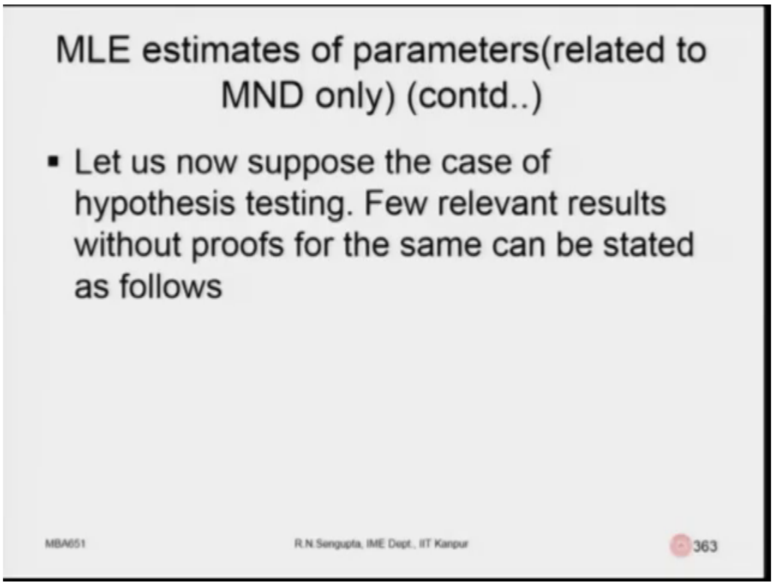
But why n minus 1 because you are losing 1 degrees of freedom summation of X i j 1 minus this is suffix 1 X bar so, I will take i S 1 to n and I am basically change this to j 1 multiplied by X i j 2 minus X j 2 bar. So, these are the average values which are already found. So, they are from here, you already found these are the values, these are the values which you are handing out. And once you find out the homeless would be corresponding to this formula would be equal to the standard header whole square, or standard errors corresponding to the fact that the mean value is not known, if the mean value is known obviously, in that case it would be and only.

(Refer Slide Time: 15:17)



You are not do this and in this case obviously, one consider this. So, this would become mu j 1 bar mu j to not at the not bar sorry mu j 1 and mu j 2 which are the expected values. And corresponding to the case the symbol generally denoted for the variance covariance estimate value would be S, the bold s because there is a peak. So, you can find them and then basically utilize them in your calculations, I went a little bit slow. So, we could basically understand.

(Refer Slide Time: 16:11)



Now, let us consider the few cases of hypothesis testing, now in the hypothesis testing again I will recap. If, you remember in the hypothesis testing and there are four different four types of problems something to do with the mean something we will do with the variance, but under each mean and variance even I am the two class. For the mean you will consider in the first case the variance is known for the population so; obviously, in that case you will use the z distribution.

In other case you will consider the variance is unknown. So, what you will do you will replace the variance of the population using the best estimate from the sample, which would be S, because S dash basically means you have some informations about the mean of the population.

So, when you use S without the dash you will use the T distribution with n minus 1 degrees of freedom, we are trying to find out the difference of two means provided the variance of both the populations are known, then you will use the z distribution without any loss or degrees of freedom. But when you have basically the variances of both the populations are unknown and, you want to find out something to do with the differences of the means you will use the T distribution. And the m minus 1 from the first case and n minus 1 from the second case. So, it will be m plus n minus 2.

