Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 36
Non-Linear Regression Modelling _ Model Transformation

Hello everybody, this is Rudra Pradhan here, welcome to Engineering Econometrics. Today we will continue with Non-Linear Regression Modeling and that to discuss with various examples. In fact, in the last couple of lectures we have already discussed this particular you know unit, that to highlights various forms of non-linear modeling starting with you know dummy modeling, interactive effect and various functional forms through which we can discuss and analyze some of the engineering problems, and come with a kind of you know decision making as per the particular you know management requirement.

So, today what we have to do, we will just take a problem and then we like to connect how actually works in the real life scenario and that with the help of a particular software. In fact, there are various softwares are there, economics softwares through which you can solve this problem and that to you know work out the engineering problems with the help of various you know forms of the models, and then again with the help of you know softwares we can help get the estimated results.

And the once we ready with you know estimated econometric results, then we will be good we will discuss the particular problem and we will work out the decision making process. The thing is that first of all we should to know what is the exact problem and what are the various variables attached with this problem. And whether the particular problem can be analyzed with a linear setup or in a kind of you know non-linear set up, if it is actually solved through linear set ups.

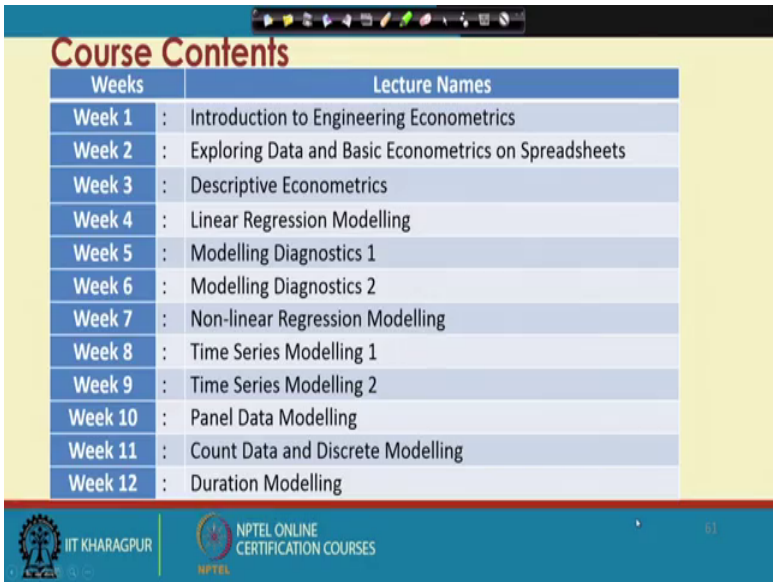
Then no point to come to the non-linear setup, but if the requirement is the non-linear then we have we are bound to use that particular form and then the decision making process will be very effective and very efficient. In fact, with the help of data visualization, you can slightly get to know what kind of you know model we can use or what kind of you know functional forms or the types of you know models we can allow to estimate the process and then you know you know predict the particular you know

requirement. Sometimes if the model is means the problem itself will you know demand that there is need of you know such kind of you know models, for instance and say the dummy modeling and the interactive effect interactive effect kind of you know modeling is the cited examples.

So, let us you know you know discuss these aspects, and briefly I will just highlight once again the various non-linear type you know econometric models then I will take examples, and I will try to you know find out solution for this a models. In fact, same data analysis package excel solve you know package we can use to have these estimated output. In say excels you know data analysis packages can help the process provided you have to do some kind of you know initial processing and bring into a kind of you know set up through which you know excel data analysis can help you to get the estimated output.

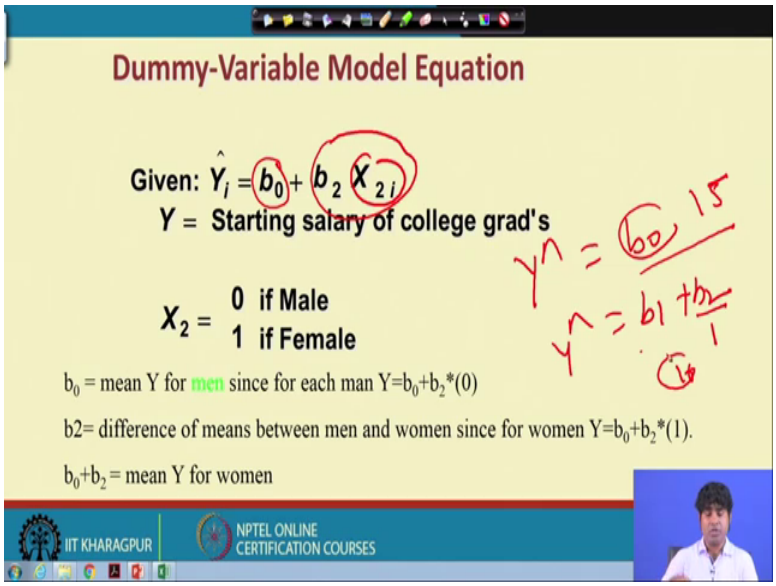
Otherwise there are so, many you know econometric softwares the softwares are there starting with you know are more if we use microfit a (Refer Time: 04:27) all the softwares can be directly used to estimate these outcomes. Let us see first various forms then connect with the kind of you know examples and the estimated outputs.

