Marketing Research and Analysis-II
(Application Oriented)
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 37
N-way ANOVA and MANOVA in SPSS

Hi friends, I welcome you all to the course Marketing Research and Analysis second part. So in this course, we have been discussing about several things using the field of marketing research. In the last few classes, we have been discussing about basically experimental research, experimental design and then followed with we stated with analysis of variance one-way and two-way and today we will be continuing with the same thing multiple analysis of variance and analysis of covariance in the coming lectures.

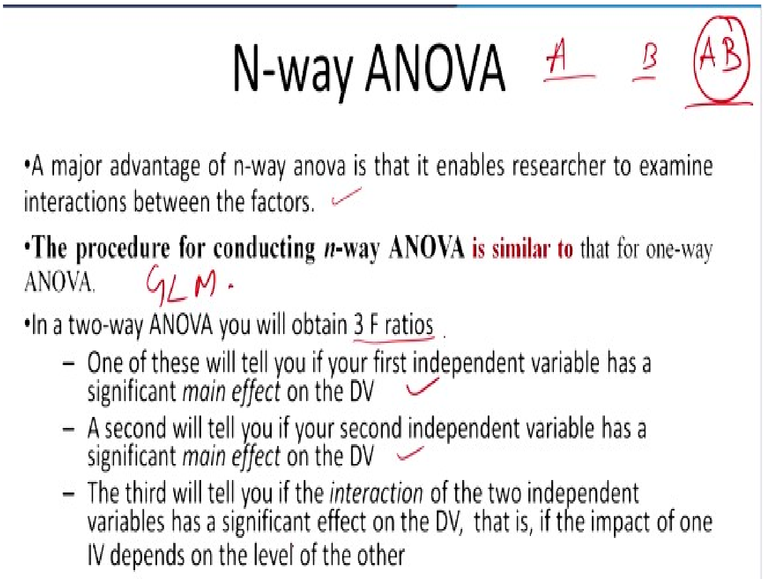
So let us begin with the same which we had stopped in the last lecture that was N-way ANOVA. So as you understand n way stands for that means where there is more than one factor, in the case of analysis of variance study with more than one factor, then we say it is n way. It could be 2, 3, 4, 5 anything, so n way, so there are n factors, N-way ANOVA. So what happens in the N-way ANOVA and what is the advantage. The advantage is that many a times in life you need to understand that there could be more than one factor which is affecting the dependent variable.

So the dependent variable is a continuous you know metric variable in our case in case of ANOVA and the independent variable is measured in a categorical scale, categorical way. So let us say I want to understand for example the influence of the kind of school, the type of school on the student's marks, the grades of students or the type of teaching method or the type of school whatever it is. So when I am saying the independent variable that means is which affects the dependent variable score is my type of school or type of teaching method.

So in there let us say I have school A which is a public sector school, public school, a private school, some schools run by cooperative as 3, so I have several schools. Same time the researcher feels that not only the type of school but the type of teaching method adopted by the schools, may be are they technology based, are they chalk and blackboard based, how they are teaching, so this also has an effect.

So when you take 2 such factors at the same time school and the type of teaching method, then a one-way ANOVA does not work and if you do more than several one-way ANOVA, then again it is not very advisable thing because it leads to higher type I errors. So how to do this? We will see. We have done in the last class through a very general primary method, so today I will show you one more problem in the n-way ANOVA.

(Refer Slide Time: 03:25)



So let us say the major advantage is that it enables researcher to examine interaction between the factors, this is the biggest advantage of the N-way ANOVA. So in one-way ANOVA you had no chance of understanding the interaction effect. So in N-way ANOVA you have 2 factors A and B, now whether A and B is there a third factor which is A and B combined together. So this effect is the most important thing that in real life it has a very huge profound impact.

So that means variables do not individually affect the dependent variable but they effect dependent variable in a combined manner. So the procedure for conducting N-way ANOVA is similar to the one-way ANOVA, it is similar. So only thing is it adopts a general linear modal method which I will show you, what it is also I will tell you. So in a two-way ANOVA, you will have 3 F ratios, what are the 3 F ratios, the first one it says the first independent variable has a significant main effect on the dependent variable.

So one of these will tell you if your first independent variable that means the first factor has a significant effect main effect on the dependent variable or not, so this is the first F. Second F will be the second independent variable or the second factor has a significant main effect on the dependent variable or not. The third will tells us the interaction of the two independent variables the factor 1 the factor 2 taken together. So that is the impact of 1 independent variable depends on the level of the other or not, so this is what is the interaction effect.

So these 3 F ratios are determined during the N-way ANOV. So let us take this example, in the last class we have done one, so today we will do one more. In this what is happening let us see.

