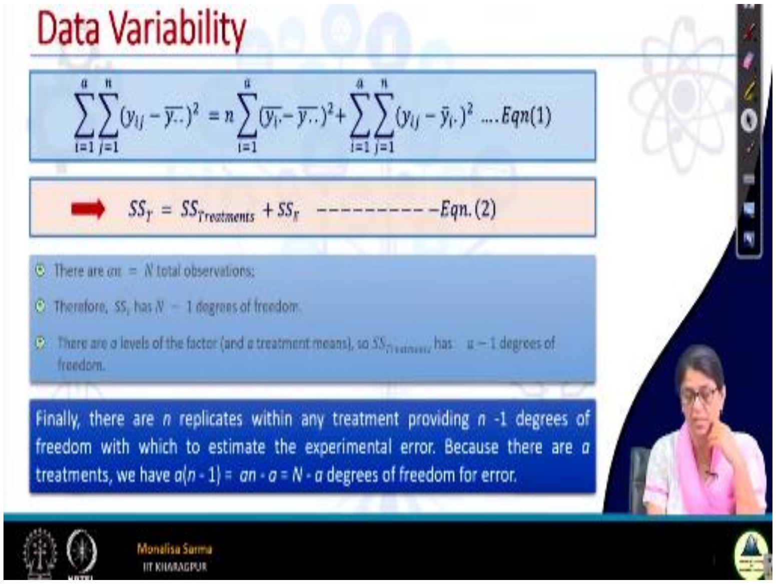
Statistical Learning for Reliability Analysis

Prof. Monalisa Sarma
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology - Kharagpur

Lecture – 35
ANOVA - IV

Welcome guys. So, we were in discussion on ANOVA we were discussing one way ANOVA basically.

(Refer Slide Time: 00:37)



Data Variability

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{..})^2 = n\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 \text{ ....Eqn(1)}$$

$$SS_T = SS_{Treatments} + SS_E \text{ ----------Eqn.(2)}$$

- There are $an = N$ total observations;
- Therefore, SS_T has $N - 1$ degrees of freedom.
- There are a levels of the factor (and a treatment means), so $SS_{Treatments}$ has $a - 1$ degrees of freedom.

Finally, there are n replicates within any treatment providing n -1 degrees of freedom with which to estimate the experimental error. Because there are a treatments, we have $a(n - 1) = an - a = N - a$ degrees of freedom for error.

And in our last class we have stopped here, we have stopped at this point. So, our SS T that is the overall sum of squares, the total sum of squares we have seen it can be partitioned into 2 parts, these are the 2 parts one is SS treatment and another is SS E. Whereas SS in treatment is the sum of treatments sum of squares basically, that means here we are trying to find out the difference of treatment average from the grand average.

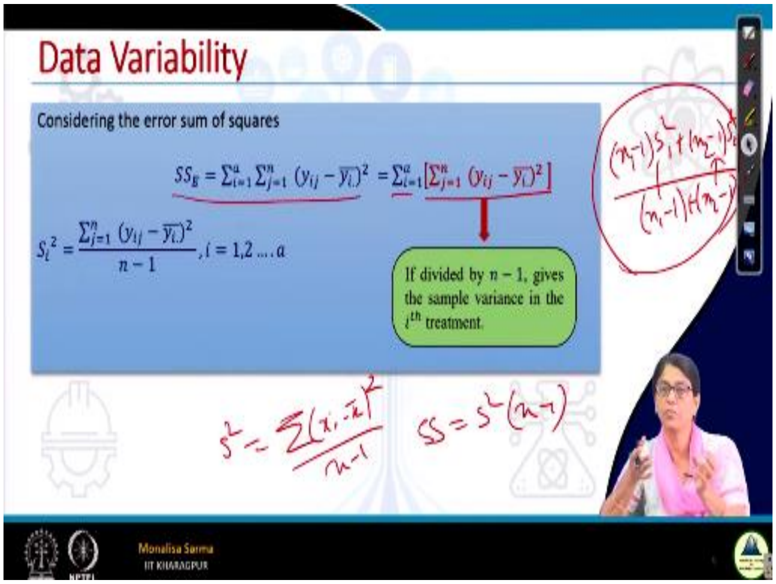
And SS E is the due to random errors where we are trying to find out now difference of all the data from the grand mean sorry, within a sample difference of data from the treatment mean. So, that is basically that is the variance within the treatment. So, that is SS E that is due to the random error. So, now, we will take pick up each and every both that terms together and the first of all, so, total how many, absorptions are there? Total there are N observations why?

Because there are total a level and each level there is small and data, so total is a into n = N. So, now, what is the degrees of freedom for SS T degrees of freedom will be because they are we are considering the overall the total so, degree of freedom will be N - 1. So, now tell me what will be a degree of return for SS treatments. So, by dividing by the degree of freedom basically sum of squares divided by the degree of freedom will get a variance.

So, what is the treatment variance to find out the treatment variance the degree of freedom for SS treatment is basically a - 1. So, now, what is the degree of freedom for SS E that is of course, when we know this term and you know this term definitely we can find out this term that is definitely one way next other way also we can also find out see, there are total n replicates within a treatment and each provides a n - 1 degrees of treatment.

In each treatment, we have degrees of freedom is n - 1 and total there a such populations. So, total degree of freedom will be a into n - 1 that is N - a is the degrees of freedom for SS E.

