Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Sharmishtha Mitra

Department of Mathematics and Statistics

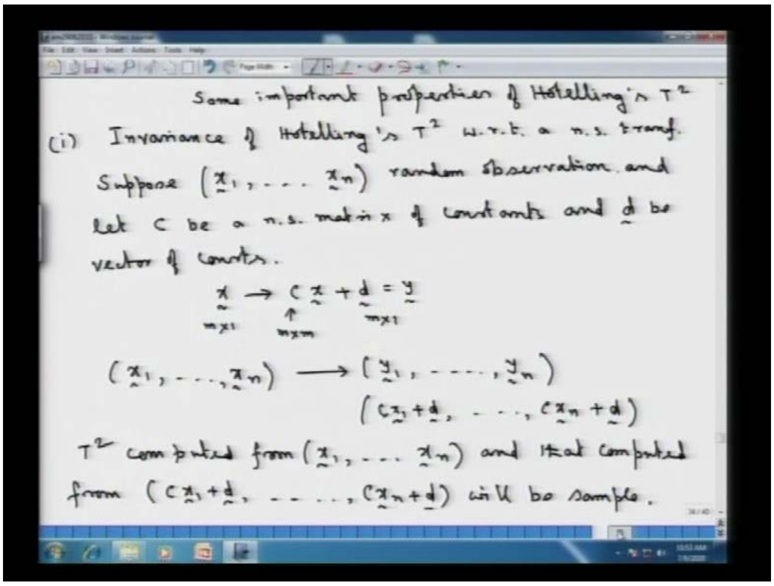
Indian Institute of Technology, Kanpur

Lecture No. # 15

Hotelling's T2 Distribution and Profile Analysis

In the last lecture, we had looked at some applications of hotelling's T square statistic, and also we had looked at how to use hotelling's T square statistic when we have two sample normal problem, and related inference regarding that. Now, what we will do today is we will look at some important properties of the hotelling's T square statistic.

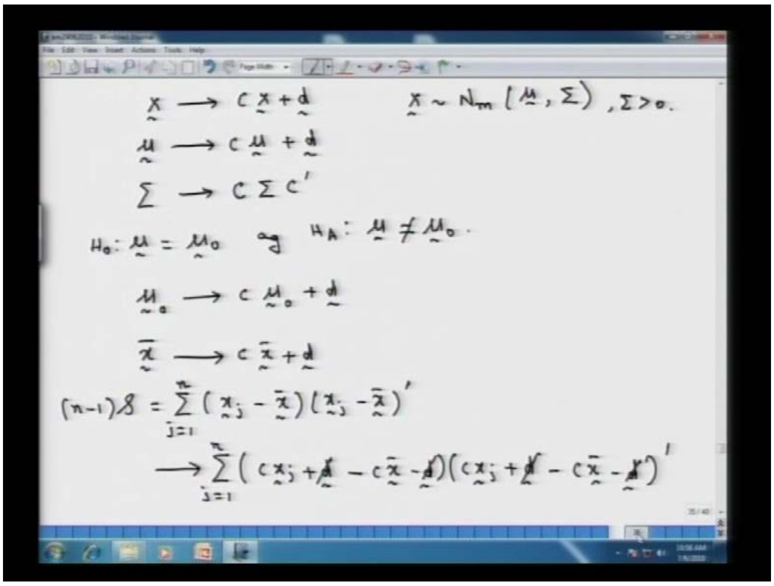
(Refer Slide Time: 00:37)



Some important properties of the hotelling's T square. Now, the first property that we are going to discuss is the invariance of the hotelling's T square statistic with respect to a non-singular transformation. So, invariance of hotelling's T square statistic with respect to a non-singular transformation. Now, what we mean by saying is that is the following that suppose we have this data x 1, x 2, x n the random observations from a multivariate normal distribution with appropriate dimensions, and mean vector and then non-singular positive definite covariance matrix. Let c be a non-singular matrix of constants, and d be a vector of constants. Now using this c non-singular matrix and d the vector of constants, we can actually make a transformation from this x to c x plus d.

Suppose all these x s are m dimensional; we will take this c to be an m by m non-singular matrix and this to be an m by 1 vector. So, this is a new vector that is derived from this x random vector. Suppose we have this x 1, x 2, x n, the original data through the transformation will be getting a new set of observations that is y 1, y 2, y n which are nothing but c x 1 plus d, c x 2 plus d and so on, and the last one is cxn plus d. Now, the T square statistic computed form this x 1, x 2, x n is going to be the same, that is going to be computed from c x 1 plus d, c x 2 plus d and c x n plus d. In that sense actually, T square will be invariant or rather I will just write that T square computed from x 1, x 2, x n and that computed from c x 1 plus d, c x 2 plus d, c x n plus d will be same.

(Refer Slide Time: 04:29)

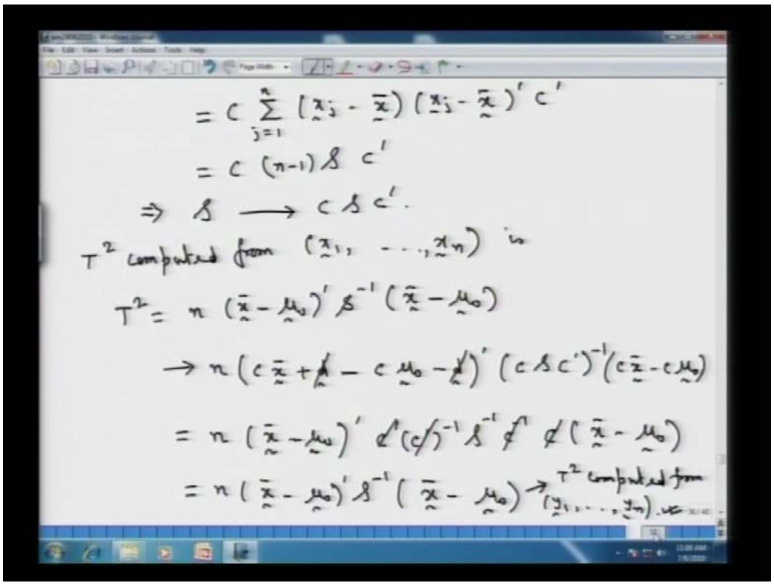


How do we claim this? We claim this particular invariance in the following way that since we are made this non-singular transformation, the x random vector the underlying random vector will get changed to c x plus d. the corresponding x to follow a multivariate normal m with a mean vector mu and a covariance matrix as sigma. Then with this transformation, this mu will be changed to c mu plus d and then this sigma matrix the associated variance covariance matrix to be positive definite. So, this sigma is going to be changed to c sigma c prime. Now, T square statistics comes into existence actually for testing null hypothesis of the form that mu is equal to mu naught to be tested against an alternate hypothesis HA that mu is not equal to mu naught.