(Refer Slide Time: 05:13)

Example 2

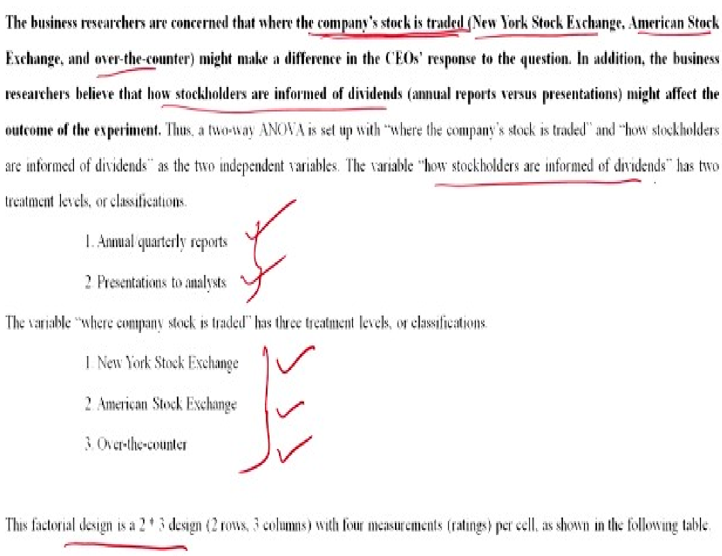
CEOs often must decide whether to award a dividend to stockholders or to make a company investment. To determine whether this factor is important, business researchers randomly select 24 CEOs and ask them to rate how important "availability of profitable investment opportunities" is in deciding whether to pay dividends or invest.

The CEOs are requested to respond to this item on a scale from 0 to 4, where 0 = no importance, 1 = slight importance, 2 = moderate importance, 3 = great importance, and 4 = maximum importance. The 0–4 response is the dependent variable in the experimental design.

CEOs must decide whether to award a dividend to stockholders or to make a company investment, they want to pay through to the dividend to the stockholders or they want to put back the money as a form of investment into the company. To determine this, researchers randomly select 24 CEOs and ask them to rate how important availability of profitable investment opportunities in deciding whether to pay a dividends or invest.

The CEOs are requested to respond in a scale of 0 to 4 where 0 is no importance that means there is no availability of profitable investment opportunity, 1 is slightly important and it goes on where 4 is maximum important. So this, the 0 to 4 response is the dependent variable in the experimental design, so this is what is basically is a thing that we are trying to see how important are they are, they really important or not so what they are doing here is;

(Refer Slide Time: 06:21)


The business researchers are concerned that where the company's stock is traded (New York Stock Exchange, American Stock Exchange, and over-the-counter) might make a difference in the CEOs' response to the question. In addition, the business researchers believe that how stockholders are informed of dividends (annual reports versus presentations) might affect the outcome of the experiment. Thus, a two-way ANOVA is set up with "where the company's stock is traded" and "how stockholders are informed of dividends" as the two independent variables. The variable "how stockholders are informed of dividends" has two treatment levels, or classifications.

1. Annual/quarterly reports
2. Presentations to analysts

The variable "where company stock is traded" has three treatment levels, or classifications.

1. New York Stock Exchange
2. American Stock Exchange
3. Over-the-counter

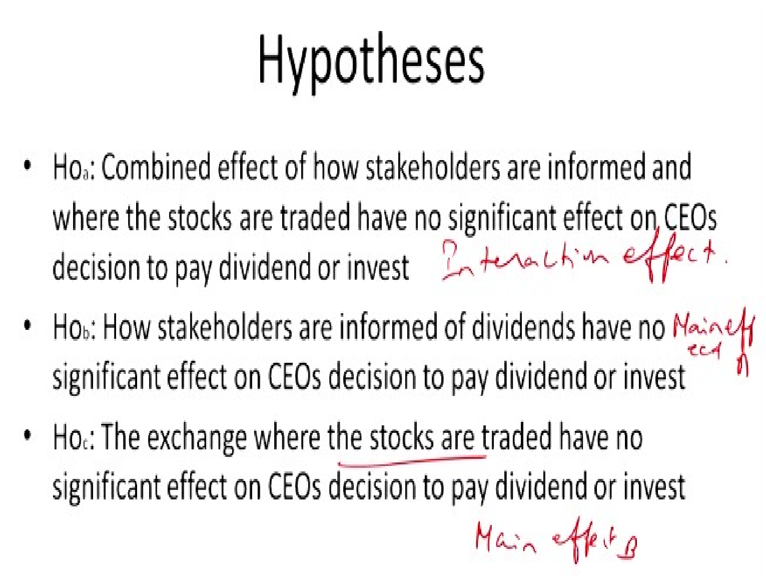
This factorial design is a 2 * 3 design (2 rows, 3 columns) with four measurements (ratings) per cell, as shown in the following table.

