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

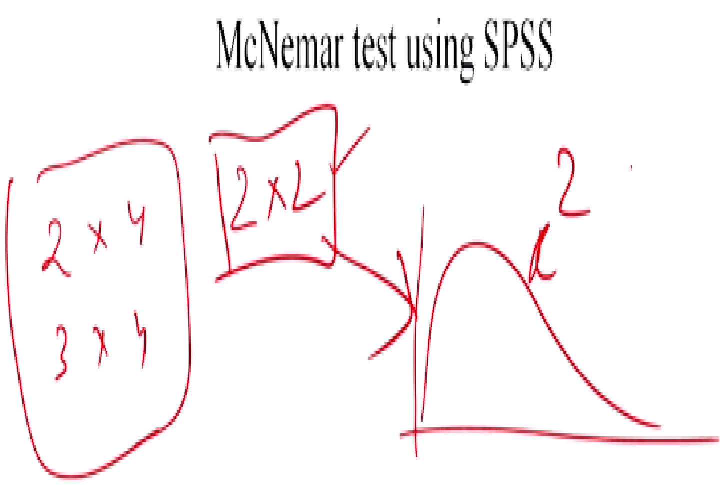
Lecture – 31
Non-Parametric Test - VII

Welcome everyone to the lecture series of the course Marketing Research and Analysis. We have been dealing with the non-parametric tests since the last 2 to 3 lectures. We are trying to understand what non-parametric test means all about and why they are used and where they are used, what are the conditions in which parametric test does not work and in place of that we tend to use the non-parametric tests. We have conducted several non-parametric tests including for example the Mann Whitney U test, the Kruskal-Wallis test.

Then we did with the signs test, the Wilcoxon signed-rank test, then we did it with the Spearman correlation test which is the similar test with Pearson correlation, so it is a test of association like the correlation test you have for the Pearson correlation, but this is for the non-parametric. So continuing the same, even we have talked about the chi-square test and McNemar test. So these 5-6 tests we have already discussed and may be today we will wind up this lecture on non-parametric, may be 1 or 2 more techniques we will cover up.

I will show you how to conduct those tests on even the SPSS and then we will wind up this lecture on non-parametric. So well we were talking about McNemar.

(Refer Slide Time: 01:50)



So what is this McNemar test? Let us go back and think about what we have discussed in the last class. In the last lecture, we discussed that McNemar is a similar test to the chi-square test which follows a chi-square distribution. You remember chi-square distribution, so a chi-square distribution is more of a right skewed distribution and the McNemar follows a follows chi-square distribution, but of course, the problem is then what is the difference between the chi-square and the McNemar test.

The difference is that a chi-square test can be done between more than 2 variables, let us say, and the variables will be nominal obviously. So let us say we can conduct a 2 x 4, a 3 x 4, so these kind of studies can be done in a chi-square test, but when you do a McNemar test, it restricted to only a 2 x 2 contingency table. So this is the limitation of a McNemar test, otherwise the distribution that it follows is similar to the chi-square distribution. So how do you do this McNemar test using SPSS, so let me explain that.

So what I will do is I will show you one of those slides which I have prepared. So McNemar test, so this is the data. So the data is between two things. **(Video Starts: 03:11)** So what is it saying, let us see. So pre-situation and post-situation as you can understand. So what is this pre-situation and post-situation? It is a test where people who are smokers and who are nonsmokers they were taken and these people were shown treatment like an advertisement which talks about the ill effects of smoking and then they were checked what is the impact of this advertisement on the people’s mind.

So did their smoking habits change or did it remain the same. So in this case, we have taken 0 as nonsmokers so who people who do not smoke are taken as 0 and the people who smoke are taken as 1. Taking this so we have created the data how does it look like pre that means before the advertisement what was the condition, so let us see the data. So in the pre-situation, you can see so 1 were smokers, so these were all smokers, smokers, smokers, smokers, these are nonsmokers, so as you go down so these are nonsmokers, 0, 0, 0, 0 all are nonsmokers.

After the advertisement was shown to them or the treatment was given to them, he is a smoker and continues to be a smoker, smoker again smoker, but somewhere here you see it is a smoker, after the advertisement or after the treatment, he or she has been converted to a nonsmoker. So similarly you can see so 1, 0, so again conversion is there, here there is a conversion, here there is a change in mind, here again.

So here they were already nonsmokers, so let us see has there been any change with the nonsmokers, so that are nonsmokers became smokers something like that, no I do not think anything has been there. So there has been no absolutely change with the nonsmokers, they have been nonsmokers, have stayed as nonsmokers as before. So now to conduct a McNemar test what we will do is, how do we test this? So let us go to analyze. Here there are two ways of doing it.

You can do with a non-parametric test going into the parametric test, but then I will show you another way which is more simpler. Now go to the crosstabs, so take any one of them as the pre-situation as may be in the row and post-situation in the column. Now what do you want? We want the McNemar test. Now if you can see here the McNemar test is there. We take this go with continue and we know that in a in a-chi square test the minimum cell size has to be 5, less than 5 it does not work. So now let us see what is the output.

