Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology- Roorkee

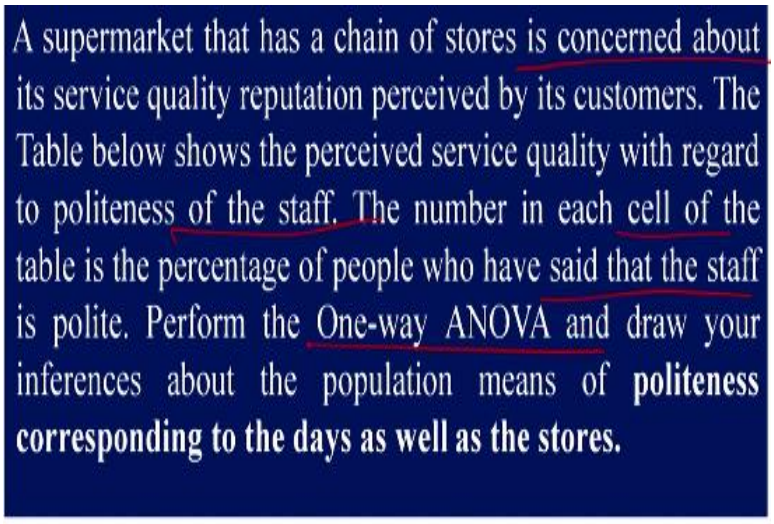
Lecture – 51
A Factorial Design – I

Hello friends, welcome you all in this session as you are aware in the previous class, we were discussing about randomized block design. This is the 2nd type of experimentation prior to this we have seen the completely randomized design and the 3rd one is factorial design which would be looking after some time. So, we will have couple of questions on randomised design as you are aware randomized design.

We have got done an independent variable with certain levels treatment levels you can have 2 levels 3 levels or 4 levels and we also have a blocking variable. Because the independent variable affects dependent variable but the dependent variable gets affected by other variables as well apart from independent variables. So, we need to control for those other variables we need to we can control those other variables by blocking variable.

So, blocking is nothing but method of controlling something called extraneous variables so let us look at this example.

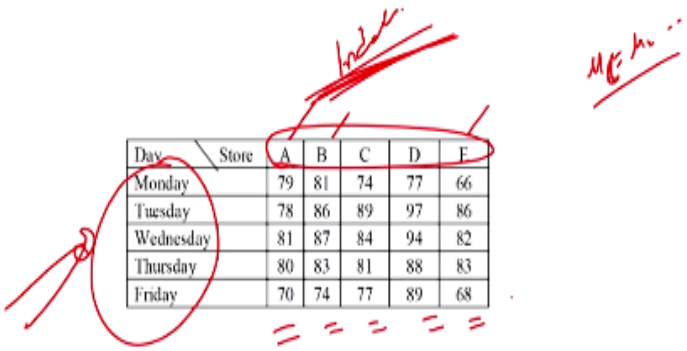
(Refer Slide Time: 01:56)



A supermarket that has a chain of stores is concerned about its service quality reputation perceived by its customers. The table next slide shows the perceived the service quality with regard to politeness of this stuff. So, the chain of stores is interested in knowing the politeness level of stuff and how that politeness level is being perceived by its customers. So, the number in each of the cells in the table is the percentages of people have said that the staff is polite.

We need to perform one way or ANOVA and draw our inferences about the population means of politeness corresponding to the days as well as these stores. Now we know that the politeness level of staff depends on several factors. It depends on let us say which day of the week the time of the week the ambiance of the and the store the qualification levels of staff the training level achieved by those staff members for which we are measuring politeness level. So, let us look at this data has been gathered.