The business researchers are concerned that where the company's stock is traded might where it is traded that means is it traded in the New York Stock Exchange or American Stock Exchange or some over the counter, whether the place where it is traded does that have an effect on the opinion, will it make a difference in the CEOs response to the question or in addition the business researchers believe that how the stockholders are informed of the dividends they might almost effect the outcome of the experiment.

So now what is happening there are two things that might affect the outcome of the CEOs. One where the company is trading the stock, second how they are informing the stockholders about the dividends how. So what are the methods, so how stockholders are informed has 2 levels, these 2, through quarterly reports through presentations; where they are traded, the 3 levels, 1, 2, 3 okay.

So this is a factorial design with 2 x 3 ANOVA, why they are saying it is a factorial design because the factorial design can take up any number of levels, any number of factors unbalanced also no issues and it allows for an interaction study with 4 measurements per cell, because there are if you see 24 CEOs and there are 6 total factors right, so each 6 cells, so each cell will have 4.

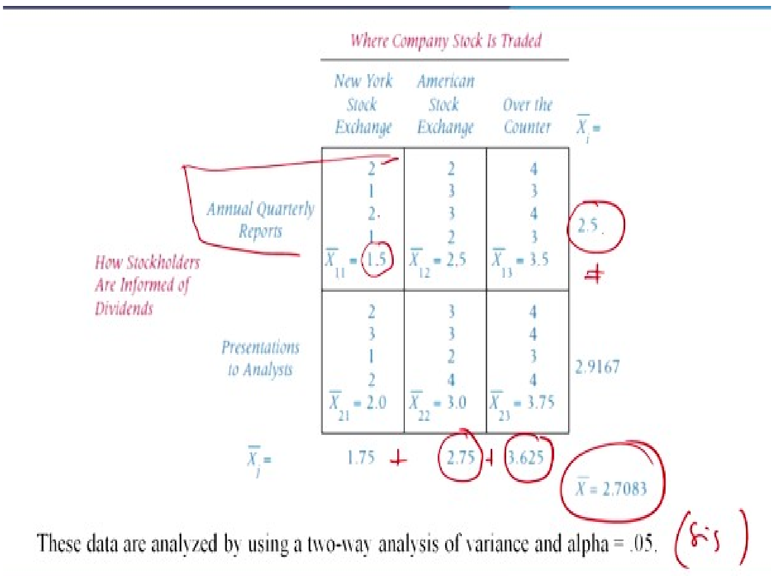
(Refer Slide Time: 07:53)



Now hypothesis, what is my hypothesis. The combined, effect of how stakeholders are informed and where the stocks are traded have no significant effect on the CEOs decision to pay dividend or invest, this is my interaction effect. Second how stakeholders are informed of dividends have no significant effects, so what is this, the significant this is the main effect of the first factor which is the way the information about the dividends is told to the stockholders.

Third, the exchange where the stocks are traded also have no significance, so this is again the main effect for the trading of the stocks. So this is main effect B, this was main effect A okay

(Refer Slide Time: 08:55)



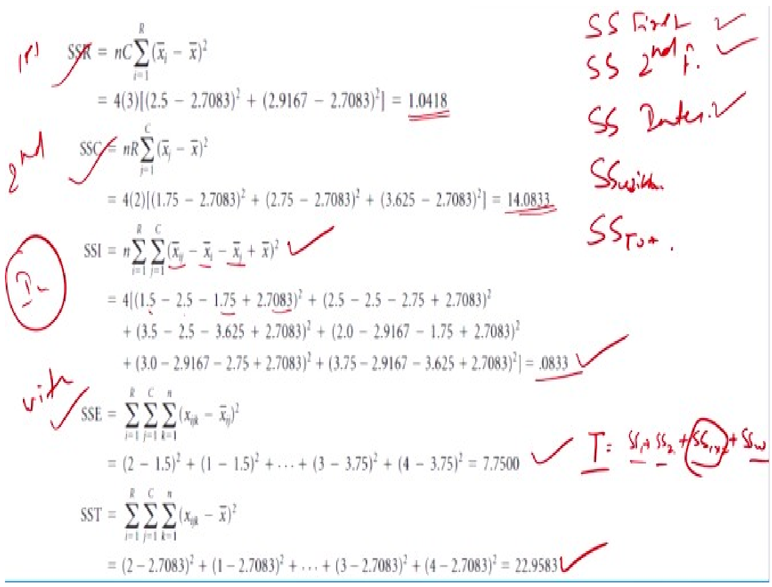
So this problem you see we are having how the information is given to the stockholders, so through the reports and through the presentation, and where it is traded, stock exchange, New

York Stock Exchange, American and over the counter. Now first let us find out individually for there are each 6 cells; 1, 2, 3, 4, 5, 6, for each cell what is the mean. So for the first cell it is 1.5, 3, 4, 5 6/4.

