Econometric Modelling

Prof. Dr. Rudra P. Pradhan

Department of Management

Indian Institute of Technology, Kharagpur

Module No. # 01

Lecture No. # 12

ANOVA for Bivariate Econometric Modelling

Good afternoon, this is doctor Pradhan here. Welcome to NPTEL project on Econometric Modelling. So today, we will discuss the component called as a anova. Anova means, analysis of variance, so before I start with this concept of anova, we will like to highlight what is the objective behind anova? So, in the last class, we have discussed the reliability of the models. In fact, anova is, one of the components under reliability of the bivariate econometric modelling.

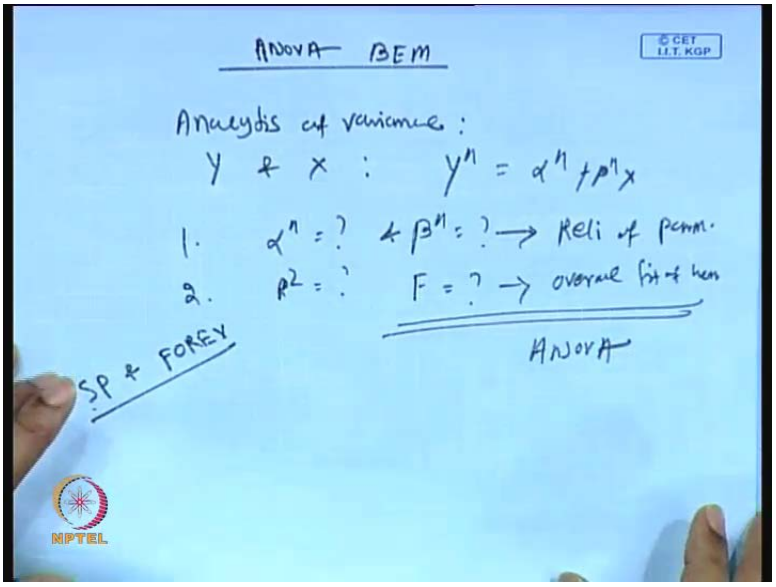
So, first of all what is the objective behind this econometric modeling? When we will go for econometric modelling, the fundamental objective is to fit the data for a specific problem and over the process; we look for a structure in a data that means, a good fit. So, there is no hard and fast rule how to get the good fit or a good structure in the existing data setups. So, there are several procedures, there are several methods, there are several techniques, so we have to apply to get a good fit.

So, this is how you need lots of knowledge, lots of skill, then, through which, we have to design a models or we have to find a structure, good structure in a existing data setup. There are, you can say lots of permutations and combinations you have to apply to get the best fitted ones or best structure in the existing setup.

So, one of such component, we can say, anova. So there are you know the basic idea behind this, you know, good fit or you can say good structure in the data set is, you start with a specific problems by using any particular techniques or any methods, then there is a certain rules and regulations through which, you can get a best fitted one.

So, the basic foundation is you start with a specific problem, so the moment you will get the estimated model then, your journey will start. So now, how to get all these details? So, now last class, we have discussed the details about the, you know, estimation of econometric models and the reliability part of the econometric model. So, today we will start with this particular specific component, anova.

(Refer Slide Time: 03:05)



So, this term ANOVA stands for Analysis Of Variance, analysis of variance the term ANOVA stands for analysis of variance. It provides information about the levels of variability when, you know regression model and a basis for test the significance, so you have to find out the reliability part of this particular model fit. So, the moment you have best fitted models, so we like to know how variability, we can have in this existing setup; so, this is what we, investigate through this component called as a anova. ANOVA stands for Analysis Of Variance, so that means, so here this particular anova with respect to bivariate econometric modelling.

So, now we have discussed for bivariate econometric modelling, we must have two variables Y and X and through which, we have fitted the model y had, equal to alpha had plus beta had X.

So, in the last lectures we have discussed, you know last couple of lectures we have discussed, several structures or several ideas behind the reliability of the estimated models; so now means, once you have the estimated model, your specific idea is to test

whether this parameters are statistically significant and in the same times, the fitness of the model must be statistical significant; so that means, we must have two specific objective, the model significance with respect to parameters and another is overall fitness of the model with respect to R square.

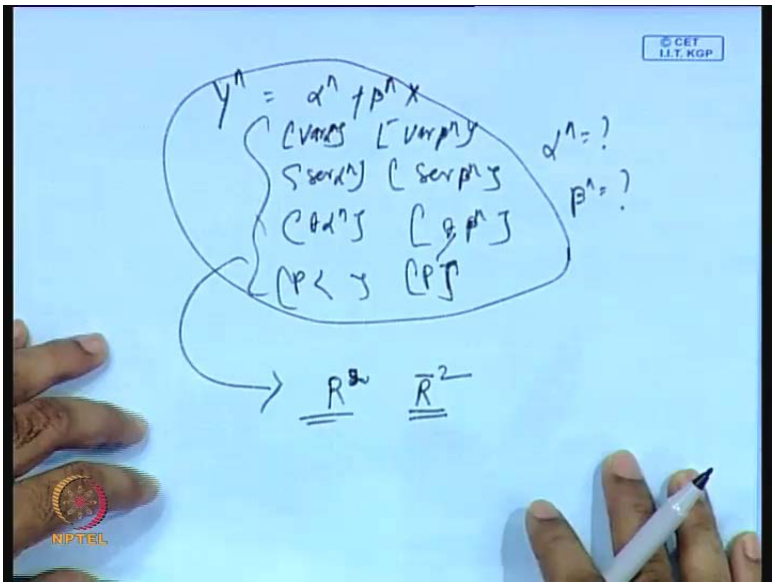
So that means, here we have two specific objectives, you like to know alpha hat should be significant, whether it is significant or not and beta hat should be significant or not and second to get the value of R square which, represents the you know degree of model fit, the degree of model fit means, with respect to best one; and the whether, this R square will be statistically significant, this is how we have to decide and through which, we have to prepare a structure best structure within the existing setup.

So that means, what we means what is our agenda? So, with the available information we have to fit a model, then, we have to look for the best ones. So, there are many ways we can fit the, you know, data with respect to existing problems even if there are two variables say, you can say you can say stock price and you can say forex for an exchange.

So, even if we will fit the models, whether stock price depends upon forex or (()) depends upon stock price. So, still you know there are many ways, you can start the modelling. So, there may be with respect to data transformation, with respect to means technique transformation or with respect to assumption means relaxation, so many things are there so, that you will get the best fitted ones.

So, you go any process, so idea must be to get the best structure within the existing setup or to get the best one within the existing system. So now so, the first part is called as a reliability of the, reliability of the estimated parameters reliability of the estimated parameters and second part is the overall fitness of the model, overall fitness of the fitness of the model; so, this overall fitness of the model test is otherwise known as called as a anova, so what is this anova structures?

