Marketing Research and Analysis-II
(Application Oriented)

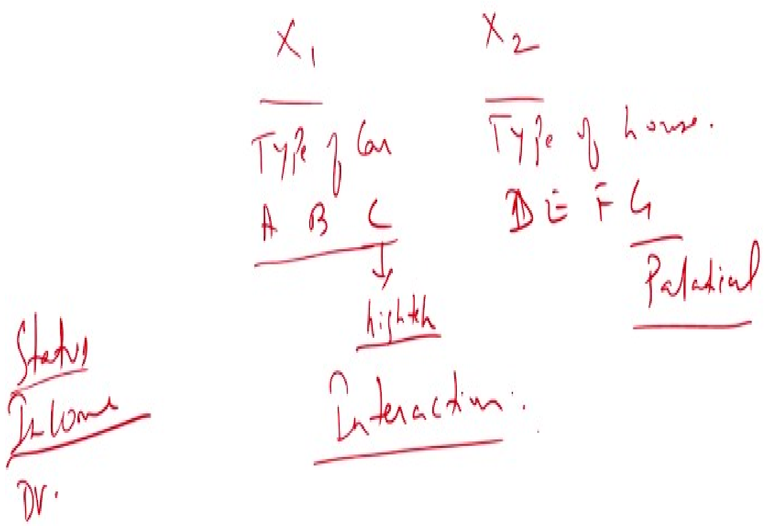
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 36
Solving n-way ANOVA - II

Hi friends, I welcome you all to the course Marketing Research and Analysis. In this lecture, we will continue with the one that we had left in the last one. So in the last class, we were discussing about the n-way ANOVA which is nothing but an extension of the one way ANOVA. So we discussed that one-way ANOVA is used when there are more than 2 levels or groups.

Then the researcher has to find a way to measure the significance of the test through a single way instead of having multiple t tests, but what would you do if you are faced with 2 factor or 3 factors or 4 factors and each factor having let us say certain number of levels n. So in such a condition, we use the n-way ANOVA which is very useful.

(Refer Slide Time: 01:31)



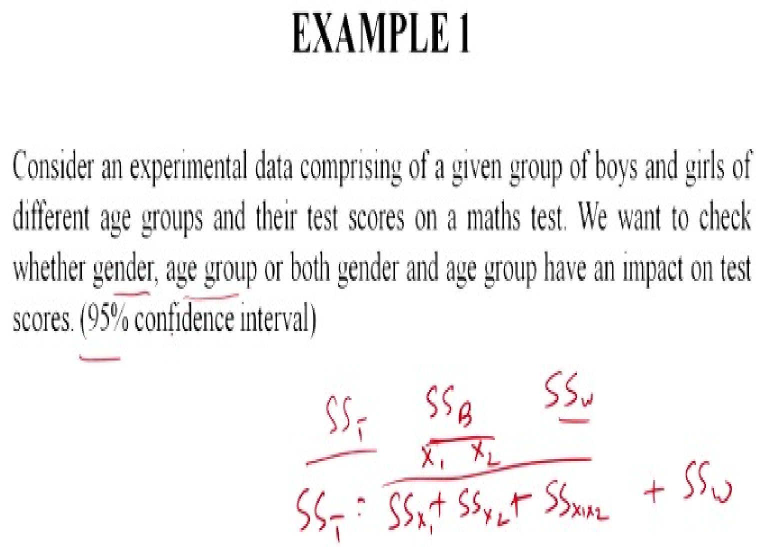
So we understood that in life there are many cases where there are 2 things, let us say X_1 and X_2, let us say this is for example the type of car and this X_2 let us say the type of house. Now these 2 things will effect the status of a person. Now status can let us say it looks like a categorical, we need to have a dependent variable which is continuous, so you may take let us say is the status, type of car and the type of house, effects does the income of a person.

Let us say income, so income is my dependent variable which is a continuous variable income and I am measuring it through 2 factors. The factor 1 is type of car where I have 3 type of cars A, B, and C and I may have 3 types of or 4 types of houses, let us say D, E, F, and G. So when I know that 2 such cases coming up, earlier in one-way ANOVA, I was looking at the effect of only X_1 the type of car that means A, B, C on the income or let us say type of house D, E, F, G on the income.

But now we are considering 2 factors at the same time and we want to see whether not only this X_1 and X_2 have a direct impact on the income and but do they have an interaction effect. That means let us say the type of car, let us say is this is a very high cost car, very very high technology, a hi-tech car and the type of house it is a very big palatial house. So when I am saying these 2 things, are they together have making some another kind of impact? So in such conditions when I would see it is called an interaction effect.

Interaction effect is that means when the value of one factor depends on the value of another factor, so this condition is called an interaction and during a factorial design, we had earlier also learnt during the factorial design, we had learned that interaction effect is one of the most important thing that need to be learned while doing a factorial design and thus we do it through a n-way ANOVA. We will see how to do that.