(Refer Slide Time: 03:28)



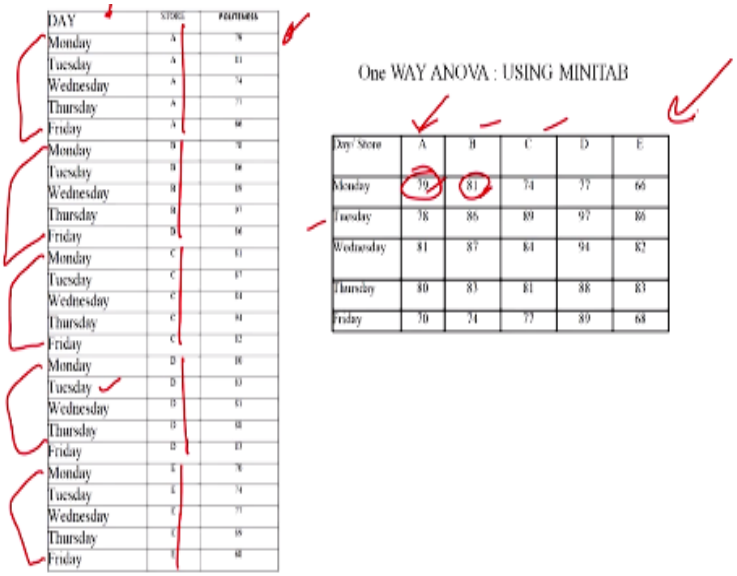
Day \ Store	A	B	C	D	E
Monday	79	81	74	77	66
Tuesday	78	86	89	97	86
Wednesday	81	87	84	94	82
Thursday	80	83	81	88	83
Friday	70	74	77	89	68

So, there are five stores A B C D E So what is the so these are let us say this is a independent variable having five different levels, right A B C D E and the blocking variables is lets the number of days or the days of the week because politeness level may vary depending upon week of the day. It is possible that on Monday the politeness level of employee in the store is higher than politeness level on let us say Friday.

So, we have done blocking of this particular blocking in this particular experiment so the blocking variable here is weekdays so we have to identify is there any significant difference in politeness level in these stores. So, initially we will say that the politeness level the same so mu1= mu2 or mu c you can call it c because we have to see the effect of these treatments on how customers are perceiving politeness level.

So, we have to see the effect of whether these means are same or different rates so these are 5 means we are not bothered about the means of rows right.

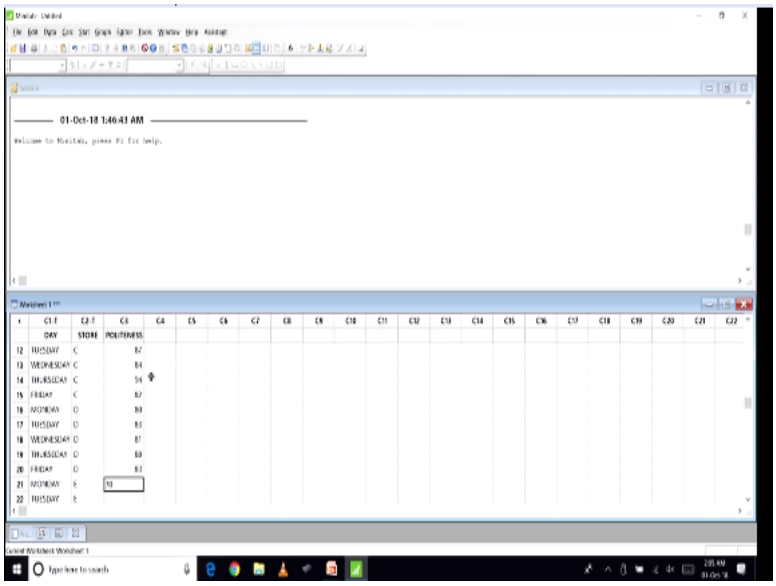
(Refer Slide Time: 05:13)



So, let us look at how to solve this question using mini tab. So, this is how you should be doing data entry so you got days over here so Monday to Friday again Monday to Friday and these are five blocks Monday to Friday and Monday to Friday. Okay because there are five stores so A B C D and E so what is the politeness level let say in store A Monday 79 let us see what it is Tuesday and store Tuesday store B.

So, you can write let us say for example store a Monday Tuesday Wednesday Monday similarly you can have let us say Monday store B so Monday store B 81 in fact, what I would suggest is this you can directly go for data entry and let us see the output from this. So, we will have data entry for this particular question so let us look at mini tab software.