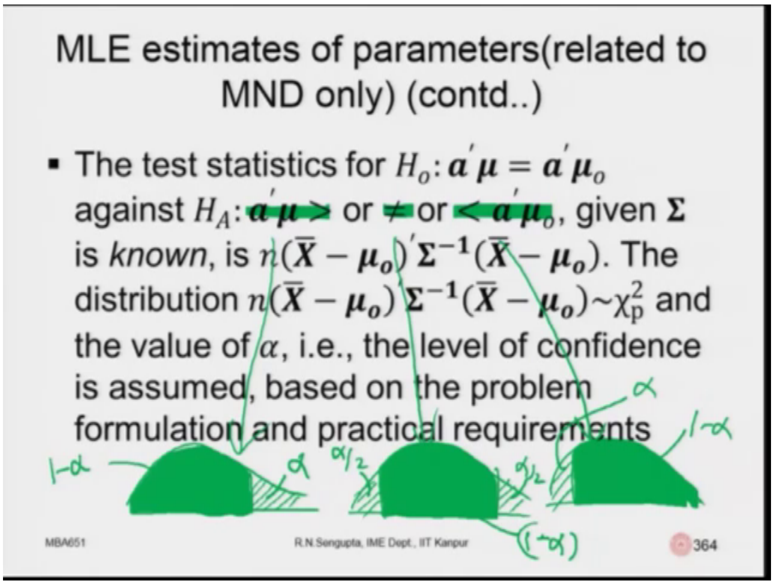
Now, when you go to do something with this the variants are two different sets there, in the variants if provided the population mean is known, then you will is use the chi square without losing any degrees of freedom, because in that case you are using the S dash, where the population mean is known. And in case and; obviously, degrees of freedom would be n.

And when you are trying to find out something to do with the for the one population, when something to do with the variance of the population giving the mean of the population unknown newly use the chi square with n minus 1 degrees of freedom. Then if you switch on to the case of finding or the ratios of two populations, there would be two different instances or two different flavours.

In the first case the flavour is both the mean values are known, hence you use the f distribution with degrees of freedom m comma n, because the freedom using in the ratios of the variances. But, the mean values of them you will use the f distribution with m minus 1, 1 losing 1 degrees it from the first case next one losing 4 1 degrees of freedom from the second case.

Now, if we remember these distributions which you have considered on the counterpart in the multivariate case, were the student T distribution for the multivariate case and weiser distribution for the chi square. So, these concepts which we have used in the univariate case would be utilized in the multivariate case also. So, let me continue reading, let us now suppose the king few relevant results without proofs for the same can be stated as follows.

(Refer Slide Time: 19:20)



Now, our case for so, there would be multiple cases, but we will go logically. So, we want to state test the statistic the alpha betas the pi and the type 1 type 2 it as concept would be there, the test statistics for H naught given that here in this case you will consider mu, under H naught I am H A which is the alternative one you can have greater than less than not equal to so; obviously, if its greater than you will consider the right hand side, I am looking from my side, less than you look at the left hand side not equal to you look at the on the middle range; obviously, there is the upper control limit.

And the lower control limit and remember again the same concept that for the upper control and the lower control, when it is basically greater than less than you will always use the suffix for the distribution, whether T or chi or f it will be 1 minus alpha and alpha. For the left and the right, when it is two sided, when you are using not equal to it would not be 1 minus alpha and alpha, it would be 1 minus alpha by 2; that means, alpha by 2 not the total value and in another case it would alpha by 2; such that the whole of area which you are going to cover from the leftmost to the right most would be 1, as per the concept that the c d f value is 1.

The test statistics is as I mentioned for the h naught is a prime, you know a prime a transpose a is a vector, a transpose into mu is equal to a transpose into mu naught on the h naught. And, on the under the alternative 1, it is a transpose into mu a which can be greater than a transpose into mu equal to H, a transpose mu naught and will be less than equal to a transpose mu naught so, three cases So, if we have for the greater than type you have I am just drawing the distribution on the right hand side, not equal to and creep less than type this is for this, this is for this is for this.