So if you go to the output, we will see the pre-situation versus post-situation, there are 40 sample respondents and there is no missing value. Pre-situation was that nonsmokers who remained nonsmokers were 20, nonsmokers who became smokers 0, so no nonsmoker has become a smoker. Smokers who became nonsmokers are 5 and smokers who remained smokers is 15. So is this data telling us some significant change, at the moment if you see from the McNemar test here the value, now we can see that it is coming 0.063.

Let us assume our significance level that we have taken is, usually we take it at a 5% significance level. So if you take a 5% significance level, what do you infer form here. You infer from here that the null hypothesis is not to be rejected because this p value is more than 0.05. Now let us make small changes in the dataset. Let us go back to the dataset and make small changes. So let us make some changes and see in the dataset if our results are varying now. So let us go back to the dataset again. So what I will do is I will change some of the data.

For example, let me change this as 1. Let us assume he was also a smoker earlier right and you can also convert some of these as smokers let us say, but then we do not have to have the adverse effect, we do not want to see that, let us assume still if you want we can. So some nonsmokers have become smokers I mean to say that, so this is another let us say smoker. So let us take this much and check has there been any change in our result. Now if you go by again this and you run the test keeping everything as same, now you see because you converted 2, 2 smokers, so nonsmokers who became smokers is 2 now.

So it is a very dangerous effect which I had shown and smokers who were earlier smokers and now they are nonsmokers is 8 and smokers who remained smokers are 15. So it has become more poor, the significance value has become more poor. So instead of this if you do one more thing, instead of this, we will not make it we will keep it 0 as it is. So let us make it 0, 0, and this is also 0, we do not want. So this is 1, this is 1, 0, so this has become some of them more let us say take one more and we make this also 0, so I think okay it is done.

So now let us run it, let us see has there any change. So if you go by this and you check, so now if you see what I have done by only changing two things, I have this is just an experimentation, you can try several things while you do. So now just what I have done you see, the nonsmokers, 0 are nonsmokers 1 is smokers, so nonsmokers who remained nonsmokers after the advertisement is 17, nonsmokers who became smokers 0. Smokers who became nonsmokers now we increased from earlier, I think it was 8 or something, we made it to 10 and remaining smokers also has gone down **(Video Ends: 09:47)**.

As a result, you see the McNemar test says that there is a significant change in the people's smoking habits after they watched the advertisement or went through the treatment. So this is what you have to write when you write the result of the test, this is a significant test and the null hypothesis is rejected that the advertisement has got no effect on the smoking habits, rather it has got a smoking impact on the smoking habit of people right, a positive impact. So this is one of the things that we were doing. Now let us go back to the ppt okay.

(Refer Slide Time: 10:24)

•If the statistical significance level (i.e., *p*-value) is less than .05 (i.e., $p < .05$), you have a statistically significant result and the proportion of non-smokers before and after the intervention is statistically significantly different.

•Alternatively, if $p > .05$, you do not have a statistically significant result and the proportion of non-smokers before and after the intervention is not statistically significantly different (i.e., the proportion of non-smokers does not change over the course of the intervention).

•In our example, $p = .027$ (using the exact *p*-value), which means that the proportion of non-smokers is statistically significantly different after the intervention as compared to before.

•Put another way, the change in the proportion of non-smokers following the intervention was statistically significant.

So we have covered this. In our example p is equal to the significance level is 0.027, this means that the proportion of nonsmokers is statistically significantly different after the intervention which I am saying as the advertisement as compared to before or put in another way, the change in the proportion of nonsmokers following intervention was statistically significant. So this is how you write, so that means what, that at 0.05 level it was significant and you got it was significant at 0.027. So this is how you write it in the research reports and all.

Today we will do one more test, but before I do one more test if you remember we had done one test which I said I would do it later on the analysis for it. So that was the Wilcoxon test, you remember the Wilcoxon test. So Wilcoxon signed-rank test, this was the test which is basically like a paired sample t test. If you remember by any chance, a paired sample t test is a test where a before and after situation happens. So what happens is how does a person react before and let us say after some treatment.

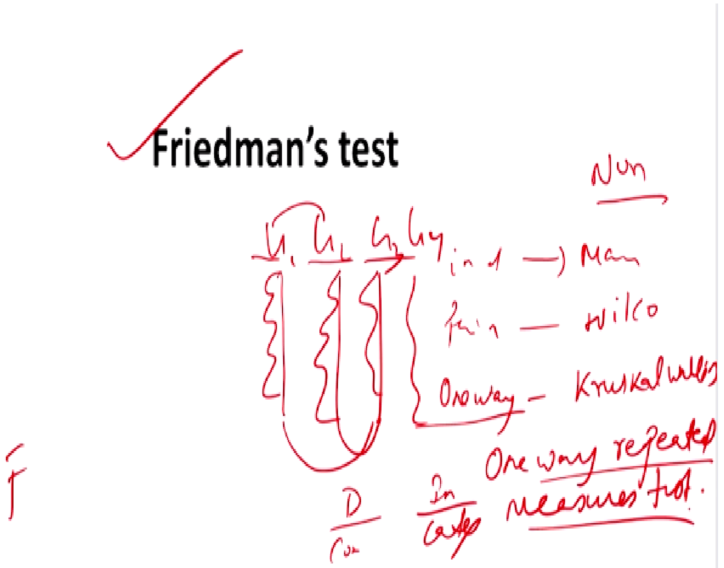
Let us say you have been given some medicine or some kind of a method of teaching, some kind of a method of let us say some experimentation has been done and what is the effect of that medicine as we did earlier also. So we will see **(Video Starts: 11:59)** what is the change in this case? So Wilcoxon signed-rank test, this is a test in which the pain of people was measured, same people.