(Refer Slide Time: 07:00)

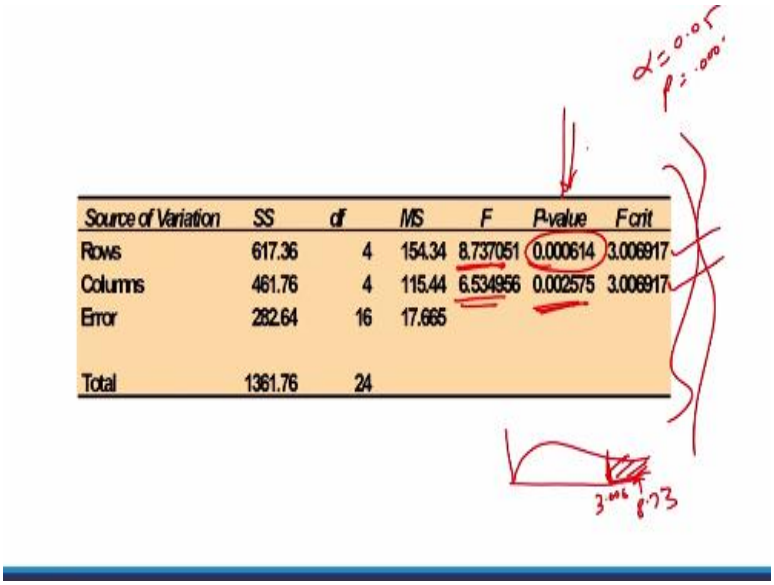


So, here you have got days and you have got store and then politeness level so politeness level is dependent variable which is getting affected by different types of stores. The employees of different stores so Monday Tuesday Wednesday Thursday and Friday night so you can now just copy all this you just keep pasting in other cells. So, once you are you just copy this just paste 10 15 20 25 so 5 days 5 stores so total 25 experiments.

So, store A B then up to 15 C then D up to 20 then E now you need to enter data for politeness level right Monday A 79, Tuesday A 81, Wednesday A. So, here we are writing politeness level of all these stores of all A stores on week days similarly politeness level in store B throughout the week Monday to Friday then for C then for store B Friday then for store E so once you are done with data entry.

So, we will not be for solving we will go back to our question and see the output of this particular question using meters. So this is the output.

(Refer Slide Time: 09:44)



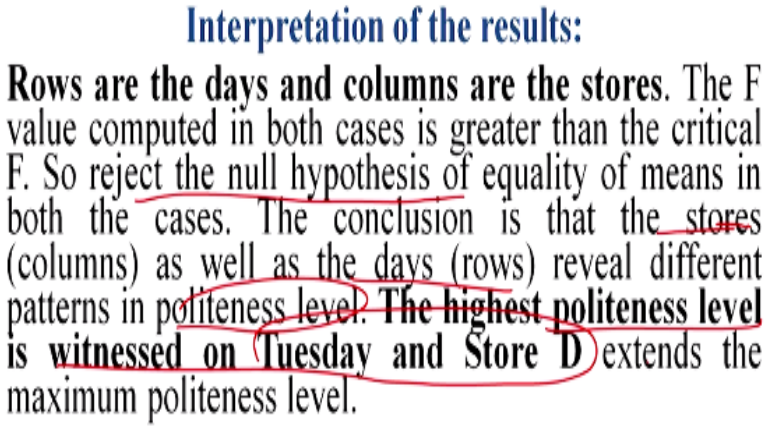
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	617.36	4	154.34	8.737051	0.000614	3.006917
Columns	461.76	4	115.44	6.534956	0.002575	3.006917
Error	282.64	16	17.665			
Total	1361.76	24				

Now what conclusion would you draw by looking at this table so in fact you need to look at the significance level else also in this question though it is not there but you can have a significance level of let us say 95%. Let us say alpha = 0.05 and p value here is .000 p values is < alpha, we will reject null hypothesis. It means and similarly for this as well so we will reject the null hypothesis and will say that the column means are different as well as rows means are different.

Okay in fact you can find out crit F critical values and these are calculated values so if you draw distribution like this so this is your critical value 3.006 and this is 8.73 so we will reject null hypothesis that the conclusion can be drawn from p value as well. Okay so what finally we will say politeness level is more or less right so we have that the politeness level is different amongst these 5 stores.

Politeness level of this staff which has been measured on 100 pounds scale and this measurement was done on customers so we asked customers what was the politeness level of staff of this store. So, if you look at the F value we are rejecting the null hypothesis.