(Refer Slide Time: 06:51)



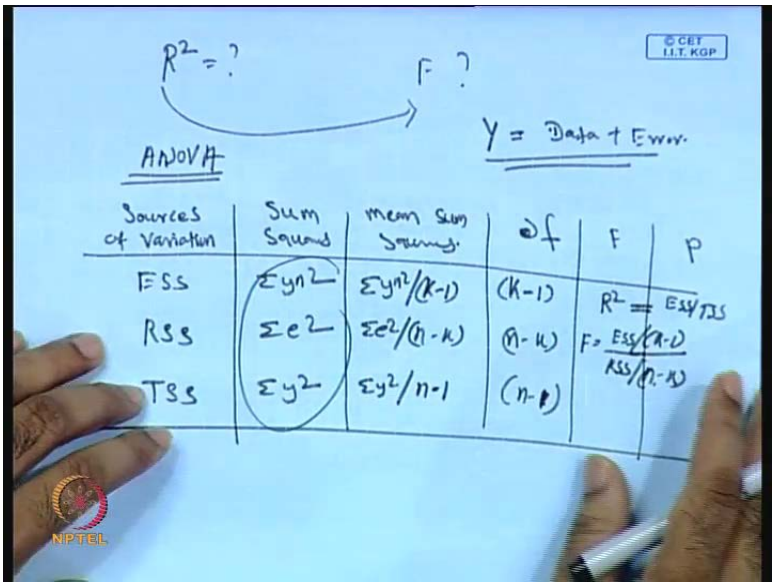
So now, once you have this, you know estimated models, so the moment you have estimated models. So, y hat equal to alpha hat plus beta hat x then model model information must have, you know, the structure we like to have is nothing but, variance of alpha hat then, variance of variance of beta hats. So, then standard error of standard error of alpha hat, then standard error of beta hat and you know, t of alpha hat and t of beta hats and probability level of significance and probability level, probability level of significance, this is for beta.

So, this particular setup is, mean to know, whether alpha hat, alpha hat is significant one or beta hat is significant one. So, within the existing information, we like to know or another component called as a R square and adjusted R square. R square represents the, R square represents the percentage of variation explained by the regression that means, what is the influence of the independent variable to dependent variable in percentage term.

So, that is how, we calculate, like the ratio between explained sum square by total sum squares, so explained sum square will represent in terms of the influence of independent variable and total sum square is nothing but, the influence of dependent variables. So, we like to know the, you know degree of influence through which, through the independent variable to dependent variables. So, that is how, we have represent the R square and

adjusted R square. Adjusted R square is just to you know incorporate with R square and that to degrees of freedom, so we will discuss detail what is the existing setup.

(Refer Slide Time: 08:53)



So, what is all about this, you can say, anova all together, so anova specifically deals with two things, so let us start first with this particular you know, information tables, so what we look for; so, we look for R square statistics and we look for F statistics, so, that means, we like to know whether, this overall fitness of the model is statistically significant. For this we need to have a, prepare the anova that is, an analysis of variance tables.

So, what is this anova, analysis of variance tables? So now, we like to have sources of variations, sources of variations, this is sources of variations, then sum squares, sum squares, then mean sum squares, mean sum squares, mean sum squares, then, degrees of freedom, then F statistics and the probability level of significance. This is how you have to prepare the tables. So, the table information contents, what is the sources of variations; sum squares means, sum squares; degrees of freedom and F statistics and probability level of significance.

You see the idea behind the, idea behind this particular problem is so, when we will look for, you know, econometric models, so we have two different structures, so that means, we have Y equal to, you can say x component, that is, the data plus you know error component. This is what we will call it error component. So, the existing setup, the

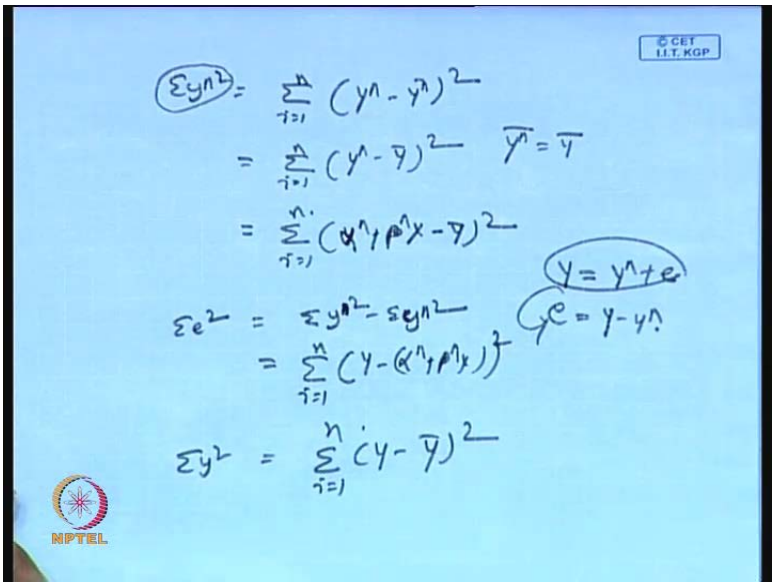
existing setup is you know, with respect to, means what is the influence of X information on Y and what is not captured by X that we will represent in terms of, you can say error terms.

So, we like to know what is the percentage impact on a X variables, which is explained to us and which is the not explained, that is called as an unexplained which is, represent by error component. So, obviously, there are two sum squares. So, first sum is with respect to explained sum squares. This is called as explained sum squares, then, there is term called as a residual sum squares, then, there is term called as a total sum square, explained sum squares, so, which is nothing but, summation y hat squares.

So, this is summation e squares and this is, summation y squares. So now, so, so far as, so far as a mean sum square is concerned, it is with respect to degrees of freedom so, that means, here is y hat square divide by you can say, k minus 1 or it is k divided by k minus 1. So, this is nothing but, summation e square by n minus k and this is nothing but, summation y square by n minus 1, n minus 1.

So now the degrees of freedom, accordingly, this is k minus 1, this is n minus k and this is n minus 1, this is n minus 1. So now, we need to calculate f statistic with the help of above information. So now, f statistic is nothing but, the ratio between explained sum squares divided by residual sum square, provided, it is the, you know, weightage factors of k minus 1 divide by r minus k means, it is n minus k. So, this should be followed by r square statistic, which is nothing but, ESS by TSS, ESS by TSS. Let me exactly highlight what is this issue about these three, these three, is you know nothing but, like this.

(Refer Slide Time: 12:38)



So now, our idea is, here is, our idea is, here is, summation y hat square. So, which is nothing but, summation i equal to 1 to n.

Now, y hat minus y hat bar whole squares. So, this is how summation y hat square because, this is in small deviation format, so now, if we will simplify further. So, this is nothing but, summation i equal to 1 to n, then y hat minus Y bar, Y bar whole square because, we know Y hat bar is equal to always Y bar. So, this we have discussed long back.

