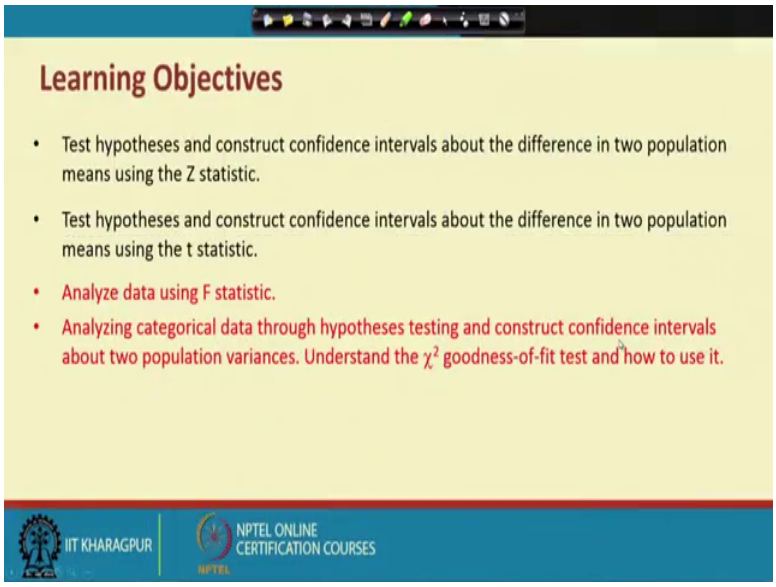
Business Analytics for Management Decision
Prof. Rudra P Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 25
Inferential Analytics Part – 2 (Contd.)

Hello everybody, this is Rudra Pradhan here, and welcome to BMD lecture series. And today we will continue the concept inferential analytics. And in fact, in the last couple of lectures we have already discussed this issue and it is the case where we are dealing with the business problems relating to multiple sample case. And in the last couple of lectures we have solved problems by using z statistic, t statistics and f statistic.

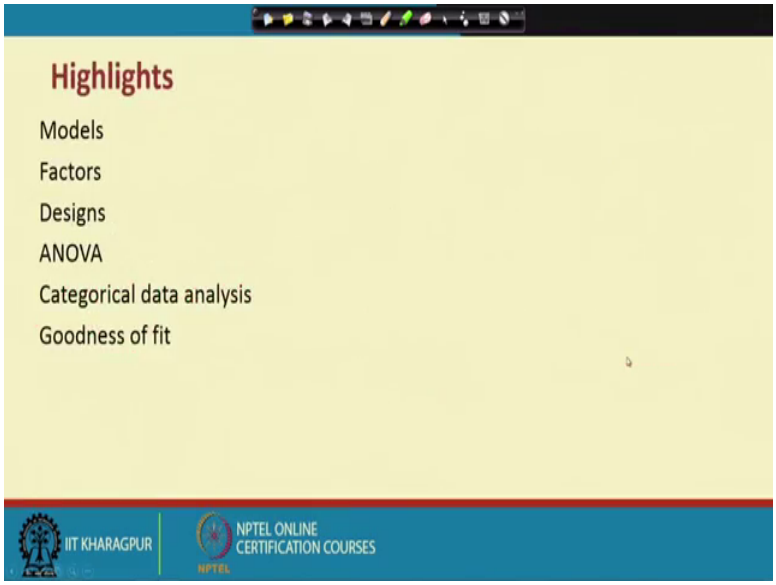
And today we will continue the similar kind of you know problems by using f statistics and typically we will focus on the concept called as you know ANOVA that is the analysis of variance. And today we will continue the ANOVA by connecting one way structures, two-way structures and three way structure. And finally, it will be continue the concept called as you know MANOV that is multivariate analysis of variance. And the concept is so interesting that if you have a multiple sample case so how we can actually check the kind of within the difference and between the difference, and with the different kind of you know layers or different kind of you know factors, in typical it is a kind of you know multivariate kind of structures.

(Refer Slide Time: 01:42)



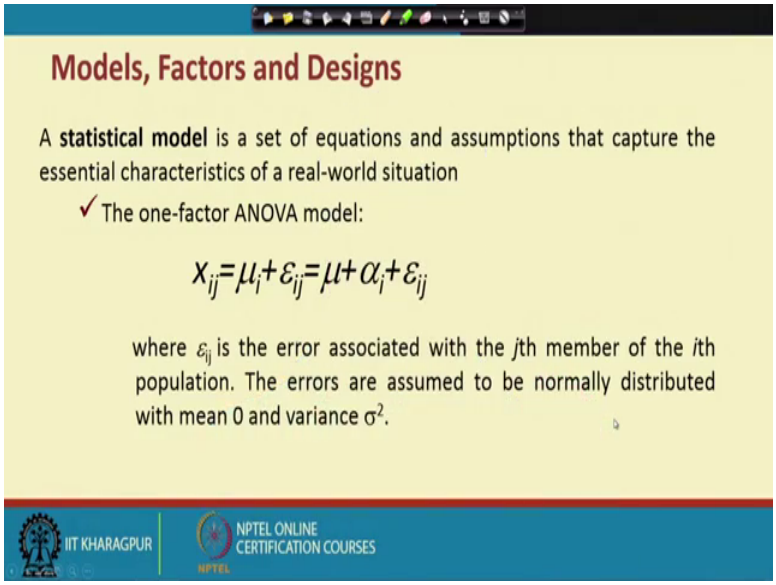
So, let us see how is this particular concept? And then, we will discuss. So, this is what the kind of you know discussion today.

(Refer Slide Time: 01:46)



The highlights of this discussion or you know will be the ANOVA models. And followed by factors then designs and a we will connect with the concept that is ANOVA. And we will dealing with the problems, categorical data analysis and finally, the goodness of fit.

(Refer Slide Time: 02:08)



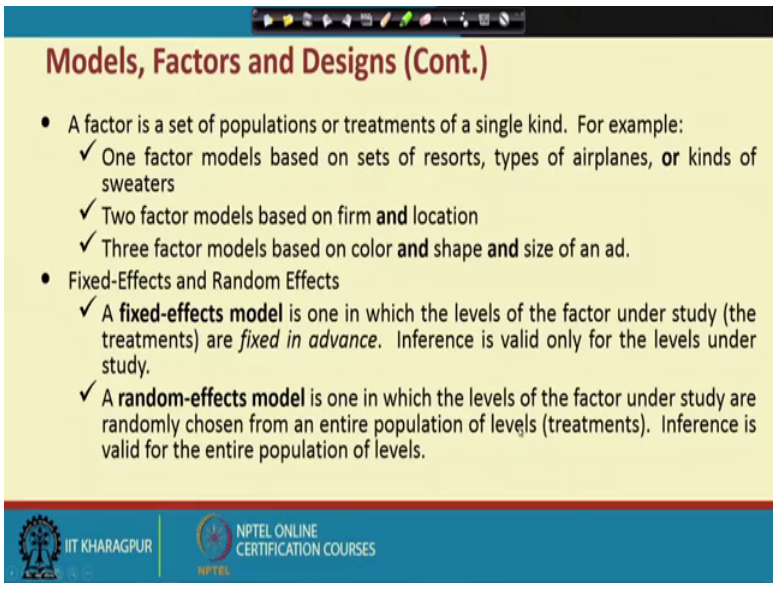
So, in the last couple of lectures, we have already discussed the concept typically the f statistic issue and in a kind of you know structure multiple sample cases. And here so we

will start with the same concept, but with a kind of you know different structure altogether. And let us start with a concept called as you know models. So, you know since it is a kind of multiple sample case or multivariate kind of you know structures, so there is a kind of you know typically you will finding relationships and that relationship is a kind of you know a complex kind of you know structure.

So, obviously, since there are a couple of variables and that to multiple sample case, so you cannot actually simply you know analyze the particular structure and get some kind of you know inference as for the business requirement. So, that is why we need to develop a models and through the models we have to analyze the particular you know problem and then we will get some kind of you know inference.

So, now as we have already discussed typically statistical model is a set of equations and assumption that capture the essential features of the real world situations. And the structure is you know we start with here kind of you know concept called as you know one-factor ANOVA model and the model will be like this you know x ij equal to mu i plus epsilon ij. And so mu is the kind of you know concept called as you know sample means. And then epsilon ij is the error associated with the jth member of the ith populations. And we are assuming that you know these errors are normally distributed with the mean 0 and unit variance that is the sigma square.

(Refer Slide Time: 04:00)

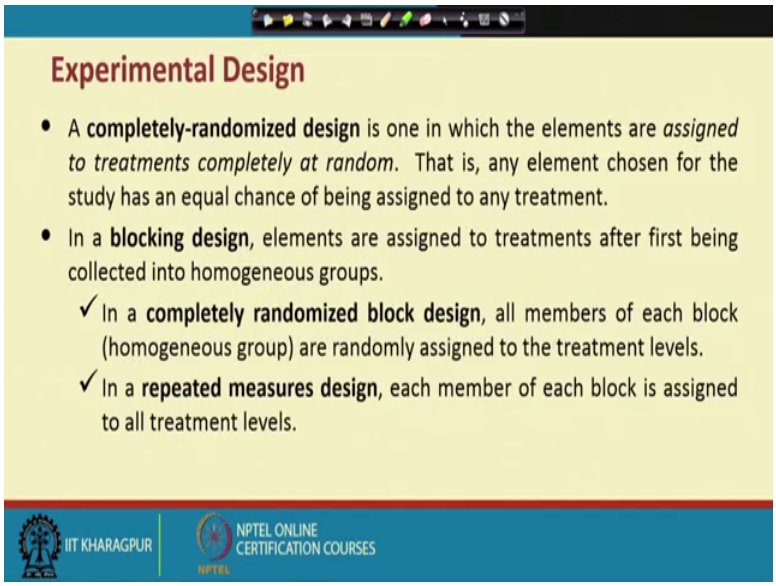


And in the other side, a factor is a set of you know populations or you know kind of know treatments of a single kind. For examples, we have actually one-way structure, two-way structure, three-way structure. And in the case of you know one way structures that is called as you know one way typical you know factor models was done you know examples like you know hotels type of you know at aeroplanes and the kinds of you know colors, in a particular product. And two factors models based on business firm and the kind of you know site location; and the three factors models can be like you know color shape and size of a particular you know service or you know kind of you know product. But typically you know analysis will be on the basis of you know the concept called as you know fixed effect models and a random effect models.

In fact, this particular you know ANOVA is well connected with your concept called as you know panel data and a in the last couple of lectures we have already discussed this data issue. And then you know data can be actually connected or you know using a kind of you know business environment or you know business analytics in the form of time series cross sectional and you know penner. So, ANOVA is a kind of you know cluster like you know panel data structure all together. So, we have a separate lectures about the panel data modelling. But in the mean times we will analyze the similar kind of you know problems in a kind of you know structure called as you know of factor models that is one way factor model, two-way factor model and three way factor model depending upon the kind of you know structure of the business problem.