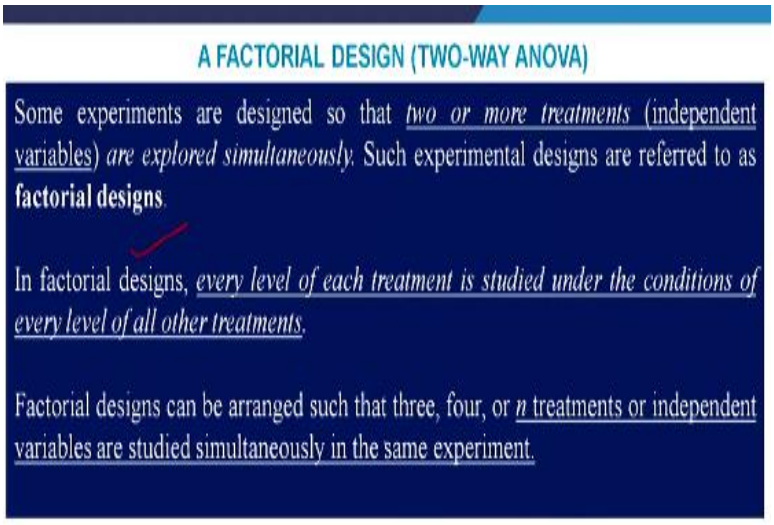
(Refer Slide Time: 11:53)



Of equality of means so conclusion is that the store the stores that is the column as well as the days and rows reveal different pattern in politeness level. So, it means the politeness level is different on the different days as well as different on different stores so the highest politeness level is witnessed on Tuesday and store D extends the maximum politeness level. We will say that the highest politeness level is on Thursday and store D.

How did we arrive at this conclusion? Just look at this look at all these values. The highest is this one so we will say that on Tuesday in store D the politeness level is highest so this is how you can work out an example of randomized block design so let us look at the 3rd type of experiment.

(Refer Slide Time: 12:59)

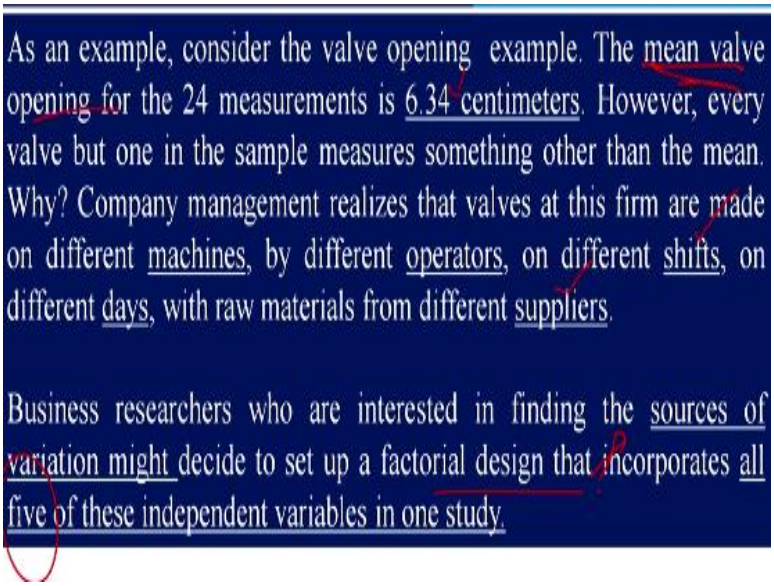


It is called factorial design so far we have seen one independent variable but some experiments are designed so that do are more treatments 2 or more independent variables are explored simultaneously. Such experimental designs are referred as factorial design so what we are trying to say here is whenever we see the effect of 2 independent variables on the dependent variable is called factorial design experiment.

So, what is factorial design? So every level of each treatment is studied under the conditions of every level of all other treatments. So, let us say if you ever two independent variables and each has got two levels so it would be called 2/2 design or 2/2. It is also known as full factorial design so factorial design can be arranged such that 3 4 or n number of independent variables can be studied simultaneously.

So there is the plus point of factorial design in completely randomized design we see the effect of independent variable on the dependent variable in completely in an in randomized design we see the effect of independent variable by having some blocking variable while in case of factorial design you have got more than t2 independent variables and we can see the effect of those independent variables simultaneously on dependent variable. So let us say we have worked out this example it is valve opening example.