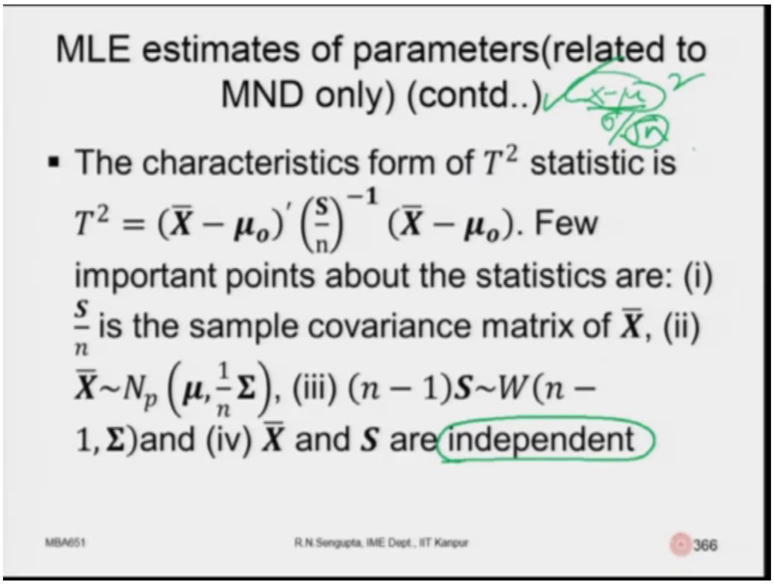
So, given the variance covariance matrix; obviously, covariance is variance matrix is known, then we need to find out the distribution of n into X bar into mu not into the inverse of transform the inverse of the variance covariance matrix. That is equal to and multiplied by X bar which is the sample mean minus the mu naught random under h naught that is chi square distribution with p degrees of freedom and the value of alpha would be decided accordingly which is the level of Corfu confidence.

So, in this case here and use a different colour I will use the highlighted. So, this area on this area and finally this area so these areas is equal to 1 minus alpha. So, this is alpha this is alpha this is alpha by 2, this is equal to 1 minus alpha, this area is equal to alpha this is equal to the whole area remains constant as 1. So, in the value of alpha there is the level of confidence for formulation and the practical recommend and you will basically solve the problems accordingly.

Now, you will consider that test assist statistic considering the same thing H naught value of a transpose mu is equal to a transpose mu naught and a H naught and under H A, ordinary I this is your three different flavours. First one is a transpose into mu is less than, the second case is a transpose mu is greater than and the third case is a transpose mu is not equal to with respect to the H naught; so, less than not equal to greater than r with respect to the H or not.

Now, in this case the variance covariance matrix is not known. So, if it is not known you basically use the S matrix. So, it may stick we have just done it. So, basically you lose one degrees of freedom in each of this reading when you are trying to find out, the standard deviation square for x 1 or x 2 or x 1 and so, forth. So, that test statistic which is highlighted here is T distribution this multivariate distribution with T squared is 1 with p degrees of freedom. And the value of in the practicalities and we solve it accordingly.

(Refer Slide Time: 25:26)



Now, the characteristics form of the T square statistic this is not the T distribution T square statistic and by the fact that, if you remember it we take basically the case, where it is not 1 to 1 similarly, but I am trying to basically give you some hint. So, in generally what you do you basically find out X minus mu by square root of n so, y square root of n.

So, that would basically be and same square sorry. So, base basically when I have the sample is becomes this. So, now I will consider this. So, this so in this case I am taking the square so; obviously, it becomes X bar minus mu transpose into X bar into mu so, you are basically going squaring this, which is right. And in other case this ended deviation which goes up so; obviously, minus 1, but we have is the square root of n in the case with; obviously, that would also go up in the case it becomes with the minus 1 power.

So, the stress statistics few important points about the statistics are sample covariance matrix and, technically the properties are X bar N which is the sample distributed with mean value mu and the covariance of the covariance matrix by N, which is the sample size. And if you remember we did discuss about the chi square distribution in the same way n minus 1 into S would be wizard distribution which is the distribution was the counterpart for the chi square distribution, with n minus 1 and the variance covariance matrix of the parameters.

And finally, which is very important for the normal distribution that for the univariate case also as well as the multivariate case, the sample mean and the sample standard deviation will be independent, this is not always true for the other distributions. So, I know that we are going a little bit slow considering the theoretical one and try to basically wrap up, with lot of examples in the later half. For any queries whatever you have please send us the in the forum the queries and we will try our level as far; have a nice day.

Thank you very much.