(Refer Slide Time: 04:10)



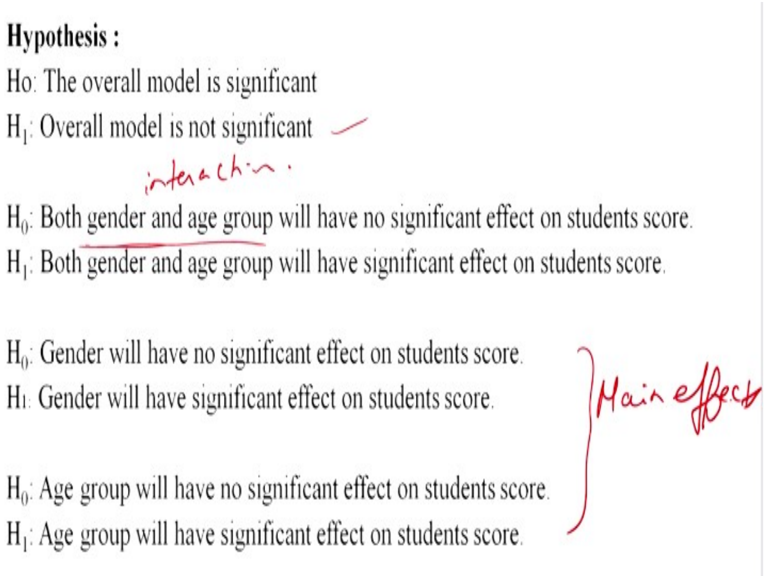
So in the last lecture, I had explained you the steps what to do. It is very very simple, just remember there are 2-3 things you need to find out, the sum of square total, the sum of square between the groups, the sum of square within the groups or which is called error also. So these 3 things are there, but only thing is that here the between becomes little elongated or extended when you have 2 factors.

So suppose there are 2 factors now X_1 and X_2, so it will become sum of square between for X_1, sum of square between for X_2. Then if there is an interaction effect, if see, please do not expect that if I am having 2 factors then is bound to be an interaction effect, no first logically think whether these 2 factors is there any effect is there any possible logically any effect on each other or if they are not connected at all, then there is no point in making interaction effect.

But generally we see in real life in marketing research and other psychological things that there is an interaction effect. So in that case, we find out the sum of square of X_1 and X_2 together. These 3 things together will be added up with the sum of square within to make the total. So this is what we had said, so this plus this, so this is a condition where an experimental data comprising of a given group of boys and girls of different age groups and the test scores on a Maths test has been checked.

We want to check whether gender and age group or both gender and age group have an impact on the test scores and you want to test it at a 5% level of significance or 95% confidence level right. Now how do we proceed first? Let us see.

(Refer Slide Time: 06:08)



First we have to build the hypothesis. What is the hypothesis now? I have made it very very simpler for example for you. Now first we will see that the overall model should be significant, that means if I am taking 2 factors and there is a dependent variable with 2 independent variables of different levels, I am saying the overall model should be significant, my null hypothesis is that and alternate is it is not significant.

The next is I am saying both gender and age group will have no significant effect on students' score, that means here I have taken the interaction effect. Similarly I am saying gender and age group will have significant effect on the students score. Third now we know that first we have to check the overall model, then you have to check an interaction effect, and if your interaction effect is not significant, then only you should check the main effect.

So now suppose it is insignificant the interaction effect, then we are going for the main effect. What is it saying? Gender will have no significant effect on students score, gender will have a significant effect on the students score. Similarly age group also will have no significant effect and age group will have. So now these 2 are basically we are looking at the main effects. Now let us proceed.

(Refer Slide Time: 07:34)

Gender	Age group (years)	Test score
Boy 1	10	4
Boy 2	10	6
Boy 3	10	8
Boy 4	11	6
Boy 5	11	6
Boy 6	11	9
Boy 7	12	8
Boy 8	12	9
Boy 9	12	13

Gender	Age group (years)	Test score
Girl 1	10	4
Girl 2	10	8
Girl 3	10	9
Girl 4	11	7
Girl 5	11	10
Girl 6	11	13
Girl 7	12	12
Girl 8	12	14
Girl 9	12	16

Now this is the gender, so there are 2 groups, one boy and one girl, so boys and girls, so male female and age group which has been divided into 10, 11, 12. So there are 3 age groups, children or students of age group of 10, 11, and 12. The scores are given to you. So there is a boy 10 years of age he scored 4, another boy 10 years score 6, another boy 10 age he scored 8.

Remember all these boys and girls I could have written 1, 2, 3, 4, 5, 5 6, I should have in fact, so because to avoid confusion for many people, because remember they are not repeated, they are always different, similarly this is also 1, 2, 3, 4, 5 we do not repeat, they are not repeating here. So this is all the scores you have got.

(Refer Slide Time: 08:37)

Independent variable 1: Gender (two levels= boy and girl)

Independent variable 2: Age (three levels= 10, 11, and 12 years)

Dependent variable: Test scores

This factorial design is a 2 * 3 (2 rows and 3 columns) with six measurement cells.

So now what is the independent variable? Gender. How many levels? 2; boy, girl. What is the second independent variable? Age, how many levels? 3; 10, 11, and 12 right. What is my dependent variable? Test score right. So this factorial design is a (2 x 3) with 6 measurement cells.

