Applied Multivariate Analysis

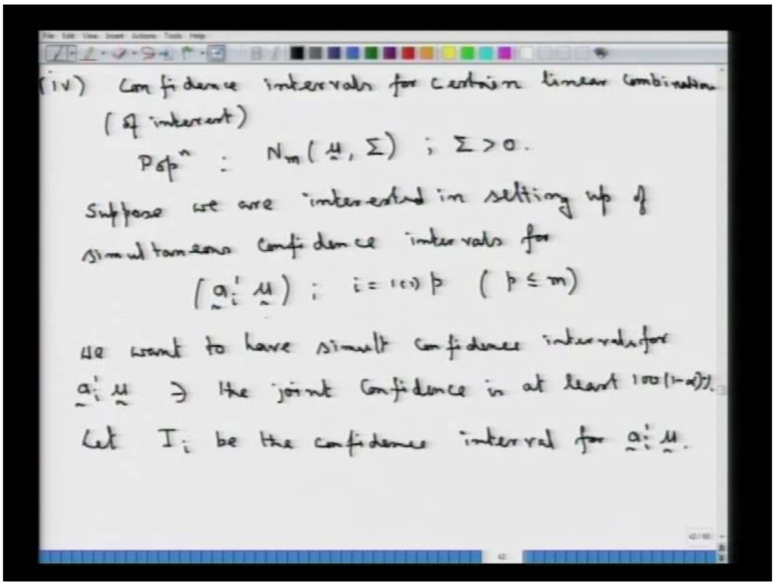
Prof. Amit Mitra
Prof. Shramishtha Mitra
Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur.

Module No. # 01
Lecture No. # 16

Profile Analysis – I

(Refer Slide Time: 00:26)



(iv) Confidence intervals for certain linear combination (of interest)

Pop^n : $N_m(\underset{\sim}{\mu}, \Sigma)$; $\Sigma > 0$.

Suppose we are interested in setting up of simultaneous confidence intervals for

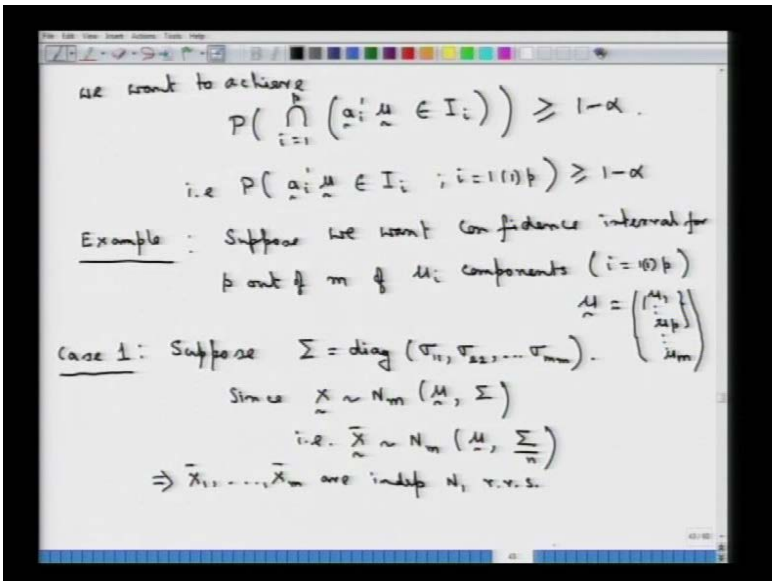
$(\underset{\sim}{a}_i' \underset{\sim}{\mu})$; $i = 1(1)p$ $(p \leq m)$

we want to have simult confidence intervals for $\underset{\sim}{a}_i' \underset{\sim}{\mu}$ $\ni$ the joint confidence in at least $100(1-\alpha)\%$.

Let I_i be the confidence interval for $\underset{\sim}{a}_i' \underset{\sim}{\mu}$.

In the last lecture, what we were discussing was confidence interval, setting up of confidence intervals for certain type of linear combinations of the elements of the mean vector mu, coming from the multivariate normal population. So, we had an underline multivariate normal population; multivariate normal m dimensional with the mean vector mu, and covariance matrix sigma which is which was actually assume to be positive definite. And we had the following discussion that we were looking at these linear combinations a i prime mu quantities for i equal to 1 to p, where p is less than or equal to m. And we were trying to find out simultaneous join confidence interval for all of these linear combinations, such that the coverage is atleast 100 into 1 minus alpha percent.

(Refer Slide Time: 01:02)



we want to achieve

$$P\left(\bigcap_{i=1}^{p}\left(\underset{\sim}{a}_i'\underset{\sim}{\mu} \in I_i\right)\right) \geq 1-\alpha .$$

$$\text{i.e } P\left(\underset{\sim}{a}_i'\underset{\sim}{\mu} \in I_i \ ; i=1(1)p\right) \geq 1-\alpha$$

Example : Suppose we want confidence interval for p out of m of μ_i components $(i=1(1)p)$

$$\underset{\sim}{\mu} = \begin{pmatrix} \left.\begin{matrix}\mu_1 \\ \vdots \\ \mu_p\end{matrix}\right\} \\ \vdots \\ \mu_m \end{pmatrix}$$

Case 1 : Suppose $\Sigma = \text{diag}\,(\sigma_{11}, \sigma_{22}, \ldots \sigma_{mm})$.

Since $\underset{\sim}{X} \sim N_m(\underset{\sim}{\mu}, \Sigma)$

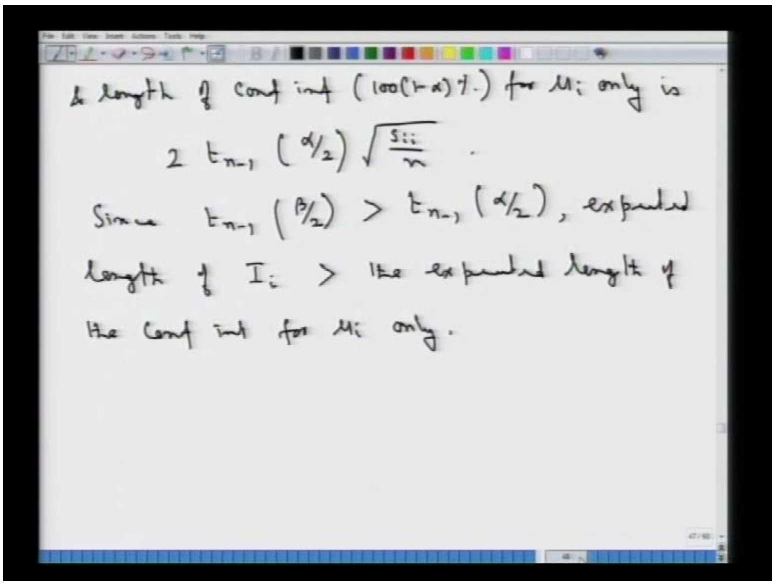
i.e. $\bar{\underset{\sim}{X}} \sim N_m\left(\underset{\sim}{\mu}, \frac{\Sigma}{n}\right)$

$\Rightarrow \bar{X}_1, \ldots, \bar{X}_m$ are indep N_1 r.v.s.

What we had seen in the last lecture was, we had taken a specific example of these linear combinations a i prime mu terms. And we had taken those linear combinations as p components. mu 1 mu 2 mu p without loss of generality, we take the first p of them mu 1, mu 2, mu p and we were looking at two different possibilities; two different cases, under which such simultaneous joint confidence intervals may be derived.

Now, the first case what we had seen in the last lecture was that, we had assumed that sigma is a diagonal matrix. So that, the components of x multivariate random vector becomes independent. And under such a circumstances, we had derived the confidence simultaneous confidence interval of covering mu 1 mu 2 mu p and we had also that discussions about that.