(Refer Slide Time: 02:44)



Data Variability

Considering the error sum of squares

$$\underline{SS_E = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \overline{y_{i.}})^2} = \underline{\sum_{i=1}^{a}}\left[\underline{\sum_{j=1}^{n} (y_{ij} - \overline{y_{i.}})^2}\right]$$

$$S_i^2 = \frac{\sum_{j=1}^{n} (y_{ij} - \overline{y_{i.}})^2}{n-1}, i = 1, 2 \ldots . a$$

If divided by $n - 1$, gives the sample variance in the i^{th} treatment.

$$\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \qquad SS = s^2(n-1)$$

Monalisa Sarma
IIT KHARAGPUR

So, now, first thing, so, here there are 2 terms SS treatment + SS E So, we will take one, one term. So, first let us take SS E. So, this was my SS E terms is not it? Now, here see in this term, you just see this term which are market with red if this term if I divide by simple n - 1 what do I get this term y ij - y i dot bar that means, within a level, I am trying to find out the difference between each data from the mean is not it? So, I will get the variance of that treatment variance of the ith treatment.

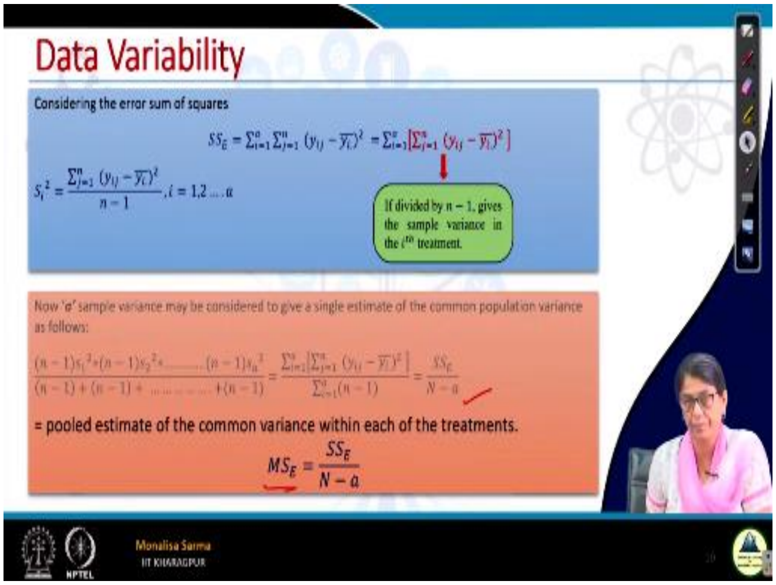
So, now, this is again ∑ i = 1 to a. So, if I tried to sum of all the ath level basically what I will get that is nothing but the same thing what we have got a pooled variance estimate for the two population remember our two population pooled variance estimators what was that if you remember n - 1 s 1 ^{2} + n - n 1 - 1 n 2 - 1 s 2 ^{2} / n 1 - 1 + n 2 - 1 this is the degrees of freedom for first one this is the degrees of freedom for the second one, remember this was our pooled variance estimate.

And if you forgot how do you how did you get this simple variance formula S ^{2} is equals to this (x i - x bar)^2 / n - 1. So, if this is sum of squares, this is nothing but some of square that is sum of square is nothing but s ^{2} into n - 1 is not it? So, this is the pool variance estimate. So, similarly, this term if we divided this one time if you divided by just one and minus one I get the variance of the ith level.

Similarly, if this of all these levels if I divide by the n - 1 correspondingly then basically it is nothing but a weighted average of weighted sum of all the variances of all the levels. So, that is nothing again the pooled variance estimate because here see, we have considered the initially only we have assumed that the population variances are same. So, population variances are same. So, we need to find out what is the population variance. So, how did you do that?

We found out the sample variance from the sample variance, we come out with the pooled variance estimate of the population.

(Refer Slide Time: 05:20)



Data Variability

Considering the error sum of squares

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2 = \sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2\right]$$

$$s_i^2 = \frac{\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2}{n-1}, i = 1,2 \dots a$$

If divided by $n - 1$, gives the sample variance in the i^{th} treatment.

Now 'a' sample variance may be considered to give a single estimate of the common population variance as follows:

$$\frac{(n-1)s_1^2 + (n-1)s_2^2 + \dots\dots (n-1)s_a^2}{(n-1) + (n-1) + \dots\dots\dots + (n-1)} = \frac{\sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2\right]}{\sum_{i=1}^{a}(n-1)} = \frac{SS_E}{N-a}$$

= pooled estimate of the common variance within each of the treatments.