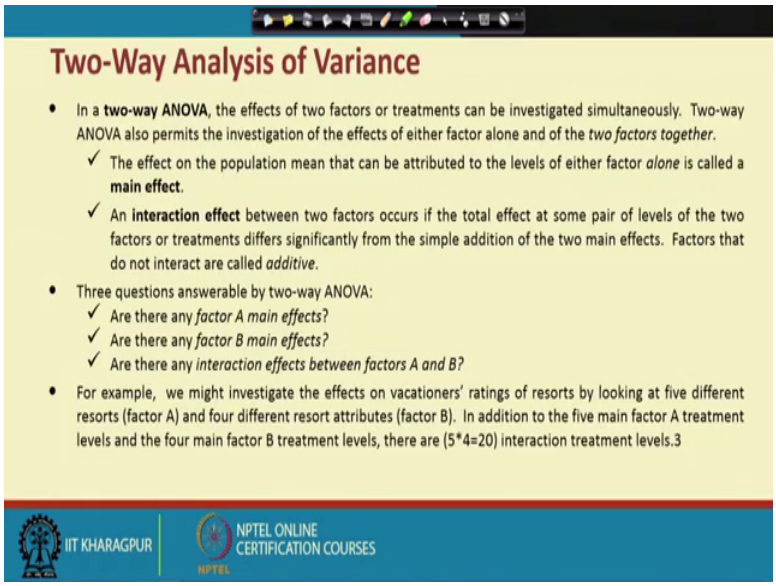
And here is we have actually typically two different you know models which you can connect with these particular you know situations. So, the first one is called as a fixed effect model, and the second one is called as a random effect model. In the fixed effect model the levels of the factor under study are fixed; and in the case of random effect models, the level of factors under the study are randomly chosen from an entire populations and that is the treatments. And the inference is valid for the entire populations for this particular you know study or analysis.

(Refer Slide Time: 06:20)



And the third component is with respect to experimental design, and basically ANOVA can be connected with the design of experiments that typically called as you know DOE. And here a completely randomized design is one in which the elements are assigned to treatments completely at a random basis, and that is any element chosen for the study has an equal chance of being assigned to any treatment. And in the case of you know block designs elements are assigned to treatments after first being collected into homogeneous groups. So, in a completely randomized block designs, all members of each block that is the homogeneous groups are randomly assigned to the treatment levels. And in the case of you know repeated measures designs, each member of you know each block is assigned to all treatment levels.

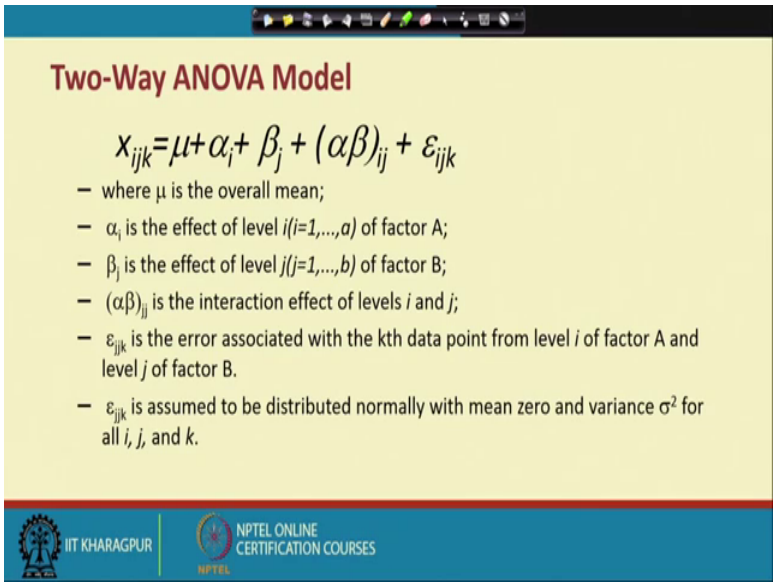
(Refer Slide Time: 07:17)



In the case of you know two analysis of ANOVA typically the effect is with respect to two factors or it is with respect to treatment two treatments, so that means, typically actually we have a concept one factor model, two factor model, three factor model. In the case of one-factor models, so we have a one factor then the kind of you know different sample variation will be there; and then in the two factor models we have actually two different factor through which the investigations can be you can say gently you know investigators. Then you know in the case of three factor models, we have a three factors through which these simultaneous interaction can be investigators.

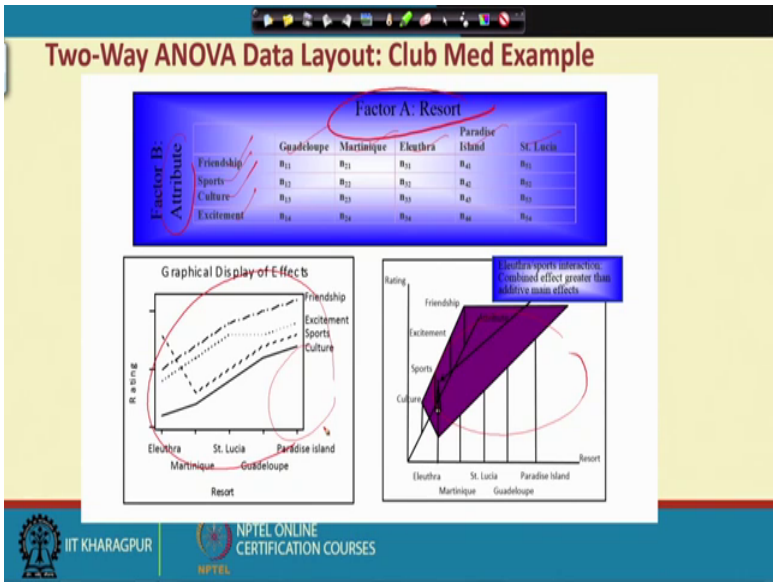
So, now in the case of you know two factor models, so we have actually a main effects and then the kind of you know interactive effect. So, for instance, if there are two factors A and B then the main effect will be effect A and effect B and the interactive effect will be a you know AB together so that means it is a intersection between A and B. So, in the kind of you know a problem, so multivariate problem is the two factors. So, we defined actually kind of you know a cross correlation or you know kind of you know a concept that need to be connected while you are investing the kind of you know interrelationship among the variables. So, typically in the two ANOVA case, we have a three different effects. So, the main effects that is with respect to two factors and they are you know interactions.

(Refer Slide Time: 08:51)



So, let us see how is this particular you know structure. And this is how the two-way ANOVA model and x ijk is the kind of you know a total effect and that is a connected with a mu that is the population mean and subject to alpha i plus beta j and plus alpha beta ij and epsilon ijk. So, that means, mu is the overall you know mean and alpha is the effect of level i for factor A; and beta is the effect of level j for factor B; and alpha beta is nothing but the interactive effect. And the a epsilon ijk is the kind of you know error associated with this kind of you know link because we are actually structuring with respect to it in two different factors. So, obviously, there is a kind of you know cross correlation and then you need to have a you know interactive effect and along with the you know error terms. And here is there are error terms are you know distributed normally and with the assumption that you know zero mean and again unity variance with respect to i, j and k.

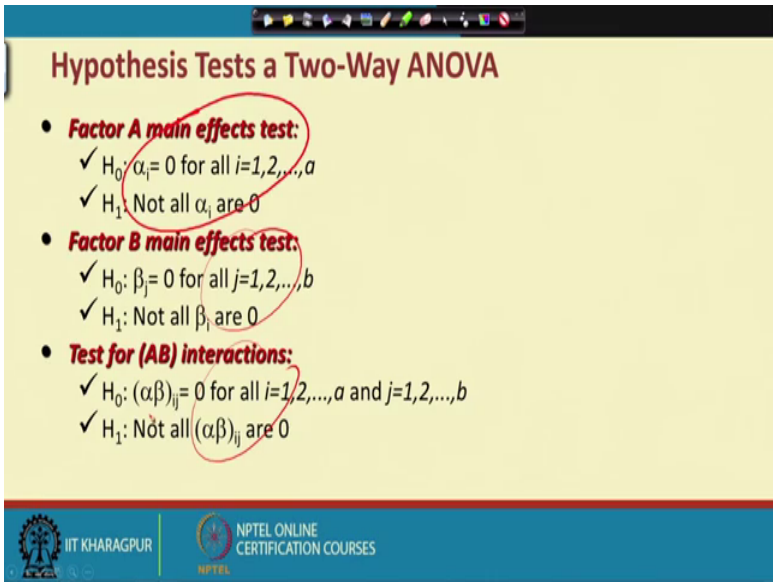
(Refer Slide Time: 10:00)



Typically, two-way ANOVA problem is like this. So, let us take a case you know factor A, so we have an information here the kind of you know let us say this is a kind of you know problem say A resort and then factor B attributes. So, in this particular you know problems we have actually four attributes friendship, sports, culture and excitement, and then we have it you know five different results. And again so we like to check you know how is the kind of you know a rating, a rating of a particular you know hotels and with respect to these you know four attributes.

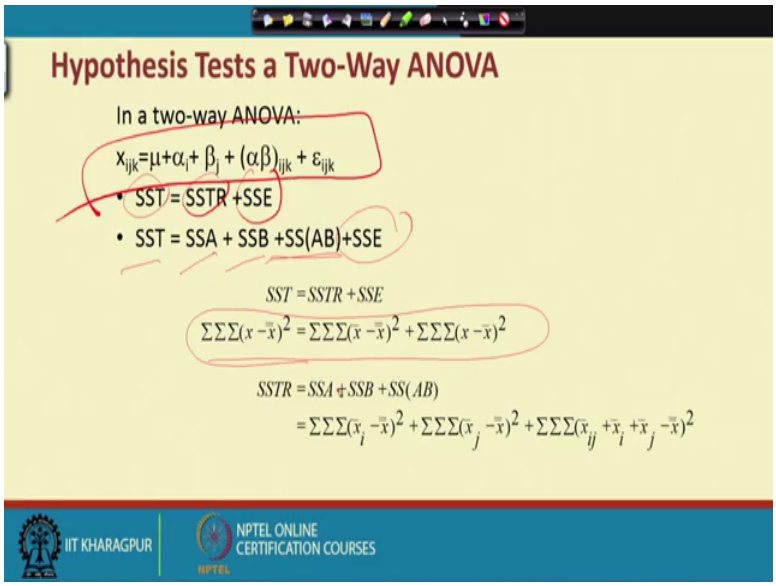
So, you know it is a kind of you know a hotel business, you will find you know the customers, and they like to enter and they like to leave. And a during these times we usually collect the information and that information will we know will be very helpful to analyze the problem and then you can actually use for some kind of you know improvement or some kind of you know management requirement. So, in this case, so the rating are recorded for different hotels with respect to different attributes. And the usual typical structure of the effects will be like this. So, it is a kind of you know multiple kind of you know plotting and with respect to rating and that too with respect to five different you know resorts. So, the kind of you know interactions, so this is the graphically visualization about this particular problem, and we need actually the kind of you know quantification the total effect or total quantifications how they are you know interacting each other so far as you know the particular problem is concerned. And accordingly we will see how is the particular you know structure then we will discuss.