(Refer Slide Time: 04:38)



So, all together we have discussed you know dummy kind of you know modeling.

(Refer Slide Time: 04:39)

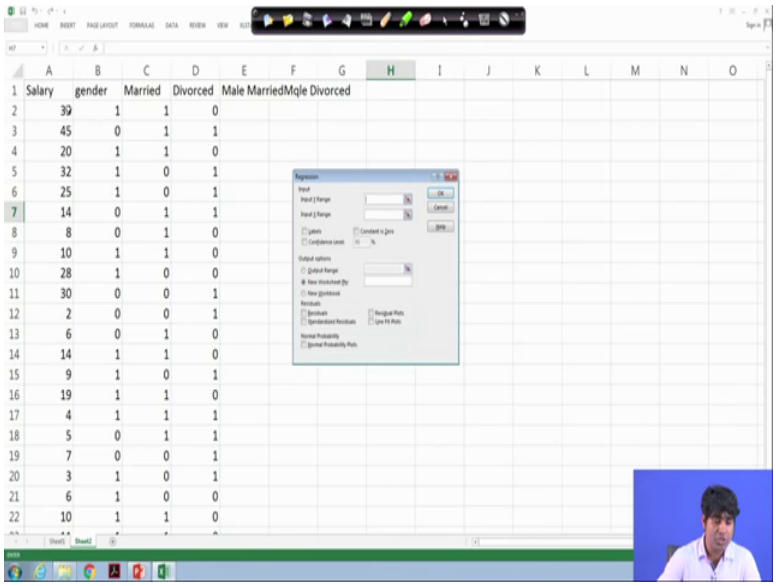


So, where we have a dependent variables, that is that you know the information which you would like to have this process is numeric and then it is connected with a independent variable which is a dummy type. For instance why is here a salary structure of the college students those you are just enter to the a job market after the you know college degree. And we are interested to know here is what is the difference between you know male income and you know female income.

So in fact, we have already highlight this models earlier I am just once again connecting this and then I will take you to the software to see how is the kind of you know actual difference through software results or estimated results. So, in this case X 2 is the treated as a independent variable, that to the kind of you know dummy. Dummy means it is a proxy and X 2 is here gender, and usually gender information for gender firsthand information is the kind of you know male and female which we can actually take care through software that is why we need a transformation.

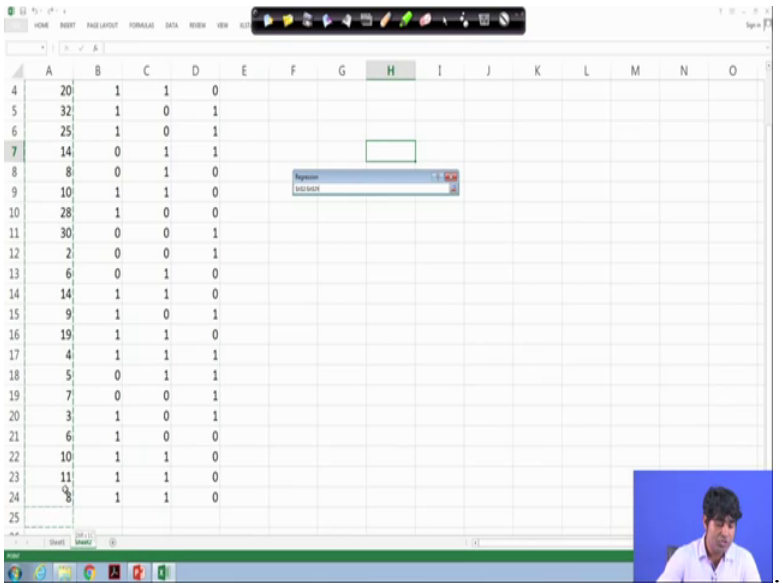
So, we put you know 0 for male and by default 1 is a kind of you know female. Then the entered data you know information basket will be transfer into salary structure and corresponding to the gender male female and that will transfer into 0 and 1 form. So, ultimately end of the day your spreadsheet will be having numeric informations that to salary structure with you know numeric content, and against the kind of you know gender with you know coding either in the form of 0 or in the form of 1.