Similarly for the second cell it is 2.5, third cell 3.5, fourth, fifth, and sixth so individually for each first you find out the means. Then for annual quarterly report what is the average mean, the average mean is if you take 3, how many are there 1, 2, 3, so 3, 2.5+1.5, 4+1.5 = 7.5/3 is 2.5. Similarly for the presentation it is 2.9167. For New York Stock Exchange if you go vertically, then what is the mean 1.75, for American Stock Exchange what is the overall mean 2.75, for over the counter what is the overall mean 3.625.

My grand mean is the summation of this plus this plus this divided by 3 or this plus this divided by 2 okay, this comes to 2.7083. We have taken a significance level of 5% okay. Now let us solve this problem.

(Refer Slide Time: 10:26)



So first we need to find out 5 things, what are they? Sum of square for the first factor, sum of square for the second factor, sum of square of the interaction between the first and the second. then sum of square within which you say also error, then sum of square total. If you find any 4 of it also, then the fifth one is automatically we can find out. So let us see the first one, sum of square for the first factor, so the first factor right.

So many first factor, we are talking about this one, so how many cells are there here 3 right, so $4(3)[(2.5-2.7083)^2+ (2.9167-2.7083)^2] = 1.0418$. Let us find out for the second factor, so what is the n, again 4, how many cells are there this time for example we are trying to see the cells for each 1, 2; 1, 2; 1,2, so 2 cells, so 1.75 let us go, so $(1.75 - 2.7083)^2$+ $(2.75 - 2.7083)^2$+ $(3.625 - 2.7083)^2$.

Now if you do this, you get a score of 14.0833. Now we are finding the third, this is done, this is done the interaction. So what is the interaction, interaction is saying 4 x how do you find out x_{ij} right so let us see, so this value x_{ij} is x_{11} or x_{12}, this value minus the corresponding mean, so this - this square + this - this square + this - this square + this goes on. So 1.5-2.5-x_{ij}-x_i first one then x_j, so i and j, so i first row and j is my column, so column wise if you go is 1.75, so $(1.5-2.5-1.75+2.7083)^2$.

So similarly, now go to the next one, it goes (2.5-2.5) so $(0-2.75-2.7083)^2$ so it gives us square plus similarly $(3.5-2.5+3.625-2.7)^2$. So as you do this, so this is the formula for you can remember it also, so x_{ij} mean of the first cell minus the total of that column the row similarly minus the x_{ij} + the grand mean. So if you use this formula we are getting 0.0833. Now once you have got this, then you have to calculate the sum of square within or this. Now how do you calculate the within, within means each cell minus that particular mean.

So 2 observed value these $(2-1.5)^2 + (1-1.5)^2$ it goes on so + goes on till you reach to 3 minus let us say you see up to this. So individually, so this-this this-this + this-this, this minus this, this minus this, this minus this goes on for all till you reach $(4-3.75)^2$, so this total overall total is equal to 7.75, sum of square within and what is my total. Now if I add these 4, interaction, within, if I add these 4 also I will get total or else simply you can do one thing if you get confused with let us say interaction or any one of it.

First one is very simple, second one is very simple, you may not remember the formula, just find out the total. How to find the total? Every cell minus the grand mean. Now $(2-2.7083)^2$+1-this, this-this, this-this, and this-this, this goes on for all right. When you do the square, this gives you the total, so this is equal to 22.95. Suppose I have my total is equal to I know that is SS of first factor + SS of second factor + SS of interaction, interaction of 1x2 + SS of within.

So suppose I have total, I have this this and this, I need not calculate this, I can also find out from here just by subtracting right, so anyway.

(Refer Slide Time: 15:53)

$$MSR = \frac{SSR}{R-1} = \frac{1.0418}{1} = 1.0418$$

$$MSC = \frac{SSC}{C-1} = \frac{14.0833}{2} = 7.0417$$

$$MSI = \frac{SSI}{(R-1)(C-1)} = \frac{.0833}{2} = .0417$$

$$MSE = \frac{SSE}{RC(n-1)} = \frac{7.7500}{18} = .4306$$

$$F_R = \frac{MSR}{MSE} = \frac{1.0418}{.4306} = 2.42$$

$$F_C = \frac{MSC}{MSE} = \frac{7.0417}{.4306} = 16.35$$

$$F_I = \frac{MSI}{MSE} = \frac{.0417}{.4306} = 0.10$$

So from here what we do next, we find the mean sum of square. So mean of sum of square for the first factor is SSR/R-1, so that is how much so you had for the first there are 2, so 2-1 so that is = 1.04. For the second factor you have how many 3, so 3 are American Stock Exchange, the New York Stock Exchange and over the counter, so if you divide it then it comes to 7.04.