Now why I am saying it is like paired sample t test because the sample remains the same, so the sample there is a group of sample of let us say 25 people and their pain score was recorded, before some medicine was given, the pain score was recorded. So after recording, can we say that because of this medicine there has been a change in the pain among the patients, so this is the test. So Wilcoxon signed-rank test is nothing but a non-parametric test for the paired sample t test that you were doing earlier in the t test. So how do you do this.

So you can see Wilcoxon signed-rank test. There are two related samples. So let us go to week 1, so this is one, week 4, so this is the one. So we take we want the descriptive, we continue and okay. So let us go to the result and check the result. So what is the result saying. Now as per the result if you see, if you look at the test statistics, the z value is -3.706 something and if you look at the significance it says there is a significant change in the pain of people in comparison to the week 4 and the week 1.

So that means the medicine has worked wonders. So week 1 pain score mean was 6.8, week 4 pain score was 4.8, so that means the pain has significantly reduced and this is proven from here. So this is another test which is very very important and I have said the Wilcoxon signed-rank test is a more powerful test in comparison to the sign test right, which we have already seen earlier. So these are the 2 tests which I was talking about.

(Refer Slide Time: 11:05)



Today we will move into one more test the Friedman's test. So what is this Friedman's test. This Friedman's test is nothing but another test which is similar to the analysis of variance test. Whenever you have two groups, so you use an independent sample t test or a paired sample t test. So on the non-parametric you have for that the Mann Whitney or you have the

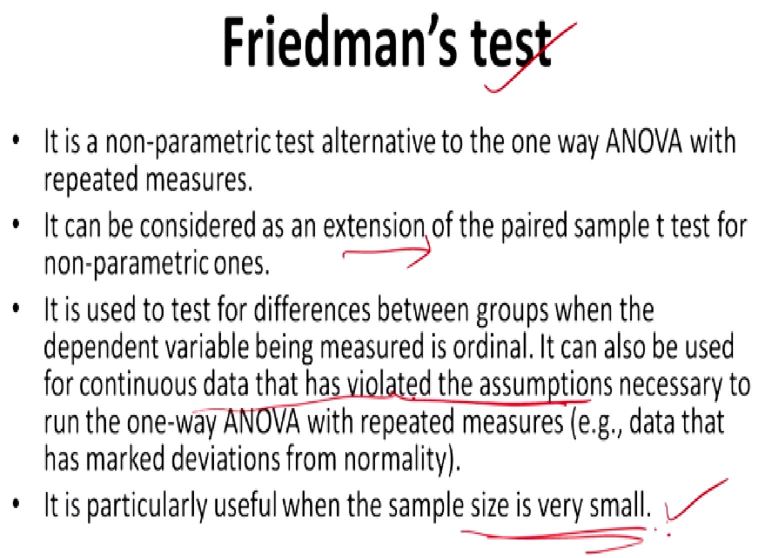
Wilcoxon for example, but if you have more than let us say 2 groups, let us say you have 3 groups, group 1, group 2, group 3.

Now there if you do an independent sample t test that means between group 1 and group 2, group 1 and group 3, group 2 and group 3, so there are 3 tests and if you do 3 tests, we have learned earlier also that the type I error increases and the inflation the alpha gets inflated. So to avoid that, you do a one-way ANOVA correct. So to have a one-way ANOVA in the non-parametric, you have a similar test called Kruskal-Wallis which already we have discussed, but what in the one-way ANOVA, the sample is repeated again and again.

That means in a test in the one-way ANOVA you have the dependent variable as some kind of a continuous variable let us say and the independent variable as a categorical variable, but suppose we have a repeated measure that means the same sample is getting repeated more than 2 times right you are checking for the same sample, the change in the value of the sample 3 times, 4 times, 5 times, so then each of these groups are having the same people but only the experimental values are being recorded, may be 3 times or 4 times if it is let us say something like this.

So in such a condition you use a technique called, in the parametric case you use the one-way repeated measure test, but here similarly in the non-parametric case you have a test called the Friedman's test or this test which is similar to this one-way repeated measures test, ANOVA this is right. So let us see what it is saying.