(Refer Slide Time: 01:51)



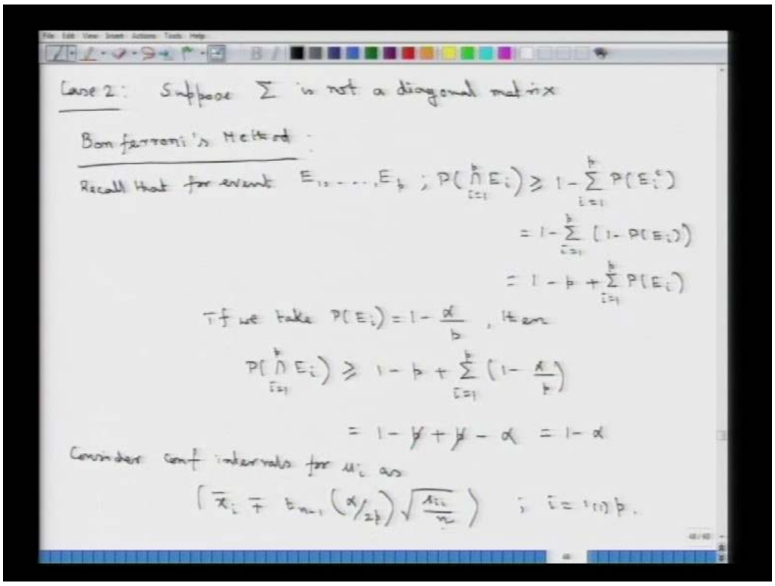
& length of conf int $(100(1-\alpha)\%)$ for μ_i only is

$$2\, t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}}\ .$$

Since $t_{n-1}(\beta/2) > t_{n-1}(\alpha/2)$, expected length of I_i > the expected length of the conf int for μ_i only.

Let us now look at, the more general case, which is the case where, case 2 where we have these… Suppose this sigma is not a diagonal matrix that the most more general case rather is not diagonal matrix.

(Refer Slide Time: 01:57)



Case 2: Suppose Σ is not a diagonal matrix

Bonferroni's Method:

Recall that for event $E_1, \ldots, E_p$; $P(\bigcap_{i=1}^{p} E_i) \geq 1 - \sum_{i=1}^{p} P(E_i^c)$

$$= 1 - \sum_{i=1}^{p}(1 - P(E_i))$$

$$= 1 - p + \sum_{i=1}^{p} P(E_i)$$

If we take $P(E_i) = 1 - \frac{\alpha}{p}$, then

$$P(\bigcap_{i=1}^{p} E_i) \geq 1 - p + \sum_{i=1}^{p}\left(1 - \frac{\alpha}{p}\right)$$

$$= 1 - \cancel{p} + \cancel{p} - \alpha = 1 - \alpha$$

Consider conf intervals for μ_i as

$$\left(\bar{x}_i \mp t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}}\right) \quad ; \quad i = 1(1)p.$$

So, we do not have independence of the components of x is x 1, x 2, x p and hence we cannot also infer about the independence of the components in x bar; that is x 1 bar, x 2 bar, x p bar or x n bar to complete it. They are not necessarily independent; they are not independent actually.

So, sigma is not a diagonal matrix. Under such circumstances, what is used is what is referred to as the Bon Ferroni’s method for construction of this simultaneous confidence interval for these linear parametric functions or the linear transformations a i prime mu terms for i equal to 1 2 up to p.

Now, we may recall that for events E 1, E 2, E p; p events; probability of intersection of E i terms; i equal to 1 to up to p; this is greater than or equal to 1 minus summation i equal to 1 2 up to p probability of E i complements. Now, this is equal to 1 minus summation i equal to 1 2 up to p 1 minus probability of these E i terms. So then, this is equal to 1 minus p plus summation probability E i terms

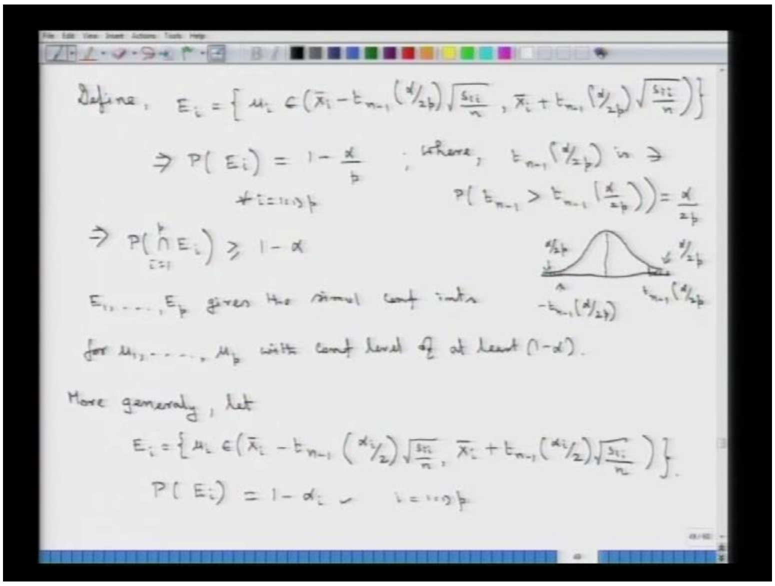
Now, suppose we take, if we take this probability of E i to be equal to 1 minus alpha by p, then what happens in the above inequalities are following? Then probability of intersection of these E i terms; i equal to 1 2 up to p; this is greater than or equal to.. Now, here we have summation i equal to 1 2 up to p. So, what will be having is 1 minus p plus summation I equal to 1 to up to p 1 minus alpha by p. So that, this term is equal to 1 minus p plus p minus alpha. So, this p term cancels out, and what we will have this is equal to 1 minus alpha. So, this basically this basically is derived from bon Ferroni’s inequality and hence the name of construction of the simultaneous confidence intervals, what is given as the bon ferroni’s method.

Now, when we have seen that, for such p events E 1 E 2 e p intersection of this is greater than or equal to 1 minus alpha. If we can choose probability of E i to be equal to 1 minus alpha by p, this leads us to the thought of construction of the simultaneous confidence intervals, such that the coverage would be greater than or equal to 1 minus alpha.

So, if we can choose these E i events as the events that mu i is belonging to a certain random interval, such that the probability of each one of those E i is… If we can make that equal to 1 minus alpha by p, then we will be able to achieve the simultaneous coverage of such linear functions. Here, we have just mu 1, mu 2, mu p in terms of simultaneous confidence intervals with the coverage of at least equal to 1 minus alpha right.

Now, if we take that particular clue, consider confidence interval intervals for each of these mu is as, say x i bar minus plus t n minus 1 alpha by 2 p; this multiplied by s i i by n; this is for i equal to 1 to up to p. So, if we take confidence interval of this particular form x i bar, so the lower confidence limit is x i bar minus t n minus 1 alpha by 2 p into root over s i i by n. And the upper confidence limit is x i bar plus t n minus 1 alpha by 2 p root s i i by n.

(Refer Slide Time: 06:52)



Define, $E_i = \left\{ \mu_i \in \left(\bar{x}_i - t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}} \,,\, \bar{x}_i + t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}} \right) \right\}$

$\Rightarrow P(E_i) = 1 - \frac{\alpha}{p}$; where, $t_{n-1}(\alpha/2p)$ is $\ni$

$\forall i = 1, \ldots, p$ $\qquad P\left(t_{n-1} > t_{n-1}\left(\frac{\alpha}{2p}\right)\right) = \frac{\alpha}{2p}$

$\Rightarrow P\left(\bigcap_{i=1}^{p} E_i\right) \geq 1 - \alpha$

$E_1, \ldots, E_p$ gives the simul conf intn for $\mu_1, \ldots, \mu_p$ with conf level of at least $(1-\alpha)$.