So, where does this mu naught go to? mu naught is a known vector. So, this known vector mu naught will be shifted to c mu naught plus d. Now, when we have such transformations and place what happens to this x bar quantity? Now, x bar is the sample mean vector that is obtained from x 1, x 2, x n. So, this x bar is computed from x 1, x 2, x n, the data. So, this will be changed to c x bar plus d and n minus 1 s say with a divisor n minus 1; that is given by x j minus x bar x j vector minus x bar vector transpose j equal to 1 to up to n. Now, where does this gets changed to when we are looking at the transformed observations, this x to be replaced by c x j plus d.

Now, what happens to x bar? x bar is c x bar minus d in to the transpose of that. So, it is c x j plus d minus c x bar minus d transpose. So, this d vector cancels out from both the quantities and what happens is, the following we will have here c is a non-singular matrix. So, this c can be taken out from the left, c transpose can be taken out from the right and what will be having is the following that, the quantity that I had written is going to be, c times summation j equal to 1 to n x j minus x bar then x j minus x bar transpose c transpose and this term is equal to n minus 1 times s with the divisor n minus 1.

(Refer Slide Time: 07:13)



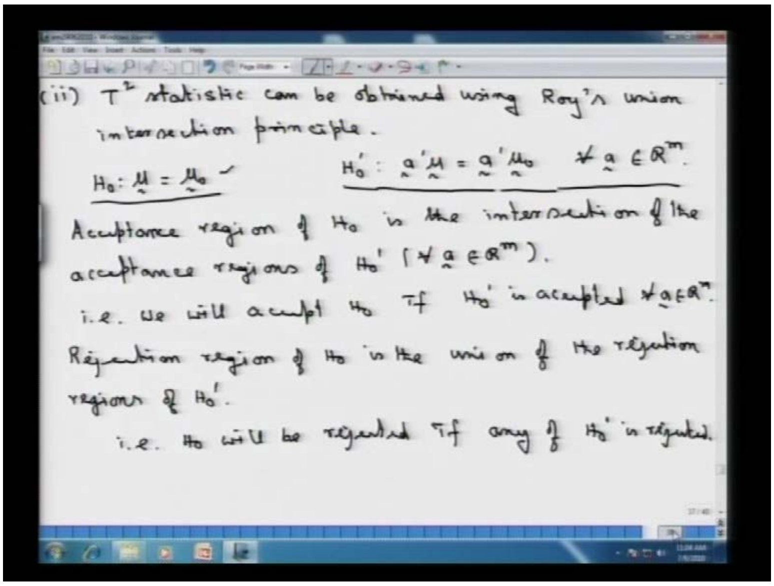
So, what we have seen is the following that this quantity out here that n minus 1 s. Now, with this non-singular transformation is going to c times n minus 1 s c prime. This would imply that this s the sample variance covariance matrix is going to c s c transpose. Now, once we have these things in place the T square computed form x 1, x 2, x n the original set of observations is given by T square is equal to n times x bar minus mu naught transpose s inverse x bar minus mu naught. So, we had seen this and again this is basically that hotelling's T square when we are looking at x 1, x 2, x n; n observations from a multivariate normal distribution for testing the null hypothesis mu equal to mu naught against mu not equal to mu naught.

Now, where does this go to? We will have this under the non-singular transformation that we are discussing; this x bar is now c x bar plus d. Where is mu naught? mu naught is c times mu naught minus d transpose. Now s is going to c s c transpose. So, here what we have is this is just c s c transpose whole inverse and then that is multiplied by the transpose of this particular quantity what would remain? This d is cancelling out and we will have c x bar minus c mu naught on this side. Now, this is equal to n times; now this c can come out side with a transpose; this goes here and what will be having here is this x bar minus mu naught.

So, this c with a transpose comes here and then we will have this term here. So, which is c transpose inverse s inverse c inverse and then we will take c from this side as well; this is c x bar minus mu naught. So, this c into c inverse will give us an identity matrix; this c transpose into c transpose inverse will also give us an identity matrix. So, what will be having is n times x bar minus this into s inverse times x bar minus mu naught. Now, what is this quantity? This is the T square statistics which is computed from the y observations; because this is nothing but y bar; this is nothing but the mean corresponding to the y random variables.

This is the sample variance covariance matrix that is computed from the y observations that is what we had seen out here; that we have these under the transformation here; that we are making that x 1, x 2, x n is transformed to y 1, y 2, y n. Then what happens to the corresponding means? This is the mean vector corresponding to the y observations. This is the variance covariance matrix corresponding to the y observations and what we have proved is that the T square computed from the x 1, x 2, x n which is equal to what is this quantity. This is the T square computed from the y observations computed from y 1, y 2, y n. So, what we have proved is that the T square statistics remains invariant with respect to this non-singular transformation.