(Refer Slide Time: 16:25)

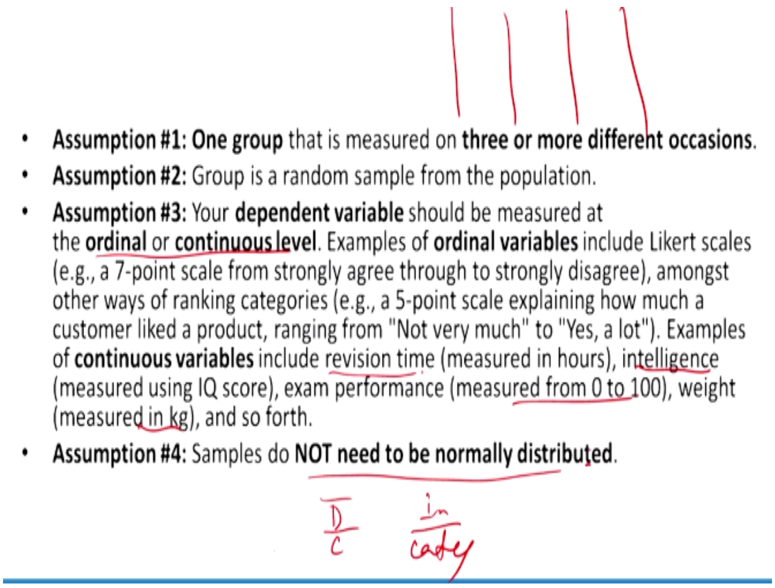


It is a non-parametric test alternative to the one-way ANOVA with repeated measures that was I said just now, so it is being repeated, the same sample is getting repeated. It can be considered as an extension, yeah this is true, if you have a paired sample t test, what was happening in the paired sample t test, in the paired sample t test, the sample respondents were being checked 2 times, may be once before and once after, but here it is an extension of that. That means it is may be not 2 times, but 3 times 4 times or more than that.

It is used to test for differences between groups when the dependent variable being measured is ordinal. It can also be used for continuous data that has violated the assumptions, so it is a powerful test so it can also work for continuous data that has violated the assumptions, but the question is if it is continuous data, generally one will go for a one-way repeated measure ANOVA, but then the condition here it is saying it has violated the assumption of the normality, so that is why one-way ANOVA will not work, repeated measure ANOVA will not work.

So in that case, we use this Friedman's test okay. It is particularly useful when the sample size is very small, now that is very interesting. So if you have a one-way ANOVA, you do some kind of a parametric test. If the sample size too small, then it does not work well, but this test has a capacity to run smoothly, to run function well, even when the sample size is small.

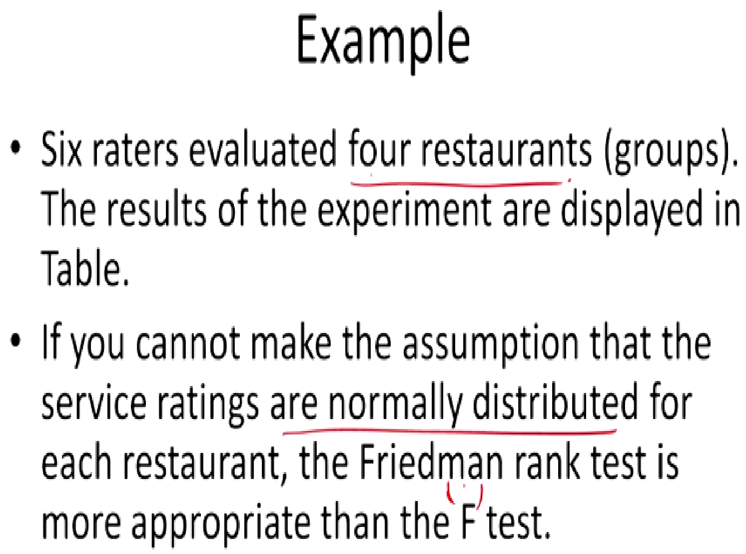
(Refer Slide Time: 18:00)



So let us see a test, I will show you. There are assumptions first, let us discuss. One, one group that is measured on 3 or more different occasions. I said one group, that is the sample respondents are the same are measured once, twice, thrice, may be four times. The group is a random sample, so there is no bias of selection. Your dependent variable should be measured at ordinal or continuous level. So in ANOVA if you remember I had said, just now also I said, in the ANOVA what was happening, the dependent variable was continuous, so it was continuous and independent variable was categorical.

So examples of ordinal variables include Likert scales, amongst other ways of ranking categories a 5-point scale explaining how much a customer liked a product, so these are some of the examples. Continuous variables include revision time, how much time measured in hours, intelligence may be in IQ, exam performance measured from may be scale of 0 to 100 what is your percentage in the exam, weight measured in kg, etc. So if the dependent variable is in this kind of a data, no issues, it is okay. Samples do not need to be normally distributed, fantastic, obviously, otherwise what is the use of a non-parametric test.

(Refer Slide Time: 19:19)



Now let us take this example. Six raters, raters means there are 6 people who are experts, they are they are evaluating 4 restaurants okay. The results of the experiment are displayed in table, next table I will show you. So what are these people doing. They are rating the restaurants and trying to see which restaurant is better than the other. If you cannot make the assumption that the service ratings are normally distributed for each restaurant, then the Friedman test is appropriate.

However if it would have been normally distributed and the sample size would have been larger a bit, then you would have gone for the F test right, but this is not the case, our sample, our ratings whatever they have given actually are non-normal okay.

(Refer Slide Time: 20:13)

- The null hypothesis is that the median service ratings for the four restaurants are equal.
- The alternative hypothesis is that at least one of the restaurants differs from at least one of the others:
- H0: M.1 = M.2 = M.3 = M.4
- H1: Not all the medians are equal.