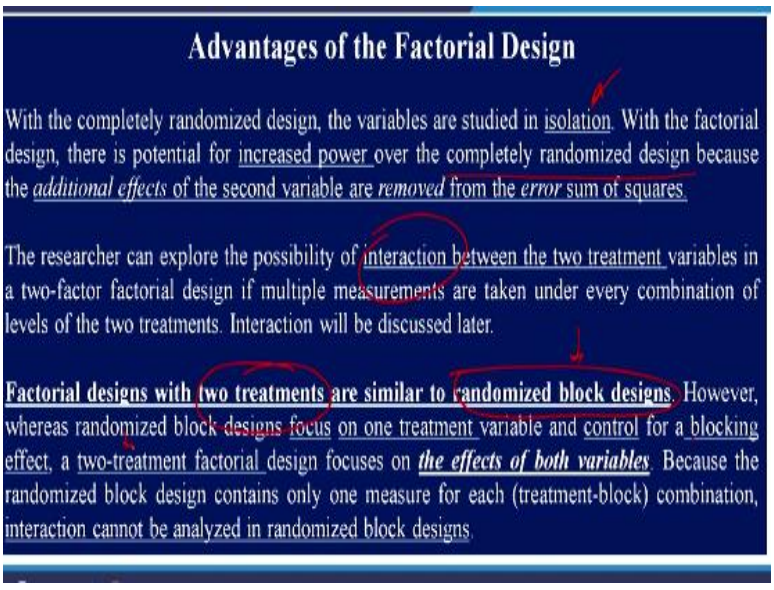
(Refer Slide Time: 14:48)



So, we have seen that the mean was this much so we know that the mean value of the valve depends on different types of other independent variables namely the machine you have got different types of machines different types of operators. You can shift day shift night shift and different days and you are getting raw material from different suppliers so let us say you want to find out the sources of variation in the and diameter of let us say this valve opening question.

And decide to set up a factorial design so that that will incorporate all these 5 independent variables which are those 5 machine operator shift days and suppliers. So, there are 5 day independent variables affecting the diameter of valve right.

(Refer Slide Time: 15:53)



So, why we should use factorial design and what is wrong with randomized design of course if they are two variables then you may use randomized design completely randomized design you would be requiring 2 setups. So, if 2 dependent variables are analyzed by using completely randomized design the effect of each variable are explored separately it takes 2 completely we randomized design.

To analyse the effects of these two independent variables so you will have to perform explain two times which you would be time consuming and which would be expensive as well. So, by using factorial design a researcher can analyse both variables at the same time and it saves time and efforts. Some business researchers are researchers use the factorial design as ways to control confounding or concomitant variables in this study.

So, rather than having blocking variable for controlling confounding and concomitant variables it is good to have those variables in the experiment itself. So, by building variables into the design the researcher attempts to control the effect of multiple variables in the experiment. So that is the benefit of factorial design so with completely randomized design the variables are studied in isolation.

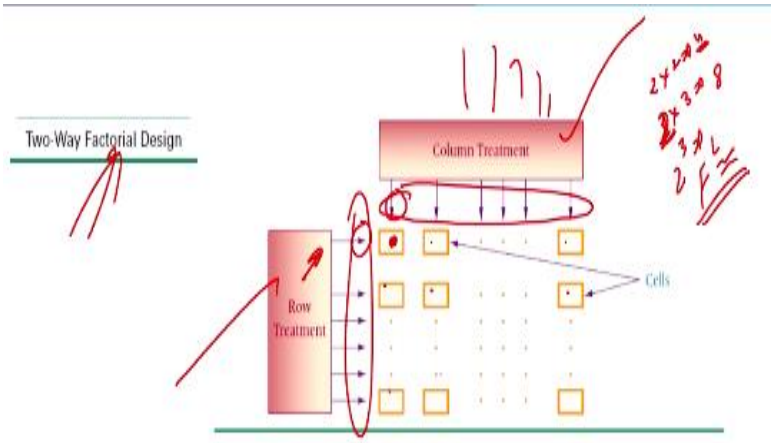
With factorial design there is potential for increased power over the completely randomized design. Why? Because the additional effect of the 2nd variable are removed from error terms so that is why factorial is then completely randomized design. The researcher can explore the possibility of interaction between two treatment variables this is really good one. This is one of the most important benefits you are getting when we use factorial design.