(Refer Slide Time: 11:38)



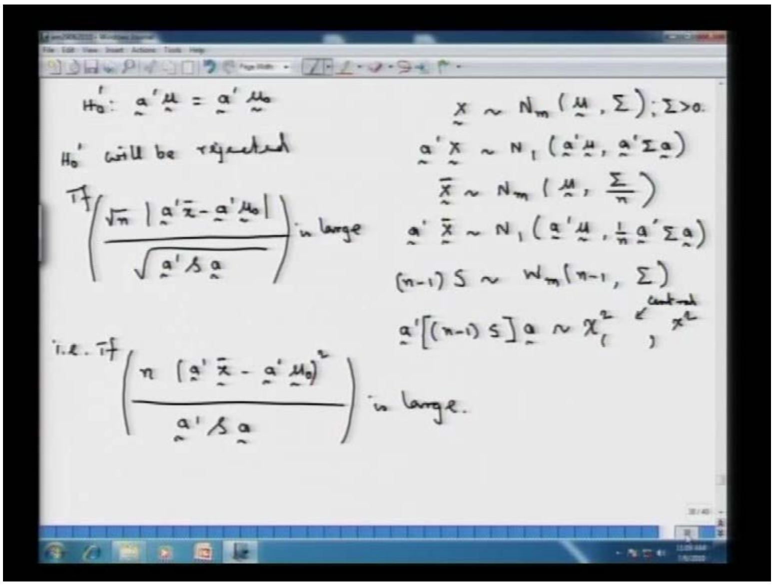
So, that was the first important property that, we can say about this hotelling's T square. Now, the second thing is that the T square statistic can be obtained using Roy's union intersection principle. Now, what is that? We have this null hypothesis H naught that mu equal to mu naught. Suppose we take this alternate type of hypothesis H naught prime, which is a prime mu equal to a prime mu naught. This is for every a belonging to the appropriate dimension R to the power m; because we have got a multivariate normal which is m dimensional. So, what is the correspondence between the two hypothesis H naught and H naught prime? Now, the acceptance region note that this is a single hypothesis here and there are a number of hypothesis for possible choices of the a vector belonging to R to the power m on the right hand side.

Now, the acceptance region the relationship that is what we are talking about; the acceptance region of H naught is the intersection of the acceptance regions of H naught prime for every a belonging to R to the power m. Now, what does that mean? That means, this H naught null hypothesis is going to be accepted. If we have this H naught prime hypotheses to be accepted for every value of this a vector; that is, we will accept H naught. If H naught prime is accepted for every a belonging to R to the power m and that is quite obvious; because if for some a this null hypothesis H naught prime hypothesis is rejected, we cannot take this hypothesis to be accepted.

Now, this is about the acceptance region. So, the acceptance region of this would be the intersection of the acceptance region of H naught prime. What about the rejection region? The rejection region has the following relationship between the H naught and H naught prime hypothesis. The rejection region of H naught is the union of the rejection regions of these H naught prime set of hypothesis; that is H naught will be rejected if any of H naught prime hypothesis is rejected. So, rejection of any one of these H naught prime hypothesis would leading to the rejection of the main hypothesis; that is H naught.

Accordingly, the rejection region of H naught would be the union of the rejection region of H naught prime hypothesis for a varying in R to the power m. So, this basically is what we talk about the union intersection. It is called the union intersection principle; because the rejection region of this H naught hypothesis is the union of the rejection region of these set of hypothesis. The acceptance region of this H naught is going to be the intersection of the acceptance region of this H naught prime set of hypothesis. Now, what we will do is that we will basically look at when is this H naught prime hypotheses going to be rejected? On what sort of theory, we are going to actually base our rejection region. So, H naught prime is a prime mu equal to a prime mu naught.

(Refer Slide Time: 16:12)



Suppose, I have take this one hypothesis here; H naught prime is a prime mu equal to a prime mu naught for particular a vector. Now, in order to test this particular hypothesis, what we are going to make use of is the following. Now, we have x to follow a multivariate normal m dimension with a mean vector mu and a covariance matrix sigma positive definite. So, we will have sigma to be positive definite out here. So, x has got this particular normal distribution. So, what happens to the distribution of a prime x? a prime x will have a multivariate normal distribution which is a prime mu and a prime sigma a as its variance.

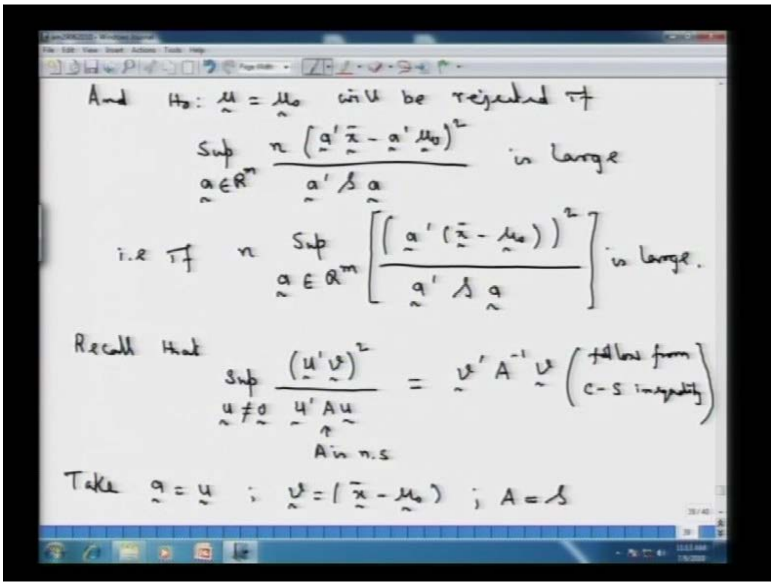
Now, from here what we can also say is that this x bar has got a multivariate normal distribution with a mean vector mu and a covariance matrix as sigma by n. So, this would imply that for this a prime vector, this a prime x bar is going to have this N is going to be one. This is N1; because this is that particular 1 by n dimensional vector. Now, this also is a univariate normal random variable with mean as a prime mu and a covariance matrix 1 upon n a prime sigma a. Now, the variance covariance matrix n minus 1 times s the sample variance covariance matrix; this has got a wishart m n minus 1 times sigma.

From the previous results, what we can say is that this a prime n minus 1 s times a. This is going to follow a central chi square that is what we had seen earlier that this has got central chi square on what degrees of freedom n minus 1 minus m plus 1 degrees of freedom and it is going to be a central chi square. So, this is a central chi square on the degrees of freedom which is going to given by n minus 1 minus m plus 1. So, using these facts actually we are in a position to test this particular null hypothesis. How we are going to frame that? H naught prime is will be rejected, if the following quantity is large; if root n absolute value of a prime x bar minus a prime mu naught divided by under root of a prime s a is large.