The null hypothesis in this case, you should actually kindly take a pause and start writing the null and alternate here, anyway, so null hypothesis is that the median service ratings for the four restaurants are same or equal. So that means median M1 = M2 = M3 = M4, so all the median values are same, so M1, M2, M3, M4 right. What is my alternate, at least not all the medians are equal, at least one of them is unequal or not equal. So may be median 1 and median 2 are same, median 2 and median 3 are same, but median 3 and median 4 may not be same or any one of these.

(Refer Slide Time: 20:53)

Runner	A	B	C	D
1	70	61	82	74
2	77	75	88	76
3	76	67	90	80
4	80	63	96	76
5	84	66	92	84
6	78	68	98	86

So let us take these, so there are 6 raters as I had said, 4 restaurants. Restaurant A, B, C, D. So the first rater gave a rating of 70 to A, 61 to B, it is between 0 to 100, 82 to C, and 74 to D. Second rater gave 77, 75, 88, 76 right. Third rater gave 76, higher the score better is the performance, 67, 90, 80. So the third rater has given a higher score of 90 and 80 also. The fourth one gives A 80, C 96, and others are 63 and 76. Similarly these are the values. For fifth one, 92 has been given for C and sixth one 98 has been given for C and 86 for let say D.

So now this is a small table you can even see, but when you have to test it and say whether it is a significant different or not, for that statistical test may have to be conducted. Now what we have done is we have tried to make a rank.

(Refer Slide Time: 22:09)

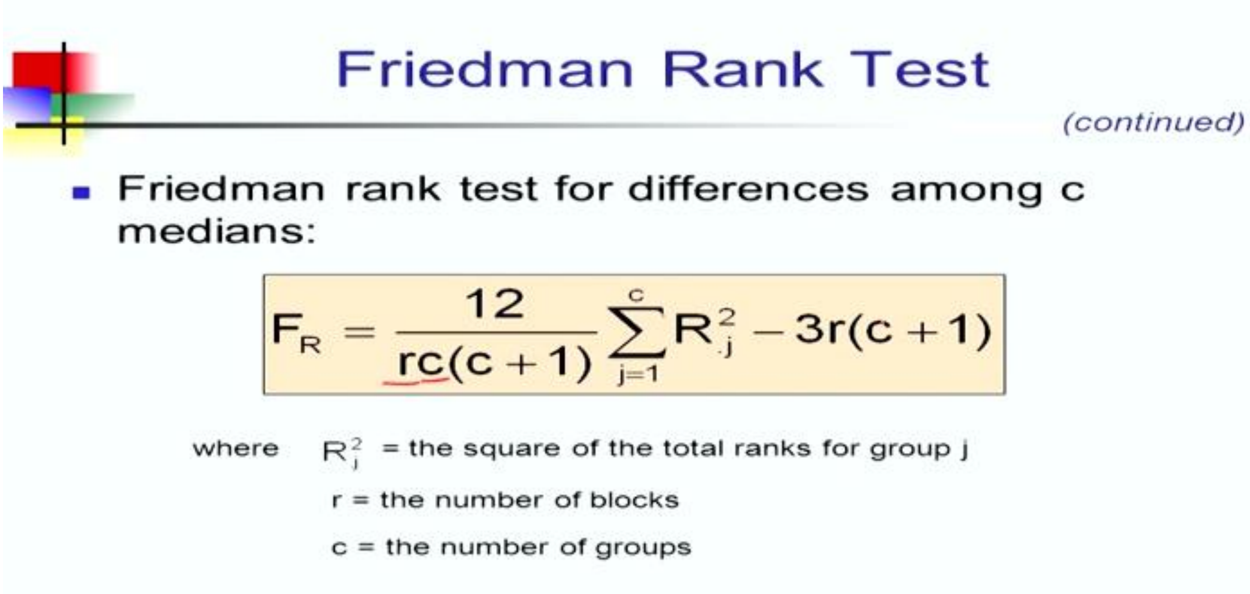
Runner	A	B	C	D
1	2	1	4	3
2	3	1	4	2
3	2	1	4	3
4	3	1	4	2
5	2.5	1	4	2.5
6	2	1	4	3
R	14.5	6	24	15.5

Now let us see what is the rank? Now in this first row, just look at the first row, which is the lowest value, the lowest value is 61. Now so 61 gets a 1, rank 1. Rank 2, 70. Rank 3, 74. So is it okay, so 1, 2, 3, 4, so now we will see here. So which is the first rank among these 4, 75 first, then second, then three, then 4, so 3, 1, 4, 2, is it there, 3, 1, 4, 2 correct. Suppose sometimes just remember even I have told you in some other tests if there is a suppose a same tie is there in between two, let us say this would also have been 75 let us say, so in this case what would have happened.

Now one and let us say this would have been let us say what is the rank, 75 is 1, so this would have become 2, so 1 and 2, that means now the rank would have 1.5 and 1.5, why, now 1+2, because since this is 76 more than 75, this would have become 3 automatically. So 1+2/2 so that is equal to 1.5 each. If it is tie in that case, you should have done that change okay. So other things also you can do for similarly 1, 2, 3, 4 right, so 1, 2, 3, 4, goes on. So we have created.