More generally, let

$E_i = \left\{ \mu_i \in \left(\bar{x}_i - t_{n-1}(\alpha_i/2)\sqrt{\frac{s_{ii}}{n}} \,,\, \bar{x}_i + t_{n-1}(\alpha_i/2)\sqrt{\frac{s_{ii}}{n}} \right) \right\}$.

$P(E_i) = 1 - \alpha_i$ $\quad i = 1, \ldots, p$

So, if we take that, then, if we define this E i event to be the event that, mu i is belonging to the random interval, that this is x i bar this is x i bar minus t n minus 1 alpha by 2 p; this into under root of s i i by n to x i bar plus t n minus 1 alpha by 2 p into under root of s i i by n.

So, this is an event that mu i is belonging to this random interval. So, this would imply that, probability of this E i is going to be equal to what? That mu i is belonging to this particular interval; this in terms of the t distribution is what is going to lead us to 1 minus alpha by p. Now here, of course, where we have this t n minus 1 alpha by 2 p is such that probability that t distribution on n minus 1 exceeds this point t n minus 1 alpha by 2 p point.

So, the probability that t distribution exceeds this given point; this is going to be equal to alpha by 2 p. So, we will have alpha by 2 p to the right of this; alpha by 2 p to the left of, negative of this particular term. So, the area on the right and the left of the positive t alpha by 2 p and to the negative of alpha by 2 p would just sum up to alpha by p, and hence the area between them. This is, what I am trying to convey.

So, if this is the t distribution pdf, we have a point here which is our t n minus 1 alpha by 2 p. So, the area to the right of this is alpha by 2 p; it is a symmetric distribution; and hence the area to the left of negative of t n minus 1 alpha by 2 p, this area to the left of this point is also alpha by 2 p. So, the two add up to alpha by p and then the area between these two points is 1 minus alpha by p, and that is what we have area because that is what is going to be given from this expression out here

So, this would imply that, if we have such random events E i which are basically coming from the respective confidence intervals for the mu i parameters; this would imply that, probability of the these intersection of E i s i equal to 1 2 up to p. So, these are basically the events associated; E i is the event associated with the mu i component. And then, this intersection probability is going to be greater than or equal to 1 minus alpha. As what we had derived out here that, if we can choose probability of E i to be equal to 1 minus alpha by p, then we will have probability of intersection of E i to be greater than or equal to 1 minus alpha. And we have chosen, actually those random intervals in such a way for each of these parameters mu i. We have chosen random interval in this particular manner, and so that we will be having probability of each of these E i events to be equal to 1 minus alpha by p, and then from that, we will be having probability of intersection of E i to be greater than or equal to 1 minus alpha.

So, this basically is going to give give this is giving us, the simultaneous confidence interval. So, this E 1, E 2, E p gives the simultaneous confidence interval intervals for this parameters mu 1, mu 2, mu p with confidence level of at least 1 minus alpha. So, we have ensured that, with such random intervals, we are able to have a coverage of at least 1 minus alpha. So, this is what it tends out to be. Now, one can actually have more general approach to constructing such confidence interval.

So, more generally, we can actually look at… Now, note that, when we are looking at these E i terms; E i events here, for the parameter mu i, what we are doing here is that, we are choosing this E i sets in such a way that probability of each of the E is… So, this probability of E i equal to 1 minus alpha by p; this is true for every i equal to 1 2 up to p. So, we have the probability of coverage of each of these mu is to be exactly equal to 1 minus alpha by p for all the is.

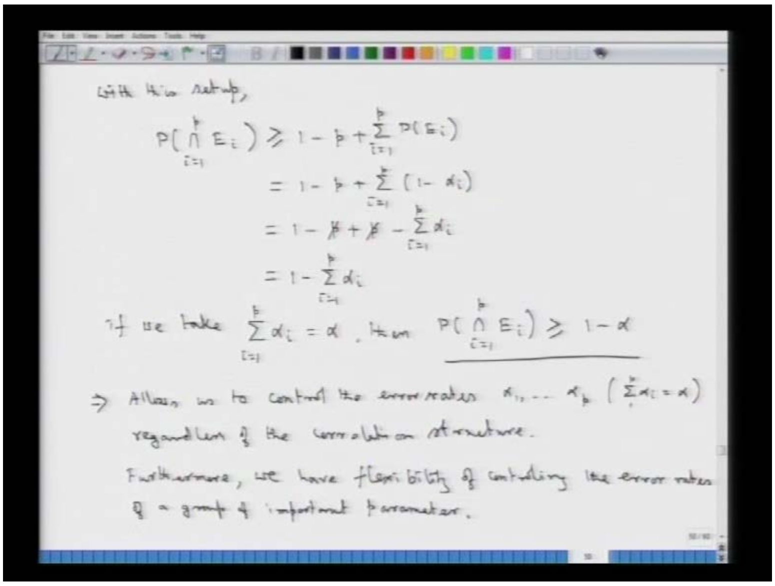
Now, we can actually play around with this probability, because not all mu is may be equally important to us; some of the mu is may be more important to us; some of the mu is may be some of the mu is may be less important to us. And hence we can take care of that event using this particular approach. So, what we can do is that, more generally let E i be… Now, the set that mu i belongs to interval similar to what we have considered, but we will actually not be having alpha by 2 p at all for all the mu is. We can make this as alpha i by 2. So that, it depends on mu i; this level of error depends on the particular mu i, that is chosen here; this into s i i by n; this is same as what we had previously.

This is x i bar plus t n minus 1 alpha i by 2; now, this t n minus 1 alpha i by 2 has similar interpretation as to what we had noted out here that, t n minus 1 alpha by 2 p is the upper alpha by 2 p cut off point of a t distribution on n minus 1 degrees of freedom. So, similarly here, this t n minus 1 alpha by 2 point is the upper alpha i by 2 percent point of a t distribution on n minus 1 degrees of freedom.

So, that this, now turns out to be that this is s i i; this divided by n. So, this is what is the random interval. Now we are considering now what would happen here? What is the probability of such E i terms? Now, alpha 1, alpha 2, alpha p actually varies with respect to mu 1, mu 2, mu p. So, the probability E i, by the similar logic would be 1 minus alpha i right. So, this is probability of E i, because we have the cut off point here alpha i by 2 percent point. So, alpha i by 2 percent point is on the right alpha; i by 2 is on the left of minus t n minus 1 alpha i by 2. And hence the total area to the right of t n minus 1 alpha i by 2, and area to the left of minus t and minus 1 alpha by 2 would be alpha i, and hence the area in between the two cut off points would be 1 minus alpha i. And this is what is thus giving us this particular quantity; this is for i equal to 1 to up to p.

Now, alpha 1, alpha 2, alpha p are different different with the restriction. Now, with this with this set up, with this general set up,

(Refer Slide Time: 15:11)



With this setup,

$$P\left(\bigcap_{i=1}^{p} E_i\right) \geq 1 - p + \sum_{i=1}^{p} P(E_i)$$

$$= 1 - p + \sum_{i=1}^{p} (1 - \alpha_i)$$

$$= 1 - \cancel{p} + \cancel{p} - \sum_{i=1}^{p} \alpha_i$$

$$= 1 - \sum_{i=1}^{p} \alpha_i$$

If we take $\sum_{i=1}^{p} \alpha_i = \alpha$, then $\underline{P\left(\bigcap_{i=1}^{p} E_i\right) \geq 1 - \alpha}$

$\Rightarrow$ Allows us to control the error rates $\alpha_1, \ldots \alpha_p$ $\left(\sum_i \alpha_i = \alpha\right)$ regardless of the correlation structure.