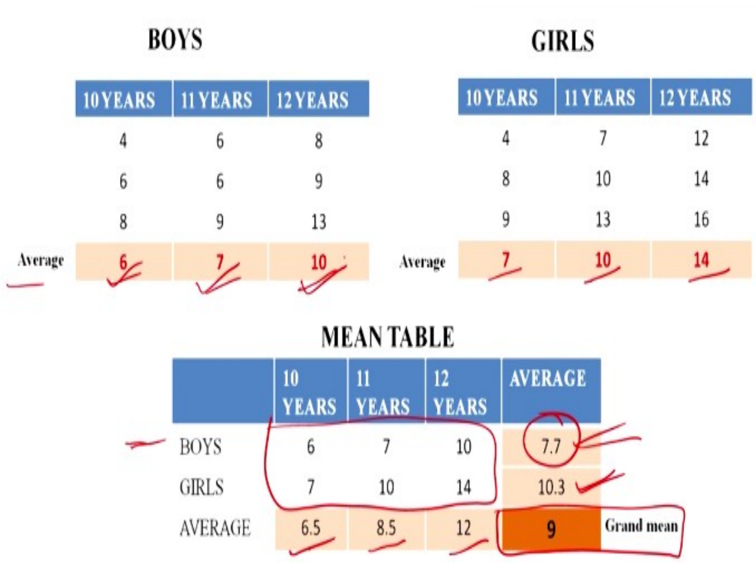
(Refer Slide Time: 09:04)

Factorial design of 2 (2 rows) * 3 (3 column)

		Age group (2nd factor)		
		10 years	**11 years**	**12 years**
Gender (first factor)	BOYS	4 6 8	6 6 9	8 9 13
	GIRLS	4 8 9	7 10 13	12 14 16

So how it would look like? Like this right. So age group you see of 10, 11, 12; gender, boys and girls. Now when I am taking these total cells you see, so boys in the age group of 10 are 3, boys in the age group of 11 are 3, boys in the age group of 12 are 3. Similarly girls also 3, 3, 3. So how many is the total? If you say the total N is equal to how much, 3+3+3+3+3+3+3+ 3+3 = 18, so this is total is 18 okay.

(Refer Slide Time: 09:38)



BOYS

	10 YEARS	11 YEARS	12 YEARS
	4	6	8
	6	6	9
	8	9	13
Average	6	7	10

GIRLS

	10 YEARS	11 YEARS	12 YEARS
	4	7	12
	8	10	14
	9	13	16
Average	7	10	14

MEAN TABLE

	10 YEARS	11 YEARS	12 YEARS	AVERAGE	
BOYS	6	7	10	7.7	
GIRLS	7	10	14	10.3	
AVERAGE	6.5	8.5	12	9	Grand mean

Now this is let us find out first the mean, the average of the mean. So 10 years, the average for the boys in the 10 year age category is how much 4, 6, and 8, so much it is 6. Second for 11, it is 6+6+9 = 21, 7; 13+9 +8 = 30, so 10. Similarly for the girls, so the 10 year aged girls, the average mean score was 7, 8+4+9 = 21/3 = 7, so this is 10 and this is 14. Now let us write the mean table at one go. So how does it look, so boys girls, 10, 11, 12, so just imaging now this much only for at the moment.

So boys 10 years 6, 11 years 7, 12 years 10; girls 10 years 7, 11 years 10, 12 years 14. Now find the mean like you were doing the ANOVA, similarly find the mean for the rows and the columns, so what is the column total here 7+6 = 13/2 = 6.5, this is 8.5, this is how much 24/2 = 12. What is the column average, now the column average is 6+7+10 = 23/3 right, 23/3 is, 7.7 okay; similarly this is 10.3. Now finally after this you find the grand mean.

So this is the individual means, the column means and the row means, now the grand mean. The grand mean is how much 6.5+8.5 is how much 15+12 = 27/3, so 9, so now we have go the value right. From here, we will go to measure the sum of square. Now what is it?

(Refer Slide Time: 11:45)

Sum of squares total = Sum of squares first factor (gender)

\+

Sum of squares 2nd factor (age group)

\+

Sum of squares within (error)

\+

Sum of squares both factors (age x gen)

First it says we want to measure the sum of square total. So what is the sum of square total? I had explained earlier. Sum of square of the first factor which is gender + sum of square of the second factor which is the age group + sum of the square of both the factors let us take this one first sum of the square of gender, age group, plus both the factors okay age and gender, then sum of square of the error. So these 4 things together will be the sum of square of total.

(Refer Slide Time: 12:16)

Sum of squares first factor (gender)

Now, we will compare the boys and girls average with total grand mean.

Boys average= 7.7
Girls average= 10.3
Total grand mean= 9

BOYS

Test score	(Boys mean – grand mean)^2	
4	$(7.7-9)^2$	1.69
6	$(7.7-9)^2$	1.69
8	$(7.7-9)^2$	1.69
6	$(7.7-9)^2$	1.69
6	$(7.7-9)^2$	1.69
9	$(7.7-9)^2$	1.69
8	$(7.7-9)^2$	1.69
9	$(7.7-9)^2$	1.69
13	$(7.7-9)^2$	1.69
	Sum of squares	15.2

Girls

Test score	(Girls mean – grand mean)^2	
4	$(10.3-9)^2$	1.69
8	$(10.3-9)^2$	1.69
9	$(10.3-9)^2$	1.69
7	$(10.3-9)^2$	1.69
10	$(10.3-9)^2$	1.69
13	$(10.3-9)^2$	1.69
12	$(10.3-9)^2$	1.69
14	$(10.3-9)^2$	1.69
16	$(10.3-9)^2$	1.69
	Sum of squares	15.2

Total Sum of squares of first factor (gender) = 15.2 + 15.2 = 30.4

Now let us do it one by one. Sum of square for the first factor, gender. Now we will compare the boys and girls average with total grand mean. So here you see the boys case is the score 4, 6, 8, 6 6, 9, 8,9,13. Now what is the boys average 7.7. So the sum of square is (7.7-9), that means the mean of the boys at the age group, let us go back, let us me just show here, so what is this the boys right, so the mean of the boys or the average of the boys is how much 7.7. Now what we are doing is $(7.7-9)^2$, 9 is my grand mean.