Why is that so? It is simple to see that; because we have a prime x bar to have a normal distribution. This a prime x bar minus a prime mu naught that divided by this variance out here; that is going to have a normal 0 1 distribution; but this sigma matrix is unknown to us. And hence, we also need to use this distribution chi square here. Eventually, what we are going to have? This distribution is going to be a tedious distribution; because that would be ratio of standard normal distribution to that of a central chi square random variable. So, we are going to reject H naught prime, if this quantity is large; that is if this square of it is large. So, there is a reason why we are looking at this square of that particular quantity; this a prime a mu naught whole square that divided by a prime s a is large.

Now, this is as far as rejection of this one single hypothesis, H naught prime for a given a prime belonging to R to the power m. So, the large quantity of this particular observed thing based on x 1, x 2, x n is what is going to lean us to the rejection of H naught. Now, what is the relationship between this rejection of this hypothesis for a particular a belonging to R to the power m and the rejection of H naught hypothesis; that is, mu equal to mu naught and what we will be having is the following. This H naught the null hypothesis mu equal to mu naught will be rejected, if we have the supremum over a of these quantities, which is n times a prime x bar minus a prime mu naught whole square that divided by a prime s a.

(Refer Slide Time: 20:54)

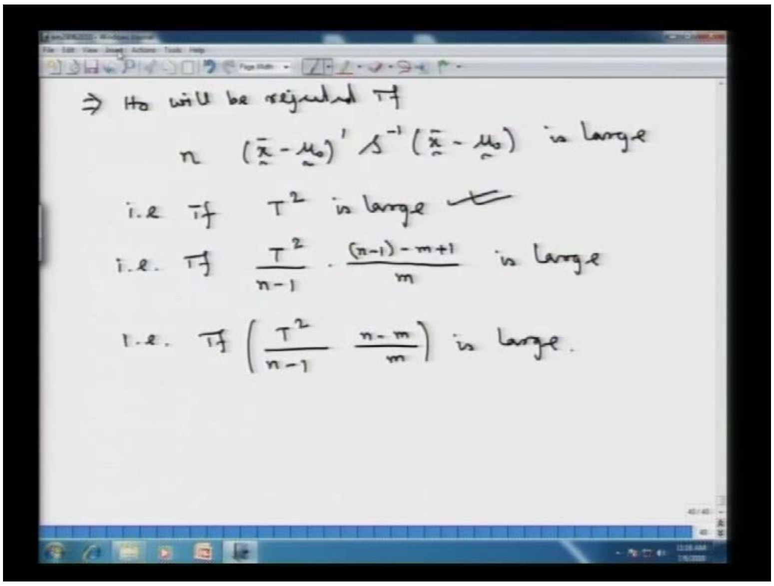


So, this is basically the supremum over every a belonging to R to the power m; that is, for all possible hypothesis H naught prime, when a varies in R to the power m. If we have supremum of these quantities for varying a is large, now we need to look at what is this; that is we are going to reject H naught, if now n is constant; n just sits outside. So, it is the supremum over a belonging to R to the power m of this particular quantity. Now, how do we write this particular quantity out here? I will take a prime outside here. This is a prime x bar minus mu naught the whole square of that divided by a prime s a is large. Now, in order to find out the supremum of this particular quantity, supremum over this term out here; we recall that we have something called a cauchy-schwarz inequality.

So, we may recall the following result; recall that supremum over u not equal to zero. Ofcourse, we also will take a not equal to zero; because a equal to zero does not mean anything; because we are going to test null hypothesis that 0 equal to 0; that does not make any sense. So, it is over all vectors which are non-null of this u prime v whole square divided by u prime Au, where A is non-singular. This is going to be given by v prime a inverse v. Now, this follows from the cauchy-schwarz inequality straight forward. So, we have this general result that for a non-singular matrix a, we will have the supremum of this particular quantity to be given by this. Now, here we will use this result, in order to find out what is the supremum of this.

So, we can take here, this a vector to be equal to this u vector; the v vector to be equal to this x bar minus mu naught vectors; this A to be equal to this s matrix. Now, if sigma is positive definite then with probability 1, this sample variance covariance matrix s is non-singular and hence, this s that we are talking about ofcourse, is going to be non-singular with probability 1; because, we have chosen sigma to be positive definite matrix and thus, this would imply that this H naught will be rejected, if we have this quantity n times; now I will plug in the supremum value of that is going to be given by, this x bar which is v transpose; then a inverse is s inverse x bar minus mu naught is large.

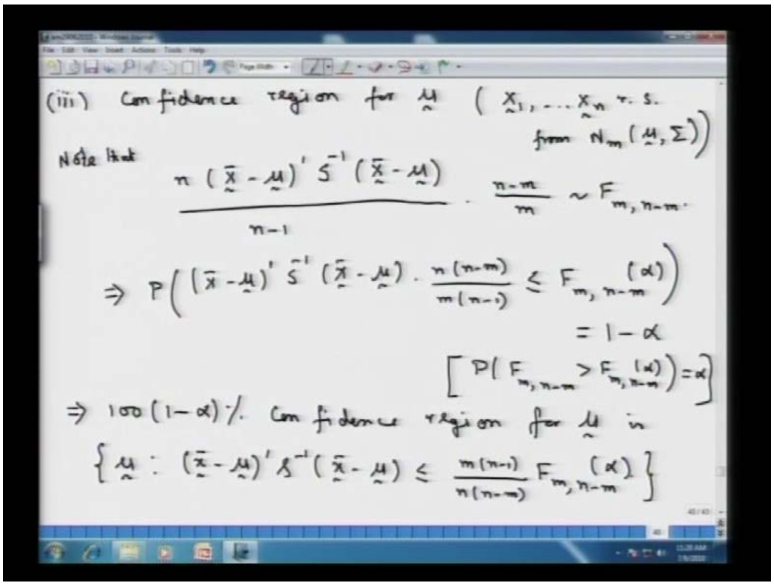
(Refer Slide Time: 24:26)



So, this follows from this equation that this is going to be rejected. This is coming through the H naught prime set of hypothesis and then that is going to be rejected, if this is large. Or in other words, if this is large; that is this term is equal to the T square statistics only; that is if T square is large. Or in other words, we can plug in the constant; that is if T square by n minus 1 into n minus 1 minus m plus 1 divided by m is large; that is, if this T square by n minus 1; now this one cancels out and you will have this here as n minus m by m is large. Now, we know what the null distribution of this particular quantity is.