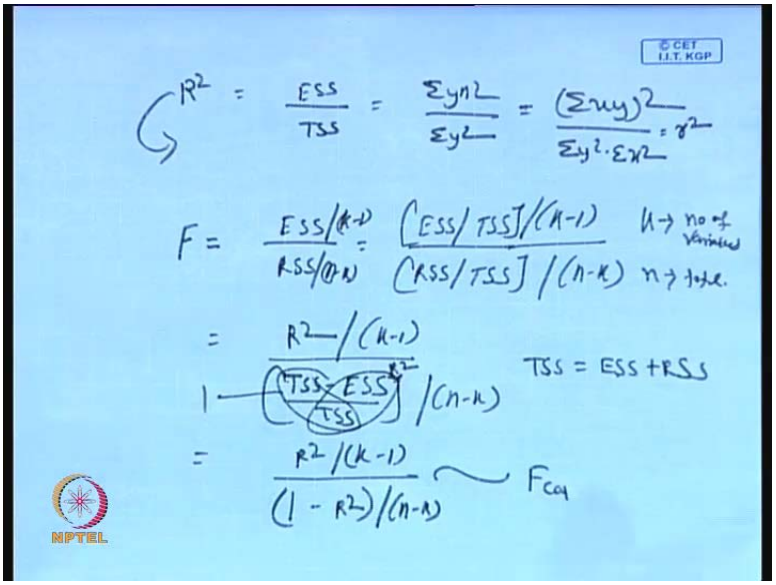
So, if we will simplify further, then this is nothing but, summation alpha, alpha hat minus sorry plus beta hat X minus, Y bar you can say whole square, i equal to 1 to n. Similarly summation e square, summation e square that is, unexplained sum or residual sum is equal to summation, you can say, Y hat squares, summation Y hat squares, summation y hat square, summation e square equal to summation Y square minus summation Y hat square, summation Y hat square. So, this is the difference between these two.

So, that means, what we will, how we will write, its means e is the devia…, e is nothing but, y minus y hat because, the overall structure is Y equal to Y hat plus e. So, this is how error error component is defined. So that means, it is nothing but, Y minus Y hat. So obviously, when we will go for deviation, then it is nothing but, Y minus alpha hat minus beta hat X, so, to the power whole squares i equal to 1 to n. So, third component is summation y square which is nothing but, summation Y minus Y bar whole squares. So,

i equal to 1 to n. This is how, the entire structure of this particular, you know, analysis of variance.

So, analysis of variance deals with three specific components to check the reliability part of the model or to check the best fitted models. So, first component is explained sum squares then, second component is residual sum square and third component is total sum squares. So, we like to know, what is the ratio between explained sum squares to total sum squares and we like to know what is the ratio between explained sum squares by residual sum squares.

(Refer Slide Time: 15:31)



So, the ratio between explained sum by total sum square is represented by r square component that is otherwise called as coefficient of determinations. This is otherwise known as coefficient of determination. That means, we look for two things first is r square which, is nothing but, explained sum square by total sum squares and this is nothing but, what is what is ESS? ESS is nothing but, summation y hat square divide by summation y square. If we will simplify, then, the entire structure will be coming summation x y whole square by summation y square into summation x squares.

So, this is nothing but, you can say, this is nothing but, you can say R squares, correlation coefficient. So, then, corresponding to R, R we like to know whether this particular component is statistically significance. So, that is how we have to require use F statistics. So, F statistic is nothing but, the ratio between explained sum square by

residual sum square. So, what we have to do. So, we will divide explained sum square by total sum square and we will divide residual sum square by total sum squares. of course, there is degrees of freedom here, k minus 1 and here, degrees of freedom is n minus k.

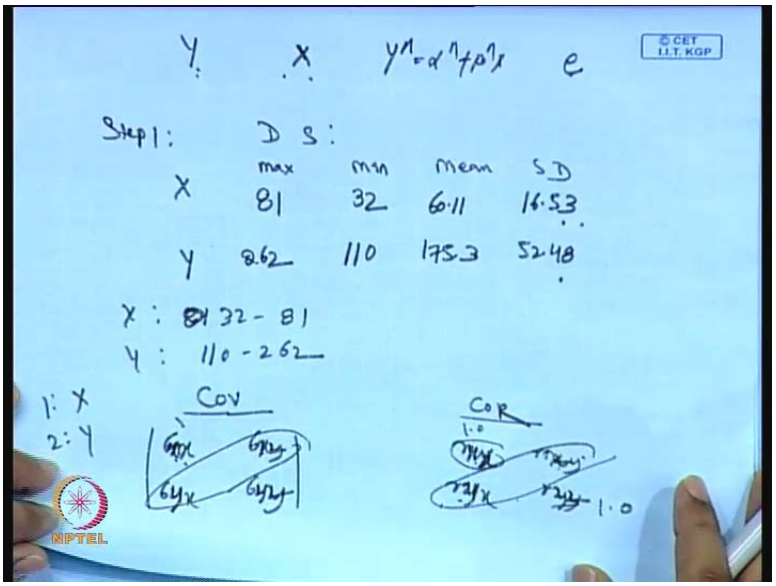
So now, what we will do here. So, this is, this is one component provided, there is degree of freedom. So, k minus a and this is, this degree of freedom is n minus k. So, k represents number of number of variables in the systems or number of parameters in the system and represents total number of observations, its number of total number of observations in the systems.

So, now this e s ESS by TSS is nothing but, r squares. So, it is nothing but, r square by k minus 1. So, divide by RSS, RSS is nothing but, e a, TSS minus you can say ESS because TSS is equal to ESS plus RSS. So obviously, RSS equal to TSS minus ESS divide by TSS and whole divide by n minus k. So, this is nothing but, R square upon k minus 1 divide by this particular term is equal to 1. So, this means 1 minus R squares because, this is one component and this is another component, this is R square and this is equal to 1.

So now, this is 1 minus R square. So divide by, divide by n minus k. So now, this particular structure is called as a F calculated, F calculated. So, we like to go for you can say again F tabulated value with, you know, proper level of significance, means probability level of significance, then like you know, reliability of the parameters. So, we like to judge the overall fitness of the models.

So now, I will take the similar problems, then, I will highlight, how we will go for this, you know, overall fitness test, whether this particular means model is good fitted or you can say best fitted. So, we have to start with this particular, you know reliability of the estimated parameters, then we have to integrate with the analysis of variance, because, most of the components, we have to derive from this parameter value only. So, so, what is the technical procedure of all these issues?

(Refer Slide Time: 18:56)



So now, we have two variables, so Y Y variables and X variables. So now, over over the time frame the moment will integrate Y and X. So, we will get the estimated model y hat equal to alpha hat and beta hat X. So, then the moment you will get Y hat equal to alpha hat, alpha hat plus beta hat X then, with we can create another variable say Y hat which is means sorry we we will create another variable e which, is the difference between Y minus Y hat.