(Refer Slide Time: 11:48)



So, what we have already discussed that you know in the two-way process of you know ANOVA, so we have actually three tests. So, the since it is the two factors, so the main two effects are you know factor A and factor B, then finally, the interactive AB, so that is connected with the alpha i beta j and alpha beta ij. So, now accordingly, so the kind of you know a null hypothesis will be alpha equal to 0 for all i again the counterpart alternative hypothesis will be not all alpha ir equal to 0. And similarly we can set for you know bj, so this is for second factors and then finally, there is a interactive effects, so that means, alpha beta j ij equal to 0 means we assume that there is no interactions, but again the alternative will be so at least one interaction is there. So, likewise you know this is how these typical you know structures and that need to be tested with this particular you know concept.

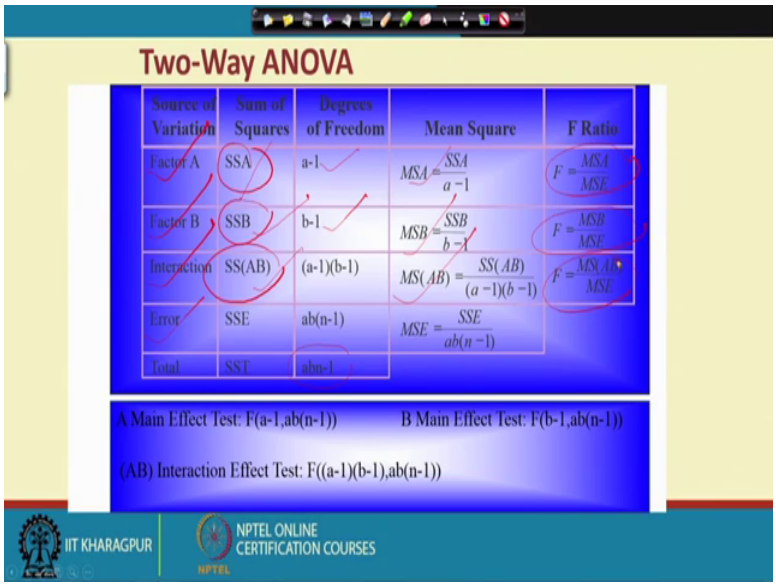
(Refer Slide Time: 12:51)



So, let us see how is this particular structure. So, it since it is a kind of you know ANOVA that is the analysis of variance. So, we like to check this particular you know effect with respect to variance, because it is a variance statistics. And as a result the typical model is here is this is what the typical models. So, it has a two parts and total SST means total sum of square that is the total variance; and then it is divided into two parts that is the error part and the kind of you know sum square totals the kind of you know factors. So, as a result so the sum square total will be so sum square you know total variance from factor A variance from the factor B that is actually main effect and then this is the interactive effects sum squares through the interactions and then this is the error component.

So, you know the procedure of you know getting this kind of an outcomes like this which we have already discussed in the previous lectures in the case of you know or one-way ANOVA. So, now you know with the help of this data, so you have to find out the variance for a factor A, factor B and for the kind of you know interact interactions and then the error component. And if you add or all together then you will get the sum square totals or else you know yeah you know a sum square totals if you can calculate and SSA and SSB and SSB then by doing some kind of you know simplifications, you can get these sum square errors. But typically so our requirement is a what is SSA, SSB, SS AB and SSE. So, these are the items through which you have to check the effect you know that is the main effect and then the kind of you know interactive effect.

(Refer Slide Time: 14:45)



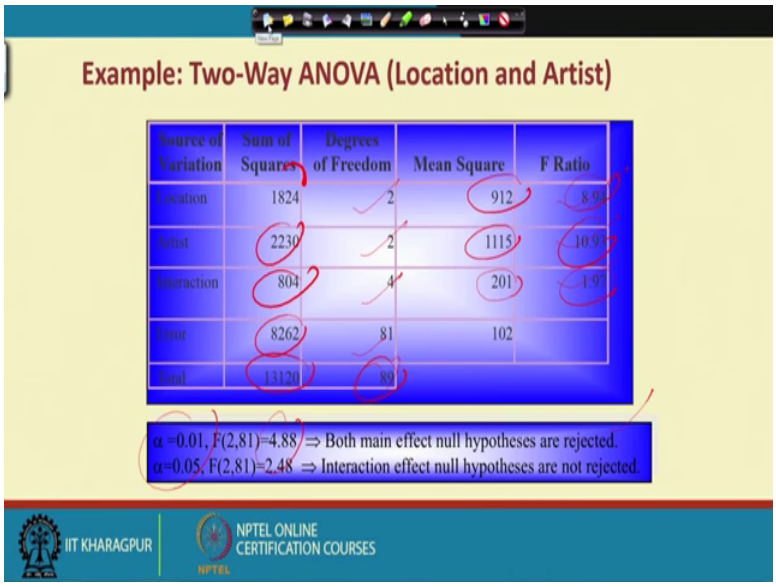
So, since we have already gone through this ANOVA that is one way ANOVA. In this case it is the similar kind of you know structure and in this case so this is for factor A, this is factor B, and this is interactive effect and then the errors. Since we are dealing with you know two factors and the interactions by default some error will be there in the systems. So, as a result so for factor A so sum will be sum squares for factor A SSA. And for factor B, so it is SSB; and then interactivity is sum squares of you know interactive AB. And for factor a, so the degree of freedom is a minus 1 because we are allowing the factor samples for you know up to a, so that is the size of the samples and by default the degree of freedom will be a minus 1. And similarly for factor B, it will be b minus 1, and the interactive effect will be a minus 1 into b minus 1s.

And the error component accordingly will be a b into n minus 1. And the total for the total that is sum square total, so this is the combined a samples that is a b ns and that to the degree of freedom will be a b n minus 1s. And we typically get to know so what is the mean square of you know factor A, mean square of factor B and mean square of you know interactive. So, this is nothing but actually sum square of factor A you know divided by the degree of freedom and it is followed by factor B and also it is followed by interactive effect.

So, accordingly we have actually three different f statistics for two-way ANOVA, one is for factor A, one is for factor B, and one is for you know interactive effects. So, we like

to know with respect to the total variations which factor is more effective and then whether the interactive effect is statistically significant or not. And in order to justify this, so we have to take and you know live examples.

(Refer Slide Time: 16:45)



So, let us take this examples and then we will see how is the kind of you know structure. So, this is the kind of you know with a particular samples and that with the two kind of you know structures location and artist, and which is you know with kind of you know data analysis. So, we have a SSA here and that to SSA here and SSB here, and then this interactive effect, and then the SSE that is sum square errors.

And if you add all these three, so you will get actually the total effect that is called as you know sum square totals. And the corresponding degree of freedom is 2, 3 and then 4, and that is for error 81 and the total samples for this particular you know data is nothing but 89 and that means, 90 and then this degree of freedom will be nine 89 for the SST. And accordingly, so this will be mean square for factor A, this is mean square for factor B, and this is mean square for interactive.

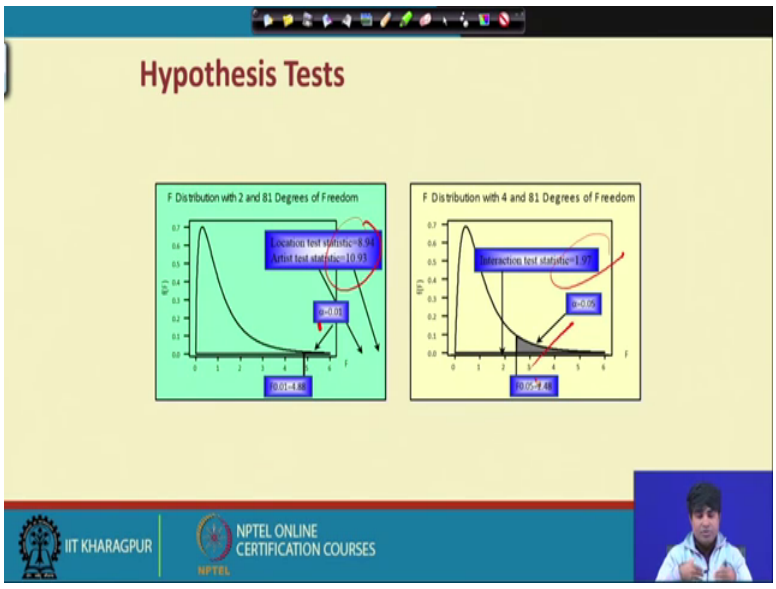
And then corresponding to this, this is the D calculated for factor A, F calculated for factor B, and F calculated for interactions. And then so corresponding to this you know degree of freedom and then you will go to the F critical. And check you know whether you know this F calculated will overtake the F critical depending upon the kind of you know alpha which we have actually fixed here at a 1 percent and 5 percent so that means,

typically actually even if you do not fix actually alpha. So, first you have the calculated value depending upon the sample information and the problem.

And then you have actually a critical value depending upon the kind of you know structure which you can actually fix or you know allow. For instance, you can fix alpha equal to 1 percent, you can fix alpha equal to 5 percent, you can fix alpha equal to 10 percent, so accordingly so you can get you know three critical values. So, now this calculated value can be compared with a 1 percent, 5 percent and 10 percent; typically if it is actually starts really significant at 1 percent that is the highest weightage you can give into these kind of you know difference followed by 5 percent and you know 10 percent.

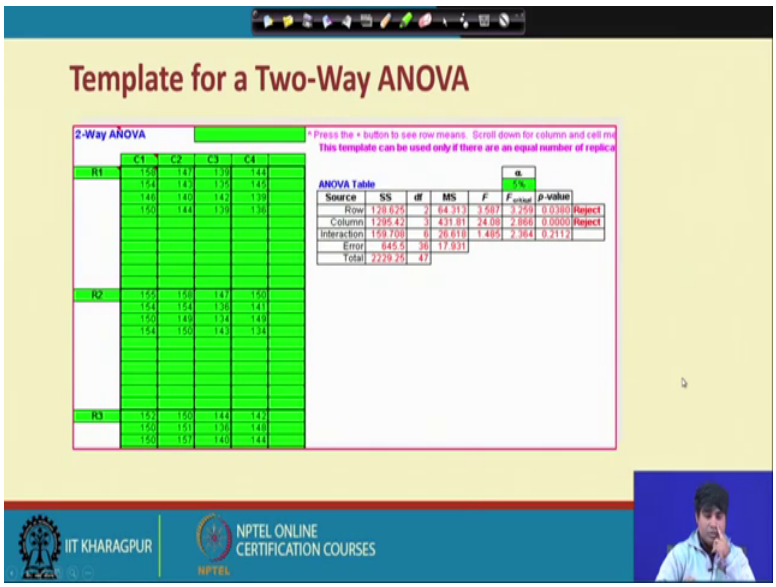
So, in this case let us see what is happening. So, the critical value will be 4.88 and then it is actually 2.48. So, in the both the cases actually and the calculated value is over taking the tabulated values. So, as a result you are rejecting the true null hypothesis, but in the case of you know interactive effect, it is actually coming you know less that is less than two critical value. So, we cannot actually reject in this case so that means, in this particular process, so the main effects are you know statistically significant while the interactive effect is not statistically significant, so that is how the kind of you know structure.