The null distribution under the null hypothesis H naught mu equal to mu naught. This is going to have an F distribution; a central F distribution on what degrees of freedom; n minus m degrees of freedom and hence, statistic is equivalent to what we have already seen. So, that is why one says that we have actually shown that the T square statistics for testing H naught mu equal to mu naught can alternatively be obtained through this union intersection principle; wherein you consider this set of null hypothesis H naught prime, which is a prime mu equal to a prime mu naught. And then, the rejection region of that actually leads us to seeing that this basically based on the T square statistic itself. Now, next what we are going to talk about is something about confidence intervals.

(Refer Slide Time: 26:51)



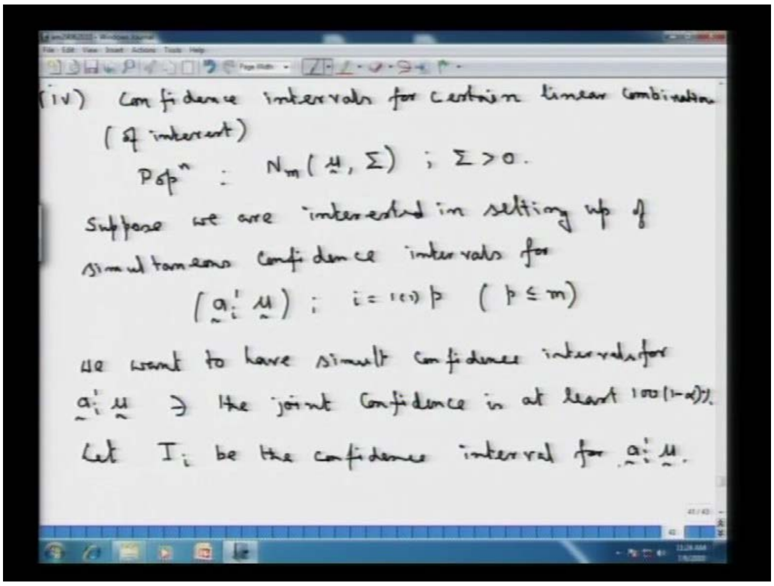
So, we will have these terms here that let me first talk about confidence region for the mean vector mu. Now this is what we have that we have x 1, x 2, x n random sample from a multivariate normal m mu sigma where sigma is a positive definite matrix. So, we are interested actually in giving confidence regions for the unknown mean vector; that is mu. Now, how are we going to do that? We know that this T square by degrees of freedom n minus 1 n minus m by m; this follows an F distribution. I will just write the full form of it. So, that it becomes easy to frame the confidence region. Note that, we will have this n times x bar minus mu transpose; then we have s inverse x bar minus mu this term here.

This is the T square statistic divided by degrees of freedom which is n minus 1, that multiplied by n minus 1 minus n plus 1. So, that is n minus m divided by m, this will follow an F distribution; now this F distribution has degrees of freedom m n minus m. Now, if we have this particular term to hold true; this would imply that the probability of this x bar minus mu transpose s inverse x bar minus mu; this multiplied by all these constants out here; n times n minus m divided by m times n minus 1; this less than or equal to Fm n minus m alpha.

What is this probability going to be equal to 1 minus alpha, wherein this particular term is upper alpha percent point of a central F distribution on m n minus m degrees of freedom? So, the area to the right of this particular point is alpha and hence, the area to the left of it is 1 minus alpha. So, I will just write it that probability of an F statistic on m n minus m degrees of freedom greater than this Fm n minus m alpha; this is a given point. So, this right tail probability is equal to alpha and hence, since this has got an F distribution on m n minus m degrees of freedom. The probability that this less than or equal to 1 minus alpha.

So, this would imply that a 100 into 1 minus alpha percent confidence region for this unknown vector mu is going to be given by the set of all mu values such that we will have this x bar minus mu transpose s inverse x bar minus mu. This is less than or equal to m times n minus 1 divided by n times n minus m times F m n minus m times alpha. So, what we have is ellipsoid; ellipsoidal region actually is giving us a 100 into 1 minus alpha percent confidence region for every mu; that is satisfying this particular condition that it is within this particular boundary region here. We will have that ellipsoid to lead us to 100 into 1 minus alpha percent confidence region for this mean vector mu.

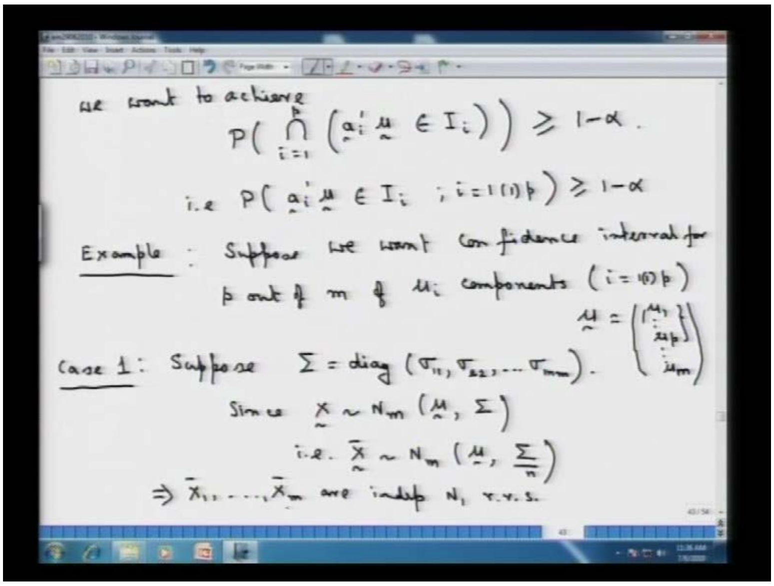
(Refer Slide Time: 31:25)



Now, let us move on to simultaneous confidence intervals for certain linear combinations of interest. Now, the underlying population still is a multivariate normal. So, the population is a multivariate normal n dimensional with a mean vector mu and a covariance matrix sigma, which is assumed to be positive definite. So, we have this sigma matrix to be positive definite. Suppose we are interested in setting up of simultaneous confidence intervals for quantities of the form that it is a i prime mu; this i is for 1 to up to p; where this p is less than or equal to m. So, we have p such linear combinations a i prime mu and these are linear combinations of the parameter. Say for example, we are interested in setting up of simultaneous confidence interval for mu 1 minus mu 2 and mu p minus mu p minus 2 something.