(Refer Slide Time: 06:43)



For examples this is a spreadsheet here and we have taken actually couple of variables and first thing corresponding to this models we have here 2 variables; salary structure and the kind of you know genders right. So, what we will do here is you see here is in the salary structure, we start with you know 35000 or 45000, 20000, 32000 or we got salary in thousands and it is a kind of you know cross sectional observation.

(Refer Slide Time: 07:07)



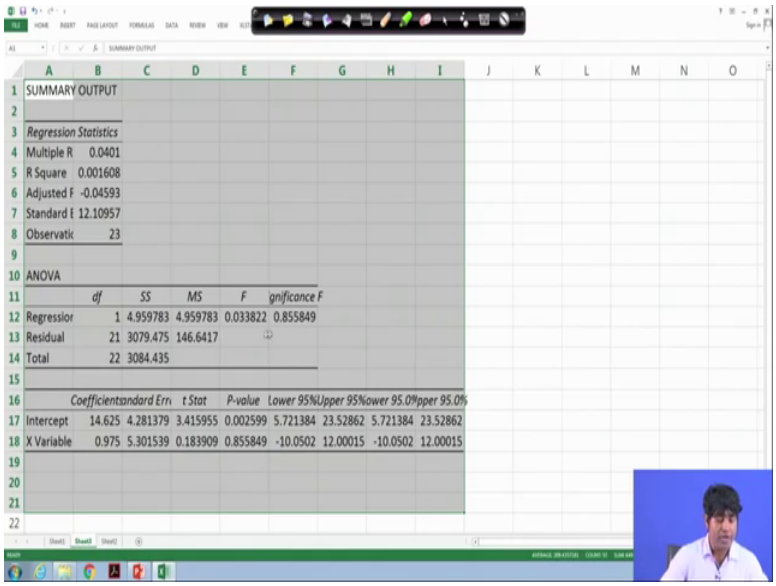
So; that means, technically we have 23 samples and then out of each a some are you know male and some are in females. So, we like to know what is the average income

difference between male, and female and this is not exactly hard core non-linear you know kind of you know model, but it will give you some kind of you know experts exposure to the non-linear setup. By the way so, having this you know data structures. So, we can you know come to the data analysis package.

So, we are here in the data analysis, and then we will go to regression package and just put and then you will find the screen like this we the way we solve the earlier couple of you know bivariate problem and multivariate problem. So, here we just like to check you know gender impact on you know you know student salary and we like to check the difference, and by the way. So, let us start with you know dependent variables. So, we can give indication about the dependent variable here.

So, this is what the dependent variable indications. So, then corresponding to the dependent variables, we can give indication about the in independent variables that too here and then we allow the software to have the estimated output and just put and it will get the estimated output here.

(Refer Slide Time: 08:36)



This is what the estimated output and you will find the a yes the model outcome is not so, good in fact, since the sample size is very small and it is kind of you know arbitrary data and what will a you know you know see here means what is more important here is the estimated output and how this dummy modeling can be interpreted, that is more important here. Then the reliability of the model and fitness of the models can be

improved by adding samples you know more sample points or you can actually you know get the actual data points and then have the estimated output.

So, that is the part of the game and because reliability process we have already gone through it, until unless you get you know the model will be clear through reliability and diagnostics we cannot use for the predictions. But what is more important here after getting the reliability and the diagnostics how we can interpret finally, that is what the beauty of this you know dummy structure.

And here is this intercept is very important. So, the model by default will be Y equal to alpha you know beta X for instance like this here, in this case here is y estimated is equal to b 0 b 2. So, now, b 2 is the slope coefficient and b 0 is the intercept coefficient so, now, if we the look here. So, the intercept is actually 14.63 and the X 2 coefficients is coming 0.98.

So, now 14 and approximately 1. So, 14 1 so, slope coefficient is approximately 1 and intercept app approximately 14.6 means P 15. So, it is the gain between 15 and 1. So, now, you come here and here the interpretation is if actually X 2 equal to 0 then you know we have Y cap equal to simply b 0 and that too you know the average salary average salary for female only.