You can also know the effect the interaction effect between two independent variables on the dependent variable. So, factorial design with 2 treatments if there are 2 independent variables that would be similar to randomized block design however what happens in randomized block design we have got one independent variable and one blocking variable so the focus is more on independent variable.

So, factorial designs with 2 treatments are similar to a randomized block designs however in the randomized block design focus on treatment variable and we control for blocking effect but the 2 treatment factorial design means the two independent variables. Factorial design focus is on the effect of both variables so if there are let us say 2 independent variables you can have randomized block design 1st and then again one more set up of randomized design.

But in case of this one factorial design but focus on both of them simultaneously and more importantly as I have said interaction effect can also be analyzed. So, this is the set up for 2 way factorial design.

(Refer Slide Time: 20:31)



So, you have got column treatments this is one independent variable in 2nd independent variable and these are different levels of independent variable one. These are levels of independent variable 2 and these are the values of dependent variables right now you can help let us say independent variable one and treatment level one independent variable 2 treatment 1 so this is the output.

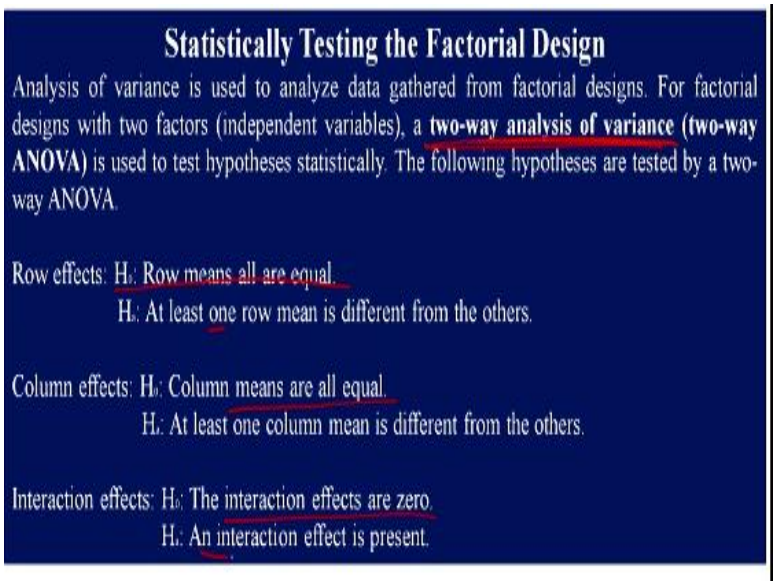
Now you can have output either just 1 one output or you can have more than 1 replicates. So, you can either take just one reading or let us say 3 readings if you take 3 readings you just average how those readings so the purpose of having more applicants is that you are removing

some noise if there is any so it is good to take more replicates. But the point is when you take more replicates the time to perform experiment would be high and cost would also be very high.

So, if you have got let us say 2*2 designs so you need to perform 4 experiments. Is not it? 2 factors and 2 levels so you will have 4 experiments but let us say if they are 3 factors and 3 levels then you will have all I said 2 factors and 3 levels so you will have 8 output right 2 to the power 3 so it is vector to the power level these many numbers of experiments you will have to perform. Is not? So as you increase number of factors number of levels.

You will be being lots of setup right for me experiment and in that case you just go for one replicate one observation. But if it is let us say 2*2, 2*3 and it also depends on how much time it takes to take reading on 1 setup is not it? So this is the setup of two way factorial design.

(Refer Slide Time: 22:58)