So, we are interested in such p linear combinations of the unknown mean vector mu and we are trying to setup confidence intervals in such a way that we want to have simultaneous confidence intervals for these a i prime mu; such that the joint confidence or the joint coverage actually is at least 100 into 1 minus alpha percent. So, that is what is our objective; we have p such linear combinations. We are trying to put up simultaneous confidence intervals for this; such that the joint confidence is at least 100 into 1 minus alpha percent. Now, what we are going to do is that let I i be the confidence interval for this a i prime mu component. So, if we have this I i to be the confidence interval for a i prime mu what we are trying to achieve is the following.

(Refer Slide Time: 34:46)



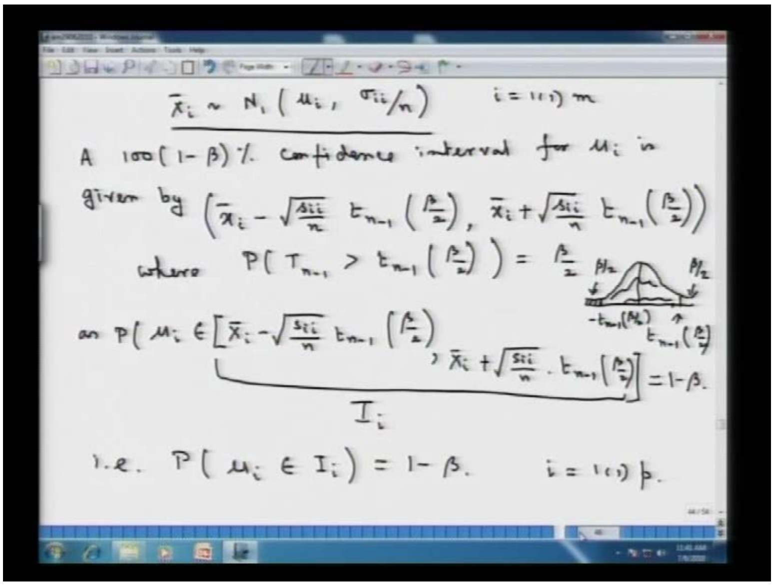
We want to achieve the following probability statement. This is intersection of each of these i equal to 1 to p that a i prime mu; this belonging to I i. So, this is an event that a i prime mu belongs to I i. So, that I i is what we are defined to be the confidence interval for particular term out there and then the intersection of all such events this probability we required that to be equal to 1 minus alpha. So, we will show I we can achieve under different setups such a probability statement that the joint probability of all these a i prime mu quantities belonging to the respective intervals that we are going to create. The joint probability intersection of all such events is going to be greater than or equal to 1 minus alpha.

So, this is the statement that we are going to make that this basically is the joint probability statements. So, this alternatively can be written in the following form that it is probability that this a i prime mu is belonging to this I i interval; this is for i equal to 1 to up to p for all these linear parametric functions a i prime mu; that is greater than or equal to 1 minus alpha. Basically, this is going to give us the confidence region for each of these p linear combination of these mean vector mu and we will ensure that, this probabilities at least 1 minus alpha. Let us take an example and then, try to illustrate how this type of problems, simultaneous confidence intervals is achieved? Now, suppose we want confidence interval for some p out of m of mu i components without loss of generality, we take that we are interested in the first p of those components.

So, we will have this as i equal to 1 to up to p. So, these are mu i components; we had this mean vector mu was mu 1, mu 2, mu m. So, we are looking at the first p components; suppose these p are important once and we are looking at setting up simultaneous confident interval for mu 1, mu 2, mu p; such that the joint probability of mu i being contained in that particular random interval; that joint probability is greater than or equal to 1 minus alpha. So, this is just an illustration; this can be anything other than these mu components also. Now we can encounter two different cases. The first case is a very simple case. Case one is suppose sigma is equal to diagonal sigma 1 1, sigma 2 2, sigma m m. Now, this sigma matrix is diagonal matrix.

Now, sigma matrix being diagonal matrix implies that since we have got x to be multivariate normal with mean vector mu and a covariance matrix sigma; in other words, this x bar has got a multivariate normal m with a mean vector mu and a covariance matrix sigma by n. This would imply that x 1 bar, x 2 bar, x m bar which are the constituent elements of this x vector, which is the sample mean random vector. So, this x 1, x 2, x m are independent N 1 random variables; because we have sigma to be diagonal matrix. Since sigma is diagonal matrix, the half diagonal entries which are going to give us a covariance terms for the components of this x vector. They are going to be zero and since the joint distribution is multivariate normal, we will have these components to be independently distributed.

(Refer Slide Time: 39:32)



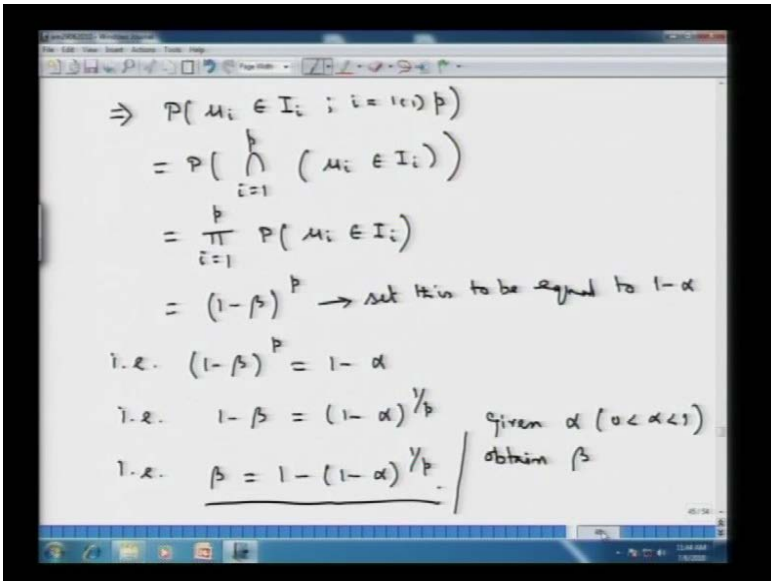
Since we have this x 1 bar, x 2 bar, x m bar to be independent N 1 random variables; in particular, what we can write is that this x i bar would follow a univariate normal distribution with mean as mu i , which is the corresponding component in the mean vector and this variance has sigma I i divided by n. This is true for every i equal to 1 to up to m and in particular, this would be the case if we are looking at any p of the components which are there in mu and the corresponding components in x bar random vector. Now, from this statement what we can write is the following; a 100 into 1 minus beta percent confidence interval for this mu i is given by x i bar.