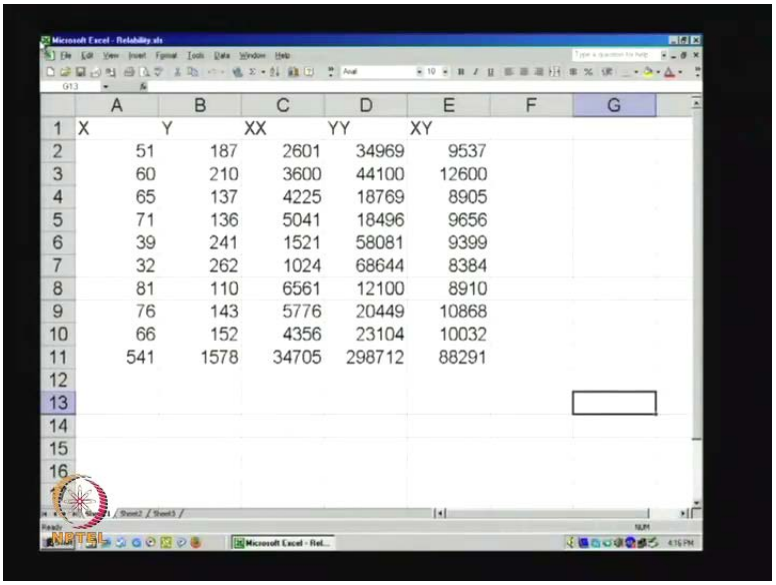
So, that means, here we start with y u y and x. So, we will get Y hat which, is nothing but, alpha hat plus beta hat X. So, then this is first variable, this is second variable and we will get another variable Y hat and with help of Y and Y hat we will get the error component e. So, all together we start with two variables then, ultimately we end with four variables, so Y, X, Y hat and e. So, there are lots of gaps between Y and X then Y Y hat then again you know Y hat e.

So, we like to know how, whatever you know, variables you will use or how these way you will use, ultimately, the objective or the agenda is to get the best fitted models. So, that means so, whatever structure we have created so, that structure should be best structure for you, so, so far as a data fitting is concerned. So, we are just, we have a problem actually. So, we have information that is what we represent in the form of data.

So, the data has to be you know, properly applied to analyze that particular problem and the the way you will fit the data to analyze that problem. So, that fit must be best fit. So,

that is how, we are doing all these jobs. So now so, we Y , X and you have got y hat and e then, ultimately, so we will take the similar problem, here is the problem with respect to X and the problem with respect to Y.

(Refer Slide Time: 20:59)



So, in the last class we have discussed, so far as a reliability of the parameters that is alpha test and beta test. So now, we take these similar problems, then we will integrate this particular, this significance of the parameters to significance of the overall fitness of the models.

The reason is that if one is, you know, in favor of you and another is not in favor of you, then the model reliability is in the wrong side, wrong side. So, that is why the the accuracy will be very high, model accuracy will be very high or reliability is very high, when all the parameters are statistically significant at the higher levels and in the same times your overall fitness of the model also significance at the higher level. If this is correct and this is not correct or this is correct or this is not correct, then, it will go go against the model fit.

So, that is why, we have to redesign, restructure, you know. reframe, so that, both the objectives can be go for a value. So that means, your parameters should be statistically significant and in the other side, your overall fitness of the model should be statistically significant. So now, taking this particular problems, so, here X contains nine observations and Y contain nine observations, so that means, the first part of the

modelling is sample observations are uniform and you can proceed because, with two variables and nine samples, it is possible to estimate. In fact, if the model information will be very high then, the model fitness will be also very high accordingly.

Since it is in the class, it is not possible to go which, u set of data because, the moment you will take u set of data then, you need statistical software. So, in the, in the beginning we should not start with any statistical software. So, whether, we start with a small problem we like to know actually or our objective is here to know what is the structure and how do we get this reliability or how to test this reliability with respect to estimated parameters and with respect to the overall fitness of the models.

So, that is why you know, artificially we have created a small data set with respect to two variables, so, that we can we can establish the concept very carefully. So now, the moment you have this much of information then, we have to proceed step by steps. So, what should be the first step and how do we go for our final objective.

The final objective is to know the reliability, means, significance of the parameters, estimated parameter that to alpha hat and beta hat and second objective is the overall fitness of the model, that is, the significance of r r square. So, for that we have to integrate with f statistics. So now the step 1 process, the step 1 process is here is. So, we like to know the descriptive statistics. So, as usual, we have calculated already descriptive statistics for X variables and Y variables. So, what is descriptive statistics? So, we are reporting a descriptive statistics is, in fact, it is a, it is a very complex and it is a multivariate in nature, because, descriptive statistics includes so, many, so, many statistics like mean, median, mode, maximum, minimum, standard deviations, skewness, kurtosis etcetera.

So, in the mean times we need not require to represents all these details simultaneously because, when we will go for hardcore reliability part, then, that time you need to have all such information. In the mean times, we we we specifically highlight few things because, these few things are very relevant so, far as a reliability check is concerned.

So now, first first item we will consider is maximum, then second item is minimum, then mean, then standard deviations. So now, we like to know for X what is the maximum and for Y what is the maximum. Similarly for for X what is the minimum and for Y what is the minimum.

So now, if we will make here, is the standard procedures is without any calculation. So, just you arrange it ascending, descending very quickly you can guess it. Otherwise, you just give a common. So, automatically there is here mathematical properties and statistical properties through which, you can get this you know statistics, discrete statistic that is with respect to maximum and minimum. So, for this X the maximum item is 81 and the, for Y the maximum item is 262. So, minimum item is for X is 32 and minimum item for Y is 110. So, that means the, for X the range is from sorry 32 to 81 and for Y the range is from 110 to 262.

So now, you have X information and Y information. We like to know how these information are perfectly, it can be fitted. So, that is our objective. So now, mean is here, for X series it is 60.11, that is sum X by n and here n represents total number of observation that is equal to 9 here. So, this is mean for X and mean for Y is nothing but, 175.3, corresponding mean we have a standard deviation here 16.53, then here, standard deviation is 52.48; that means, for Y series the standard deviation is 52.48 and for X series the standard deviation is 16.53.

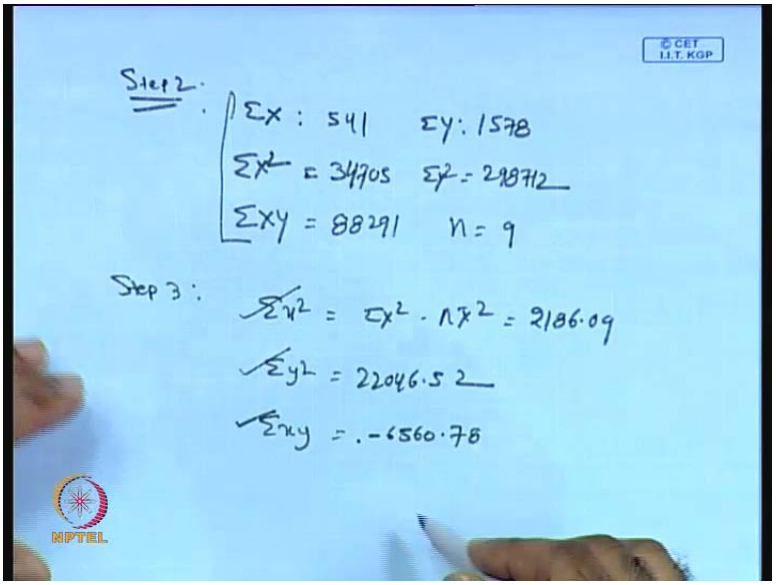
So now, with respect to, with respect to all these information you can take another, another two matrix, that is you can say covariance matrix and you can say, correlation matrix. It can give you little bit indication whether there is any association between these two variables. So, far as a covariance matrix is concerned, it is nothing but, sigma 11, sigma 12, sigma 21, sigma 22.