The interaction is sum of square of interaction which you have calculated divided by (R-1) x (C-1), so 0.41. The mean sum of square error or within is equal to this is the formula, SSE/RC(n-1), so this is giving me .43. So I am getting the F ratio for the first factor, F ratio for the second factor, and F ratio for my interaction. So these 3 once I get, then I can write it like this.

(Refer Slide Time: 16:53)

Source of Variation	SS	df	MS	F
Row	1.0418	1	1.0418	2.42
Column	14.0833	2	7.0417	16.35*
Interaction	.0833	2	.0417	0.10
Error	7.7500	18	.4306	
Total	22.9583	23		

*Denotes significance at $\alpha = .05$.

So I see from here so my F values, it tells me if my F value is you see 2.42 and for column is 16.35, interaction is only 0.10. Now how do you measure it, how do you write it?

(Refer Slide Time: 17:10)

- Overall effect F overall (5, 18), p < 0.05, the critical value from F table is 2.77. Calculated value is 7.05. 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 = {(1.04+14.08+0.08)/5}/ {(7.75/18)}

= 7.05

Given,

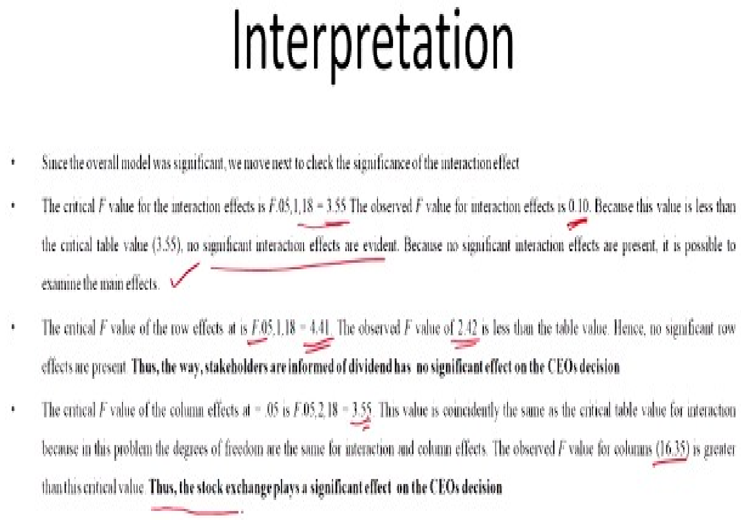
$df_n = (C_1 * C_2) - 1 = 2*3-1=5$

$df_d = N - (C_1 * C_2) = 24 - 2*3=18$

So the overall effect F at 5, 18, 5 and 18 is the degree of freedom for numerator and denominator at 5% level of significance, the critical value from the F table is this much, for this row and column if you find out the table value is this much, the calculated value is this much. So since calculated value is greater than the table value, we reject the null hypothesis. Now you see we have calculated here. So F is equal to how much 1.04+14+0.08/5 this is my upper side and then within, this is my within, so within is how much 7.75/18 degree of freedom is 18.

So overall I am getting a F ratio of 7.05. My degree of freedom is this formula is also given to you, you can check it. So from here we can infer what is our interpretation?

(Refer Slide Time: 18:17)

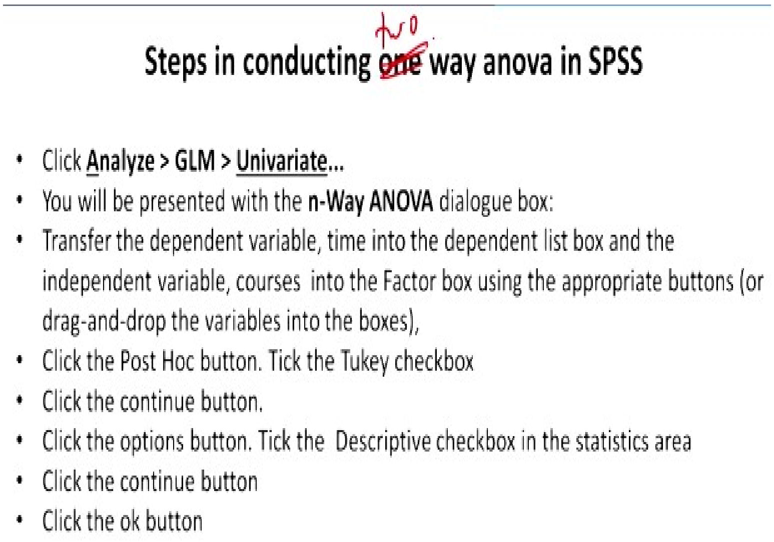


Since the overall model was significant, we move next to check the significance of interaction effect. The critical value for the interaction effect is 3.55. The observed F value for interaction effect is this much, so this is my table value, this is my observed. So since the table value is much more than the calculated value, this value is less than the table value, so no significant effects are evident. So the interaction effect is not there and remember when there is no interaction effect, the researcher should move next to the main effects.