Furthermore, we have flexibility of controlling the error rates of a group of important parameter.

so to say, probability of this intersection of E i events that what is given from the bon ferroni's inequality is what we had derived earlier. That was 1 minus p plus probability of E i. So, what does this now lead us to 1 minus p plus summation probability of E i terms i equal to 1 2 up to p.

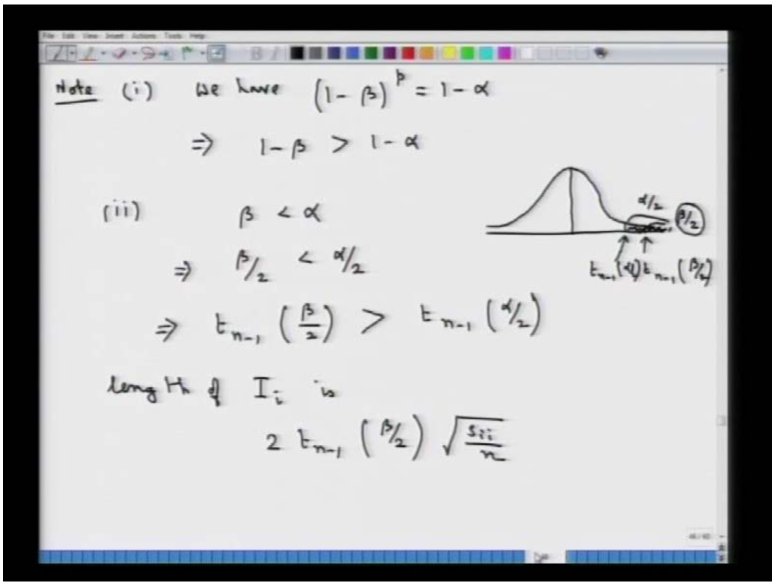
So, with probability E i given by 1 minus alpha, we will have this to be equal to 1 minus p plus summation i equal to 1 2 up to p 1 minus alpha i is what we are getting. So, it is 1 minus p plus p minus summation alpha i; i equal to 1 to up to p. So, this is nothing, but 1 minus summation of these alpha i quantities. So, if we take summation alpha i, i equal to 1 to p equal to alpha, the desired level of error. So, if we take this summation alpha i, then we will have probability of this simultaneous coverage; that is, intersection of these E i events; this is going to be greater than or equal to 1 minus alpha. So, this is the required thing or the desired thing, what we are trying to ensure that, the joint coverage of all these parameters mu 1, mu 2, mu p which are given through these E i events random events. We will have this joint coverage to be greater than or equal to 1 minus alpha, if we restrict our self to summing this these alpha i quantities to be equal to alpha.

So, this actually allows us to control the error rate of the respective mu i components in the way that, when we are looking at a particular mu i, probability that mu i lies in that particular confidence interval is 1 minus alpha i. So, load the value of these alpha i here; higher is the coverage for that particular mu i; higher is the value of alpha i; lower will be this E i.

Now, this alpha i is of course, i have to line between 0 and 1. So, with that restrictionl we are now playing around with alpha 1, alpha 2, alpha p. These are corresponding to the mu 1, mu 2, mu p terms, and hence what we are now looking at is to control the error rate with respect to various mu is. And then, we will have alpha i values lower for the mu is, which appear to us as more important. And hence we can do that quite easily, when we have this particular general set up.

So, this will imply that, this allows us to control the error rates; the error rates alpha 1, alpha 2, alpha p with this summation alpha i equal to 1 to p equal to alpha, which is the desired level of significance; this is regardless of the correlation structure; the underline correlation structure which is given by the sigma matrix. Further more, we have flexibility, actually; we have a flexibility of controlling the error rates which are alpha is error rates of a group of important parameters. right Because we can make alpha i corresponding to a particular mu I, which is of… Say more importance to us, and then we can have that being compensated by some other mu i which is not that important. And hence, we can thus look at having higher coverage for the parameters of more importance, and comparatively lower coverage in the simultaneous confidence interval set up, to have been attached to other type of contrasts.

(Refer Slide Time: 20:08)



Note (i) We have $(1-\beta)^p = 1-\alpha$

$\Rightarrow 1-\beta > 1-\alpha$

(ii) $\beta < \alpha$

$\Rightarrow \beta/2 < \alpha/2$

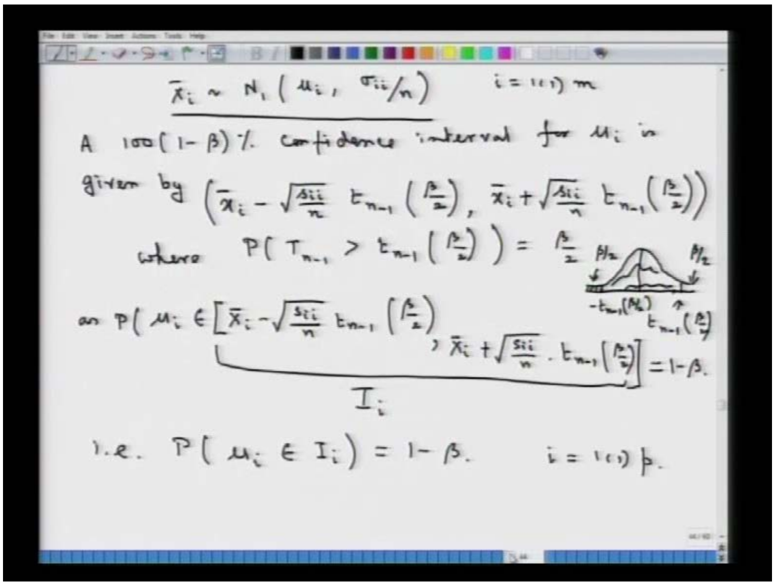
$\Rightarrow t_{n-1}\left(\frac{\beta}{2}\right) > t_{n-1}(\alpha/2)$

length of I_i is

$2\, t_{n-1}(\beta/2)\sqrt{\frac{s_{ii}}{n}}$

Now, this particular example, when we had considered to start with that, we are looking at a particular type of linear combination a i primes, where a i is vector was having one at the i th position, and 0 at all other locations.

(Refer Slide Time: 20:11)



$$\underline{\bar{x}_i \sim N_1\left(\mu_i, \sigma_{ii}/n\right)} \qquad i = 1(1)m$$

A $100(1-\beta)\%$ confidence interval for μ_i is given by $\left(\bar{x}_i - \sqrt{\frac{s_{ii}}{n}}\, t_{n-1}\left(\frac{\beta}{2}\right), \bar{x}_i + \sqrt{\frac{s_{ii}}{n}}\, t_{n-1}\left(\frac{\beta}{2}\right)\right)$

where $P\left(T_{n-1} > t_{n-1}\left(\frac{\beta}{2}\right)\right) = \frac{\beta}{2}$