Similarly, here r 11, r 12, r 21, r 22, so this particular structure is nothing but, this is with respect to, you can say, 1 transfer here X and 2 transfer here Y. So, that means, we can call it here instead of 11 we can call it sigma x x we can call it sigma x y, similarly sigma y x, then sigma y y. So, this is another way you can represent also. Similarly this is r x y this sorry this is r x x, this is r x y, this is r y x and this is r y y.

So, this particular structure is very symmetric in nature, this is very symmetric in nature and in this particular case its always equal to 1.0, this is always equal to 1.0. So, we like to know what is the association between these two with respect to covariance, with respect to correlation? So, I am not calculating all these details because, sigma axis is nothing but, variance of x. So, which, and if you will calculate, if you will square root then, you will get the standard deviation, this is how it is directly derived.

So, similarly we will get the correlation coefficient. It is nothing but, covariance of x y by variance sigma x and sigma y, standard deviation of x and standard deviation of y. So, with this, these are the basic information you need to incorporate before you go for the reliability test because, these are all essential elements through which, we can observe this issue. So, this is the first step of this particular reliability check, or you can say analysis of variance.

(Refer Slide Time: 28:32)



So, in the step 2 in the step 2, what you have to do, you report all these, you know, summary of this particular information. So, what is the summary of this particular information, so the summary of information, because we have two variables. So, the summary of summation is, summation X first this is 541, then, summary sum of Y information is 1578, so this is 1578, so, sum and this is sum X is 541, so now, we like to know what is summation X square, summation X square is equal to here 34705, 34705 and summation Y square is here, is equal to 298712, so, 298712. Similarly and there is another component summation XY which, is nothing but, 88291, 88291 and obviously, n is equal to 9.

So, this is, this is in the step, to this is first step, first step means, first process. So, corresponding to step 2 we can, we will call it is step 2, then in the step 3, we transfer this into every way you know, small letter that is in deviation format so that, this

deviation format will help you to go for this analysis of variance because, if you will go with original samples not in deviation then, it will take lots of time.

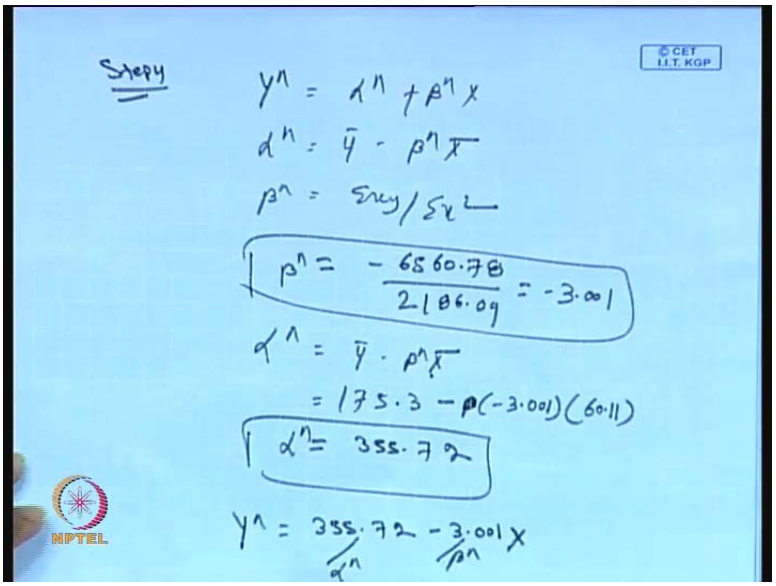
If it is classroom problem so obviously, you need to solve the problem quickly. So, as a result its better means, the test can be, in less time, if will move from this original sample to deviation problem because, it it will give you quick results. So, that is how we have to transfer this entire concept into deviation format.

What is the deviation format? So, we like to know what is summation x square, we like to know what is summation y square and we like to know summation x y. In fact, this summation x square derivation, we have discussed in the last class, summation y square we have also discussed in the last class and summation x y, this is also derived last class. So, directly we can specify here summation x square is nothing but, summation x square minus n x bar square. So, if we will simplify, we will get 2186.09, so that means, summation x square is already here. So, you just calculate the x bar then n into x bar square, if we will subtract from this summation x square you will get summation small x squares.

Similarly, summation y square is equal to here 22046.52, so similarly summation x y equal to minus 6560.78, so that means, since covariance is coming negative, so obviously, so, this particular, this particular particular you know, item represent that. So, they are negatively related to each other. So, because variance is always positive for x and variance is also positive for y. So, the covariance can be positive, can be negative and that will know the nature of relationships.

Since it is negative, so obviously, the representation is that, they are negatively related to each other. So, so this is the step 3 process of this particular analysis this is the step 3 process of this particular analysis. Now, what we will do. So, we like to know or we like to go for this you know estimated parameters because, through the estimated parameters, we will observe certain things and that will be very helpful for this you know anova - analysis of variance.

(Refer Slide Time: 32:20)



So now, what is this? This we have to move to step 4 . So, what is the step 4 ? This step 4. Idea is here is, so this step 4 idea is here is, to discuss the reliability part of the models. So, what we have to do. So, the moment you will get all these information here is so, within the information, we like to calculate first alpha hat and beta hat.

So now, step 4 you create the model, here is Y hat equal to alpha hat plus beta hat X alpha hat equal to Y bar minus beta hat X bar and beta hat equal to summation x y by summation x squares. So now, we have already summation summation x y here. So, sum summation x y here is. So, these summation x y we can put it here is. So, this is nothing but, minus 6560.78 divide by summation x square, summation x square is here is. So, 2186.09, 2186.09.

So now, if we will simplify then, we will get beta hat equal to simply beta hat equal to minus 3.001, so that, we have also calculated today morning, so now, alpha hat is equal to, alpha hat equal to y bar minus beta hat x bar. So, which is nothing but, y bar, what is y bar here? So, y bar we have already calculated so, that is what is y, y bar here is.