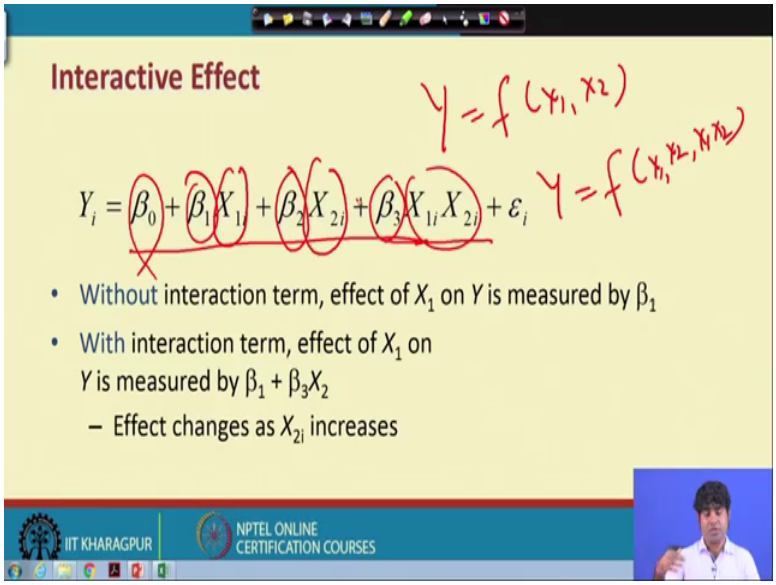
Because when we put actually X 2 equal to 0 and that too for male that particular component will be removed and; that means, technically if you put here actually X 2 0 and then the entire component will be 0. So, then finally, y cap equal to b 0 only so, b 0 coming here 15 so; that means, technically. So, when means when X 2 equal to 0 then the a Y cap equal to simply b 0 and that represents the male income male average income; that means, you know average salary of male group and here it is 15 and then when you put X 2 equal to 1 then you know b 2 is already equal to 1. So, 1 into 1 so, it is equal to 1 again.

So, then 15 plus 1 so, then that impart will be female income so; that means, technically Y cap equal to b 0 and Y cap equal to b 0 plus b 2 and here b is equal to 15 and there by the estimated you know excel output and then b 2 is equal to 1 so; that means, technically when it is 15 then this is 15 plus 1 16. So, only 1 you know 1000 you know extra is the kind of you know difference.

So, that is what the you know dummy modeling can find out the difference between or classify the entire sample into 2 groups or the same you know problems analyzing a different kind of you know a structure. In fact, it is the kind of you know requirement you know we are interested to know the gender difference income, you know difference among the you know or between the male and female. So, that is how the structure which will follow and find out, the difference whether there is a difference and against we sometimes know whether the particular difference is statistically significant or not.

Of course there are other way we can find out the difference like you know you can put simply teeth t statistic and ANOVA divide the sample into 2 parts male income and female income, then check the difference which are statistically significant or not in the regression what we do? We add all these samples create a single sample and regress it then you will find the kind of you know difference. So, or both or you know both ways you can find out the inference and in fact, this is you know more interesting because we are you know analyzing the particular sample in a once you know a single you know structure. So, likewise we will find different kind of you know structures through which you can actually understand the dummy modeling, and then the second form of this structure is a nothing, but called as a interactive effect.

(Refer Slide Time: 13:51)



ah Because I am just connecting all these forms, what we have already discussed. And every you know model means various forms of non in non-linear models and then I will

connect with the problems and the software output that is what the kind of you know requirement here. And in the interactive effect because with the one variables we can we cannot study the interactive effect and for this s or we can start with you know trivariate setup and in this case so, we have 2 variables X 1 is a independent variable and X 2 is inter in indicate in the independent variables.

So, by default the interactive effect will be X 1 into X 2 so; obviously. So, we start with actually what the means the original gamma is Y is equal to function of X 1 and X 2 then finally, the model will extend through Y equal to X 1 X 2 and X 1 X 2 that is the interactive effect.

So, as a result so, we have a 4 coefficients beta 0 is the intercept, b term is this slope coefficient for X 1 beta 2 is the slope coefficient for X 2 and beta 3 is the slope coefficient for the interactive (Refer Time: 15:10) the intersection between X 1 and X 2 so; obviously, this is what the form of the way you know interactive model now we will see what is X 1 what is X 2; and then we will see the kind of you know interactive. So that means, this is a specific types of you know models which may not you know used frequently, depending upon a particular you know engineering you know econometrics problems, we can actually specifically use.