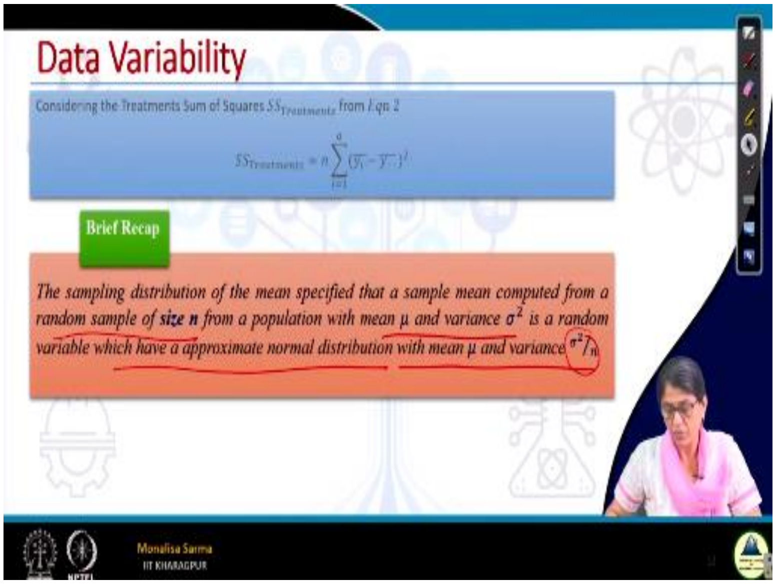
$$MS_E = \frac{SS_E}{N-a}$$

So, basically, if we do this, this particular expression from this particular expression, if we extend it, so, and then just I am dividing by this degrees of freedom of each level. So, this is what degrees of freedom of each level if I get this, this is nothing but this equal to N - a and what is this term this term is nothing but my this term. And what is this term? This term is SS E, to SS E / N - a. So, SS E / N – a, what did they get this? This is nothing but my pool variance estimate.

This is the pool estimated a common variants within each of the treatment. See here, I am going slowly SS T is equals to I got SS treatment + SS E. Now from SS E, I could see that SS E is nothing but if I just divided by the degrees of freedom, I will get the variance of all the pooled variance estimate of the population. So that is what I have done. So, that means my distance add just added one extra one that is divided by N - a.

And that becomes my pool estimate of the common variants within each of the treatment and this term SS E / N - a, I am using a different expression now, for that let me use the expression MS E. Our latest SS E now SS E divided by N - a let me use this term MS E, MS E is nothing but a pooled estimate of the common variance. So, each population has a variance of this SS E / N – a, this we got it from SS E

(Refer Slide Time: 07:00)



Data Variability

Considering the Treatments Sum of Squares $SS_{Treatments}$ from Eqn 2

$$SS_{Treatments} = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

Brief Recap

The sampling distribution of the mean specified that a sample mean computed from a random sample of ***size n*** *from a population with mean μ and variance σ^2 is a random variable which have a approximate normal distribution with mean μ and variance σ^2/n*

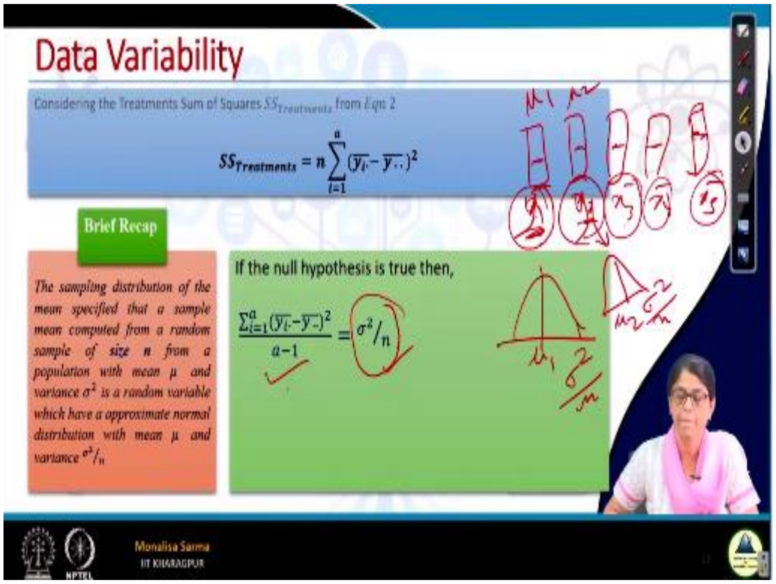
Now, we will come to SS treatment. So, this was my SS treatment formula what is this? the sum of squares of the treatment average from the granaries? Now, you if you recall the sampling distribution of the mean which we have learned when we have learned sampling distribution similar to central limit theorem only central limit theorem just talks about the normality. So, what is the sampling distribution of the mean? Sampling distribution means specified that the sample mean computed from a random sample of size n.

From a population with mean μ and variance σ^2 is a random variable whichever approximate normal distribution with mean μ and variance σ^2 / n what does this say? So, if we take a sample from a normal populations with mean μ and variance σ^2, so, the mean of the sample we have taken a sample there is a population, population has mean μ and variance σ^2 from the population we have taken a sample.

So, now, from the sample we have found out what is the mean of the sample? Now, this mean of the sample let me call it as x bar. Now, this x bar it will have a sampling distribution because x bar different sample we will get different x bar. So, it will have a sampling distribution now, sampling distribution which is approximately normal. So, normal is 2 parameters that is mean and the variance what is the mean and the variance.

So, mean of the sampling distribution is same as population mean that is μ and its variance is population variance divided by n. So, variance of this sampling distribution is population variance divided by n that is σ ^{2} / n. So, this we know from the sampling distribution of the mean now, we will be utilizing this.

(Refer Slide Time: 08:53)



Data Variability

Considering the Treatments Sum of Squares $SS_{Treatments}$ from Eqn 2

$$SS_{Treatments} = n\sum_{i=1}^{a}(\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}})^2$$

Brief Recap

The sampling distribution of the mean specified that a sample mean computed from a random sample of size n from a population with mean μ and variance σ^2 is a random variable which have a approximate normal distribution with mean μ and variance $\sigma^2/_n$

If the null hypothesis is true then,

$$\frac{\sum_{i=1}^{a}(\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}})^2}{a-1} = \sigma^2/_n$$

Now see here so, now try to think and the analogy now what we have we have total a levels a populations from this a population we have taken some sample total a samples we have taken okay each of size n now, this sample has mean μ i each of these sample has mean of μ i and each of the sample has what is the what does this sample has variance each of the sample will have variance σ ^{2} / n why? See, suppose there are 5 population we have taken 5 samples this is 1 sample 2 sample 3 samples 4 sample 5 sample we have taken 5 sample.