In this table there is similarity for this value, you see this is 1, this is 2, this is 3, and this is 4. So now this value rank when you come to the rank what it should become, 2+3, 5/2, so each gets a 2.5, 2.5, let us see, 2.5, 2.5 right, so you have done it. So now this is the total of the ranks, 14.5 for A, 6 for B, 24 for C, 15.5 for D.

(Refer Slide Time: 24:17)



Now this is the formula. I have got this formula because it is very tough to write these things, so I have copied this formula, downloaded it. So Friedman's the formula is

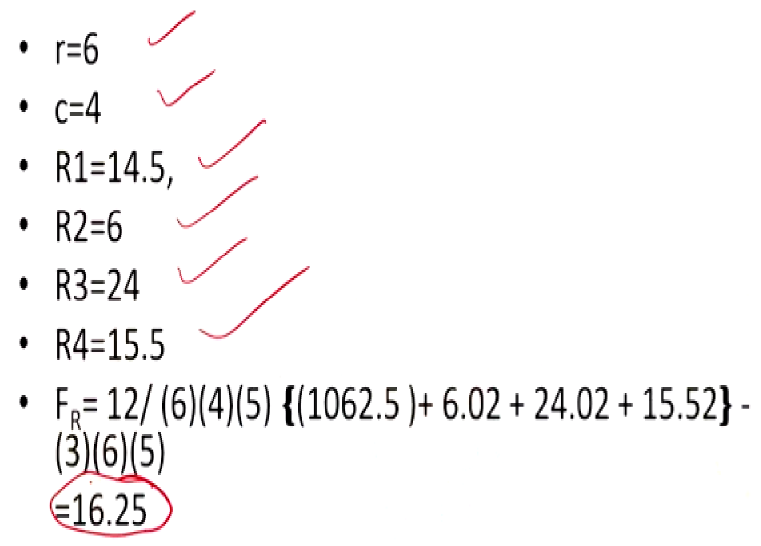
$$F_R = \frac{12}{rc(c+1)} \sum_{j=1}^{c} R_{.j}^2 - 3r(c+1)$$

Where, R_j^2 = the square of the total ranks for group j

r = the number of blocks

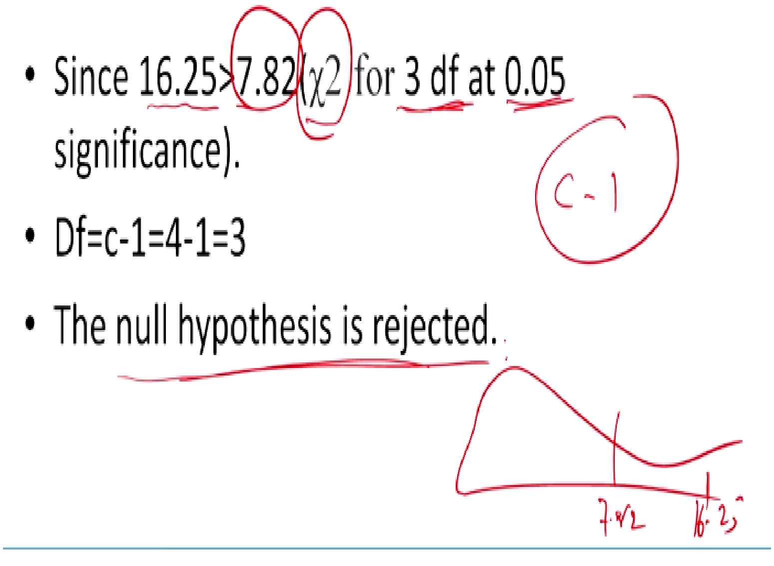
c = the number of groups

(Refer Slide Time: 24:49)



Now let us use, r = 6 in our case, 6 raters are there, 4 restaurants are there, R1 was 14.5, 6, 24, 15.5. Now putting into the formula you see if you do this then the score that comes is 16.25. Now 16.25 is the Friedman's test value.

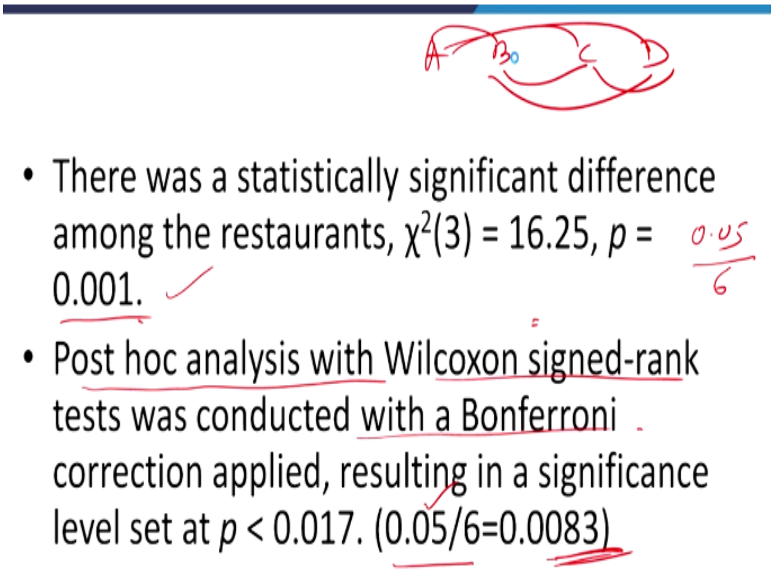
(Refer Slide Time: 25:16)



Now since you got a 16.25, but what is the table value you have to compare, the table value you have to look at from the chi-square table, you have to compare with the chi-square table. So chi-square table for 3 degree of freedom, why 3 because the degree of freedom in the Friedman's test is measured as c-1, so c was what is the c, you see the c is the number of groups you have, so were 4 groups, so it is like a within, so c-1 so you have 3. Now if you look at the chi-square table for 3 degree of freedom at 0.05 significance level, the value is 7.8.