(Refer Slide Time: 19:35)



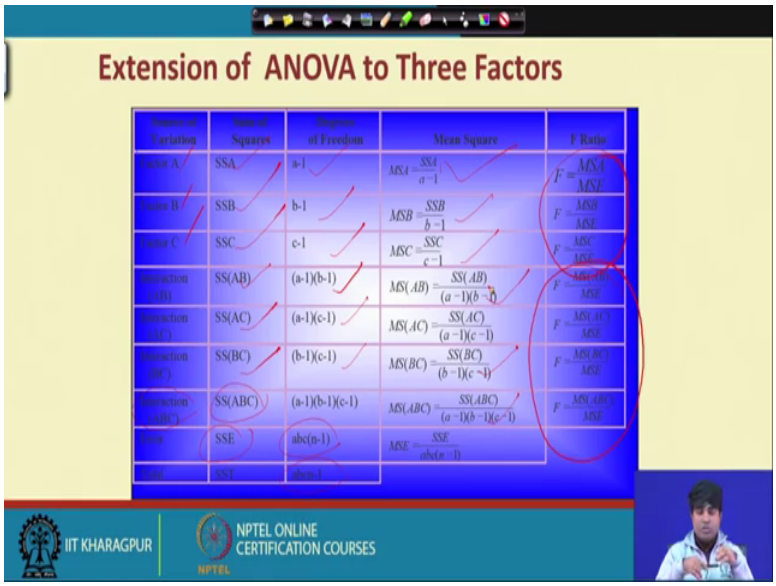
And then accordingly we will discuss this particular process and this is how the significance or you know which we have already highlighted what is the graphical kind of you know plotting and this is the rejected zone, and this is the rejected zones and this is also rejected zones. So, in the case of main effects, it is coming statistically significant, these are the two main effects and this is what the interactive effect and which is actually a lesser to critical value. As a result, we are not in a position to reject here, but in this case, your calculated value is your overtaking the critical value, and hence we are actually rejecting as a results. So, the locational effect is there, and artist effect is there, but the interactive between the, you know location and artists which is not actually statistically significant here.

(Refer Slide Time: 20:28)



And this is actually the kind of you know excel template while you analyzing two-way ANOVA. And I will discuss these problems through you know excel sheet.

(Refer Slide Time: 20:39)



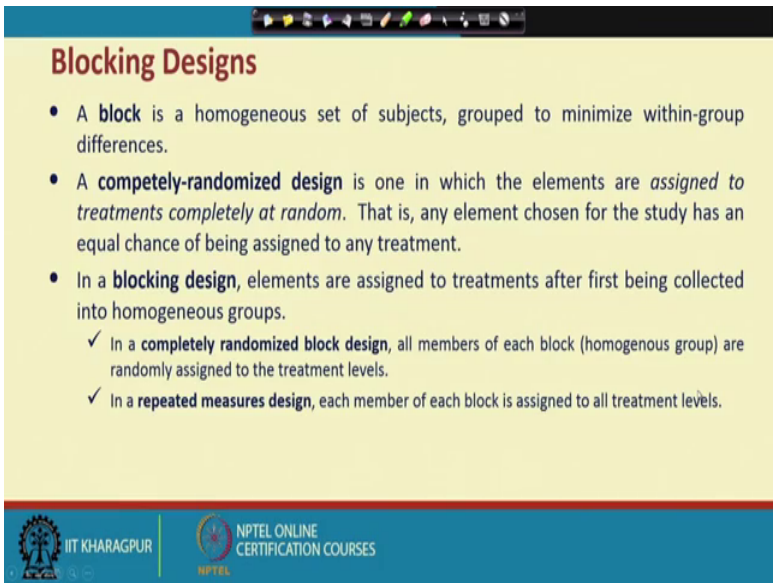
Let me first complete this one then I will go to this particular excel view. And in this case you know the a and this is actually three factor model and this is extension of you know again ANOVA. And in this case we have actually three factors. And the three factors is actually a, b and c. And in the case of you know two factors we have a main effects a and b and the interactive effect will be ab, but when there are actually three factors or four factors and five factors like this, so the interactive effect will be you know continue in a kind of you know complex process. And a typically for these three factors models. So, we have a three main effects that is SSA, SSB and SSC that is sum squared of you know factor A sum square of factor B and sum square of you know factor C. Then by default so there are you know three interactive effect, so between AB, AC and BC and then finally, the combined interactive effect which is nothing but actually ABC.

So, as a result sum square AB, sum square AC, sum square BC and sum square ABC and finally, when you are dealing with more factors and by default there will be error component and that will be represented by sum square errors. In fact, some square error was also there in the case of you know two-way ANOVA. And in that case, we have actually to two main effect and one interactive effect. Here is we have actually three main effects ABC and then three interactive effect AB, AC, BC and then combined effect that is the interactive between A Band C. So, the corresponding the degree of freedom will be a minus 1 b minus 1 c minus 1. And like the previous one, it is a minus 1 into b

minus 1 a minus into c minus 1 b 1 minus 1 to c minus 1 and then a minus 1 b minus 1 c minus 1 that is for you know joint effect of you know abc.

And finally, this is for you know error sum squares, and this is for you know sum square totals. And accordingly you can find out mean square you know for factor A, factor B and factor C and this is what you know the interactive effect between AB interative between AC and interactivity to BC and this is the interactive effect combined ABC. And accordingly we have actually in that case of you know three factor models, so these are the three main factors and then these are the three actually you know kind of you know interactive factors. So, all together, so we have actually seven different test statistic which you would like to report in the case of you know three factor models. So, while in the case of you know two factor models we have a two main effects and then one is the interactive effect. So, let us see here is, so connecting to this, so this is actually three factor model situation.

(Refer Slide Time: 23:25)

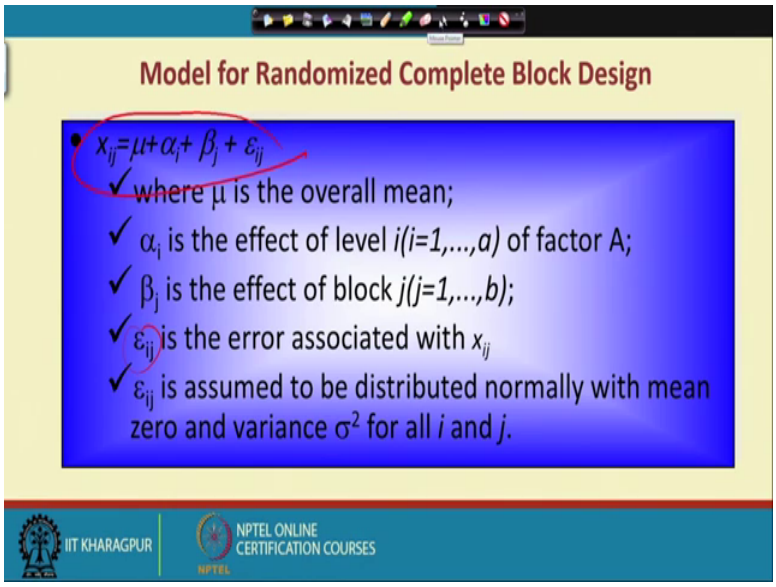


And accordingly we can connect with you know four factor model five factor model. So, when we add one after another layers then the degree of complexity and the kind of you know effect will be more and more. So, which we have already seen from the moment from one to two, and two to three like this. So, when will you add one after another factor, so then the combined effect or the kind of you know interactive effect will be complex and then and then it will be very in fact, it will be very interesting and also it

will be through complex process we have to investigate and then we can analyze the process. In fact, for you know more factors, and you know more kind of you know typical samples, so manually it is not possible through software we have actually you can deal with the kind of you know an analysis then finally, we will check the kind of you know variance.

And in the case of you know block designs, so let us see what is actually block. A block is a homogeneous a set of subjects a group to minimize within group differences and. In fact, we have what we have already immense on the completely randomized design in one which the elements are you know assigned to treatments completely at the random. And we in fact, we have already mentioned all these things completely random designs and block designs.

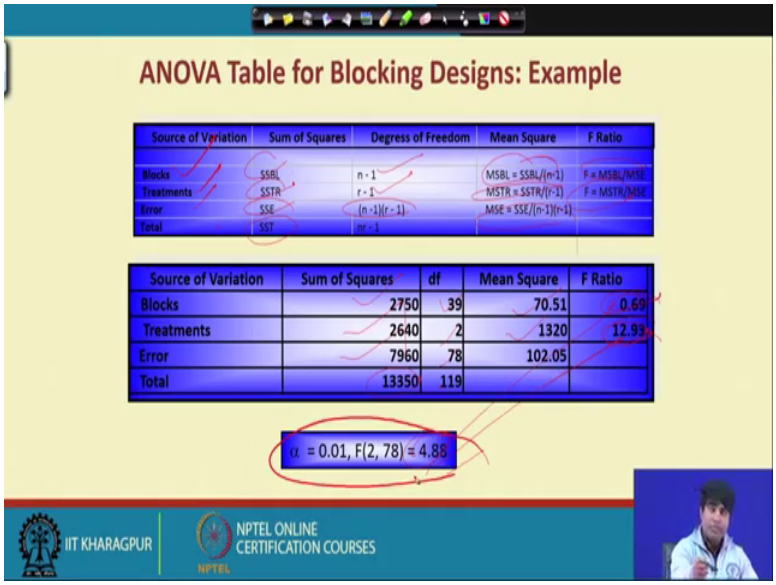
(Refer Slide Time: 24:44)



So, let me take the kind of you know a model part of this you know block designs. And like the previous you know factor models, you know a block design model usually represented like this. So, the main effect which is actually mu alpha i beta j you know summation you know alpha ij and then and so it is followed by the alpha component and that is the effect of level 1. And beta j the it is effect of you know block j, and mu is the overall mean. And like the kind of you know factor models, so a this is the error term. And again we are assuming that you know the error term will be normally distributed

with a mean zero and unit variance right and that is the homogeneous variance. So, let us see how is the particular you know cases.