So the same thing happens for all. So I am getting a sum of square of how much 15.2. Now for girls, girls what is the value let us check 10.3, so girls is 10.3, so average is 10.3, so the girls average minus the grand mean, so the average score of the girls minus the grand mean the square of it. So we are doing between the groups now, so between the gender that means, so this is giving us this much 1.69 and this total is 15.2 again. So total sum of squares of the first factor is how much 15.2 this one + 15.2 this one is equal to how much, 30.4.

(Refer Slide Time: 13:53)

Sum of squares of 2nd factor (age group)

Now, we will compare the average values of each age group with total grand mean.

Average (10 years)= 6.5
Average (11 years) = 8.5
Average (12 years) = 12

Total grand mean= 9

BOYS Test score	(age group mean – grand mean)^2		Girls Test score	(age group mean – grand mean)^2	
4	$(6.5-9)^2$	6.25	4	$(6.5-9)^2$	6.25
6	$(6.5-9)^2$	6.25	8	$(6.5-9)^2$	6.25
8	$(6.5-9)^2$	6.25	9	$(6.5-9)^2$	6.25
6	$(8.5-9)^2$	0.25	7	$(8.5-9)^2$	0.25
6	$(8.5-9)^2$	0.25	10	$(8.5-9)^2$	0.25
9	$(8.5-9)^2$	0.25	13	$(8.5-9)^2$	0.25
8	$(12-9)^2$	9	12	$(12-9)^2$	9
9	$(12-9)^2$	9	14	$(12-9)^2$	9
13	$(12-9)^2$	9	16	$(12-9)^2$	9
	Sum of squares	46.5		Sum of squares	46.5

Total Sum of squares of 2nd factor (age group) = 46.5 + 46.5 = 93

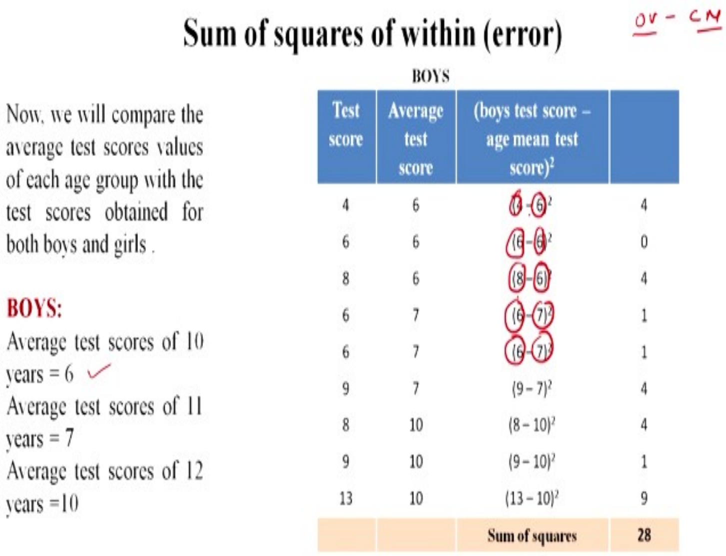
Now coming to the next for the age group. So we will calculate the sum of square for the second factor. Now let us go back to the earlier table, what is it. Age group 10 is 6.5, 8.5, and 12 is 12. So I take the first one, so $(6.5-9)^2$, so this is for the first 3, then the second the 8.5, 8.5 and 8.5 right, and then the third group which is 12 year age group it is $(12-9)^2$, $(12-9)^2$, $(12-9)^2$.

So what we have done, the age group we have taken the scores and we have the means, this 6.5, 8.5, 12 and we have divided from the grand mean and we have found out the square, so this is 46.5. Similarly you do it for the girls, how much is the score 6.5, 8.5, 12. Now for the girls, let us go back, this is the girls 6.5, 8.5, 12 one more. So when we are dividing we will see that this score is again 46.5.

So now for the boys you see so we have there are 3 age groups, so average is 10, the mean is 6.5, for 11 it is 8.5, for 12 it is 12, grand mean is 9. So now we will subtract it, so (6.5-9) and we are finding it. Similarly we are finding it across the girls also. So the total is 46.5 here

46.5 here, so total sum of square for the second factor is how much 93 for both boys and girls together.

(Refer Slide Time: 15:39)



Now after this we will find out the sum of square of within for example, now within means the error term. Now we will compare the average test scores values of each group age group with the test scores obtained for the both boys and girls. So average test scores of 10 years is 6, average test scores of 11 is 7, and test scores of 12 is 10. Now you can see this here, average test scores. So 10, 11, 12; 10 years is 6, 11 is 7, 12 is 10 okay.

Now let us do this. So what is it saying? When you calculate the within variance, that means within the group, what you have to do is you have to subtract the observed value minus the mean of that particular column. So this case the column score the mean was 6 for this 10 year old boys and 11 year old 7, so let us take it and finally 10. So the average test scores are given to you and the boys test score is now to be calculated.

So first boy got $(4-6)^2$, so $(6-6)^2$, I hope you are understanding. This is the observed value minus the column mean okay. So observed value is 4, column mean was 6, observed value is 6 minus column mean 6, observed value 8 column is 6, second case for 11 year age 6, column was 7, observed 6, 7, so it goes on. So when you do this, when I take the square of it $(4-6) = (-2)^2$ is 4. So this sum of square is coming 28.