Now since 16.25 is greater than 7.82, what is your assumption, what inference will you draw. The null hypothesis is rejected, why, because if you look at the table value is 7.82 and the calculated value is somewhere 16, so if this is, the null hypothesis is rejected. So what is the null hypothesis, the null hypothesis said that the median scores of the all the restaurants are the same, so that is rejected, so that means there is a difference between the median score of the restaurant. Now how do you write this, you see.

(Refer Slide Time: 26:38)



There was a statistically significant difference among the restaurants, so chi-square at 3 degree of freedom is equal to 16.25 and p = 0.001. The same test I will show you on the SPSS also and we will compare. So that is how we got this 0.001. Now what you can do is suppose you want to see whether this relationship has significant differences between the each restaurant is there or not. For example so A, B, C, D were there.

Now okay there is a difference, null hypothesis rejected, but what is the relationship between A and B, A and C, A and D, B and C, B and D, and C and D so you want to check it for example, you can do it. So to do that what you have to do is you do just, I think I have written yes. You have to do a post hoc analysis with Wilcoxon signed-rank test because I said Wilcoxon signed-rank is a like a paired t test, so you can do because the sample is same here. So 6 raters, the raters are same, so you can do a study between A and B, A and C, A and D, and goes on, 6 studies but remember this, this is important.

But with a Bonferroni correction applied, now what is the Bonferroni correction it says, now suppose you are taking a significance level of 5%, now you are conducting 6 tests, so you have to divide it by the number of tests. So much is remaining, now for example I have done, 0.05/6 = so the new significance level is not 5% but is 0.0083, that means it is 0.8, so this is the new significance, it is less than 1% and you have to check at this less than 1% significance level.

If your study gets significant by comparing with this, then you can say what is the relationship between may be A and B, A and C, A and D, individually you can check and see. So I will show you how to do this. **(Video Starts: 28:48)** So this is the same case 1, 2,3, 4, 5, 6. So the 4 restaurants are there, The scores have been given and let us now do it. So go to analyze, go to nonparametric test, now what will you do? This is a multiple groups, same sample but multiple times it has been taken, 4 restaurants, so 4 restaurants same people.

So which is the one I should be using, if you use your logic here, so I am using K Related samples, they are not independent, they are related, same people, so I am taking all the 4. So I have taken, now Friedman's test is the normal choice when you take several related sample the normal choice is just Friedman. So let me take the descriptive and check. So what is my assumption coming, my study saying, the mean rank of first one is 2.42, second one is 2, third restaurant gets a 4 the mean rank is 4, so it is quite high, restaurant 4 is 2.58.

Looking at the test statistics, it is significant at what level, 0.001, it is significant at 0.001 **(Video Ends: 30:05)**. So that means what, we can easily say from here that there is a significant difference between the service offered by the restaurants as per the 6 raters okay. If you want to compare it individually you can make a comparative study between restaurant 1 and 2, restaurant 1 and 3, restaurant 1 and 4, similarly 2 and 3, 2 and 4, and 3 and 4 by making individually a Wilcoxon signed-rank test, but only thing you remember you have to take your significance level not 5% or 10%.

Whatever you take it has to be divided by the number of tests, that is in this case 6. So if there are 3 groups how many tests, if there are 3 groups there will be only 3 tests. Suppose A, B, C, A and B, A and C, B and C 3 test. So if it is 5% for a 3 tests, it will be 5/3. In this case you have 4 groups, so there are 6 tests, so 5/6 and whatever you have to compare accordingly. So this is what is the Friedman's test.

(Refer Slide Time: 31:12)

Kolmogorov–Smirnov (K–S) test

- It is a **simple non-parametric method** for testing whether there is a significant difference between an **observed frequency distribution** and a **theoretical frequency distribution.**

- **The KS One sample test is more powerful than the** $\chi 2$ test since it can be used for small samples unlike $\chi 2$.

There is one more test which I will talk about, the last test is K-S test, this is the last test which we are discussing today. So this test is a simple non-parametric method for testing whether there is significant difference between observed frequency distribution and a theoretical. I have seen many a times when you talk about models, so there you need to understand whether there is a difference between the observed frequency and the expected frequency or not, and if there is no difference then that means model is fit or strong, even that we saw in the chi-square test.

So K-S One sample test is more powerful than a chi-square test since it can be used for small samples unlike the chi-square, this is a very important thing you need to remember okay.