(Refer Slide Time: 25:35)



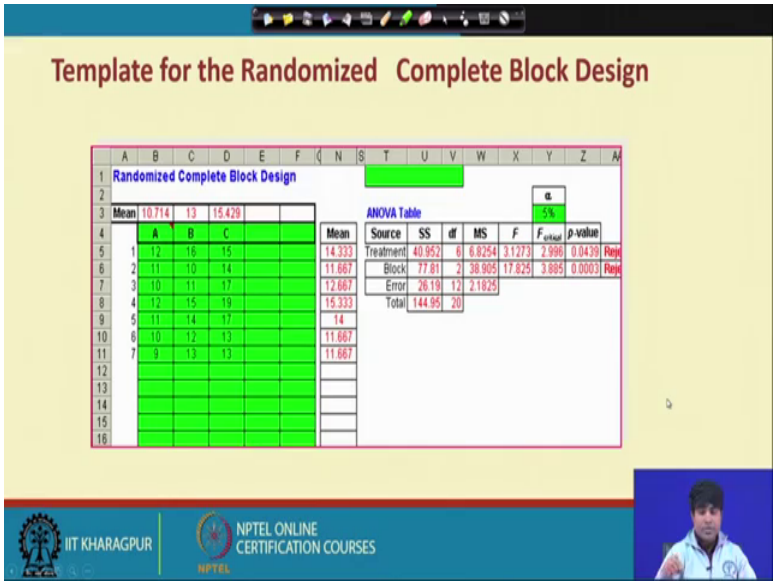
This is how the typical problems and it is it like you know or kind of typically ANOVA structures, but here the items are you know just like you know different, because it is a kind of you know block design problems. So, as a result so compared to the previous one, so the first effect will be blocks then treatments, so this will be block, statements, and the error component, and then finally total So, again similarly sum square you know for blocks and sum square you know for errors it, it means and then sum square you know for errors. And if you combine all these then you will get you know total sum of squares and corresponding to this is actually block wise is nothing but actually the kind of you know column variations and this is the row variations.

So, typically so the degree of freedom will be n minus 1 and r minus 1s and then finally, so for you know errors it will be n minus 1 into r minus 1. And like the ANOVA so we will find out the mean square for you know blocks mean square for treatments and then mean square for you know errors. And in this case we are interested about the a block effect and the treatment effect and accordingly we have a two f statistics. And with a particular you know sample problems, so we have calculated here actually sum of square for the block, sum of square for the treatments and sum square of the error. The first hand check is actually these three should actually you know if close to us total sum square and

the corresponding the degree of freedom is reported here. And the simplification you define mean square for a blocks and mean square for treatment.

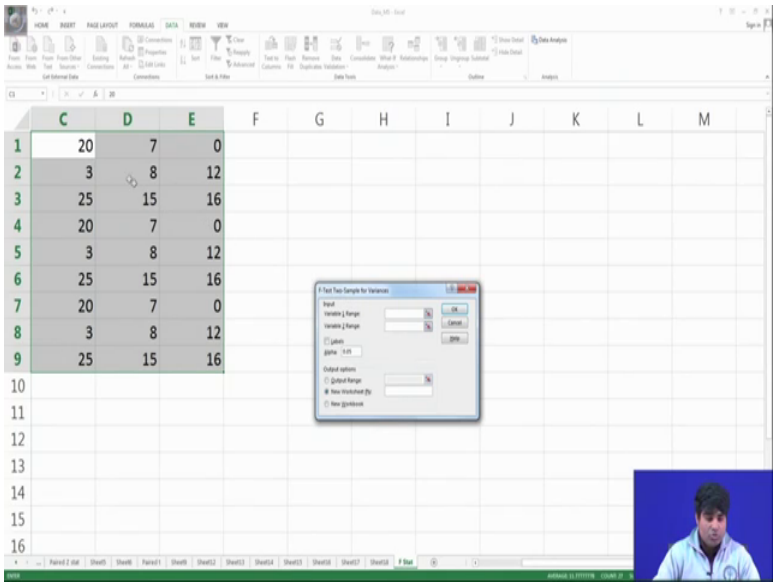
And then this is what the calculated value and again fixing alpha and then reporting the critical value for f and you can check the value. So, since actually this value is lesser to this particular value. So, we are not in a position to a reject the null hypothesis, but in this case we are actually rejecting the term, because it is over taking the kind of you know critical value.

(Refer Slide Time: 27:41)



So, this is actually again excel template how you have to do the kind of you know an analysis. And in fact so before we go to this particular process, let me highlight the kind of you know examples how you can go to the analysis and solve the problems.

(Refer Slide Time: 27:56)



So, this is actually spreadsheet and we have actually a couple of samples here and then we like to test through oh you know simple you know start with you know simple f then we will go for you know one-way ANOVA, two-way ANOVA, and in three-way ANOVA like this. And then since actually this is so go to the data analysis package and in fact, you will find here is you will find for he has these are the tests. So, we need to test first you know f statistics. So, this is actually the f statistic for two sample variance let us say. So, this is with respect to the previous class discussion.

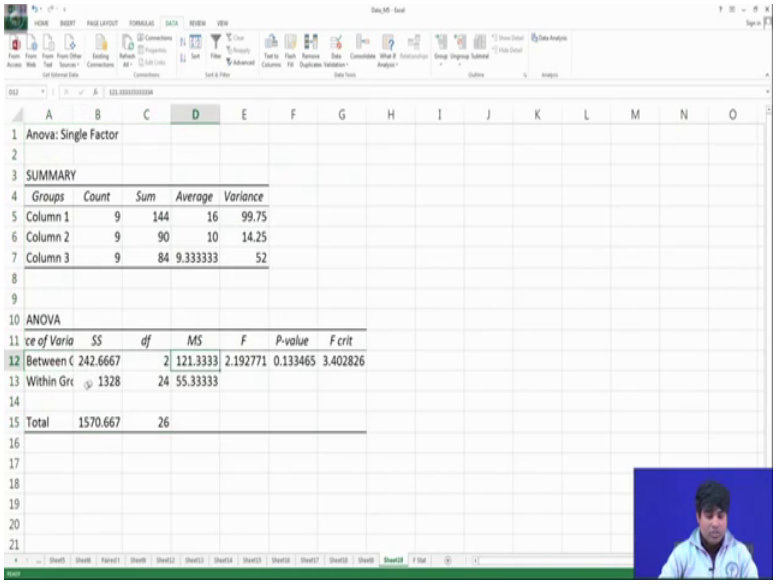
So, when we will go for you know means it is actually multi sample case. So, the multi sample case means at least you must have actually two different you know samples and then we must have some sample points to test the you know kind of you know variance. So, now accordingly you have to specify here the first sample case, and then you have to specify the second sample case, and like you know t and z so you have to just you know put. So, you will finally, this is how the f distribution this is the fk you know distribution values and this is f critical.

And then finally, this is f critical and this is what the f calculated on the basis of the samples and these are all mean reporting of these two variables and variance of these two variables, these are all counts and that is sample observation. And f reporter is this much, and the degree of freedom is this much that is actually an n minus 1. And then finally, with comparison since this f calculator is you are taking they are critical. So, it is

statistically significant. So, this is what simple f statistic through f statistic you have to analyze this problem that means there is a significance difference between these two samples and then accordingly the management decision will be a taken into consideration.

Now, again you go to the data analysis same structures. We will analyze this problem through ANOVA and you see here is another single factor models, which you have already highlighted then you know you just actually highlight again, so the entire sample is just a highlight and you know then you give the kind of ok.

(Refer Slide Time: 30:24)

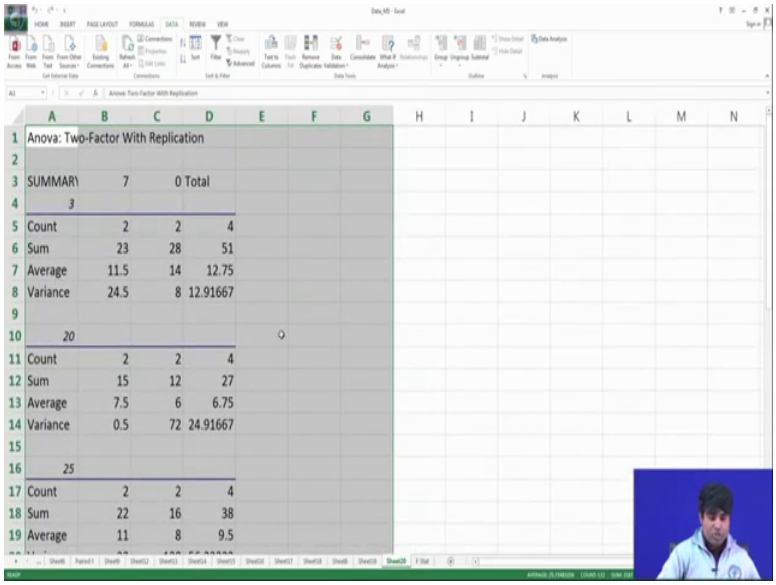


So, this is how the usual representation of this is what actually the output. And we are actually having between the difference and within the difference and the ratio between the given f value and this is what the probability which you have P and that to this is critical and this is what the calculated. And since the f critical is over taking the f calculated as a result, so this is not actually statistically significant even at actually 10 percent. And these are actually the kind of you know summary statistics for this particular you know kind of you know problem.

So, this is the actually the kind of you know output side for single factor model. And again you go to the spreadsheet and then you allow the data analysis again. So, again you allow one ANOVA two factor models. So, it is extension again. So, in this case again you

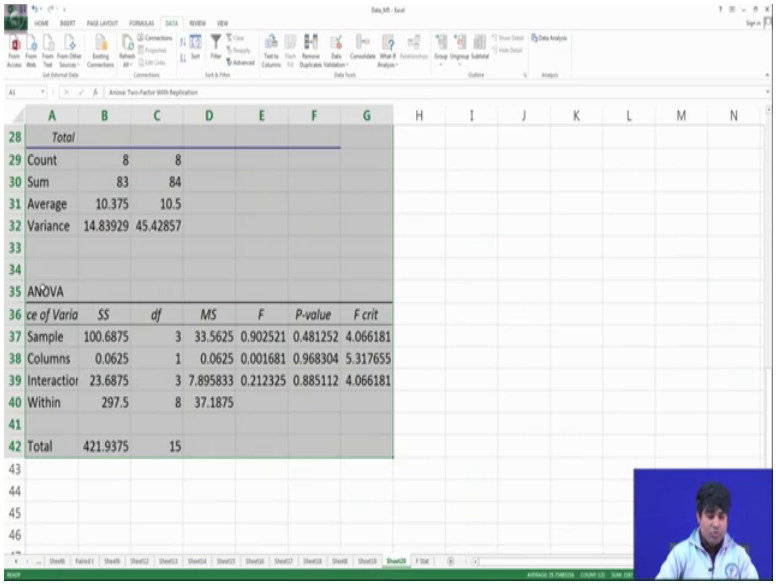
specify the range and allow that you know there are you know two since two factors. So, you allow the two factor sampling and then again put ok.

(Refer Slide Time: 31:28)



So, this is how the analysis of actually a two factor kind of you know situations, you will find plenty of you know ANOVA so that means, you is compared to the previous one. So, in the previous one, you really find here plenty of results.