(Refer Slide Time: 17:40)

GIRLS:

Average test scores of 10 years = 7

Average test scores of 11 years = 10

Average test scores of 12 years =14

Girls

Test score	Average test score	(girls test score–age mean test score)^2	
4	7	$(4-7)^2$	9
8	7	$(8-7)^2$	1
9	7	$(9-7)^2$	4
7	10	$(7-10)^2$	9
10	10	$(10-10)^2$	0
13	10	$(13-10)^2$	9
12	14	$(12-14)^2$	4
14	14	$(14-14)^2$	0
16	14	$(16-14)^2$	4
		Sum of squares	40

Total Sum of squares of within (error) = 28 + 40 = 68

Now let us find it four girls similarly. So girls it is 4 the observed value minus 7 is my mean, 8-7, 9-7, 7-10, 10-10, 13-10, you have to just refer to original table again and again, and 12-14 goes on. So this deviation score, so $(3)^2$is 9, $(1)^2$ is 1, so it go on. So this is 40. So what is my total sum of square within? Now total sum of square within in 40+28 = 68. Now you have to do one more thing, you have to calculate either the interaction or the total, so let us do. The sum of square total now we are calculating.

(Refer Slide Time: 18:35)

Sum of squares total

To calculate sum of squares total, we will compare all the test scores values of with the grand mean.

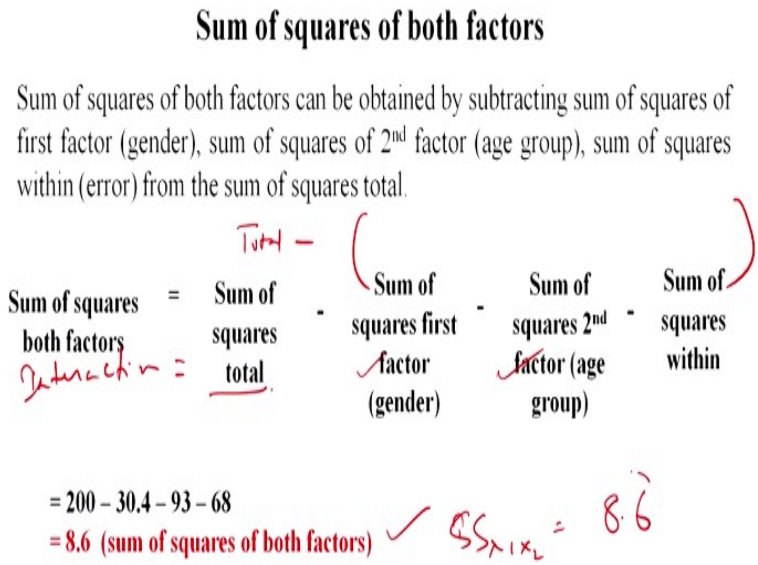
Test score	(test score – grand mean)^2	
4	$(4-9)^2$	25
6	$(6-9)^2$	9
8	$(8-9)^2$	1
6	$(6-9)^2$	9
6	$(6-9)^2$	9
9	$(9-9)^2$	0
8	$(8-9)^2$	1
9	$(9-9)^2$	0
13	$(13-9)^2$	16

Test score	(test score – grand mean)^2	
4	$(4-9)^2$	25
8	$(8-9)^2$	1
9	$(9-9)^2$	0
7	$(7-9)^2$	4
10	$(10-9)^2$	1
13	$(13-9)^2$	16
12	$(12-9)^2$	9
14	$(14-9)^2$	25
16	$(16-9)^2$	49
	Sum of squares total	200

Now what is the sum of square total, how do you do that? Simple. Each observed value minus the grand mean, the square of it. Now each observed value was $(4-9)^2$, so take all the value, both the boys and the girls together. So if you take sum of square or total irrespective of boy girl anything, so it is coming 200. So this is the total.

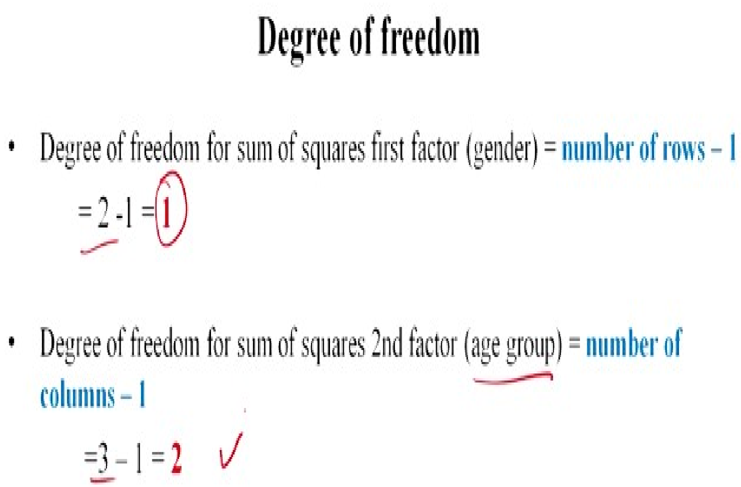
Now interaction effect you can just subtract the rest from the total and find out the interaction effect. So what is it saying?