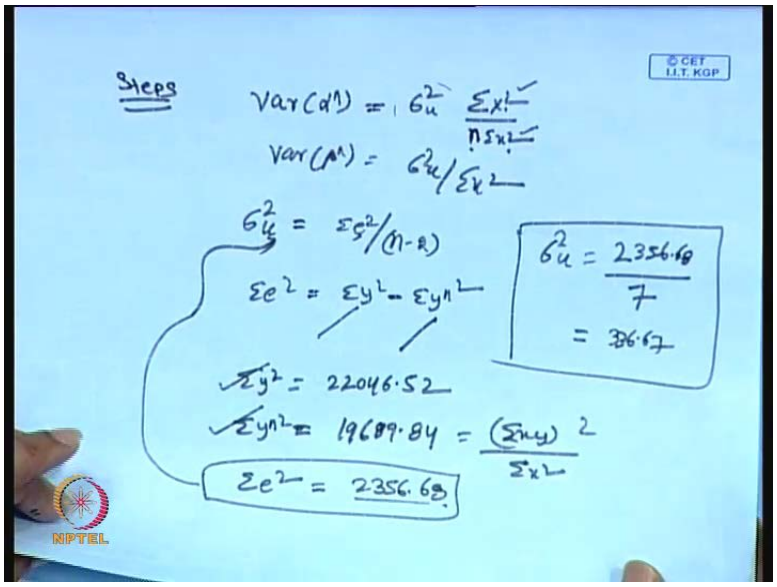
So, y bar is here is 6.11, 60.1 sorry 175 75. 75.33 minus beta hat, beta hat is minus 3 into 001 into this x bar, x bar is here 60.11, so this particular item it is, 60.11. So, this is how the alpha hat is derived. So now, with the help of alpha hats what we will get it. So, if we will simplify then, the alpha hat component will be coming 355.72 355.72, so this is alpha hat value and this is beta hat value.

So, now the estimated model will be y hat equal to 355.72 minus, minus 3.001 X, so, this is alpha hat and this is beta hat. This is alpha hat and this is beta hat, so, this is alpha hat, this is alpha hat and this is beta hat, so that means, this this particular item represents alpha hat and this particular item represents beta hat and slope is coming negative. So, by default it is negative related to each other.

So, now after getting all these things, so, next item is to check whether, this particular parameter is statistically significance, means, this particular value is significant and this particular variable significant and for that you need to you need to calculate the variance of alpha hat and you need to calculate variance of beta hat and followed by standard error of alpha hat, standard error of beta hat, t of alpha hat and t of beta hat.

The moment you will get t of alpha hat and t of beta hat then, you have to compare with tabulated value then, you can get to know whether, it is statistical significance and if it is statistical significance, at what level, that what we have already discussed in the last class. Just, I am just integrating once again, so that the foundation or you can say integration can be perfectly you know, so all right.

(Refer Slide Time: 35:55)



So, now we we have to move to step 5, we have to move to step 5. Step 5 idea is to check, find out the variance of alpha hat, variance of alpha hat is nothing but, sigma square, sigma square u summation X square divide by n summation x square. So, this is capital X, this is deviational x, small x. So, now variance of beta hat, variance of beta hat

equal to sigma square u by summation small x square. So, now, we know summation x square, we know, small summation small x square, we know the value of n, but we have no idea about sigma square u.

So, what is sigma square u? Sigma square equal to error variance which, is nothing but, summation e square by n minus 2. This is nothing but, n minus 2. Actually it is n minus k, but since it is a bivariate model, so, k represents to, k represents number of variables in the system or number of parameters in this particular econometric system. So, since it is a bivariate setup, so, k equals to 2.

So obviously, the error variance equal to sum summation e square by n minus 2. So now, what is summation e square? Now for that, so, summation e square equal to summation, summation y square minus summation y hat square, summation y hat square, summation y hat square, so which, we have already derived, how it is coming? So, what is our our agenda here.

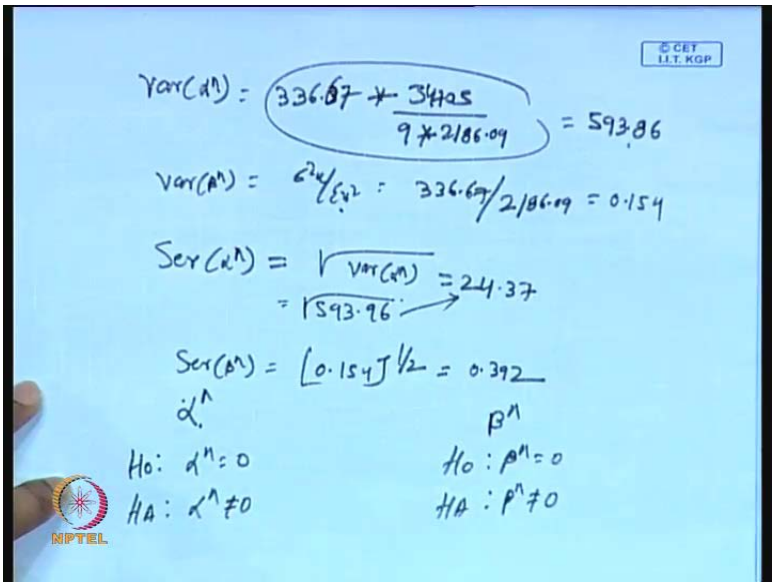
So, we like to know first what is summation y square and what is summation y hat square. So, by calculation we have summation y square is equal to 22046.52, which we have already calculated. Then, summation y hat square is nothing but, beta hat square summation x square, so, which, is nothing but, summation y hat square is nothing but, 196 a 19689.84. So, this is somewhat you can say summation x this particular item is nothing but, summation x y whole square by summation x square so, we have already derived all these things. So, this just we are reporting this value, summation y square here, summation y hat square here.

So, we like to know what is summation e square? Summation e square is nothing but, summation y square minus summation y hat square. So now, if we will subtract then, the final outcome will be a 2356.68. So, this is summation e square and that is how we have to, we have to integrate here in this particular structure because, we ultimately like to know what is the error variance?

So, ultimately now, what is error variance? So, error variance equal to sigma square u summation e square by n minus 2, so summation e square is coming like this. So, this is nothing but, 2356.68 divide by n minus 2, n minus 2 is nothing but, 7 because n is 9 and k is k is obviously, 2. So, n minus 2 represent 7, so if we will simplify then, it will be coming 336.67. So, this is how, it is the error variance. So now, with help of error

variance, you can able to get or you can get or you can able to get this alpha, variance of alpha hat and variance of beta hat.

(Refer Slide Time: 39:21)



So, now now what is the error variance of alpha hat here? So now, variance of alpha hat is equal to 33 a 336.37 sorry336.67. So, this is what we have derived right now, so, that is summation e square, sigma square summation e square by n minus 2 multiplied by summation X square by capital X square by n summations small x square.

So, summation capital X square value is 34705 divide by 9 into 2186.09, so, if we will simplify this particular structure ultimately, you will get the item called the number, called as a 593.86, so, this is variance of alpha hat. So, similarly we will get variance of beta hat. So, variance of beta hat is nothing but, sigma square u by summation x square. So this is nothing but, summation e square is 336.67 divide by summation x square, which, we have already derived, 2186.09, so, this is nothing but, 0.154.

So, this is variance of beta hat, so but, actually we need standard error of alpha hat and you need standard error of beta hat. So, what we could do, we have to do, we have to get standard error of, we need to have standard error of alpha hat and we need to have standard error of beta. Standard error of alpha hat is nothing but, variance of, variance of alpha hat and variance alpha hat and variance of beta hat.