So, from each one suppose this is x 1 bar, this is x 2 bar, x 3 bar, x 4 bar and this is x 5 bar and this x 1 bar will have a sampling distribution which is normal and what will be the mean of this distribution mean will be population mean? Suppose the mean of this population that is the first population that mean of this population is μ 1. So, it is mean is μ 1 and what will be the variance of this population? Variance of this population is population variance divided by n and what is the population variance?

Population variance is σ^2, so it is variance will be σ^2 / n where it is mean is μ 1. Similarly, for this x bar if we have a sampling distribution, what we have a sampling distribution what will be the mean of the sampling distribution population mean suppose this population mean is μ 2 and what is the sampling variance of this distribution σ^2, and here also it is σ^2 n why?

Similarly, if you take x 3 x 3 mean maybe μ 3 variance will be again σ^2 / n, x 4 also μ 4 σ^2 / n, μ 5 σ^2 / n why σ^2 / n for all the case because we have assumed all the populations we have assumed that they have the same variance that is σ^2 all the population initially only we have assumed they have the same variance that is σ^2, is not it? And we have also computed what is σ^2 from the ANOVA entity we have computed what is that.

That pool variance estimated we have computed is not it? So, pool variance estimate we have computed that is I think in the last slide we only computed, this is the pool variance estimate this one SS E / N - a what is this? This is nothing but the σ^2 that is, so, all the population has this variance. So, since, all the population has this variance, so, some variants of all the sampling distribution also will be σ^2 / n. Now, as you mean that null hypothesis is true.

Now, let us assume that a null hypothesis is true if the null hypothesis is true depends what happens all the μ are equal μ 1 = μ 2 = μ 3 is equal to say is μ if we assume that all the null hypothesis are true that means, all that means, all the population that means, all the population has same it means as if all the population that is as if we have taken a sample from a single population only that is all the population has same it is just an extended population there is no difference between the population.

If we assume that null hypothesis is true all the samples as if we have taken it from a same extended populations. So, assuming that all the population will mean will be same μ, μ 1 = μ 2 is equals to 3 to 4 to 5 say this μ. So, if in that case now, suppose the different samples that we have taken the means, now, suppose this is if considering that the null hypothesis is true that means, that means, we as if it is a single population from all the population.

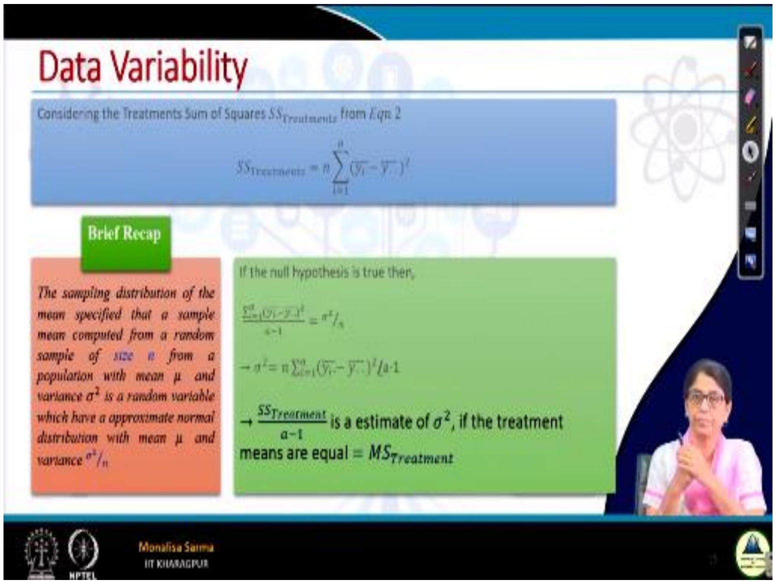
Suppose, we have taken this 5 samples and this suppose this is one is x 1 bar one value I got x 1 bar another value I got x 3 bar x 4 bar and x 5 bar. Means if from one population, we have taken some samples, 5 samples, 5 samples, I have got this data now, if I want to find out what is the variance among these samples, variance among these samples is nothing but this is the case I want to find out the variance among the samples.

So, how will I find out the variance among the sample? Instead of by x here I have written y so, so, consider it is y same thing y or x whatever it is, so, variance among the sample is y bar minus that is the grand mean is not it? How do I find out the variance? Variance is I try to find out the each and every data from the mean is not it? So, when I am taking this so, that means my mean is the grand mean so, divided by how many data total n level so, it is a - 1. So, this is nothing but this is equals to σ^2 / n.

Because all this population we have seen all the samples it has a variance of σ^2 / n is not it? So, we have taken only these samples. So, if this is the case, that means the variance if we now consider this particular nth samples and if we try to find out a variance among these a samples, so, this variance among the; a samples I got this and this is nothing but σ^2 / n is not it? Because I know the variance among this data is σ^2 / n the variance among this data is σ^2 / n.