(Refer Slide Time: 19:00)



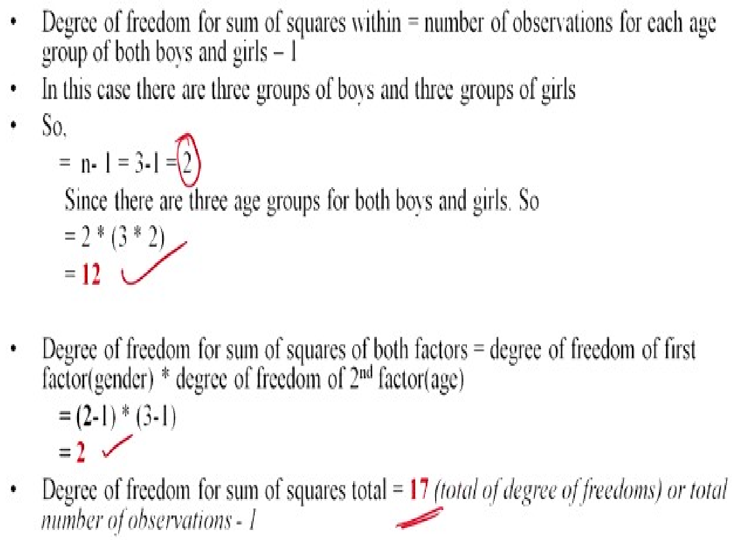
Sum of squares of both factors = sum of squares of total minus sum of square of the first factor (gender) minus sum of square of the second factor (age group) minus the sum of square within, this is interaction effect. So interaction = total - the rest, this rest right, so that is = 200-30.4-93-68, this is 40+28 = 68, so this is giving us the sum of square of the both factors of the interaction X_1X_2, this is $SSx_1x_2 = 8.6$ okay.

(Refer Slide Time: 20:02)



Now the degree of freedom for the first factor gender is how much, number of rows or the levels minus 1, so how many are there 2 levels are there, boys and girls, so minus 1 is 1. For the age group how much, 3-1= 2.

(Refer Slide Time: 20:19)

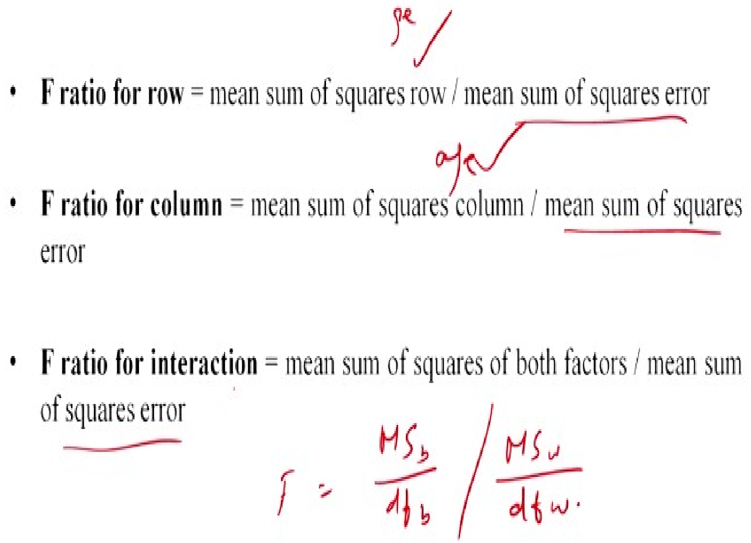


Degree of freedom for sum of squares within is equal to how much? The number of observations for each group for both boys and girls minus 1. Now you see this, in this case there are 3 groups of boys and 3 groups of girls, so n-1 is equal to how much, 3-1 = 2 because there are 3 elements in each group. Since there are 3 age groups for both boys and girls, so how much it becomes, now this 2 x (3 x 2), now that means what? Or you can simply do it like this.

Now there are 3 here, so 3-1 from here, 3-1 here, 3-1 here, so you do 3-1 here, 3-1 here, 3-1 here. So if you add 2+2+2+2+2+2 = 12, so the degree of freedom for the interaction is equal to 12. Degree of freedom for the sum of square of both the factors is equal to how much, the degree of freedom for the first factor gender into the degree of degree of freedom of second factor age, because this is interaction effect, multiplication, so much, (2-1) x(3-1) = 2.

So the degree of freedom for the sum of squares total is how much now n is 18 which you had minus 1, so sum of square of total is equal to 17 in this case.

(Refer Slide Time: 21:44)



Now we have to find out the F ratio. F ratio for the row = mean sum of squares row/mean sum of squares error this is how the formula is given. $F = \frac{MSb}{dfb} \Big/ \frac{MSw}{dfw}$

The F ratio mean sum of squares row divide by mean of squares error. Similarly mean sum of squares, because there are 2, rows and column are different, 2 different factors that is why this is being done.

For the interaction mean sum of squares of both the factors divided by mean sum of squares of the error term. Now the error is constant see in all the 3, only thing that is changing is here the row first value the column, this is the gender one, this is the age one, and this is the interaction one.

(Refer Slide Time: 22:42)

	Sum of squares	Degree of freedom	Mean square	F ratio
Sum of squares first factor (gender)	30.4	1	(30.4/1) 30.4	(30.4/5.67) 5.36
Sum of squares 2nd factor (age group)	93	2	(93/2) 46.5	(46.5/5.67) 8.20
Sum of squares of both factors	8.6	2	(8.6/2) 4.3	(4.3/5.67) 0.76
Sum of squares within (error)	68	12	(68/12) 5.67	
Sum of squares total	200	17		

Now so let us make this table. So first factor what was the sum of square 30.4, what was the degree of freedom 1, what was the second factor age, what was the sum of square 93, what is the degree of freedom 3-1 = 2, what is the sum of squares of both the factors the interaction now 8.6, the total minus we did it, and what is the degree of freedom 2 okay.