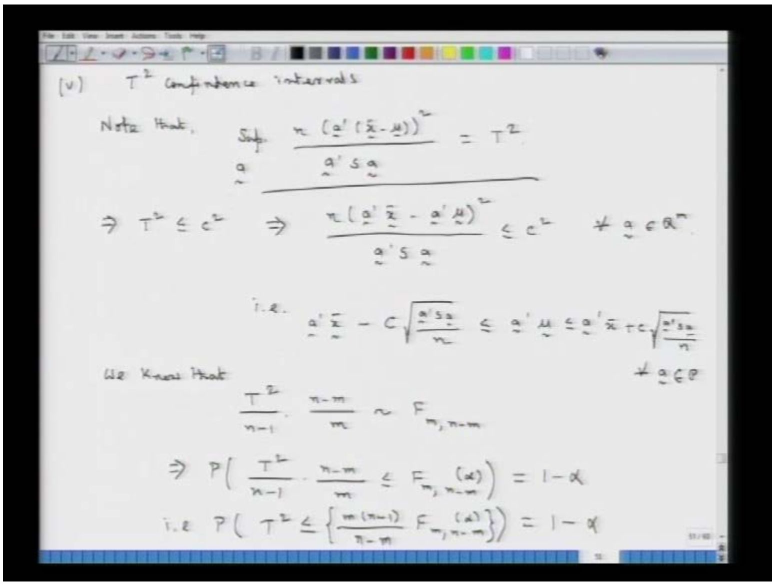
as $$P\Big(\mu_i \in \underbrace{\left[\bar{x}_i - \sqrt{\frac{s_{ii}}{n}}\, t_{n-1}\left(\frac{\beta}{2}\right), \bar{x}_i + \sqrt{\frac{s_{ii}}{n}}\, .\, t_{n-1}\left(\frac{\beta}{2}\right)\right]}_{I_i}\Big) = 1-\beta.$$

i.e. $P(\mu_i \in I_i) = 1-\beta$. $\quad i = 1(1)p$.

So, that had lead us to constructing the simultaneous confidence interval for mu 1, mu 2, mu p. We had consider two different cases; case one and case two; where case one, where sigma was a diagonal matrix. And case two where sigma was not necessarily a diagonal matrix, and have and have obtained actually all those simultaneous confidence intervals.

Now, this approach can be extended for any linear combining vector a i, any type of linear combining vector, and hence this method can easily be extended for such situations. Only thing is that, here the distributions are going to change in place of x i bar; we will be having a i prime x bar terms, and the corresponding variance term will also get replaced in place of s i i. We will have whatever will be the variance of that a i prime x bar terms right.

(Refer Slide Time: 21:25)



(v) T^2 confidence intervals

Note that, $\sup_{\underset{\sim}{a}} \frac{n(\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}))^2}{\underset{\sim}{a}' S \underset{\sim}{a}} = T^2$

$\Rightarrow T^2 \leq c^2 \Rightarrow \frac{n(\underset{\sim}{a}'\bar{\underset{\sim}{x}} - \underset{\sim}{a}'\underset{\sim}{\mu})^2}{\underset{\sim}{a}' S \underset{\sim}{a}} \leq c^2 \quad \forall \underset{\sim}{a} \in \mathbb{R}^m$

i.e. $\underset{\sim}{a}'\bar{\underset{\sim}{x}} - c\sqrt{\frac{\underset{\sim}{a}' S \underset{\sim}{a}}{n}} \leq \underset{\sim}{a}'\underset{\sim}{\mu} \leq \underset{\sim}{a}'\bar{\underset{\sim}{x}} + c\sqrt{\frac{\underset{\sim}{a}' S \underset{\sim}{a}}{n}} \quad \forall \underset{\sim}{a} \in \mathbb{R}$

We know that

$$\frac{T^2}{n-1} \cdot \frac{n-m}{m} \sim F_{m, n-m}$$

$$\Rightarrow P\left(\frac{T^2}{n-1} \cdot \frac{n-m}{m} \leq F_{m, n-m}(\alpha)\right) = 1-\alpha$$

$$\text{i.e } P\left(T^2 \leq \left\{\frac{m(n-1)}{n-m} F_{m, n-m}(\alpha)\right\}\right) = 1-\alpha$$

So, the next thing that we are going to look at is T square confidence interval; it is another important thing about the hotelling T squared statistic. This these are T square confidence intervals. What are these? Now, note that, we have proved a result in the way, that supremum over a of n times a prime x bar minus mu whole square; this divided by a prime S a.

We had in the last lecture obtained, what is the supremum of this particular quantity? And this was shown to be equal to T square right. So, this is what a given result is. So, this would imply that, T square less than or equal to some C square; this would imply. So, if we have T square less than or equal to C square, this would imply that this n times a prime x bar; let us split it; a prime x bar minus a prime mu whole square; this divided by a prime S a. This term is also going to be less than or equal to C square; this is going to be true for every vector a belonging to r to the power m, if x is m dimensional.

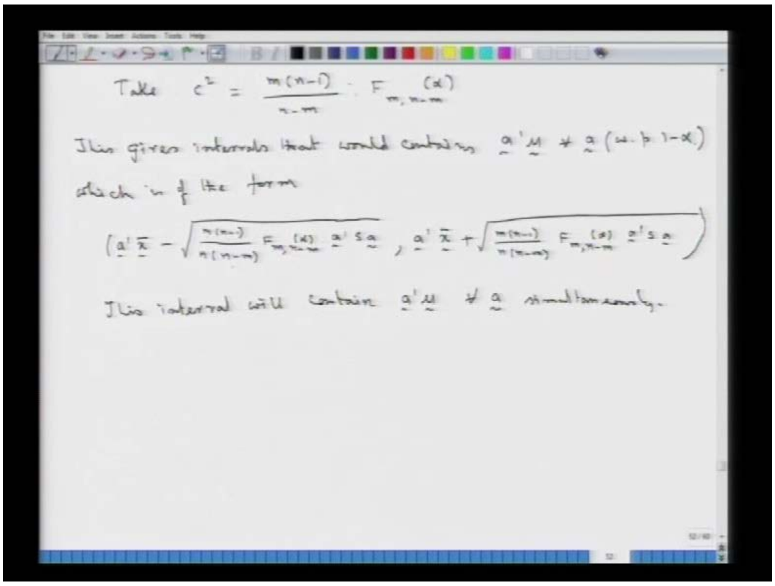
So, from this result from this result here, what we can infer is this particular line. So, in other words, because T square is the supremum of this particular quantity, and T square being less than or equal to C square would imply that, these quantities will be less than or equal to C square for every possible value of a belonging to here, r to the power m. That is, we can write this in the following way that, we will take this a prime x bar outside. This is minus C times under root of a prime S a; this divided by n; this is going to be less than or equal to… We are actually looking at type of confidence interval type scenario, and hence we are expressing this. Because once we have this less than or equal to C square, we will have the absolute value of this being less than or equal to C, and hence we will have that to lie between minus C, and plus C. And then, we will be having the expression that I am writing now

So, this is going to be given by this a prime S a this divided by n. So, the under root of course, is both in the numerator, and in the denominator; this is going to be true for every vector a belonging to r to the power m. Now, we further know that, we know that this T square by n minus 1 into n minus m by m. We have seen this time in again; this follows an F distribution m n minus m degrees of freedom.

So, this would imply that, probability that this T square by n minus 1 into n minus m by m; this less than or equal to f n n minus m alpha. So that, this is the upper alpha percent cut off point of this f distribution. So, this is going to be equal to 1 minus alpha. So, if that is the case will be able to write this T square less than or equal to m times m minus 1; this divided by n minus m; this times F m n minus m times alpha. This is equal to 1 minus alpha.

Now, why have we written this particular fact and reduced it in this form? What we are trying to see is that, what is the probability of such a such an event that T square is less than or equal to C square, that we desire that to b equal to 1 minus alpha. And hence if we choose C square to be equal to this particular point which is m times n minus 1 by n minus m into F m n minus m alpha. If this is chosen as C square, then probability that T square is less than or equal to C square is equal to 1 minus alpha. And hence that would ensure that, the joint coverage of such events for every value of a, that is going to have a coverage equal to 1 minus alpha. Because t square less than or equal to C square implies that, a prime mu for every mu belonging to r to the power m, actually has a coverage equal to the coverage of this particular event which is 1 minus alpha.