Because no significant interaction effects are present, it is possible to examine the main effects. Now what is the main effects the critical F value of the row effects at $F05,1,18 = 4.41$ and the observed value is 2.42, hence again no significant row effects are present, that means the main effect of first factor is also not present. That means you can say the stakeholders are informed of dividend, which way they are informed of dividend, has no significant effect right.

Similarly if you take the second effect, the column which was the over the counter of which kind of stock exchange. So F is equal to this much is 3.55 and the observed value is 16.35, which is much, much higher. So when it is much higher, we reject, so we say the stock exchange the type of stock exchange where the stock has been traded that makes a significant effect on the CEOs decision whether to invest or not to or to pay dividends. So this is how you do.

(Refer Slide Time: 20:10)



Now so let me show you how to do a two-way again. So let us take this so this a case where there types of cars, there are different types of car. So you can see there are cars, the type of filters and their effect on the performance. So I want to check whether the performance of my, performance means the fuel performance, is effected by the brand of the car and the kind of filter that is being used within the car right. So let us see how to go. When you have 2 factors, so that mean you have factor 1 and factor 2.

So when you have 2 factors and is one continuous dependent variable, this is a case of a two-way ANOVA. So let us do this. Now how to do? Now we go to the first general linear model and univariate, why univariate because you have only dependent variable. So performance I am taking as my dependent variable and car and filter as my independent variable. Now what I am doing is I will go to see what is the main effect and what are the interaction effects? To see that first I will take this and I will take a car into filters, so this is the interaction effect okay.

Then what is this post hoc and do we require post hoc, yes we do require, but you require only post hoc when there are more 2 levels, it is a comparison of the means right, so only car is possible because car I think there are 3 brands okay. The most effective method is Tukey method which we will use. There are 3 methods, in fact the Sidak method is also used and Bonferroni is also used when there are small cases of samples, but we will use the Tukey method.

What are the options, though I need to see the means for all, so all the 3 and I want to compare the main effects in case my interaction effect is not working is not significant, then I will use that. So these 3 things I require and I continue and I go to okay. Now let us see what kind of analysis we are getting. Now look at this, we are having 3 cars, 2 filters okay. Now the first car and the first filter, this is 1, 2, 3.

So the first car fitted with a first type of filter mean is 15, second is 18.5, and this value you can see for the third car for first filter is 15 and the second one is 12. Now first of all, we want to see the assumption of homogeneity of variance. Now the homogeneity of variance this is a very important test because it says and you need to report it when you write your thesis paper whether there is a homogeneity of variance, that means whether the groups do have equal variance or not, so that is our null hypothesis.

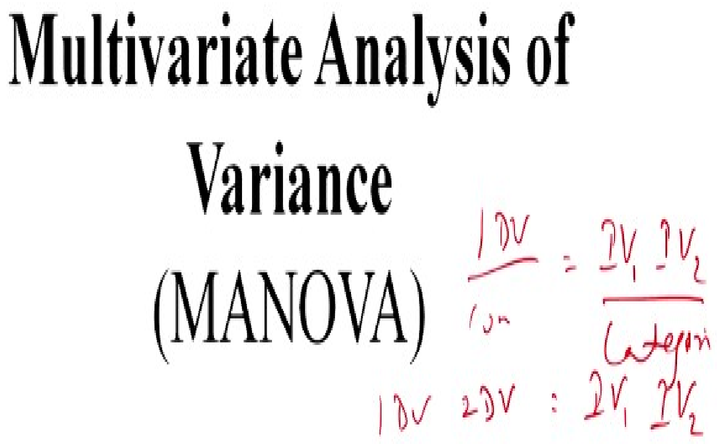
So here the Levene's test says or the homogeneity of variance test says that yes it is significant at 0.246, that means the null hypothesis cannot be rejected. So what is the null hypothesis that there is an equal variance among the groups, so it is correct, so that means there is homogeneity of variance. Next now let us look at the test of between subject effects. Now if you look at car, so first of all let us look at the F ratio 5.4 and the significance, is it significant, yes the kind of cars effect the performance.