Variance among these data is σ^2 / n, variance among these data is σ^2 / n, variance among these data σ^2 / n. if that is the case, then variance among this will also I will get σ^2 / n. So, this value is nothing but equals to σ^2 / n then by simplifying it a bit.

(Refer Slide Time: 14:57)



Data Variability

Considering the Treatments Sum of Squares $SS_{\text{Treatments}}$ from $Eqn\ 2$

$$SS_{\text{Treatments}} = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

Brief Recap

The sampling distribution of the mean specified that a sample mean computed from a random sample of size n from a population with mean μ and variance σ^2 is a random variable which have a approximate normal distribution with mean μ and variance $\sigma^2/_n$

If the null hypothesis is true then,

$$\frac{\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{a-1} = \sigma^2/_n$$

$$\rightarrow \sigma^2 = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2/a\text{-}1$$

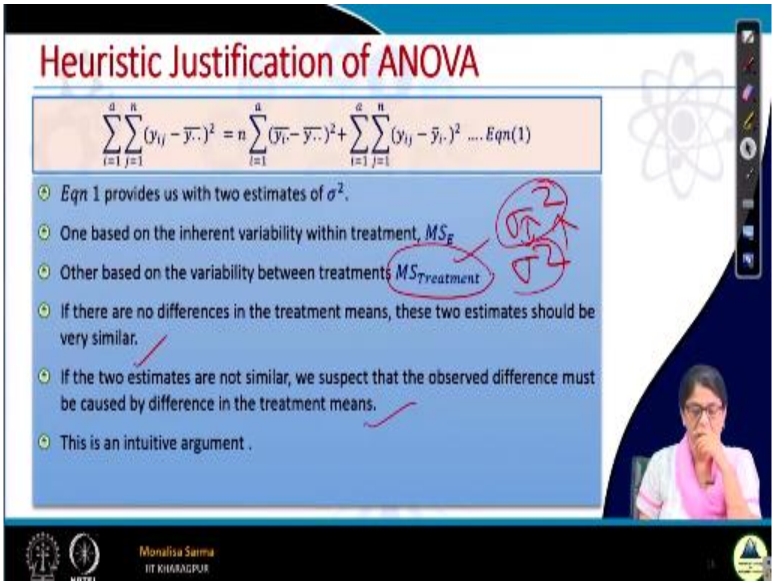
$\rightarrow \frac{SS_{Treatment}}{a-1}$ is a estimate of σ^2, if the treatment means are equal = $MS_{Treatment}$

So, what did I get σ ^{2} you see, so, my σ ^{2} is nothing but this divided by a - 1 and I have taken the side so, my σ ^{2} is this by a - 1 and how this value you see this portion this portion I this is my this portion. So, now what is the σ ^{2} that means SS treatment divided by a - 1. So, SS treatment is the estimate of σ ^{2} if the treatment means are equal so, if the treatment means are equal I have considered null hypothesis if all the treatments means are equal.

Then I have got this if all the treatment means are equal that means, there is no change there is no effect of the factor if all the treatment means are equal then my estimate of σ ^{2} is this will SS treatment / a - 1. So, see I got 2 estimate of σ ^{2} one estimate of σ ^{2} I got it from the error SS E. So, that I found it for the random error what maybe the variance that is from that also I got one estimate of σ ^{2} is not it?

That I called the pool variance estimate pooled variance estimate of all the population that is when estimate of σ ^{2} that is one estimate of σ ^{2} I got another estimate of σ ^{2} I got assuming that all the population means are same assume that all the population means are same I got another one estimate σ ^{2}, another one estimate of σ ^{2}.

(Refer Slide Time: 16:36)



Heuristic Justification of ANOVA

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{..})^2 = n\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 \text{ ....} Eqn(1)$$

- Eqn 1 provides us with two estimates of σ^2.
- One based on the inherent variability within treatment, MS_E
- Other based on the variability between treatments $MS_{Treatment}$
- If there are no differences in the treatment means, these two estimates should be very similar.
- If the two estimates are not similar, we suspect that the observed difference must be caused by difference in the treatment means.
- This is an intuitive argument .

σ^2

σ^2

Monalisa Sarma
IIT KHARAGPUR

So, equation one provides us with 2 estimates of σ ^{2} one based on the inherent variability within the treatment that is the MS E what we saw SS E divided by this one. So, SS E / n - a. So, that is one variance estimate we got another variance estimate I got is SS treatment / a - 1. So, one based on the inherent variability within the treatment that is MS E other based on the variability between treatment other another estimate of σ ^{2} I got it by finding out the variability within treatment considering that all the means are equal.

Here I have considered all the means are equal by considering all the means are equal again I got an estimate of σ ^{2} that is MS Treatment in the earlier post estimated treatment here, I was not bothered whether the means are equal not equal, I was not bothered I was just trying to find out an estimate of the variance by taking all data and the second case I tried to find out the estimate of the variance by considering that all the means are equal.