This is the interval; x i bar minus root over s i i by n, where s i i is the corresponding diagonal element of the sample variance covariance matrix. So, this s i i by n multiplied by t n minus 1 and then we will require this to be beta by 2. I will say what is this equal to this is x i bar plus root over s i i by n t n minus 1 this beta by 2 where probability that T distribution on n minus 1 degrees of freedom greater than t n minus 1 beta by 2; this would be equal to beta by 2. So, if we have the probability of a T random variable exceeding this t n minus 1. This is basically the right tail cut off point. So, we will have t distribution is symmetric.

Suppose this is a point here; this point is my t n minus 1 beta by 2. So, the area to the right of that particular point is beta by 2; this is symmetric distribution. So, we will have minus t n minus 1 beta by 2 point. Similarly, the area to the left of that would also be equal to beta by 2 and hence the area in between these two points, which is this area is going to be 1 minus this plus this that is 1 minus beta. So, we will have this particular as confidence interval; this as probability that this mu i belonging to the random interval. Now, in terms of the random interval, this is x i bar minus square root of s i i by n t n minus 1 beta by 2; this x i bar random interval plus root over s i i divided by n times t n minus 1 beta by 2.

So, this is that interval. This is the lower confidence point. This is the upper confidence limit. This probability is equal to 1 minus beta. So, this is as far as the mu i component is concerned. Now, note that these x i bar quantities are independent due to the structure of sigma matrix that we have assumed. So, let us denote this particular interval what we have here to be this I i interval that is probability that this mu i belonging to this I i. This is exactly equal to 1 minus beta. Now, similar to one mu i component here, one can take this for every I i equal to 1 to up to p.

(Refer Slide Time: 43:46)



Now, because we have the sigma matrix to be diagonal, the independence would imply that probability that mu i belongs to I i ; this simultaneously for i equal to 1 to up to p. This is going to be given by probability. This is actually intersection of i equal to 1 to p; these events are mu i belonging to this I i. Now, these events are going to be independent; because we have chosen the sigma matrix to be a diagonal matrix. Hence, this is going to be the product i equal to 1 to up to p of the respective probabilities; because, the underlying p events with which we are looking at the intersection.

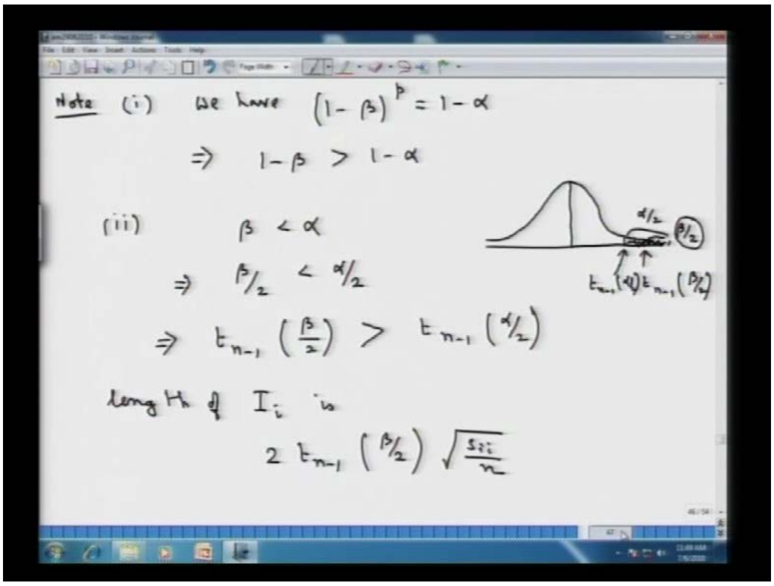
They are independent events and hence we will be having this as mu i belonging to this I i. Now, we have obtained these probabilities. I i is given by this particular random interval and what we will be having is the coverage that the probability that mu i belonging to a particular I i ; that probability is 1 minus beta. So, we will have this as 1 minus beta whole raise to the power p. Now, what do we require? We require in order to setup a simultaneous confidence interval; see here, what we had stated out here that in order to give us the simultaneous confidence interval.

We would require this statement of this type that probability that a i mu belonging to I i, for i equal to 1 to p; this is greater than or equal to minus alpha. So, for a given problem if we set this 1 minus beta to the power p to be equal to alpha, then we will able to achieve. So, we set this to be equal to 1 minus alpha and then we can solve for beta. And then that solution beta would lead us to the simultaneous confidence interval; that is, we set here 1 minus beta to the power p to be equal to 1 minus alpha; that is 1 minus beta to be equal to 1 minus alpha to the power 1 upon p; that is what we have is this beta to be equal to 1 minus 1 minus alpha to the power 1 upon p.

So, from this statement out here, what we will be able to do is given alpha lies between 0 and 1; that is associated with 100 into 1 minus alpha percent confidence region. So, given alpha obtain beta using this particular equation; because you know what is p. We know how many of these components we are actually trying to include in the joint confidence interval. So, we can obtain beta from here and then using that particular beta what we will do? We will go back to this particular equation here, where we can easily obtain what is an 100 into 1 minus beta percent confidence interval for mu i.