Filter, well no, 0.745, so it is not significant. Is the interaction effect playing an important role, yes, the interaction effect of car and filter has an important role, it plays an important role and it is significant. So when you report when you write, you have to write, and what is this partial eta square it tells you the effect size. So for example you see the effect size is 0.142 for the car and 0.145 for the car interaction effect. So now you can see the pairwise comparison, that is why post hoc test effect.

So we want to compare car 1 versus car 2, is it a significant difference, no. Car versus 1 and car versus 3, is there significant difference, yes. Car versus 2 and 3, is there a significant difference, well no; 3 and 1, yes. So the study helps us to find out which independent variables effect the dependent variable in how and what is the effect, when you compare the groups, how the groups are affecting finally the dependent variable. So this is what we are doing in a two-way ANOVA.

So you can see these for also for filter and all, but we are not interested. So car into filter, so every post hoc test you can see this, it is given to you, these values are important okay. So now I will close it, I close this also. Now I will go back to my slide.

(Refer Slide Time: 25:20)



In fact I wanted to start with this multivariate analysis of variance, I just begin. What is this multivariate analysis of variance, to continue? Multivariate analysis of variance is a technique which is a step over the ANOVA and it allows you for, in the ANOVA case, what happened we had one dependent variable continuous right with several independent variables let us say IV1, IV2 which were categorical correct this was the case.

Now the case is different now. We have multiple dependent variables let us say you have one dependent variable, then you have two second dependent variable and you have 2 independent variables or you can have whatever condition. So when I have more than 1 dependent variable, so this study is a case of a multivariate analysis of variance

(Refer Slide Time: 26:20)

Multivariate Analysis of Variance

- The multivariate analysis of variance (MANOVA) is used to determine whether there are any differences between independent groups on **more than one continuous dependent variable**.
- In this regard, it differs from a ANOVA, which only measures **one continuous dependent variable**.

So multivariate analysis of variance we will just see is used to see determine whether there are any differences between independent groups on more than one continuous dependent variable. In this regard, it differs from an ANOVA which only measures one continuous dependent variable, so that is the difference.

(Refer Slide Time: 26:37)

Types of MANOVA

- **One way MANOVA-** If the independent variable (categorical) is one, and dependent variable (continuous) is more than one. It is a case of one way MANOVA.
- **Example-** You can use MANOVA to understand whether there were differences in the **perceptions of attractiveness** and **intelligence of vegetarians** (i.e., the two dependent variables are "**perceptions of attractiveness**" and "**perceptions of intelligence**", whilst the independent variable is "**vegetarians in life**", which has three categories: "**non-vegetarian**", "**consuming both**" and "**only vegetarian**").

Now types of MANOVA, you can have one-way or again n-way. One-way MANOVA if the independent variable is only one, you have only one independent variable and dependent variable is more than one is a case of a one-way MANOVA. In the ANOVA what was happening, you had one factor then one-way ANOVA, if you had more than one factor we have 2 factors we say two-way ANOVA and 3 three-way. You can use MANOVA to understand whether there were differences in the perceptions of attractiveness and intelligence of people who are vegetarians.

That is 2 independent variables, perception of attractiveness and perception of intelligence, was measured taking independent variable as whether the person is vegetarian purely or purely non-vegetarian or a combination of both. So when I have such a kind of study, so what I am saying my attractiveness is measured as a continuous variable, intelligence is measured as a continuous variable, but the food the independent variable which is affecting this is consumed as measured as categorical 1, may be 2, may be 3. So when I have this case, it is a one way.

(Refer Slide Time: 27:55)

Types of MANOVA (continuous...)

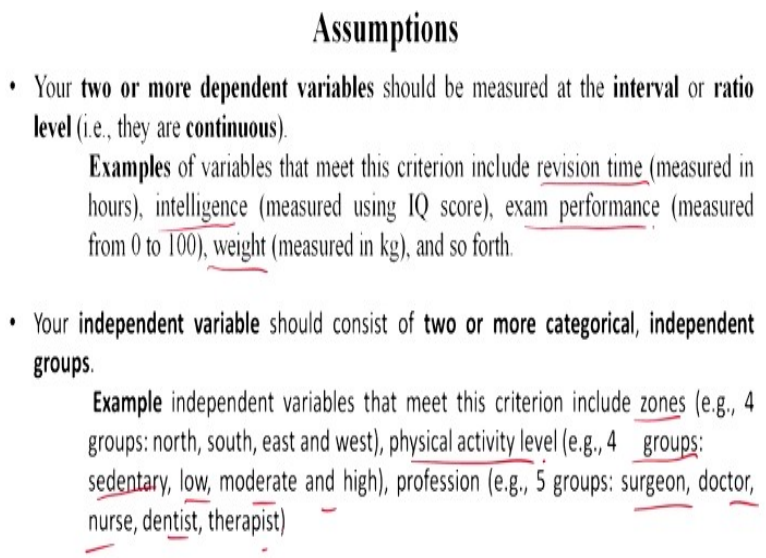
- **Two way MANOVA-** If the independent variables (categorical) are two, and dependent variable (continuous) is more than one. It is a case of two way MANOVA.
- **Example-** You can use MANOVA to understand whether there were differences in **students' short-term** and **long-term recall** of facts based on three different **lengths of lecture** (i.e., the two dependent variables are "**short-term memory recall**" and "**long-term memory recall**", whilst the independent variables are "**lecture duration**", which has three categories: "**30 minutes**", "**60 minutes**", **and** "**90 minutes**") and "**teacher's educational background**", which also has three categories: "**arts**", "**commerce**", and "**science**").