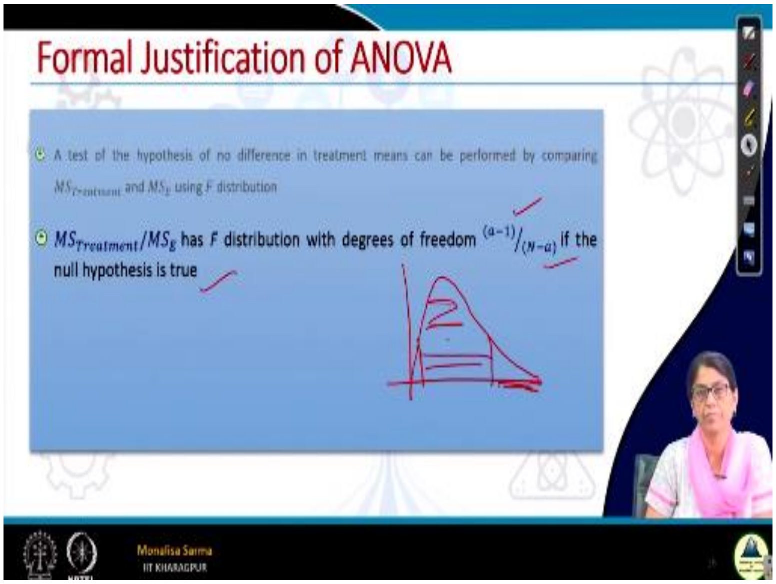
So, if there are no difference in the treatment mean these 2 estimates will be very similar. So, my second estimate of treatment σ ^{2} I have as you that is also an estimate of σ ^{2} this is also an estimate this is also a σ ^{2} this is also σ ^{2} this also gives σ ^{2} and MS Treatment also gives σ ^{2} MS E also gives σ ^{2}, what is σ ^{2} variants of the populations. Now, this σ ^{2} and the σ ^{2} will be equal when because from this I got an estimate of σ ^{2}.

How did they get an estimate of σ ^{2}? Assuming that all the population means the same. So, it may all the population means are same that means, my both the variance estimates are same if all the population means are not same, this MS treatment whatever value I will get this will not be same as MS E, so, if the 2 estimates are not similar, we suspect that the observed difference must be caused by the difference in the treatment mean, so, as you mean that both that all the means are same I got an estimate of σ ^{2}.

Now, I have assuming that all the means are same and I found it an estimate of σ ^{2} if all the means are not same that the estimate of σ ^{2} which I find that will definitely not be equal to σ ^{2} that is the pool variance estimate. So, if the 2 estimates are not similar, we suspect that the observed this difference must be caused by the difference due to the treatment means, one difference is already there due to random error.

So, if the 2 estimates are not similar that means the difference we are getting the difference because of the effect of the treatment whatever treatment we had for the music because of the effect of the music we are getting the difference in values because of the effect of the different treatment we are getting in different values, so, this is an intuitive argument we can also see those in formal approach.

(Refer Slide Time: 20:03)



Formal Justification of ANOVA

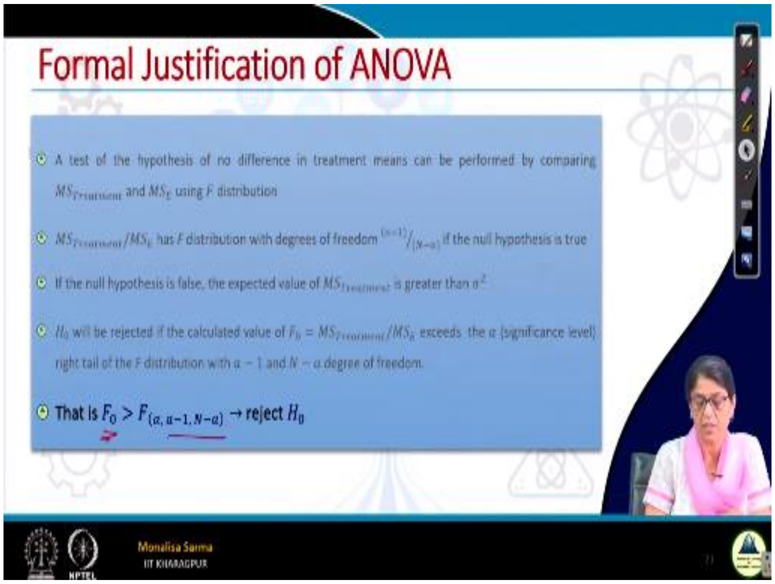
- A test of the hypothesis of no difference in treatment means can be performed by comparing $MS_{Treatment}$ and MS_E using F distribution
- $MS_{Treatment}/MS_E$ has F distribution with degrees of freedom $^{(a-1)}/_{(N-a)}$ if the null hypothesis is true

Now so, the test of hypothesis of no difference in treatment means it can be performed by comparing MS treatment and MS E. MS treatment gives us an estimate of σ ^{2} MS E gives an estimate of σ ^{2} to estimate of σ ^{2}, how do you compare 2 variance for comparing 2 variants we have seen we use F distribution. So, if Ms treatment / MS E has an F distribution, what will be the degree of freedom MS treatment degree of freedom is a - 1 total number of levels - 1 MS E degree of freedom is n – a.

We have seen why it is n - a is not it? So, if the null hypothesis is true, then this MS treatment / MS E will have F distribution. F distribution means remember the F distribution is something of this sort. So, f distribution means it will be in the acceptable range and MS treatment / MS E it will be the acceptable range. MSE is something I am trying to find out a variance of all the data. So, MS E will not change. Now, what happens if the means are not equal. So, this MS treatment that this variance I will get a very bigger value.

When they made numerator value is bigger. So, my f value f is nothing but MS treatment / MS E, so, in the numerator value is bigger I will get a very high f value that means I am well will lie in the critical region in the rejection region.