(Refer Slide Time: 31:55)

- **The null hypothesis assumes no difference** between the observed and theoretical distribution and the value of test statistic 'D' is calculated as:

$$D = \text{Maximum} \, |F_0(X) - F_r(X)|$$

Where −

- $F_0(X)$ = Observed cumulative frequency distribution of a random sample of n observations.
- $Fo(X)$ = k/n, where k=the number of observations equal to or less than X. (No. of observations $\leq$ X) / (Total no. of observations).
- $Fr(X)$ = The theoretical frequency distribution under H_0.

Now how do you do this. The null hypothesis assumes no difference between the observed and the theoretical distribution. So what is the null hypothesis saying, there is no difference, but as a researcher what do you want, there needs to be difference because that is what the alternate hypothesis talks about. So the difference D is calculated as

$$D = \text{Maximum } [F_o(X) - F_r(X)]$$

Now what are these two? $F_o(X)$ is equal to observed cumulative frequency distribution of a random sample of n observations. Now $F_o(X)$ is equal to k/n where k is equal to the number of observations equal to or less than X and divided by the n, total number of observations. $F_r(X)$ is the theoretical frequency distribution.

(Refer Slide Time: 32:39)

- The critical value of D is found from the K-S table values for one sample test.
- **Acceptance Criteria:** If calculated value is less than critical value accept null hypothesis.
- **Rejection Criteria:** If calculated value is greater than table value reject null hypothesis.

So let us this one of the problem. So this is the more or less the same thing. The acceptance criteria if calculated value is less than the critical value. If calculated value is greater than the critical value, you will reject. The same thing raised.

(Refer Slide Time: 32:51)

Problem Statement:

In a study done from various streams of a college 60 students, with equal number of students drawn from each stream, are we interviewed and their intention to join the Drama Club of college was noted.

	B.Sc.	B.A.	B.Com	M.A.	M.Com
No. in each class	5	9	11	16	19

It was expected that 12 students from each class would join the Drama Club. Using the K-S test to find if there is any difference among student classes with regard to their intention of joining the Drama Club.

So this is the study of 60 students. Each student were interviewed to join the Drama Club. So B.Sc 5 people, B.A. 9 people, B,Com 11 people, M.A. 16 people, M.Com 19 people. It was expected that 12 students from each class would join the Drama Club. So using the K-S test to find if there is any difference among student classes with regard to the intention of joining the Drama Club, let us check that.

(Refer Slide Time: 33:18)

H_0: There is no difference among students of different streams with respect to their intention of joining the drama club.

We develop the cumulative frequencies for observed and theoretical distributions.

What is the null hypothesis. There is no difference among students of different streams with respect to their intention of joining. So we want to say there is difference in the thought process of people with B.Com degree and M.Com degree or a B.A. and B.Com we want to compare. We develop the frequencies for the observed and the theoretical distributions.

(Refer Slide Time: 33:43)

Streams	No. of students interested in joining		$F_O(X)$	$F_T(X)$	$\|F_O(X)-F_T(X)\|$
	Observed (O)	Theoretical (T)			
B.Sc.	5	12	5/60	12/60	7/60
B.A.	9	12	14/60	24/60	10/60
B.COM.	11	12	25/60	36/60	11/60
M.A.	16	12	41/60	48/60	7/60
M.COM.	19	12	60/60	60/60	0
Total	n= 60				

So what it says is the number of students interested in joining observed is 5, 9, 11, 16, 19. Go back so 5, 9, 1, same thing right. What is the theoretical since there are 60 people and there are 1, 2, 3, 4, 5 so 5 groups, so we will divided it by 60, 60/5 = 12 for each. So what is my observed frequency, 5/60, 14/60; how 5+9 right, 14/60; then 14+11, 25/60; then 41/60; 60/60 okay. What is my theoretical distribution now 12/60, 24, 36, 48, 60. Now we will take the difference between these 2, this and this. So what it gives, 12-5, 7/60; 24-14, 10/60; 36-25, 11/60; 48-41, 7/60; this is 0 right.

(Refer Slide Time: 34:36)

Test statistic |D| is calculated as:

$$D = \text{Maximum}|F_0(X) - F_T(X)|$$

= 11/60

= 0.183

The table value of D at 5% significance level is given by

$D_0.05 = 1.36 \sqrt{n}$

$= 1.36 \sqrt{60}$

= 0.175

- Since the calculated value is greater than the critical value,
- hence we reject the null hypothesis and conclude that there is a difference among students of different streams in their intention of joining the Club.

So if you calculate the maximum now difference, it is where, where is the maximum difference, this one 11/60, so this is the maximum difference 11/60 which is equal to 0.183. The table value of D at 5% significance level is given by this is the formula. So 1.36 √n, which is coming at 0.175. Now you see your calculated value was 0.183, your table value is

0.175, so what will you do? Since table value is less than the calculated value, the null hypothesis is rejected.

So since the calculated value is greater than critical value, we reject the null hypothesis and conclude that there is a difference among students or different streams in the intention of joining the club. Well, I hope this is a very simple test but very powerful test it can be utilized. All these tests of non-parametric tests that I have tried to explain you and make it as clear as possible to my extent of knowledge, for my level what I can think of, but non-parametric tests are not to be taken lightly.

They have a very important role because many a times in life the data does not behave normally. So in such a condition, the best thing is to go for a non-parametric test and draw your research inferences okay. Thank you so much.