(Refer Slide Time: 26:23)



Take $c^2 = \frac{m(n-1)}{n-m} \cdot F_{m,n-m}(\alpha)$

This gives intervals that would contain $\underset{\sim}{a}'\underset{\sim}{\mu} \ \forall \ \underset{\sim}{a}$ (w.p. $1-\alpha$)

which is of the form

$$\left(\underset{\sim}{a}'\bar{\underset{\sim}{x}} - \sqrt{\frac{m(n-1)}{n(n-m)} F_{m,n-m}(\alpha)\, \underset{\sim}{a}' S \underset{\sim}{a}}\ ,\ \underset{\sim}{a}'\bar{\underset{\sim}{x}} + \sqrt{\frac{m(n-1)}{n(n-m)} F_{m,n-m}(\alpha)\, \underset{\sim}{a}' S \underset{\sim}{a}}\right)$$

This interval will contain $\underset{\sim}{a}'\underset{\sim}{\mu} \ \forall \ \underset{\sim}{a}$ simultaneously.

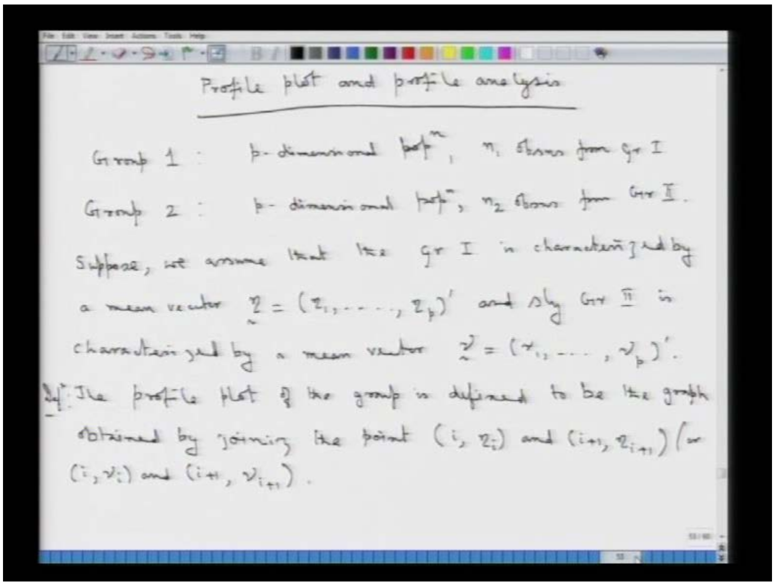
So, we will take this c square to be equal to that constant what we had; m into n minus 1 into n minus m; this into F m n minus m times alpha right. So, this gives interval that would contain that would contain this a prime mu term for every value of a with probability equal to 1 minus alpha with probability 1 minus alpha. This gives intervals, that would contain this which is of the form that is given from this previous discussion, that we are going to have this

Now, with C which is given by the under root of this particular quantity, we will have the coverage equal to 1 minus alpha. So, this interval now is given by this a prime x bar; this minus under root of this entire quantity which is m n minus 1. This divided by n times n minus m into F m n minus m alpha; this into a prime S a; this is the lower confidence limit for that particular a prime a prime mu. And the upper limit is a prime x bar; this plus the same quantity out here which is m into n minus 1, that divided by m into n minus m F m n minus m a prime S a

So, this basically is going to give us that, simultaneous confidence interval. So, this interval this interval will contain a prime mu for every a simultaneously. So, this leads us to another type of confidence, simultaneous confidence intervals which are termed as t term confidence intervals the reason why they are called T square confidence intervals is that this cut off actually F m n minus m alpha this term here along with the constant m n minus 1 and n minus m this is what is a the cut off basically this particular term here that is given from the T square hotelling T square distribution. So, that brings us to the end of this particular section.

What we are now going to see is an important application of hotelling T square statistic which is called the profile analysis now profile analysis is an important applied multivariate tool wherein we look at the following things and let me first introduce what we are now going to see this profile plot and profile analysis.

(Refer Slide Time: 29:45)



Profile plot and profile analysis

Group 1 : p-dimensional pop^n, n_1 obsns from Gr I

Group 2 : p-dimensional pop^n, n_2 obsns from Gr II.

Suppose, we assume that the Gr I is characterized by a mean vector $\underset{\sim}{\eta} = (\eta_1, \dots, \eta_p)'$ and sly Gr II is characterized by a mean vector $\underset{\sim}{\nu} = (\nu_1, \dots, \nu_p)'$.

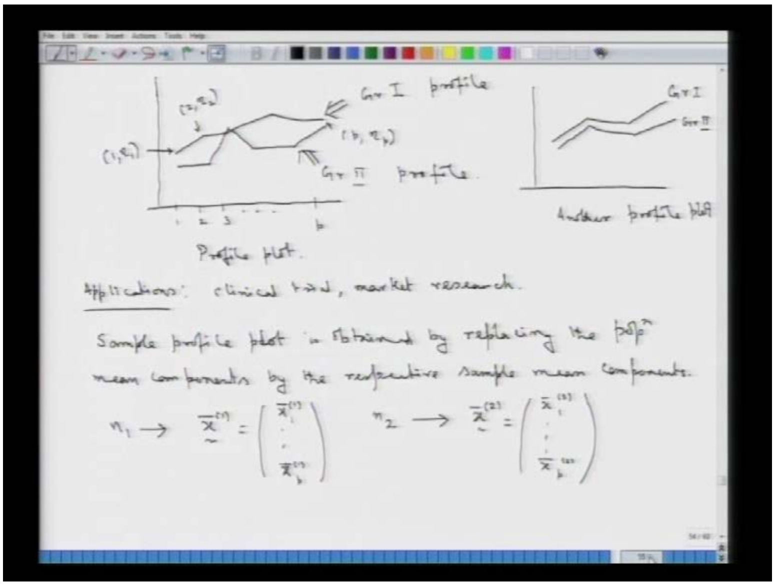
Defn: The profile plot of the group is defined to be the graph obtained by joining the point (i, η_i) and $(i+1, \eta_{i+1})$ (or (i, ν_i) and $(i+1, \nu_{i+1})$.

Now, what is the setup of this particular type of analysis we have proved two groups group 1 this is a p dimensional population p or m whatever p dimensional population and we have a group 2 second group which also is a p dimensional population now suppose we have n 1 observations from this group 1 and similarly we have n 2 observations taken from group 2.

Now, what we are trying to see is that suppose we assume that suppose we assume that the group 1 is characterized group 1 in the population is characterized by is characterized by a mean vector say eta which is equal to eta 1 there are p characteristics in that population is the p dimensional population and hence this is eta of 1 eta 2 eta p and similarly group 2 is characterized by group 2 is characterized by a mean vector which is given by say mu this is equal to mu 1 mu 2 mu p say. So, these are the two mean vectors characterizing these two populations say

Then we define the profile plot of these two groups as the following. So, this is a small definition out here the profile plot of the group is defined to be the graph it is very simple actually let me first write the definition obtained by joining the points i eta i and i plus 1 eta i plus 1 or we can similarly say that it is i nu i and i plus 1 nu i plus 1. So, this is what is called the profile plot what it is actually the definition can be given in this way.

(Refer Slide Time: 33:07)



Applications: clinical trial, market research.

Sample profile plot is obtained by replacing the pop^n mean components by the respective sample mean components.