(Refer Slide Time: 31:45)

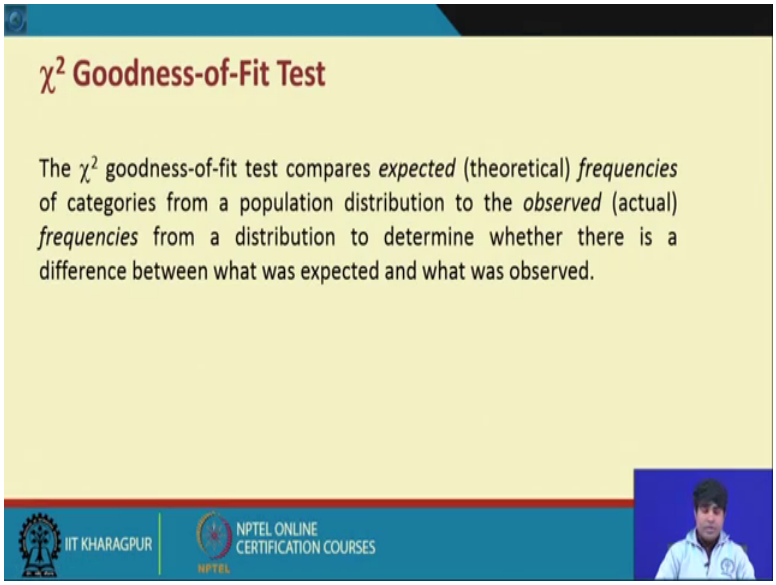


And then finally you come to this you know a final outcome. And here this is the sample the factor A and this is the factor B and this is what the interactive way; that means, this

is the SSA, this is SSB, and this is SSAB. And then it is a kind of you know a error component, and then this is means sum squares of A, and sum square of B, and this is the interactive mean sum square and accordingly. So, these are all critical values f critical values corresponding to a particular alpha levels. And this is actually f calculated. And compared to these critical, so we are not in a position to reject for the sample one that is the factor one case and in the case of factor A, B also it is also not in a position to reject. And again in the interactive, you are also not in a position to reject.

So, that means, in this particular in a sample and problem, so we do not find any significant difference. In fact, the problem is actually this may be with respect to in a small sample case, because this is actually multiple sample case means you must have actually more number of data points to validate this particular you know a problems or the kind of you know issues. And then you can get some kind of you know better inference. So, likewise you can extend this particular model for three factor models then you can go for you know completely you know four factor models and then MANOVA like the kind of you know situations right. So, typically what I like to say that you know f is a kind of you know interesting factor through which you have to discuss the particular you know problem.

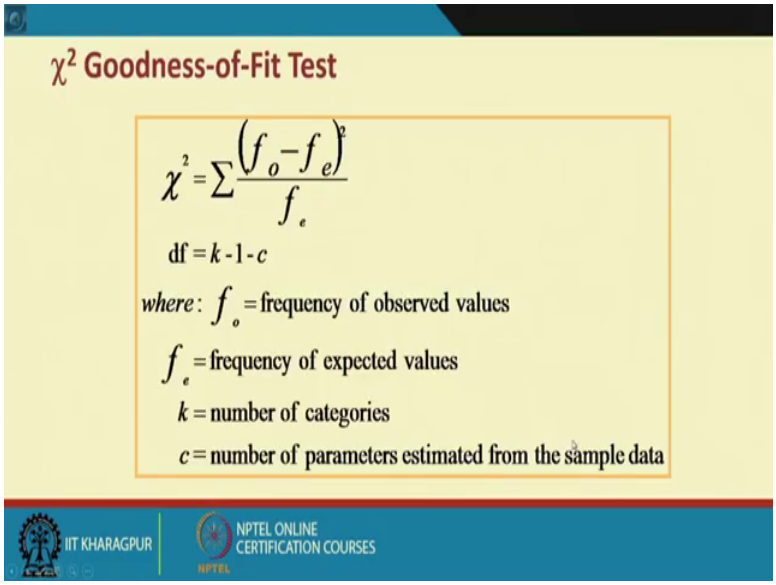
(Refer Slide Time: 33:27)



And the last but not the least in this particular you know multi sample case is the testing through chi square observations. And in fact, chi square we have already discussed with

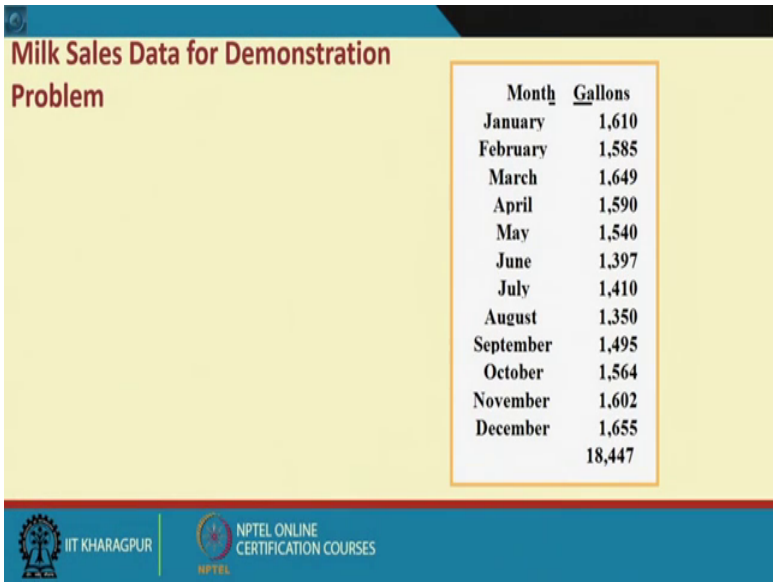
one sample case. And here is we are dealing with the chi square with one sample case and taking you to the clue actually for multiple sample case and typically for categorical variables, because variable means information may be numeric, but the variables may be kind of you know qualitative in nature. So, when a problem is having actually qualitative variables and then how you have to dealing the situations and then we will get the inference. So, chi square is the best statistic through which you can actually analyze the particular situations.

(Refer Slide Time: 34:10)



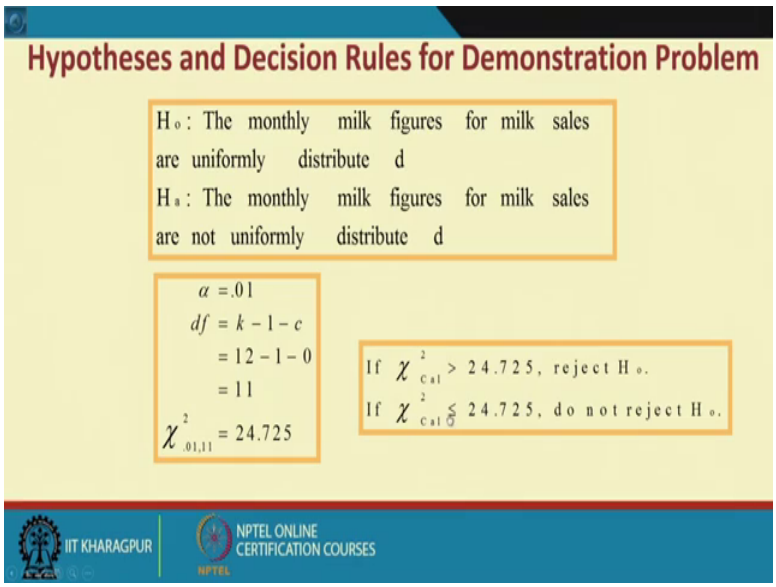
So, chi square is a goodness of fit measures and this is how the test structure through actually you have to test the procedure. It is more or less same like you know z test, t test and f test. In fact, a chi square one sample case we have already tested. So, it is actually sum total of you know the difference between observed frequency and expected frequency. And for all by the you know you will get the kind of you know calculated and then followed by the kind of you know critical and then you check whether there is a significant difference between observed frequency and the expected frequency.

(Refer Slide Time: 34:43)



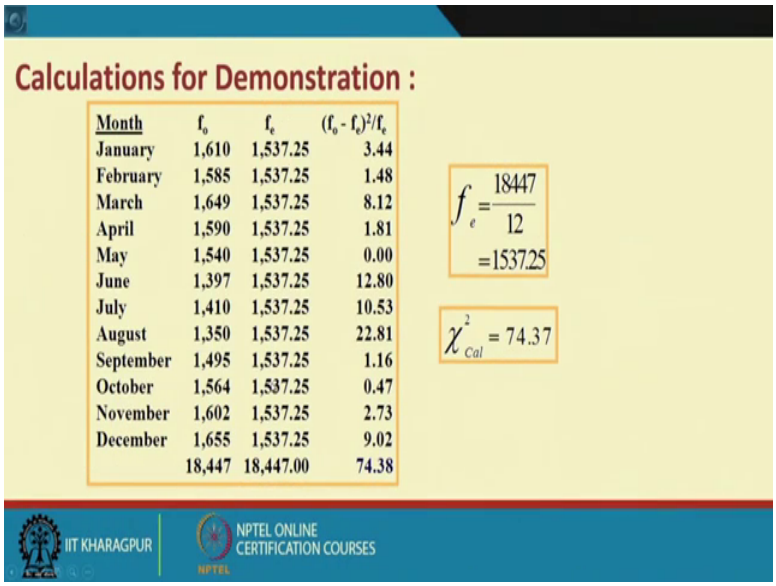
For examples, let us take example like this is you know milk sales data over the you know last twelve months of a year and we like to check whether there is a significant difference about these samples with respect to sample means. So, the total of these sales is nothing but 18,447, and as a result these are all individual sales.

(Refer Slide Time: 35:07)



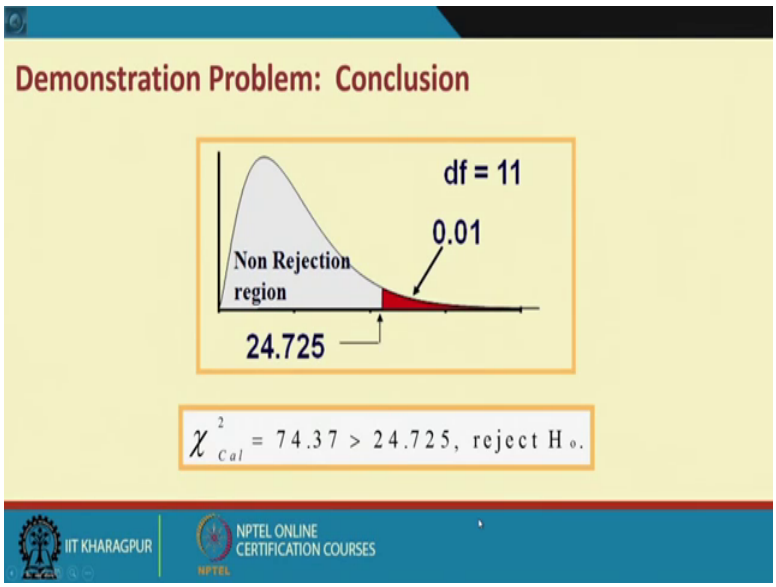
And while doing the analysis, so this is how the null hypothesis really specify and then and the type of you know the structure which you have fixed for testing. So, chi square calculated is this much, and then we will see how is the kind of you know difference.