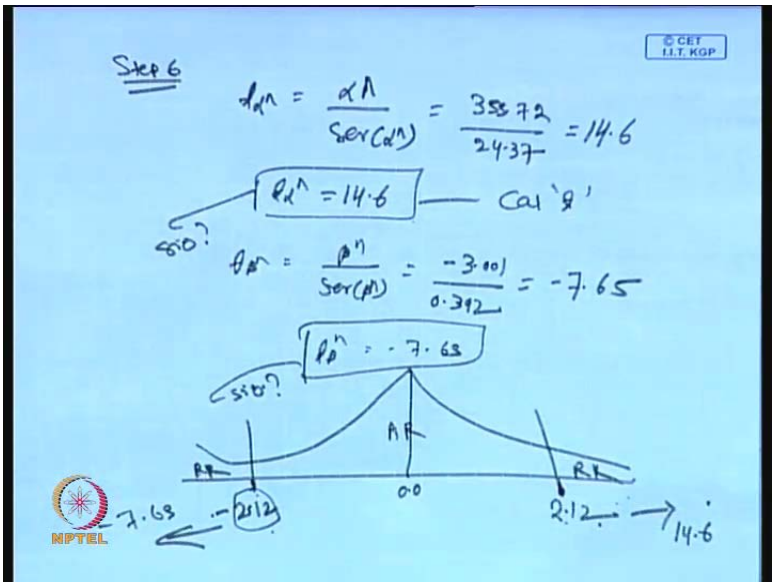
So, standard error of, standard error of alpha hat is variance of alpha hat which, is nothing but, coming 24.37. Because, that means, this particular 5, this is nothing but, 593.86 square root. So, this is, this will be coming 24.37, so, standard error of beta hat is equal to variance of beta hat, so 0.154 to the power 1 by 2. So, which is coming 0.392? So now, we have received variance of standard error of alpha hat and standard error of beta beta hat.

So now, we have to go for hypothesis testing because, we like to know, like to know whether, this particular item is statistically significant and for that we need to set a hypothesis. So, what is this hypothesis you have to set? Hypothesis for alpha and you have to set the hypothesis for beta.

So that means, we like to know whether, statistically, alpha parameters estimator alpha parameter is statistical significant or not or beta parameter is statistically significant or not. So that means, the the item will be significant if it has some value. So obviously, we will we start with that, this this item has no value then, you reject the statement. So, because last class, I have discussed, so, the moment you go for hypothesis testing then, you start with the integration of two hypothesis called as a null hypothesis and alternative hypothesis. Null hypothesis means, they, you have to assume that the two statement is wrong. So, then you have to reject this wrong statement and we will coming to the true fact.

So, for that we have to assume that here alpha hat, we have to assumed hypothesis such that null hypothesis, alpha hat equal to 0 against alternative hypothesis alpha hat not equal to 0 and similarly for beta hat H 0 you have to assume that beta hat equal to 0 and alternative hypothesis, beta hat not equal to 0. So now, the moment to test all these hypothesis, then we need to have, we need to move to, we need to move to step step 8, I am sorry step 6 . So, we need to move to step 6.

(Refer Slide Time: 43:08)



So, what is step 6. Since, since alpha hat equal to 0, so obviously, t of alpha hat equal to alpha hat by standard error of alpha hat and t of beta h which, is nothing but, how much here is, so now, 355.72 divide by 24.37, so, this is what we call as a t of alpha hat. So, t of alpha hat we will get it 14.6, 14.6, so that means, t of alpha hat this is equal to 14.6 and this this is calculated t, this is calculated t statistic.

Similarly, t of beta hat is equal to beta hat by standard error of beta hat, standard error of beta hat so which, is nothing but, minus 3.001 divide by 0.392 and if you will simplify, you will get simply minus 7.65. So now, so t of beta hat is equal to minus 7.65. So now, you have to go for statistical test. For that you have to draw the normal distributions core. So now, this is accepted region and this is rejection region, this is also rejection region, this is left tailed test and this is right tailed test and this is together we call as a two tailed test.

So, now let us assume that for a particular instance, we we actually, this particular you know intervals. So, we will get through tabulated statistics. So, let us assume that, at a particular level of sample size and particular degrees of freedom, we we we we let us assume that this dead line you can say, 2.12 and this side is minus 2.12. So, let us assume that, the t of alpha hat, the you know, target line is 2.12 and this is minus 2.12. So, then we have to compare what is the, our status of our t alpha hat and status of t beta hat.

So, we like to know what is this? Whether, this particular item is significant and this particular item significant, then this is our, you know, objective. So now, we have to place the tabulated statistics with respect to no sorry we we have to place the calculated statistic with respect to tabulated statistics. Then, then, we have to justify whether, we, means, we have to conclude that or we we have to justify that, whether, the null hypothesis is rejected or not.

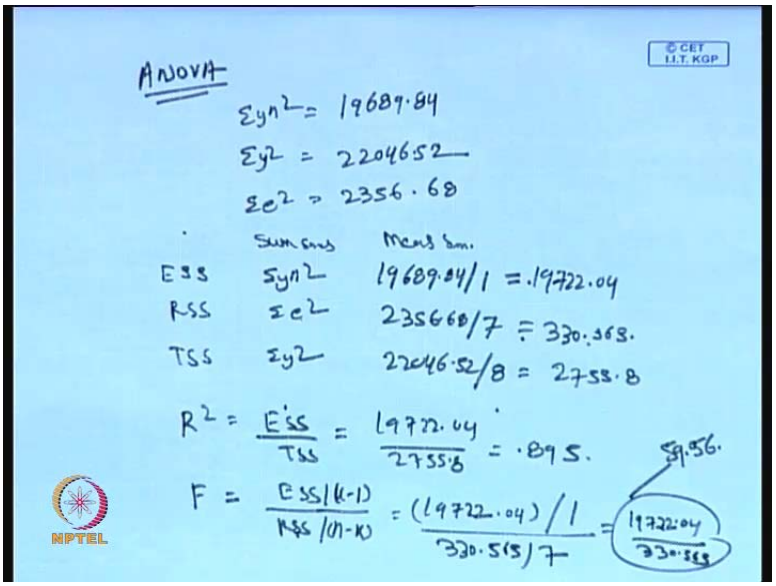
So, now what are the criteria here. So, we start with t alpha hat, so, t alpha hat is 14.6, so that means, it is coming right tail. So, this is how it is called as 0.0, this is origin. So 2.12 means, it is greater than to this item you know, interval. So, this 2.12 means so, 14.6 will coming this side.

So, this side will be coming 14.6, that means, it is in the rejection side, so that means, once it is in the rejection side, so, we are rejecting null [hypothesis] null hypothesis, that means, our statement is that alpha hat, alpha hat not equal to 0, so that means, alpha is statistically significant. So now, if it is statistically significant then, we have to check it at what level it is statistical significant. Is it 1 percent, is it 5 percent or 10? Then, we have to look into these items 2.12, so, this 2.12 is derived with respect to 1 percent level or 5 percent level or 10 percent level. Then accordingly, we have to give justification.

Similarly, come to tail; that means, we are concluding that, alpha hat is statistically significant and of course, it will be highly significant; that means, it is significant at one percent level because 14.6 is absolutely very high, all right.