$n_1 \to \underset{\sim}{\bar{x}}^{(1)} = \begin{pmatrix} \bar{x}_1^{(1)} \\ \vdots \\ \bar{x}_p^{(1)} \end{pmatrix}$ $\quad n_2 \to \underset{\sim}{\bar{x}}^{(2)} = \begin{pmatrix} \bar{x}_1^{(2)} \\ \vdots \\ \bar{x}_p^{(2)} \end{pmatrix}$

The plot actually is nothing, but the following that what we have here these are the p dimensions say 1 2 3 and so, on. So, these are the p dimensions in the two populations now here we just plot what are the means corresponding to these group. So, this point is one say for group one this point is 1 eta 1. So, we join 1 eta 1 with 2 eta 2 suppose that particular mean point is here. So, we go on joining that like this

So, this is just the point obtained. So, this point would be what this point would be suppose this is for group one this is suppose the profile plot for group one then this point here is nothing, but p eta p this point the starting point here is 1 eta 1 and. So, on this point is 2 eta 2 and. So, on. So, we are joining consecutive points i eta i and i plus 1 eta i plus 1 and this basically is the simple graph which is called the profile of the group one

Similarly, we can one can draw profile of the second group say suppose that profile is given by this it can be of any shape right. So, this is similarly the profile for group two.

So, this is group one profile this is the group two profile now another profile plot can be of the following nature that it is like this another profile plot. So, this is also a profile plot

Now, thus in profile analysis we look at what type of analysis what type of questions are interest when we have two or more such profile plots now this is for two groups that I have two profiles if we have k groups then we will have k such profiles drawn on one paper one graph here and hence we will have that figure to give us an idea about the profile of k groups under consideration this is just a simple thing to have two such groups

Now, this profile analysis has got immense applications in the field market research chemical trials applications in the areas of chemical trials by a statistics market research and various other areas now the point to be noted here is that although we are saying that this is the profile plot of group one this is the profile plot of group two similarly this will also be characterized by some characterizing feature group one group two group three and. So, on

Now, such profile plots are actually unknown to us because for all practical purpose this is these are means in the respective populations. So, this eta 1 eta 2 eta p is characterized by the mean vector in the population similarly mu 1 mu 2 mu p is the mean vector which is characterizing that second group of population and hence these are unknown quantities and hence the actual shape of these profile plots are unknown to us since we have the exact structure of this profile plots unknown to us because the respective group means in the population are not known to us that is eta 1 eta 2 eta p or nu 1 nu 2 nu p all of them are unknown quantity and hence this exact shape of the profiles plots are known to us

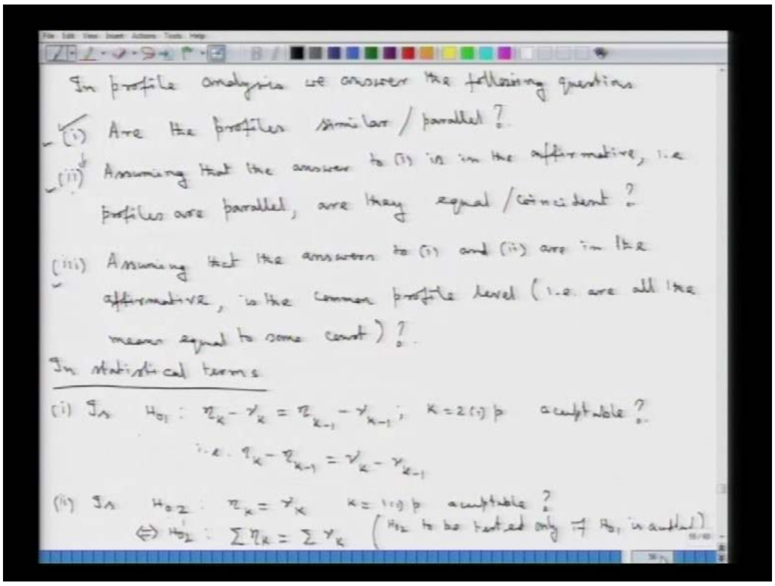
What is done is to look at the sample profile plot because we have n 1 observations taken from the first sample n 2 observations taken from the second population the n 1 observations from the first population n 2 observations from the second population and hence using those n 1 and n 2 populations one can obtain the sample means sample means random vectors for the two groups and then the sample profile plot is obtained

The sample profile plot is obtained by replacing the population mean components by the respective computed sample mean components now this is what we can actually see because from n 1 observations will be able to get to x 1 bar I say this is the mean vector corresponding to the first population. So, this will have the elements as x bar first population first component and similarly this would be x p bar first population p th component

Similarly, from n 2 observations from the second population or the second group we will be having x 2 bar vector this is computed from the n 2 observations coming from the second group and this will have the components as the following that this is x p 2 now using these values now and thus replacing these eta is by the corresponding x 1 I terms and mu is by this x bar 2 I terms will be having the corresponding sample profile plots

Now, given that we have got in the unknown population profiles of this nature profiles of this nature or some other nature there are certain points of interest in profile analysis in which one is interested in they are the following.

(Refer Slide Time: 40:08)



In profile analysis we answer the following questions

(i) Are the profiles similar / parallel ?

(ii) Assuming that the answer to (i) is in the affirmative, i.e profiles are parallel, are they equal / coincident ?

(iii) Assuming that the answers to (i) and (ii) are in the affirmative, is the common profile level (i.e. are all the means equal to some const) ?

In statistical terms

(i) Is $H_{01}: \eta_k - \nu_k = \eta_{k-1} - \nu_{k-1}$; $k = 2(1)p$ acceptable ?

i.e. $\eta_k - \eta_{k-1} = \nu_k - \nu_{k-1}$

(ii) Is $H_{02}: \eta_k = \nu_k$ $k = 1(1)p$ acceptable ?

$\Leftrightarrow H_{02}': \sum \eta_k = \sum \nu_k$ (H_{02} to be tested only if H_{01} is accepted)

In profile analysis we try to address the following points in profile analysis we answer the following questions number one is that are the profiles similar are the profiles of the groups actually are the profiles similar similar in terms of they being parallel this is the first type of point of interest in profile analysis like here these two group profiles does not appear to be similar in the sense that they are not parallel on the other hand if we look at group one and group two profile on this profile plot they appear to be parallel although there is a divergence here, but we need to see whether this divergence from parallelity is statistically significant or not. So, that type of things are what we are going to answer.

Now, the second type of question that is of interest assuming that the answer to one is in the affirmative assuming that the answer to one is in the affirmative that is we answer this particular question that they are similar or parallel we say that well they are if they are then we are further interested to see that whether these two parallel profiles are coincident or equal profiles. So, once they are in the affirmative that is profiles are parallel are they equal that is are the two profiles coincident right

So, once we assume that they are parallel we look forward whether they are actually equal or there is a significant difference between these two similar parallel profiles now the third point of interest once we have excepted this first type of hypothesis that we say that well the profiles are similar that is they are parallel and next also we say that well they appear to be equal coincident and the divergence from the two being equal is not statistically significant.

Then the third type of question that we usually try to answer is that whether the common profile is a level profile level profile in the sense that we have got all the components common in the now the first is eta for the group one the second is new for the second group in the first one we are looking at the two profiles and then seeing whether the two profiles are parallel or not

Once we except the fact that they are parallel then we move on to see whether the two parallel profiles are actually coincident profiles that there is not significant difference between the two profiles of the two groups if we actually except that particular question or rather that particular hypothesis also that the common the two profiles actually are equal that is there is a common profile.