Similarly if the independent variable are 2, there are 2 independent variables and dependent variable is also more than 1, it is a two-way MANOVA case. Example you can use MANOVA to understand whether there were differences in students short-term and long-term recall of facts based on 3 different lengths of lecture. So there are 3 different types of lectures, so the two dependent variables are short-term memory recall and long-term memory recall.

What are my independent variables lecture duration which has 3 categories, 30 minutes, 60 minutes, and 90 minutes and the teachers' educational background which also has 3 categories arts, commerce, and science. Now you see you have 2 independent variables which are categorical in nature, one the duration, so these three, and now and the teachers' background arts, commerce, and science and what is my dependent variable now there again 2, short-term memory and long-term memory recall.

So when I have this case, how do you solve this problem? So to solve this problem, we use the multivariate analysis of variance.

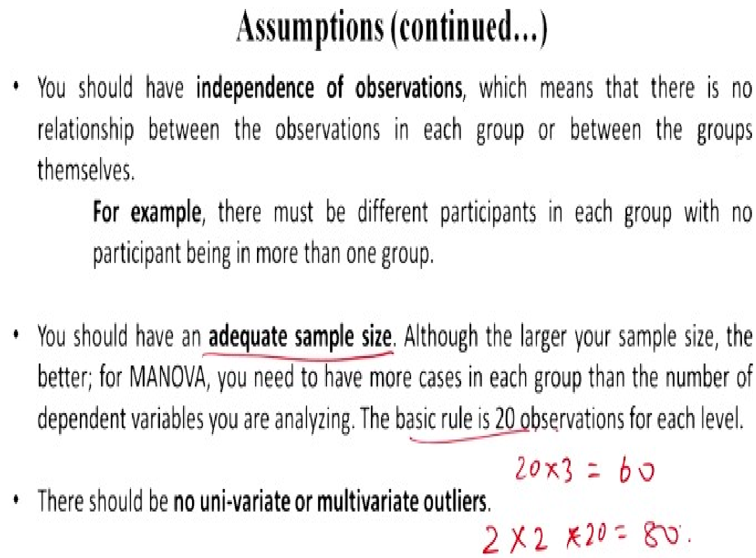
(Refer Slide Time: 29:12)



So these are some of the assumptions you see. Your 2 or more dependent variables should be measured at the interval or ratio level. What is it saying your dependent variable should be measured at interval or ratio level that is they are continuous. Examples of variables that meet this criterion include revision time, intelligence, exam performance, weight etc. Your independent variable should consist of 2 or more categorical independent groups.

What are the examples, the north zones in India for example, from which zone does the person come north zone, west zone, south zone, east zone. What physical activity does he have, so there are 4 groups; for example sedentary, low, moderate, and high. What is his profession, surgeon, doctor, nurse, dentist and therapist. So when I have this kind of a condition using my independent variables are categorical dependent variables are in a ratio.

(Refer Slide Time: 30:12)



You should have independence of observation which means that there is no relationship between the observations in each group or between the group themselves, that means they are all individually different. There must be different participants in each group with no participant being in more than one group and you should have an adequate sample size. So the larger your sample size, the better. You need to have more cases in each group than the number of dependent variables you are analyzing.

The basic rule is 20 observations for each level. So suppose you have a 1 factor with 3 levels, what should be your sample size, 20 x 3 = 60. Suppose you have 2 factors with 2 levels right each, so how many now 2 x 2 x 20, so how much is it coming for 20 x 2 are 40 x 2 are 80, so it goes on. So you can find out. So well I will just stop here. I will wind up here. We will continue in the next lecture.

Today we discussed, again we solved one more problem with analysis of variance or two-way analysis variance and we have just started with MANOVA and in the next lecture we will continue with MANOVA I will try to cover ANCOVA also, analysis of covariance which is another technique where a covariant is used and then we will proceed further. Thank you so much.