(Refer Slide Time: 35:24)



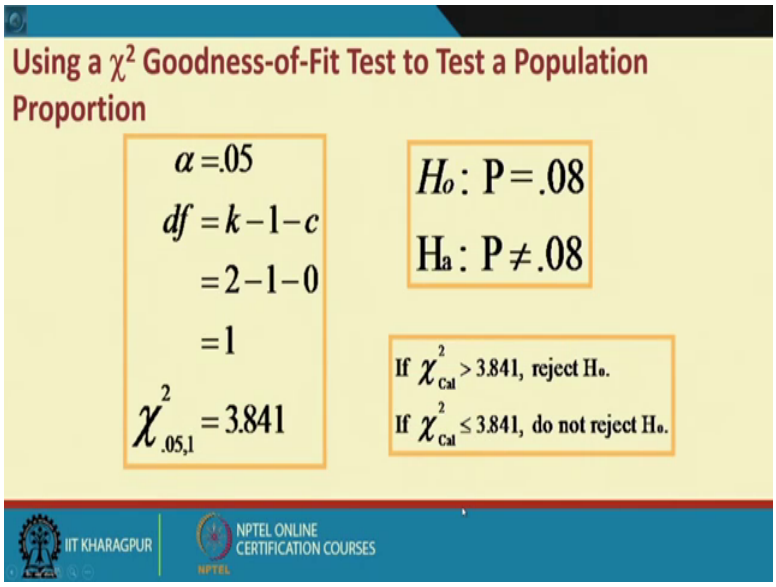
So, the procedure of you know calculation is like this. So, now, having the sample information this ones and then what will you do. So, you have to calculate the kind of you know mean and then you have to find out the difference between the individual items within a sample mean. And then with the help of you know this particular you know different squaring of these difference and followed by expected frequency, and once you have it this sum that is what is actually total chi square variance. And that need to be tested through critical and that the critical is here.

(Refer Slide Time: 36:03)



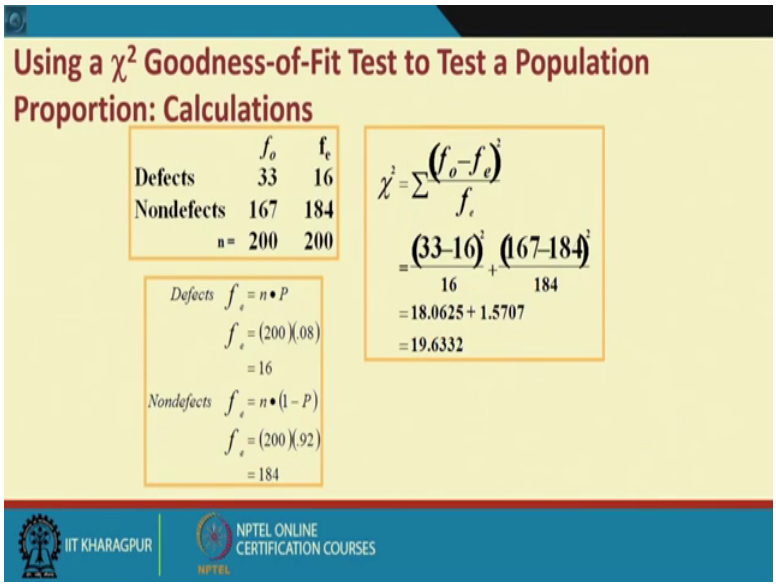
And f critical is here coming actually 24.725 corresponding to alpha at actually 1 percent. And depending upon the degree of freedom that is 11 that means, the sample size is a 12. So, the calculated value is coming 74.37. So, as a results. So, we are in a position to research so; that means, there is a significant difference between the individual items corresponding to sample you know I mean or mean of this particular you know series.

(Refer Slide Time: 36:29)



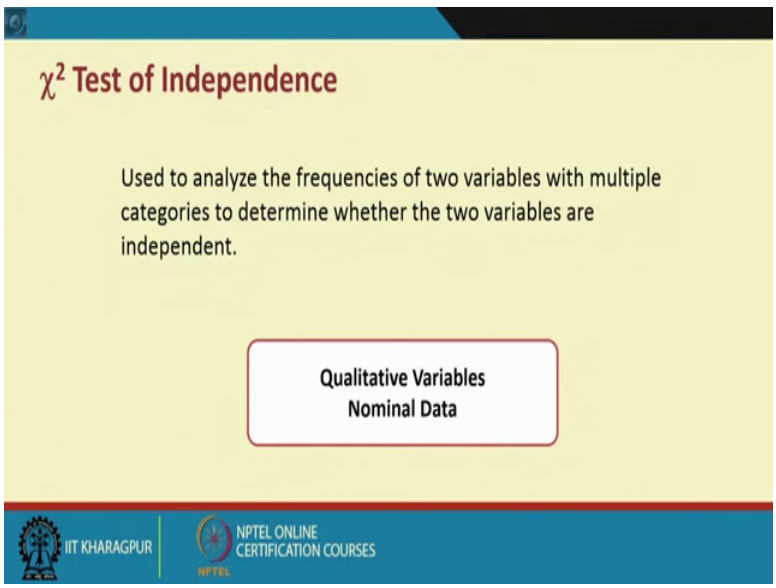
So, likewise actually you can go for you know testing for population proportions, it is more on the same.

(Refer Slide Time: 36:36)



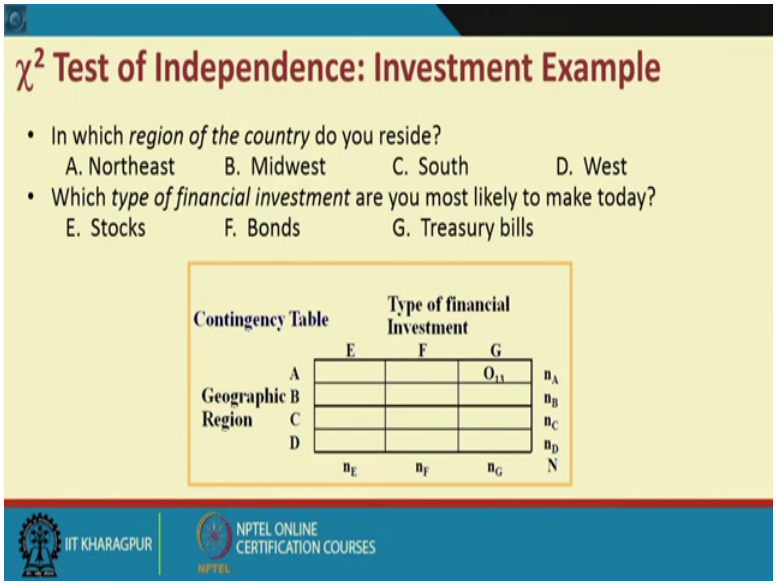
And here the problem of proper population proportion will be like this. So, this is the observed frequency then you have a expected frequency. And it is more or less same against, so you have to find out the difference. And the squaring of the difference need to be tested again compared with the you know kind of the critical value. And on the basis of this data, so we have actually calculated chi square where region 19.64 and then again that need to be tested and again which is actually in a position to reject. So, this is typically actually one sample case you know these two problems are with respect to one sample case.

(Refer Slide Time: 37:12)



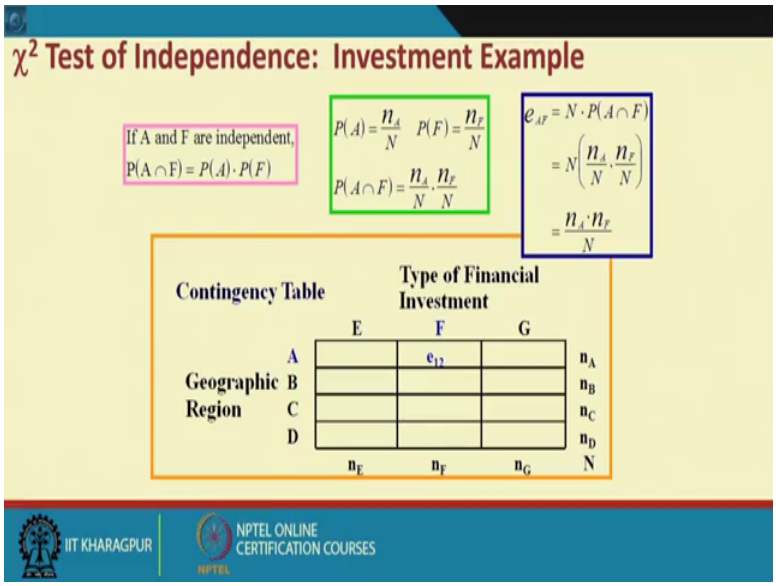
But I like to highlight you know the concept with you know qualitative variables and the nominal data structure.

(Refer Slide Time: 37:20)



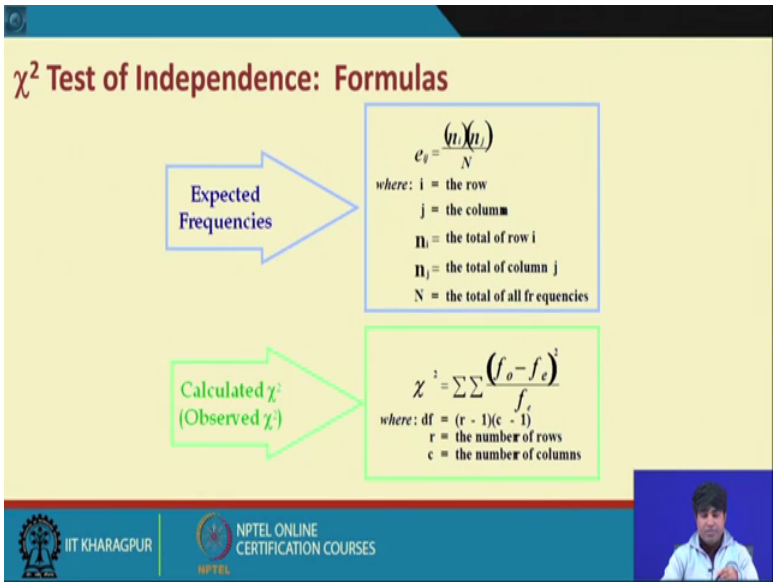
And let me stake you the problems. So, usually it is the sexually chi square test for you know test of independence and the examples will be let us say investment and then you know find kind of you know how it is varying with respect to different you know locations. It is a kind of you know multiple sample case like you know two factor model which you have actually discussed. So, here all the similar kind of you know structures. And the same problems can be analyzed with you know chi square test statistics. Let us have a look here is how is the kind of you know structure (Refer Time: 37:52) are there.