Then we had sum of squares within what is the total within is 68, what is the degree of freedom 12,each was you minus 1 from each, and total is now 217. So this gives us the total mean square, so this is you can find out the MS square, the mean sum of square. Now this value when you do have it, so this is 30.4/1, 93/2, so this is all you get. So finally you get the F ratio. Now F ratio for the first factor is how much 5.36, second 8.2, third 0.76.

(Refer Slide Time: 23:47)

- Overall effect F overall (5, 12), $p < 0.05$, the critical value from F table is 3.11. Calculated value is 4.65. Since, calculated value is grater than the table value . We reject the null hypothesis.

$$F = \frac{(SS_{x_1} + SS_{x_2} + SS_{x_1x_2})/df_n}{SS_{error}/df_d}$$

$$= \frac{SS_{x_1,x_2,x_1x_2}/df_n}{SS_{error}/df_d}$$

$$= \frac{MS_{x_1,x_2,x_1x_2}}{MS_{error}}$$

where df_n = degrees of freedom for the numerator

$= (c_1 - 1) + (c_2 - 1) + (c_1 - 1)(c_2 - 1)$

$= c_1c_2 - 1$

df_d = degrees of freedom for the denominator

$= N - c_1c_2$

MS = mean square

Calculation:

F = {(30.4 + 93+ 8.6)/5}/ {(68/12)}

= 26.4/5.67

= 4.65

Given,

$df_n = (C_1*C_2)-1 = 2*3-1=5$

$df_d = N- (C_1*C_2) = 18- 2*3=12$

Now from here we will find out, we will go step by step. So we have calculated all the values. So overall effect of F at 5 and 12, what is this 5? 5 is the numerator because there were 6 groups if you remember, so 6-1, and 12 is 18-1 from each, so 18-6 that is why 12, p is at a less than 0.05, 5%, critical value from the F table is 3.11, I have the F table I think, okay it is not here, so is 3.11, and the calculated value we have got is how much check here 4.65, we have calculated here..

So how do you calculate, you see the formula is given to you. So $F = \dfrac{SSx1+SSx2+SSx1x2}{dfn} \Big/ \dfrac{SSerror}{dfd}$

So this gives us how much SSx_1 was 30.4+93+8.6/5 again divided whole is by the error which is 68/12, so this gives us 4.65, so this is what we have said. Since this calculated value is greater than the table value, we reject the null hypothesis.

So what is my null hypothesis, that the overall model there is no impact or is not significant. So now the overall model is found to be significant, the null hypothesis is rejected and how do you calculate the degree of freedom. It is given to you (C_1 x C_2) -1, without this also you can do, I have already told you how to do, so this is done, so 6 groups, (3x2) -1, similarly here.

(Refer Slide Time: 25:49)

- For interaction effect $F_{gender*age\ group}$ (2, 12), $p< .05$, the critical value from F table is 3.89.Calculated value is 0.76. Since, calculated value is smaller than critical value. So, we fail to reject null hypothesis. Thus, we move to the next step i.e. Measure the main effect.
- For F_{gender} (1, 12), $p< .05$, the critical value from F table is 4.75. Calculated value is 5.36. Since, calculated value is greater than critical value. So, reject null hypothesis.
- For $F_{age\ group}$ (2, 12), $p< .05$, the critical value from F table is 3.89. Calculated value is 8.20. Since, calculated value is greater than critical value. So, reject null hypothesis.

Now the interaction effect of gender and age group is how much, at degrees of freedom gender 2, age group you have 12, gender to age group the total is (2x12), just if you see 2 was your number of and age group so it is coming 12, so 12 is because of this we have the overall this is the one we are talking about degree of freedom denominator. The critical value of F table is 3.89.

So now the interaction effect you see at (2, 12), 0.05 is from the F table the value is 3.89 and calculated is 0.76, do you remember where we calculated, 0.76 here right. So 2 and then we had 4.3, this value divided by 5.67. So when you divide this 5.67 also I hope you remember when we got this 5.67, the 5.67 we got it from the yes this value, so sum of square within error.

So when you do this so we fail so since the calculated value is smaller than the critical value, critical value is how much 3.89, calculated value is 0.76, so we fail to reject the null hypothesis, thus we cannot say, that means the interaction effect is found to be insignificant, not significant. Thus we move to the next step that is measure the main effects. Now coming to the main effects first we will calculate for gender only, So (1, 12), so 1 and 12 is the numerator and denominator, degrees of freedom.

The critical value from F table is 4.75 and the calculated value is 5.36. Now let us again go back so 5.36 and you have 4.75 I think. So calculated table value is 4.75, so the null hypothesis is rejected, that means what the null hypothesis is rejected means that it says that the gender has a significant effect on the scores obtained by the students. Second we will check for age group, the second main effect at (2, 12), 2 is because 3 levels were there so (3-1) and 12 is the denominator.

So the critical value is again 3.89 from the F table we see and calculated value is 8.2 if you check back the table. Since the calculated value is again greater than the critical value, the null hypothesis is again rejected, that means what age group also plays a vital role in the performance of the students okay. So what is the interpretation?

(Refer Slide Time: 28:47)

Interpretation

- Gender will have significant effect on students score.
- Age group will have significant effect on students score.
- Both gender and age group will have no significant effect on students score.

The interpretation is that gender will have significant effect on students score, age group will have significant effect on students score, both gender and age group will have no significant effect on the students score. What I will do is to instead of going for another problem, I will try to show you how to do this in the SPSS, for example let me take a problem. Now I will show you through this problem how to go for a two-way ANOVA test.