So, now minus 7.65, minus 7.65 means, this will coming this side, so, left left side, that means, minus 7.65 is coming this side. So, this is again rejected region, so that means, we like to know, we have to reject here the null hypothesis. The moment will reject the null hypothesis then, obviously, we will conclude that this particular item cannot be zero, it will be statistically significant. Then again regarding probability level we have to look what is the, what is this value at what level? So accordingly, we will conclude that this particular item is statistically significant for the time being. So that means, alpha parameters and beta parameters are statistically statistically significant.

(Refer Slide Time: 47:48)



So now, we have to look for this, you know, we have to look for its, you know, analysis of variance. So, now what we have to do. So, we have to go for analysis of variance or now, the second part of the problem since, alpha parameter and beta parameters are statistically significant. So now, we have to integrate with the overall fitness of the model, so, overall fitness of the models looks for three three items. So, summation y hat square and summation summation y square and summation summation e square, summation e square. So, what is summation y hat square?

So, summation y hat square, so, let me highlight here what is summation y hat square? Yes summation y hat square, is here, summation y hat square is equal to this is 19689.84 and this is summation y square is equal to, this is equal to 22046.52 and this summation e square is equal to, summation e square is equal to 2356.68, 2356.68.

So now, so that means, I will, I will write it again, once again, that is better to very brief anova . So, summation y hat square equal to 19689.84. So then, this is summation y square which, is equal to 22046.52 and summation e square equal to this minus this, which is equal to2356.68. So this is how, we have received these items, we have received these items.

So, now we look for, we look for the anova tables. So, what is this anova tables? So, we need explained sum square. So, mean of explained sum square, then residual sum square, then, total sum square. So, this is, this is sum squares, then mean of sum squares, mean

sum squares then, mean sum square then we have to, obviously, there will be degrees of freedom.

So, explained sum square, obviously, explained sum square equal to summation y hat square, this is summation e square and this is summation y square. So, mean square is, obviously, it is nothing but, so, that means, y hat square is this much. So, 19689.84 divide by k minus 1, it is k minus 1. So, k minus 1 means, 2 minus 1, it is 1. So, this much of value, we have to calculate.

So then similarly, summation e square, summation e square, summation e square is nothing but, what is this summation e square 2356.68 divide by residual sum square n minus k, n minus k means it is k equal to 2 here n equal to 9, so, it will be 7, summation e square. So, summation e square means this is 22046.52 divide by n minus 1.

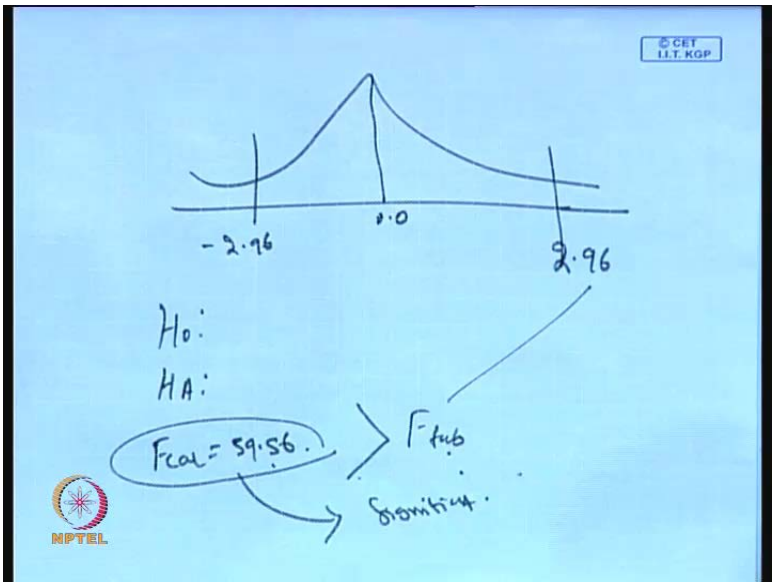
So, n minus 1 means it is 8. So, we have to find out its statistics, how much it will be coming? So that means, the picture will be like this. So, once again, I will just highlighting the detail exact item, so this will be coming, this will be coming, explained sum square. So, this will be coming 19722.04 and this will be coming residual sum square 330.565.

So, now we know r square statistics. So, r square will be, you know, summation TSS, so, this TSS will be coming how much, TSS will be coming. So, this is summation y square, summation y square by n minus k. So, this will be this will be exactly equal to 2 equal to 22046.52 divide by 8.

So, it will be coming 275 this is 2755.8, so now, your r square equal to ESS by, ESS by TSS. So, if we will simplify this one, so, this is ESS is 19722.04 divide by 2755.8. So, it will be coming, it will be coming, what is R square value here. So, the R square value is here is, this will be coming 0.895. So, this will coming 895.

So, now we have to calculate F statistics. F is the ratio between ESS by RSS followed by degrees of freedom k minus one and n minus k. So, this ESS is here, 197222 is and 0.04 and divide by degrees of freedom 1. So, corresponding by RSS, RSS is 330.565 divide by 7. So that means, we can directly just make the ratio between this and this, we will get this results. So, 19722.04 divide by 330.565, so, if we will simplify this one, then F will be give you 59.56 59.56 59.56.

(Refer Slide Time: 53:46)



So now, so, the moment will get F, so, then, you will again go for the you know, statistical significance of this particular test. So now, we have to assume that this you know, this particular item is equal to 0. So, again you have to fit the null hypothesis and you have to fit the alternative hypothesis. Then we have calculated statistics, so that means, F calculated is equal to 59.56.

So obviously, this this is 0 levels, 0.0. So, this side, let us assume, that the the calculated [value] tabulated value is say 2.96 and this side minus 2.96. So, obviously, we have to check with respect to degrees of freedom and you know sample size, we have to, we have to get the tabulated picture of the left side and the right side, then, you can get to know where is the position. So, obviously, in fact, for F statistics it will obviously, will be coming in the right side, there is no question of left tailed test in this particular instance, because, everything in square. So, it will be always in positive.

So that means, this particular structure will be like this. So, we just like to know what is F calculated and what is F tabulated, F tabulated. So now, let us assume that if F tabulated is 2.96; obviously, we will calculate with respect to degrees of freedom. So, since, it is greater than to, greater than to, tabulated statistics then, we can conclude that it is statically significant, it will calculate that, it is statically significant.

So that means, in the F F in the case of F statistic, so, the component will be always positive and the positive value, whatever you will get it, then you have to calculate the a

corresponding you can say, corresponding tabulated value, then, you have to compare this calculated value to tabulated value, if it is greater than to that tabulated value then, you have to justify that it is statistically significant and if it statistical significant then you have to look at what levels, is it 1 percent? Is it 5 percent? Is it 10 percent? Then according to, accordingly, we have to conclude that the model is the most reliable one because, our parameters are highly statistically significant and also overall fitness of the model, which is represented by R square and is tested by F statistic, is also statistically highly significant.

So that means, the model can be considered as the best models. So that means, our the data structure is the best fitted one. So, this is how, we have to go for the analysis of variance or the second part of the reliability. With this, we will conclude this particular component called as a reliability of the modelling with respect to estimated parameters and analysis of variance that means overall fitness of the models.

So, for this, we have to conclude this session. Thank you very much. Have a nice day.