(Refer Slide Time: 37:52)



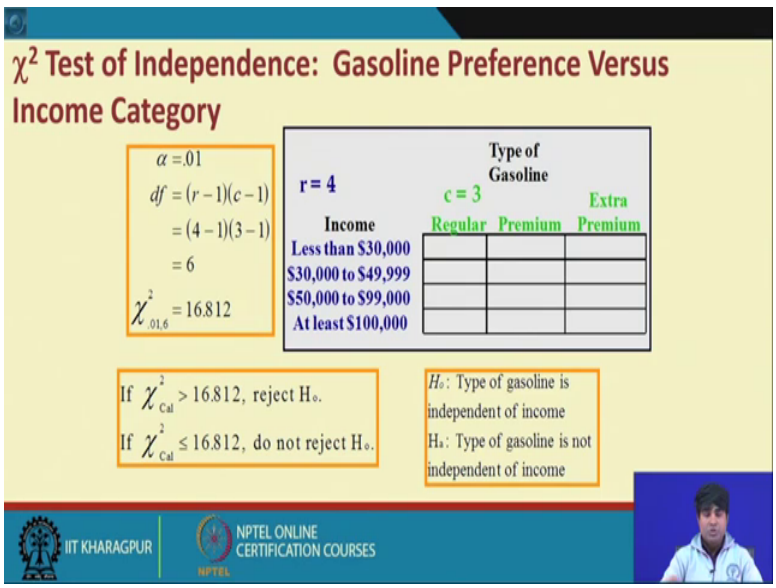
And if we add up all these column wise sums, and row by sums then we will find actually combined you know samples and then it will be tested. So, the typical idea actually like this all right. So, this is actually the kind of you know structure. And then what do you like to do here, you have to find out the again the difference between the actual of positions and the expected positions. And the expected position that means, actually in this case it is actually three into four matrix and that means, typically each entry is the kind of you know actual kind of you know observation or actual kind of you know effects. And then we will find out the expected effect and procedure to calculate the expected effect is like this.

(Refer Slide Time: 38:40)



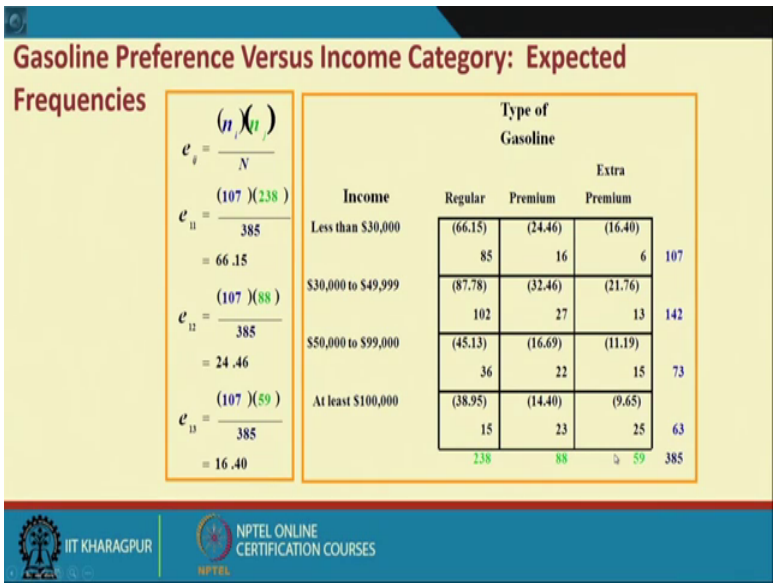
So, it is the actually row sum into column sums divided by the combined sum and that is the expected point. So, that is why n i into n j means it is i th columns and that is the i th total and the nj means it is a j th column and that to j th totals. So, the multi multiplication of that two components divided by combined means will give you the position of that you know and that to expected frequency of that particular samples. Likewise for every samples, you can find out the expected frequency and then finally, you have to find out the chi square value which is the difference between the observed frequency and expected frequency and that needs to be again tested.

(Refer Slide Time: 39:28)



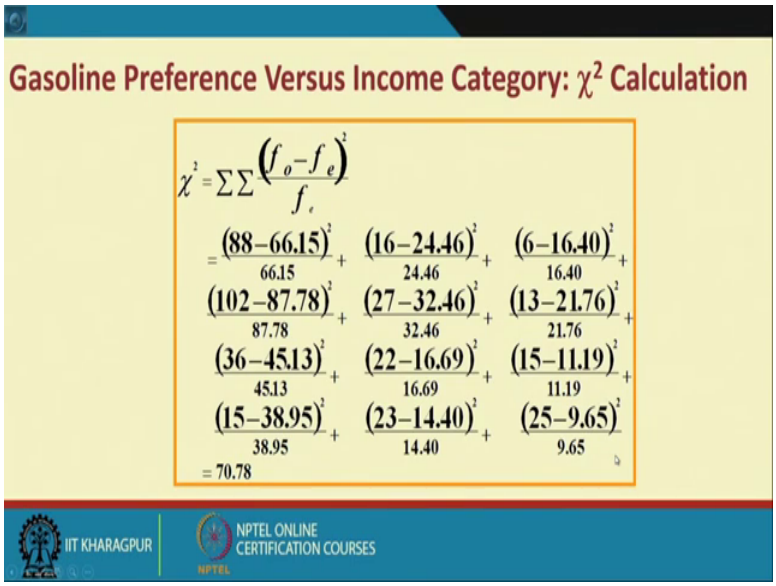
And this is the sample problems and that is with respect to income category and the type of you know gasoline. And there are three different types regular, premium and extra premium and then subject to income variations. There are four income variations. And this is how the develop the types of gasoline is independent of you know income that is the kind of you know problems which you need to investigate with these examples. And the corresponding alternative hypothesis that you know the gasoline is not in dependent of you know income.

(Refer Slide Time: 40:00)



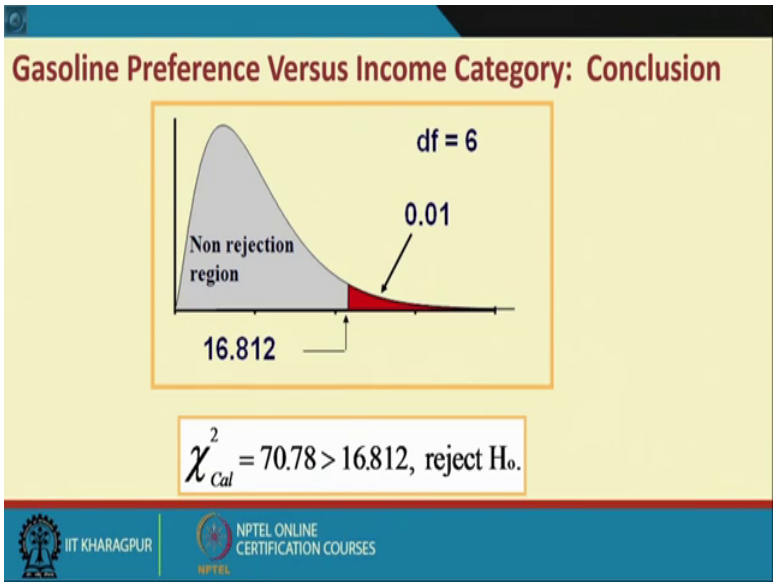
So, accordingly the sample points are like this. So, corresponding to this income and you know types of gasoline, these are you know observed frequency and accordingly with this you know some quarter or row sums and column sums for a particular positions, you have to find out the expected frequency. And then accordingly so this is how the procedure through which actually the all the bracketed figures are actually the expected frequency. And then so you will find here is for observed frequency there is a expected frequency and then you will find the difference and that is the square of the difference will you know need to be tested. So, whether they are statistically different and corresponding to expected frequency.

(Refer Slide Time: 40:45)



And the since it is actually variant statistics. So, the procedures same procedure we are following in this particular case and accordingly. So, after simplifications you are getting the total chi square is equal to 70.78.

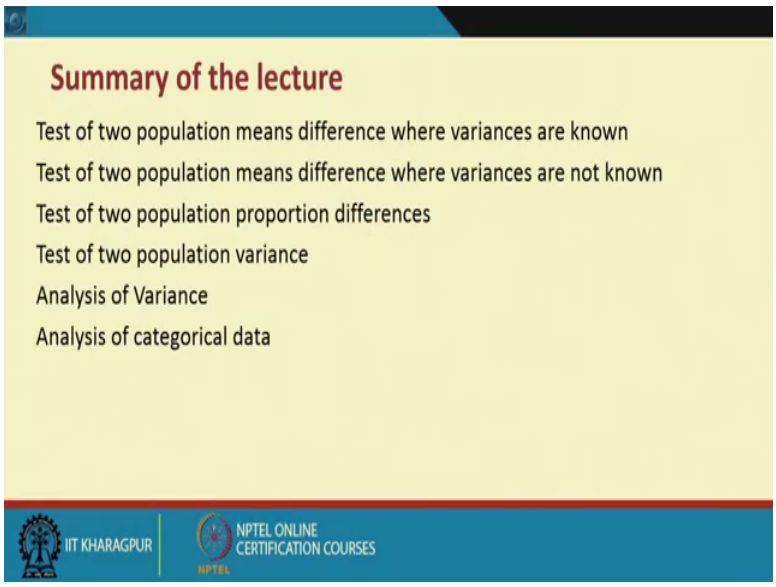
(Refer Slide Time: 41:03)



And in the case of you know critical, the critical value g is showing here 16.812 and that to four degree of freedom six that is r minus 1 into c minus 1 and that is 3 minus 1 into 2 minus. So, that is coming actually it is 4 minus 1 into 3 minus 1 that is coming actually 6 and at 1 percent you know you know alpha levels, so you will find critical value is g

16.818. So, here the calculated is 70.78 and the critical is 16.812. So, as a result so what is happening, so we are in a position to reject the true null hypothesis so that means, the hypothesis you have said here is that you know the types of gasoline is independent of income so that means they have the kind of you know you know dependencies. So, since we are you know rejecting so that means, the conclusion will be or the inference will be like that you know there is a kind of you know interdependency between these gasoline which subject to the income variations. So, likewise we have actually discussed couple of problems.

(Refer Slide Time: 42:10)



So, in this particular you know unit that is inferential analytics two, so we have actually solved some of the business problems, relating to multiple sample case and we have checked how is the kind of you know inference by using the t statistic, z statistic, chi square statistic and f statistic. And typically the variation in the kind of you know analysis of variance one factor model, two factor model, three factor models and the kind of you know ANOVA. So, that means, actually in a kind of you know a real life scenario or the kind of you know dynamic kind of you know business environment, you will find a some of the problems you know can be investigated with the multiple sample case. And again still find you know lots of you know interactive or interrelationship among these variables or the kind of you know samples. And the inferential analytics will give you the kind of you know exposures, how you have to understand these problems, and how you could you have to pick up a particular you know test statistic. And what is the kind of

you know optimum sampling, or you know the kind of you know or the kind of you know sample size which you can fix, and then how much or what is the kind of you know confident confidence interval we can build among these particular samples.

And these are the things we are actually you know supposed to know from this particular you know in inferential analytics too. So, you have to see what are the business problems. And after knowing all these techniques and the kind of you know concept the kind of you know connections, so I am very serious you know this kind of problems can be investigated very nicely or easily and then you can get some kind of you know inference as per the kind of you know business requirement. So, that means, technically you can you may be in a position to come with a some kind of you know management decision corresponding to a particular you know problems where we are dealing with a problem with multiple sample case. That means, typically a multivariate situations depending upon the kind of you know number of samples or you know number of variables, you can actually apply these techniques. And then solve some kind of you know business problems. With this, we will stop here.

And thank you very much, Have a nice day.