(Video Starts: 29:19) Now this is a case where there are 2 things, one is the day on which day the newspaper is coming and this is the type of paper. The type of paper could be a daily magazine, daily newspaper, economic, weekly or something, some kind of. There are 3 types of papers 5 days in a week and it has been tried to see whether the advertisement hits is it related to the day and the type of paper.

So that means what, what it is saying simply is the advertisement the hit or the popularity of the advertisement given in the paper does it depend on which day this paper is coming and in which type of paper the advertisement has been shown. So to do this, you have 2 categorical variables and 2 factors, this is one independent variable, this is another independent variable and this is my dependent variable.

So to do this, I will do a two-way ANOVA, now how to do this? First got to analyze. Now through a compare means you cannot do it because it only allows you for a one-way ANOVA and it does not allow you for an interaction effect or for 2 factors, more than one factor. So we use a general linear model this is used when you are doing a two-way ANOVA or n-way

ANOVA. So what you do is univariate, so way univariate? Because only one dependent variable we have okay.

So what is the dependent variable I am taking is now my hits. What are my independent variables? Independent variables are called my factors, fixed factors right so I am taking. Had I had some covariate, that means some continuous variable here, I would have put in the covariate. Now let us check. You want to check the plots you can check, so you want to check add, you want to see the effect of paper you can check.

Now I can also see the effect of both day and paper together, so that is what I am trying to see an interaction effect okay, continue. You can see the post hoc effect, now what is the post hoc effect? Post hoc effect tells you suppose you have found out yes there is a significant difference among the 5 papers the 5 days, but then can you say which day has got the highest impact or which day has the lowest impact? No.

So to understand such kind of things where we want to measure between the days or between the type of paper, we will use a post hoc test. So generally we use the Tukey test or the LSD or if it is small sample size we use a Bonferroni okay. Now I am going for some options. Now options are for example I want to check for the overall model, the means for the day, paper and this thing and I want to compare the main effects

I want to go through 2 different things, one is the descriptive statistics I want to see, I want the homogeneity test. Now homogeneity test in the beginning of the class I had said that it has to be very important that the homogeneity of variance between the groups is not significant. That means the null hypothesis which says that there is no difference should be accepted in case of a homogeneity of variance test.

That means the researcher has to say that the groups, there is no difference in the variance between the different groups that is what homogeneity of variance or Levene's test does and it is an important assumption. Now let us run this. So let us go to the output table. So if you go to the output what it says there are 5 five days, 12 each respondents you have taken, 3 papers so the same, 60 is divided by 3, 20, 20, 20.

Now let us look at the mean. Now day 1, it is day 1 paper 1, day 1 paper 2 and look at the mean so day 1 paper 1 the mean is 8.25, the hit the popularity, standard deviation is given to you, For day 1 paper 2, 11.5; day 1 paper 3 is only 4.5, that means the advertisement hit is very low here and overall average is 8.083. Similarly for second day paper 1, 10.5; second day paper 2, 8.75; second day third paper, 6.5 and goes on.

So you can check for all. So the overall total is also given to you for paper 1, paper 2, paper 3. Now as I had said first of all we have to check for the Leven's test of the equality of variance. Now since it is above 0.5, that means we can easily say that you cannot reject the null hypothesis, the null hypothesis is to be accepted, and this is what we wanted. So what it says to test the null hypothesis, the error variance is equal across groups, which is found to be true okay.

Now coming to the test of between subject effects, so what is it? Now the overall model we have seen is significant, now intercept leave it. Now coming to the day is the day in which the advertisement is given and that results in popularity or not is that significant or non-significant, now we have seen yes it is significant. So the F ratio is 20 and it is significant. Does the type of paper is it significant? Yes, it is significant. Is interaction effect significant? Yes, it is significant.

So luckily this time, you got all the 3 significant. In many cases I have seen you may get only one of them significant or only the interaction effect significant and none of the main effect significant of only one of the main effect and the interaction effect significant, rarely we get all the 3, generally I have not seen many of the cases. Then we can check it individually. Now coming to the post hoc test, this is important.

Now here what you can do is now you can compare you have now seen whether the papers are okay significant, the day is significant, but let us make a comparison between the days. Now the first day and second day, is there a significant difference between first and second day? Let us look at the difference and the significance, no, it is not significant. First and third day, is it significant? No. First and fourth day, yes it is significant. First and fifth day, yes it is significant.

Similarly second day and the let us say the third day we have not compared, so not significant. Second and fourth, yes. Second and fifth, again yes. So when you check this, it helps to tell you whether there is a significant effect or the significant difference between the different days on which the advertisement is given. So by doing this, so see this is what we are doing, you should see this table, the Tukey, HSD, this is the post hoc test.

So you can see the fifth day one all are significant mostly and what is not significant, third day and second day, no significant difference and first day and third day, no obviously not right. So this is what, for paper also you can see, now it is here. For the day we did it, for paper you can compare among the papers 1 and 2, is it significant? No; 1 and 3, yes significant; 2 and 3, yes significant.

So this helps to tell you and you can even check the profile plots also that is there for you, how it looks like. So this gives a very clear picture of how to do a two-way ANOVA test **(Video Ends: 37:33)**. Well what we can do is we will continue this may be to the next class with some other kind of test which are there. So thank you for the day. I hope you understood what is two-way ANOVA.

Kindly when you follow my class, try to stop it in between and take a pause and think what I am saying and in case you have doubts, you can always call back on me, but I am sure if you listen to it and have given a patient hearing for once or two times, you will be completely clear on what is n-way ANOVA and how to do it. Thank you very much.