(Refer Slide Time: 21:41)



Formal Justification of ANOVA

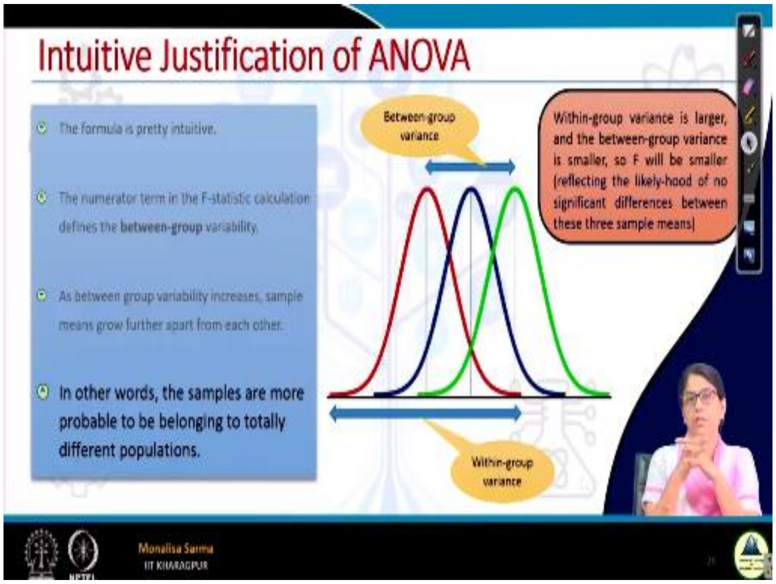
- A test of the hypothesis of no difference in treatment means can be performed by comparing $MS_{Treatment}$ and MS_E using F distribution
- $MS_{Treatment}/MS_E$ has F distribution with degrees of freedom $^{(a-1)}/_{(N-a)}$ if the null hypothesis is true
- If the null hypothesis is false, the expected value of $MS_{Treatment}$ is greater than σ^2
- H_0 will be rejected if the calculated value of $F_0 = MS_{Treatment}/MS_E$ exceeds the α (significance level) right tail of the F distribution with $a-1$ and $N-a$ degree of freedom.
- That is $F_0 > F_{(\alpha,\, a-1,\, N-a)} \rightarrow$ reject H_0

So, here is the null hypothesis is told false the expected value of MS treatment is greater than the σ ^{2} we will get a high value first. So, H 0 will be rejected if the calculated value of F 0 exceeds the α significance level that is the right tail of the distribution here we will be using only the right t because our F value will be because it is our F is what am sorry MS treatment / MS E F will never be smaller than MS E it will never returned it is not possible to be smaller.

Why because already whatever variance is there it is there how can we get less than that. So, only a MS treatment will be more either variance within the treatment if it is more. So, Ms value will be higher value. So, MS value is a higher value that means with my acceptance rate my rejection region will only be in the lower tail that means in this region, so that is why it is always one way it is and that is we are trying to test this if f is greater than this value, this is all on this one tailed test and it is greater than a particular value.

We will not check for lesser than a particular value. So, the expected value of MS statement is greater than σ^2. So, if F 0 is greater than this if the value of F 0 what we calculate if it is greater than f (α) for this degree of freedom a - 1 and N - a, then we reject H 0 mind it, it is a one tailed test with the significance level α.

(Refer Slide Time: 23:31)



Intuitive Justification of ANOVA

- The formula is pretty intuitive.
- The numerator term in the F-statistic calculation defines the **between-group** variability.
- As between group variability increases, sample means grow further apart from each other.
- In other words, the samples are more probable to be belonging to totally different populations.

Within-group variance is larger, and the between-group variance is smaller, so F will be smaller (reflecting the likely-hood of no significant differences between these three sample means)

Monalisa Sarma
IIT KHARAGPUR

This formula as we as I told you, it is pretty intuitive. The numerator term that is the in the F calculations, the statistic defines the numerator terms it defines the between group variability okay. So, this is the between group variability this is this between group variable numerator turns into dedicates its decisive between group variability and this term this is denominator trances within group variability, variability within the group.

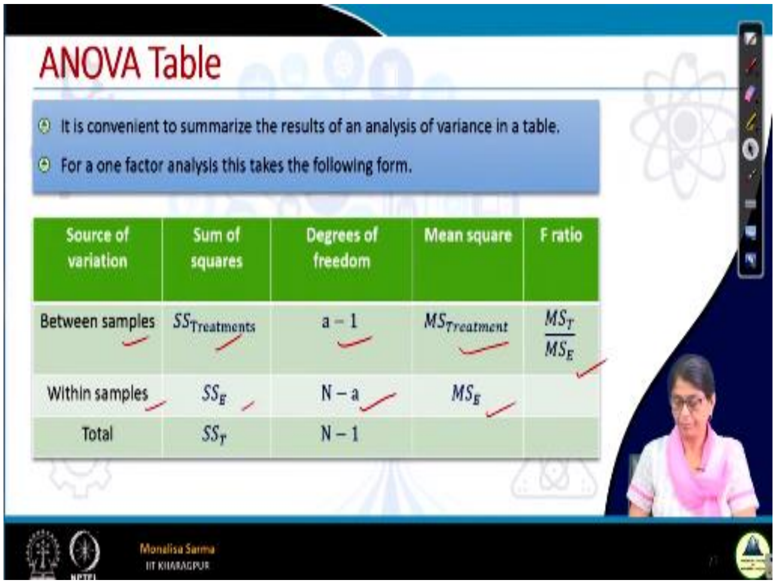
So, this if this variability between group variability is bigger that definitely between good variability is bigger that means what that means, it does not belong to same population that means it belongs to different population that means the population means are different. So, within group variance is large and between groups variance is smaller. So, f will be smaller reflecting the likelihood of no significant difference between these 3 sample means.