Then we try to look at the third point of interest whether the common profile of the two groups is level profile level in the sense that all the components are same right. So, that the third point of interest let me write it first then I will explain that assuming that the answers the answers two one and two both are in the affirmative r affirmative is the common profile because that is what we have excepted in two is the common profile level that is are all the means equal to some constant. So, that is the third type of question that is usually addressed to in the profile analysis I hope this is this is clear what is a sequence of the testing or rather sequence of answering these questions

First we look at this particular question to be answered now if at this particular stage we say that well the profiles by the statistical testing procedure that we are going to frame the two profiles does not appear to be similar or parallel then we will not proceed further we will not proceed to check whether they are equal or coincident because if the two profiles of the two groups are not even parallel then there is no question of looking at them being equal or coincident.

So, if the first question is answered as yes that profiles appear to be similar that is they are parallel and the divergence from parallelity is not statistically significant then what we do is we proceed to this second question if the answer to the first question is in the affirmative that is they are parallel we check whether they are equal or coincidence profiles now if at this particular point after acceptance of the parallelity of the profiles we come here and say that well the profiles may appear to be parallel, but they are not equal if we answer in that way that they are they do not appear to be equal then we do not proceed to test or to answer this particular last question whether they are having a level profile because if they are not at all equal then there is no question of having a common profile of the two groups and hence we do not move on to the third question here

Now, if the answer to this second question that whether the profiles are equal or coincident once again is answered in the formative that we accept first that the profiles are similar we done except that the profiles are equal is a level profile now this is all in literature terms we need to translate this three type of questions in statistical terms what are these questions in statistical terms

In statistical terms we are actually looking at the following the first question that the profiles are similar or parallel is going to be answered through testing of this particular hypothesis we will frame an hypothesis which will say that this eta k minus mu k is equal to eta k minus one minus mu k minus one this is for k equal to 2 to up to t whether this hypothesis is this hypothesis acceptable.

Now, how is this going to related with the parallelity of the profiles because if you look at the profile plot these basically the different. So, we will have this as the profile of group one this as the profile of group two. So, we take this point as eta 1 this point as mu 1. So, this is the difference eta 1 minus nu 1. So, this is the difference and this is the difference eta 2 minus nu 2.

Now, if all these differences are same then we will have the two profiles naturally to be parallel and hence when we are trying to answer the question that the profiles of the groups are parallel we are feeling or rather translating that particular question in terms of statistical hypothesis testing in terms of this H 01 hypothesis that eat k minus nu k is equal to eta k minus 1 minus nu k minus 1 this is for every k from starting from 2 to up to p

So, all the differences what we have as this eta 2 minus nu 2 is equal to eta 1 minus nu 1 that is equal to eat 3 minus nu 3 and So, on. So, all those goes different in the profile they are all same against that they are on that all of them are not same.

Now, this hypothesis also one sometimes write it writes it as eta k minus eta k minus 1 this difference is equal to nu k minus nu k minus 1 this is in the equivalent form. So, this is the first hypothesis that is going to test the parallelity or similarity of the two profiles.

Now, the second hypothesis is what we are going to frame which is going to look at equality or coincidence of the two profiles given that the first hypothesis is excepted. So, the second type of hypothesis that we are going to frame for testing equality or coincidence nature coincident nature of the profiles is that is H 02 that we will have equality which is eta k equal to nu k this is for k equal to 1 to up to p is this hypothesis acceptable or not.

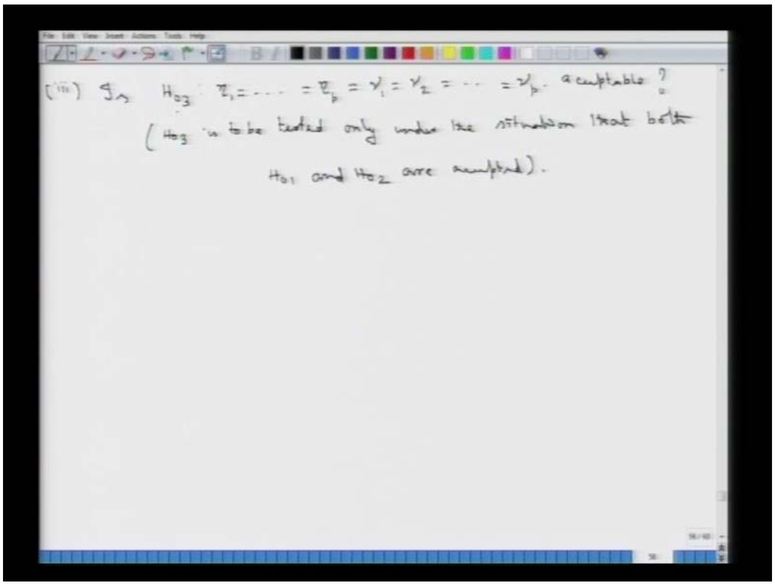
Now, note that this is equivalent to H 02 prime which is going to be summation eta k this is equal to summation of this nu k terms and an important thing to be noted is that here is that we will only test H 02 if H 01 is accepted H 02 to be tested only if H 01 is accepted then only this equality of the profiles make any sense otherwise not. So, if at the first stage H 01 is rejected that we say that the two profiles are not similar or parallel then we would not proceed further to for testing H 02.

Now, under the situation that h naught one is accepted and all these differences that we were talking about which actually leads us to parallelity acceptance of the profiles those differences being equal this hypothesis that eta k equal to nu k would be equivalent to this summation of eta k to be equal to summation of nu k

So, this is basically going to take care of this particular hypothesis this type of questions this statistical hypothesis H 01 is going to take care of this question number 1 what about question number 3 question number 3 is what is looking at that answer to one and two are in affirmative is the common profile level.

So, in terms of this statistical hypothesis how we are going to frame we are going to frame it in the following way that if H 01 and H 02 sequentially in that particular manner are accepted then we move on to test H 03 hypothesis which is going to be the hypothesis which is going to actually answer the last question that is about that level profile thing.

(Refer Slide Time: 52:16)



(iii) Is $H_{03}: \eta_1 = \cdots = \eta_p = \nu_1 = \nu_2 = \cdots = \nu_p$ acceptable?

(H_{03} is to be tested only under the situation that both H_{01} and H_{02} are accepted).

So, we are at the last question here we will have the last question being answered in terms of the following hypothesis H 03 which is going to be that hypothesis which is going to test that eta 1 equal to eta 2 equal to eta p this is equal to nu 1 is equal to nu 2 is equal to nu p right

So, this is whether this hypothesis is acceptable or not. So, these are sequentially the three type of hypothesis whether this is acceptable that is first we answer next we answer whether this H 02 is acceptable and this is to be H 02 is to be tested only if H 01 is accepted.

Now, similarly when we talk about H 03 it is looking at whether all the components eta 1 equal to eta p these are equal now this H 03 to be tested H 03 is to be tested only under the situation under the situation that both H 01 and H 02 are accepted.

So, only under such a situation we are going to test this H 03 and that H 03 under the situation under the condition that both H 01 and H 02 remember H 01 is the hypothesis for testing similarity of the profiles H 02 is the hypothesis for testing equality or confident nature of the profiles and H 03 is the last hypothesis which is going to test whether the profile is a level profile that all the components are basically same and the sequencing of the testing is that first H 01 has to be tested if H 01 is rejected we stop at that particular point if H 01 is accepted then from after testing H 01 we move on to testing H 02.

If H 02 is accepted we move further to test H 03 if H 02 is rejected we stop at that particular point and we say that well the profiles were parallel, but they are not equal if all the three hypothesis H 01 H 02 H 03 in that order are excepted we will say that the profiles first are parallel the profiles next are equal and the profiles the common profile of the two group is level profile

In the next next lecture we are going to see how these H 01 H 02 2 and H 03 are tested using a hotelling T squared statistic we are also going to see some numerical examples of looking at how this profile analysis is going to be carried out

Thank you