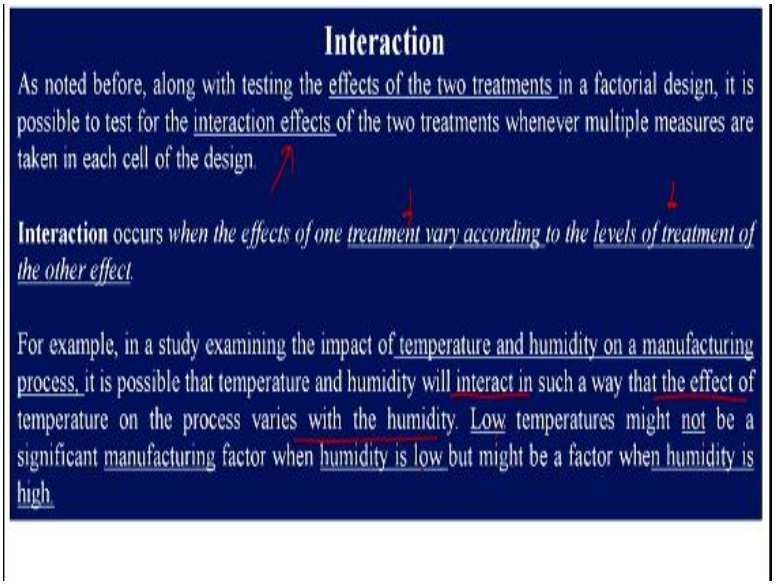


So, how do we proceed in a factorial design so as we have already said that the two factorial in factorial design two factors you will have let us say you have two levels then you total experiment would be 4. So, you can use two way ANOVA instead of one way ANOVA why we are calling it two way ANOVA because they are two independent variables otherwise in block design in randomized design we have just said it one ANOVA.

Because there was one independent variable the other was blocking variable so null hypotheses row means are equal alternative is not equal at least one of the mean is different from the others column effect column means are equal otherwise the alternative hypotheses one of the column mean is different from others and interaction effect so the interaction effects are zero we will see what is the meaning of interaction effect.

And the alternative hypotheses interaction effect is present so either interaction effect is present or it is not present.

(Refer Slide Time: 22:58)



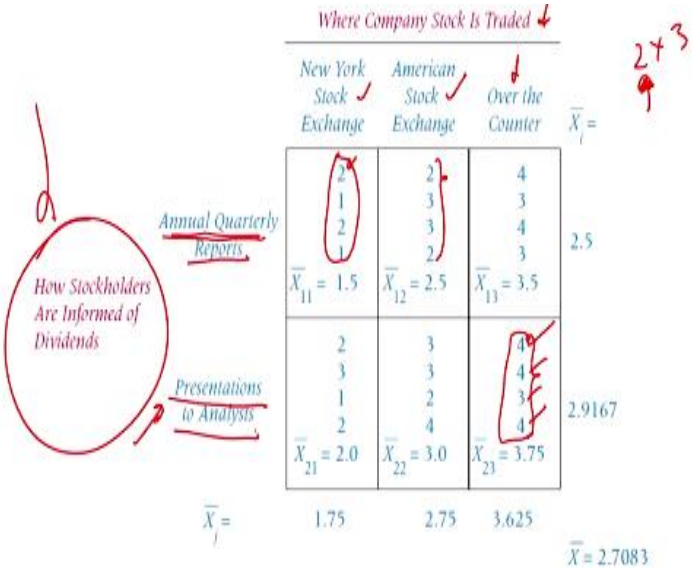
So, let us look at interaction effect what is interaction effect? So, a factorial design we find out the effect of two independent variables on the dependent variable and in factorial design we can also study interaction effect so interaction occurs when the effect of 1 treatment vary according to the level of treatment of other effect so let us take example let us say the manufacturing efficiency gets affected by temperature and humidity.

So, let us say temperature can be set up at two levels high level and low level similarly humidity high level and low level. So, we want to know how these two independent variables at their levels affect manufacturing now it is possible the temperature and humidity will interact in such a way that the effect of the temperature on process varies with the humidity so low temperature might not be might not affect manufacturing when humidity is low.

So, there is a possibility that when temperature is low the manufacturing efficiency is not effected much when humidity is low but the moment the humidity is high and temperature is low the manufacturing efficiency may get effected. So, how the level of one independent variable is changing the output when we change the level of other independent variable so that this interaction effect.

So, let us look at this example we want to know how the stock's value change with respect to whether they are being traded.

(Refer Slide Time: 26:59)



Which stock in which stock exchange and how stockholders are informed about the dividend or how the companies are informing their customers about the dividends so the stock can be treated let us traded in any of these ways so New York Stock Exchange American stock exchange over the counter and how the executives are informing the shareholders about the dividend whether it is through annual or quarterly reports are through making presentations.