So, between group variance is smaller than what happens I will get a very smaller f value smaller value means it will lie in my acceptance region. So, the ath indicates that there is no difference between the population means of all the populations. As between whom were a group variable increase sample means grow further apart from each other in other words, the samples are more probable to be belonging to totally different populations, if the between group variable is quite bigger as if it belongs to a different population.

So, understood that means, from the what is the ENOVA entity? ENOVA entity gives us the 2 variance estimate. One is variance between the group, another is variance within the group. So, when I have to compare this to 2 variants, always I will be using t distribution. So, it might between group variability which is in the numerator, if that variability is a bigger value, if the between the 2 variable is a bigger value, then definitely I will get an F a very high value and if the value f false in the critical region or rejection region that means what?

That means when between group variability is very high, that is why I am getting such a high value and then we cannot say that the populations all the population are mean are same that means, because of the treatment, there is effect on the population. Now, because of the different treatments the population means of the populations are different.

(Refer Slide Time: 26:14)



ANOVA Table

- It is convenient to summarize the results of an analysis of variance in a table.
- For a one factor analysis this takes the following form.

Source of variation	Sum of squares	Degrees of freedom	Mean square	F ratio
Between samples	$SS_{Treatments}$	a − 1	$MS_{Treatment}$	$\frac{MS_T}{MS_E}$
Within samples	SS_E	N − a	MS_E	
Total	SS_T	N − 1		

Monalisa Sarma
IIT KHARAGPUR

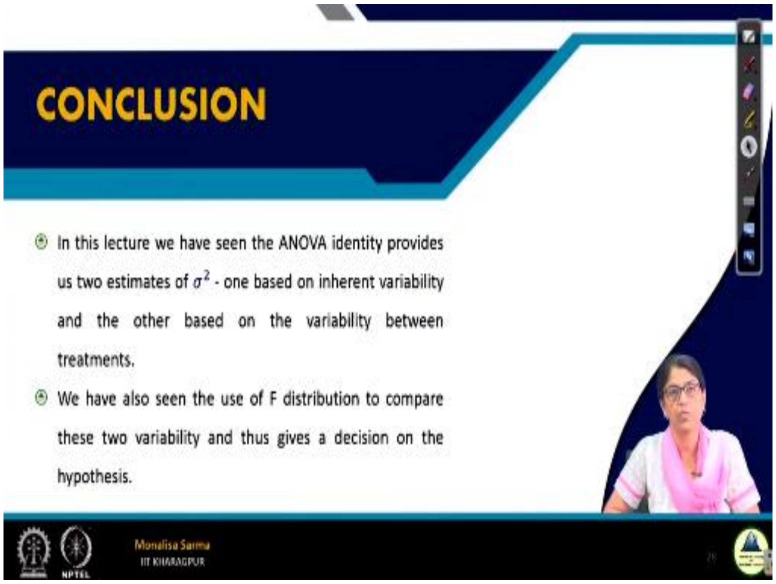
So, it is convenient to summarize the results of analysis in a table for a one factor analysis we can just give it in a table form. So, source of variance between sample that is a SS treatment with degrees of freedom is a - 1. So, then from that we can find it this MS treatment, what is MS treatment is MS treatment / a - 1 gives us MS treatment, then within sample it is SS E. And SS E / n - a we get MS E and F is nothing but MS T / MS E.

So, basically now to conclude so, when we are trying to do ANOVA, exactly what we are trying to do, we are trying to do 2 things, we are trying to find out the variability between different groups. We are trying to find out the variability between different groups that are one then second is we are trying to find out the variability within group. So, variability between groups, I am telling you this MS treatment variable the between group I am terming it as MS E.

Now, if both the terms are equal in both the terms are almost equal not maybe same, but if there is no much of difference in both that terms are similar MS treatment and MS E if they are similar, then we can say that the treatment did not have any effect on the populations. So, whatever variants we are getting there, then the variances may be due to the random error, but if both the terms are very different, that means if MS treatment is quite high, that means the variability between groups is very big.

When the variability between groups is very big that means it definitely belongs from the different population that means because of the treatment, we are having some effect and that is why the population means there is a change in a population mean.

(Refer Slide Time: 28:13)



CONCLUSION

- In this lecture we have seen the ANOVA identity provides us two estimates of σ^2 - one based on inherent variability and the other based on the variability between treatments.
- We have also seen the use of F distribution to compare these two variability and thus gives a decision on the hypothesis.

Monalisa Sarma
IIT KHARAGPUR

So, in this lecture, we have seen that provided data provides us 2 estimates of σ^2 one based on the inherent variability and other based on the variability between the treatments. We have also seen the use of app distribution to compare these 2 variability and does gives a decision on the hypothesis.

(Refer Slide Time: 28:32)

REFERENCES

Statistics, Witte, R.S., Witte, J.S., 2013 - Wiley

Probability and Statistics, 2nd edition, Morris DeGroot, 1986 Reading, MA: Addison-Wesley

Freund, R.J. & Wilson, W.J. Statistical Methods, Academic Press, San Diego, 1997

Monalisa Sarma
IIT KHARAGPUR

So, these are the references and thank you guys