So, we can use that beta in this statement out here and we will be able to find out the confidence interval corresponding to that particular mu i components. Once we have the confidence interval corresponding to one mu i component, we can use that in the statement out here and get a 100 into 1 minus alpha percent simultaneous confidence interval for p of these quantities here. Now, this p can also be all the m quantities, all the m components. Now, make a note of the following observations.

(Refer Slide Time: 47:25)



Now, since we have 1 minus beta to the power p is equal to 1 minus alpha, this would imply that 1 minus beta is going to be greater than 1 minus alpha; because both alpha and beta lie between 0 and 1.So, we have 1 minus beta a quantity raise to the power p that is equal to 1 minus alpha and hence, we would require 1 minus beta to be greater than 1 minus alpha; that is straight forward. Number two is an important thing is to look at the comparison between this simultaneous confidence interval that we are setting up for each of this p component here.

And then, we would like to compare this confidence interval that we are obtaining through this simultaneous approach to that of the confidence interval 100 into 1 minus alpha percent confidence interval for a particular mu component only. So, the difference here is that in one confidence interval we are setting up a 100 into 1 minus alpha percent confidence interval for mu i only and in the second, we are obtaining a 100 into 1 minus alpha percent simultaneous confidence interval for p mu i components in which, mu i is 1 of them. So, what does the intuition? Basically, the intuition will say that the confidence interval, where we are concentrating only on one mu i component.

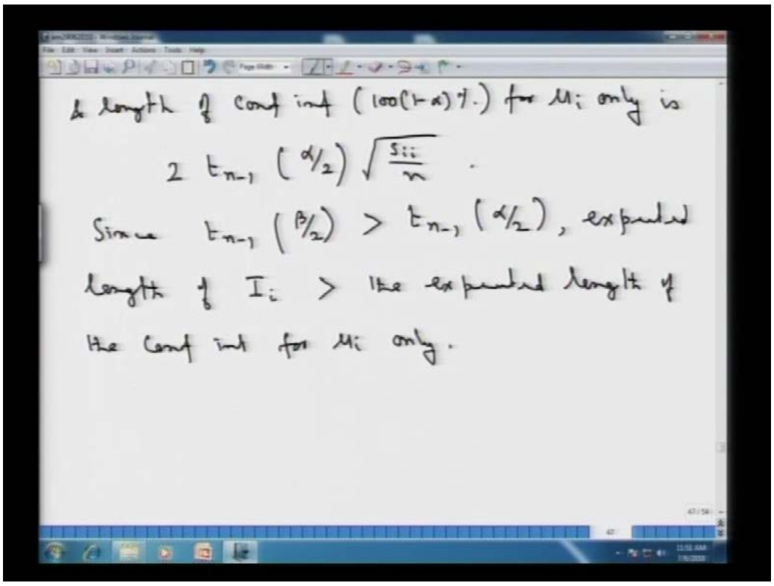
And then, setting up 100 into 1 minus alpha percent confidence interval for that is going to be shorter than the simultaneous confidence intervals for p such components and we are going to have 100 into 1 minus alpha percent simultaneous coverage for all those p components. So, the intuition would suggest that the expected length of the confidence interval, when we are looking at the simultaneous 100 into 1 minus alpha percent confidence interval; that is going to be larger than the one, when we are concentrating on one such mu i component. Let us see how we prove that intuition of ours.

So, in order to prove that we start with this particular equation here; that 1 minus beta is greater than 1 minus alpha; this would imply that this beta is less than alpha. Now, if beta is less than alpha this would imply that beta by 2 is less than alpha by 2. Now, this would further imply that t n minus 1. Now let us try to find out the logic behind this particular relationship between these two cut off points. How does this relationship between these two cut off points come across? Suppose, this is the t distribution p d f which is symmetric around 0.0 we have.

Here a point, now what do we have beta by 2 is less than alpha by 2. So, we have two points here; beta by 2 is less than alpha by 2 and these are going to be two cut off points. So, suppose I take this particular region here, beta by 2 is less than alpha by 2. So, I take this region here to have an area which is beta by 2 and then the area which is actually coming from the right of this particular point. So, this entire point here is alpha by 2. So, we will have beta by 2 which is to the right of this particular point here; having an area to the right beta by 2 and the area to the right of this point here, first point which is alpha by 2.

So, since beta by 2 is less than alpha by 2, we will have this point as t n minus 1 beta by 2 and this point is t n minus 1 alpha by 2. So, we will have this t n minus 1 beta by two cut off point to be greater than t n minus 1 alpha by 2. So, we will have this particular relationship. Since we have this particular relationship, it is now easy to see what is the expected length of the two types of confidence intervals that I was talking about. So, this is going to be the following that length of I i that we have already constructed is given by 2 times n minus 1 beta by 2. This is s i i by n and length of confidence interval 100 into 1 minus alpha percent for mu i only is going to be given by similarly 2 t n minus 1.

(Refer Slide Time: 52:06)



Since we are looking at mu i only, this is going to be given by t n minus 1 alpha by 2 and multiplier is exactly the same as this. Since, we have t n minus 1 beta by 2 as we have seen t n minus 1 beta by 2 to be greater than t n minus 1 alpha by 2 expected length of the confidence interval. Expected length of this I i is going to be greater than the expected length of the confidence interval for mu i only. So, this basically justifies our intuition that we said that our intuition suggest that; if we are trying make a 100 into 1 minus alpha percent confidence interval for mu i only, then that is basically taking care of one mu i component.

If we are going to have a simultaneous confidence interval for p such mu i's and trying to ensure that all those p components, the coverage that mu i belongs to respective I i interval that the simultaneous coverage probability is 100 into 1 minus alpha percent. Now, that is going to look at p such components and this is going to look at only one mu i component and hence, we would require in the simultaneous confidence interval set, a larger interval; larger in the sense of having the expected length of that I i to be higher.

So, this basically tells us that if we are looking at such simultaneous confidence intervals, the expected length of that confidence interval is expected to be higher. So, we will stop here today. We will look at in the next lecture what happen if sigma is not necessarily a diagonal matrix. So, we will consider a general positive definite matrix and then look at how to construct such simultaneous confidence interval for mu i components or in general for linear combinations p such linear combinations. Thank you.