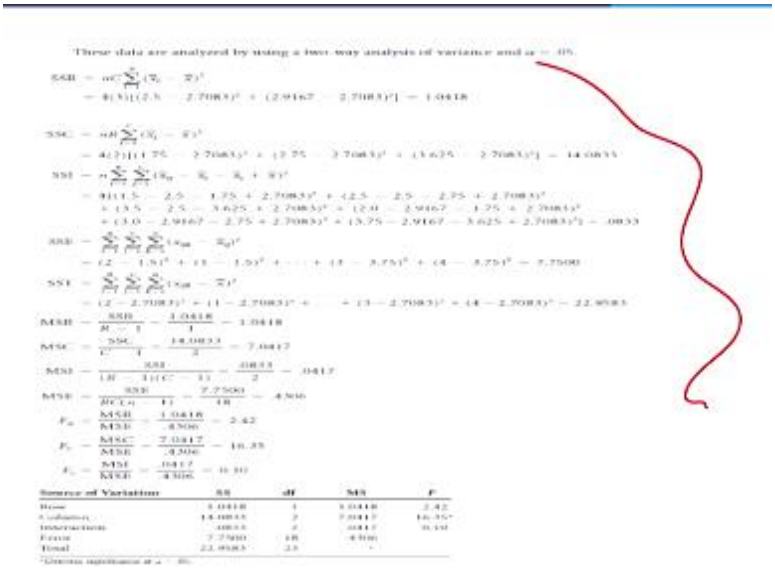
So, there is the question so this is the case of 2*3 design this is one independent variable having 2 levels the 2nd independent variable having 3 levels. So, this is the 1st reading and let us says the opening value of stock in New York Stock Exchange when the shareholder was informed

using annual report or a quarterly report so there was there were total four readings taken so we will say that there were four replicates.

Okay similarly 2nd stock exchange similarly if you look at these 4 so over the counter the values are the let us say opening value of a stock 4 4 3 4 so 4 readings were taken and the level here is presentation was made and to the shareholders. So, we want to find out is there any significant difference between how these share prices change with these different you know the stock exchanges and how they are being how they given.

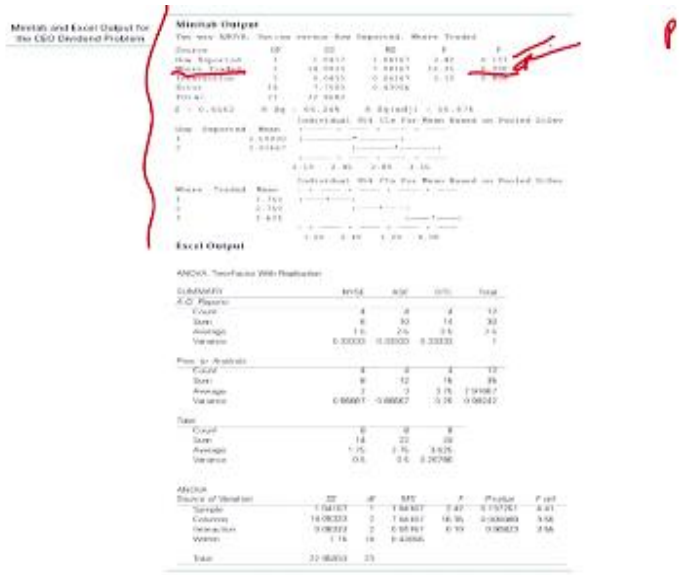
How shareholders are being informed about the dividend right so when you solve this question using.

(Refer Slide Time: 29:39)



On this particular formula over here so this just look at the F value over here right, so you need to have the p value as well for this particular question.

(Refer Slide Time: 29:52)



Let us look at this so this is the minitab output so p value here is 0.137 but for where it is traded so p value < reject null hypothesis. So, there is a significant difference in the value of the stock as far as the stock exchanges are concerned. It is not the same but if you look at how this divided are reported they are one and the same thing. So, how share holders are being informed about the dividends by the CEO of the company has no effect on the value of the stock.

But where it is being treated so there is a significant difference amongst those 3 places so this the excel output is one and the same thing. So, in fact you can see once again here so is there any interaction effect? In fact, there is no interaction effect because p value is 0.90 right at alpha = 0.05. So, with this let me stop here in next class we will work out these examples using two way ANOVA. Thank you very much.