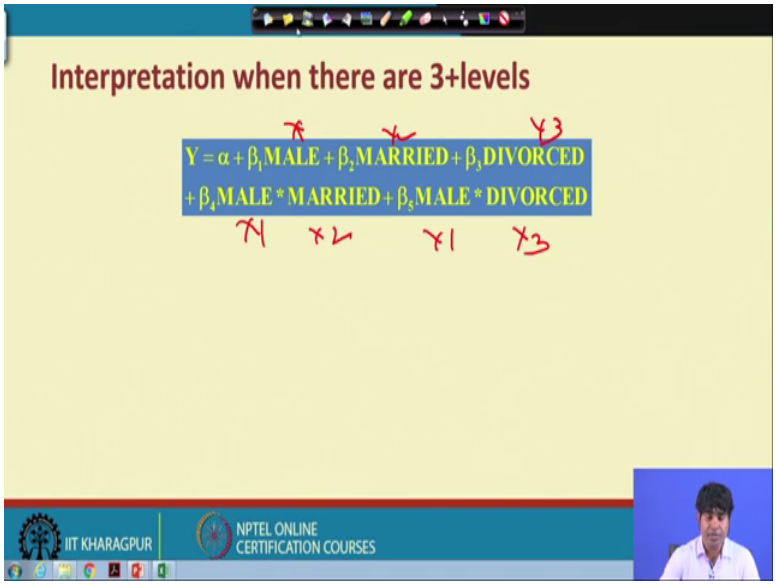
This is very special kind of you know models, of course all the you know non-linear regression models are special in a special character or in a special kind of you know models, but this is you know very special. In fact, for that you know you must you know theoretically understand that the interactive the you know interactive effect is a you know additional variable, which can be actually logically created and theoretically justified.

Then after that you can add these variables and check the kind of you know not necessarily every time you can include even theoretically you know sound or logically supported, still you see whether there is a significant contributions. If so, then you can incorporate otherwise you can just check and you know remove. Then finally, we need actually the estimated model which is actually absolutely for the decision making process.

So, now for that we can go to again excel sheet and in this case. So, we can go to the data sheet first and then here the data sheet, and before that actually. So, what I can do? I will

just connect with this you know models. So, what we can do. So, let us take an examples ok. So, X 1 X 2 and the interactive effect.

(Refer Slide Time: 16:56)



For that actually we take actually X 1 equal to male and X 2 equal to married and against we put another variable X 3 divorce.

So; that means 3 independent variables here and then we like to see the interactive effect because this model which you have already discussed in the previous you know lectures, there is some you know sample models to describes the interactive effect in the same model. So, I am just using to you know connect with you know software outputs how it works and ultimately what kind of you know results you will get and how you can interpret that is more important here.

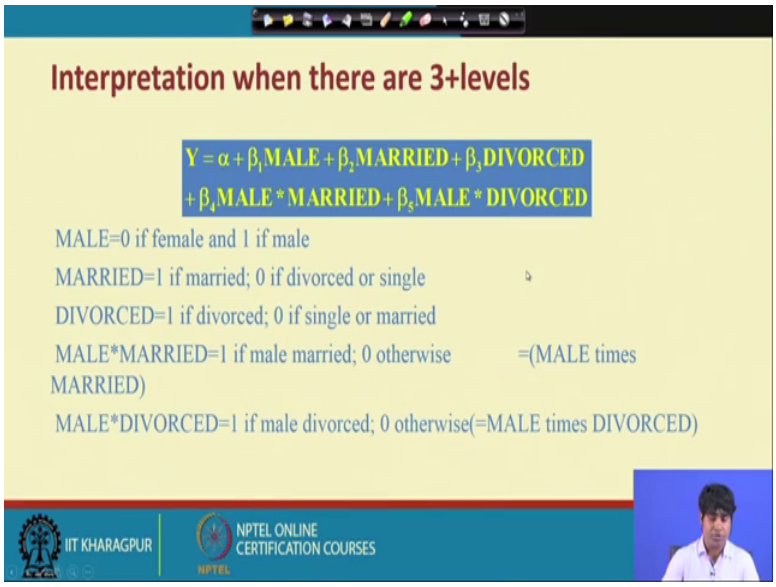
And we have here 3 variables, this is in fact, X 1 this is X 2 and this is X 3 and this is X 1 versus X 2 and this is X 1 versus X 3 . So, male married and divorce then male married and male divorce so, that is the kind of you know. So, every time the target is actually gender and that to specific target on male sides. So, by default the first variable will be the gender.

So, the information in that particular variable will be in the form of 0 1 and again married the answer will be simply yes or no and that will again transferred into 0 1 and again the third variable is divorce. So, that will be against yes no and the final input will

be again 0 1 and then we you know we have a now 3 variables all together male married divorce all are you know dummy type, but for interactive effect not necessarily variable should be interactive or you know a dummy type and you know it may dummy type it may not dummy type, but in this specific examples it is the mandatory that we should have you know dummy type, because the for the for the type of variables is like that only. And then simply what we can do male merits that is the interactive effect 1 interactive effect then male divorce that is the another interactive effect.

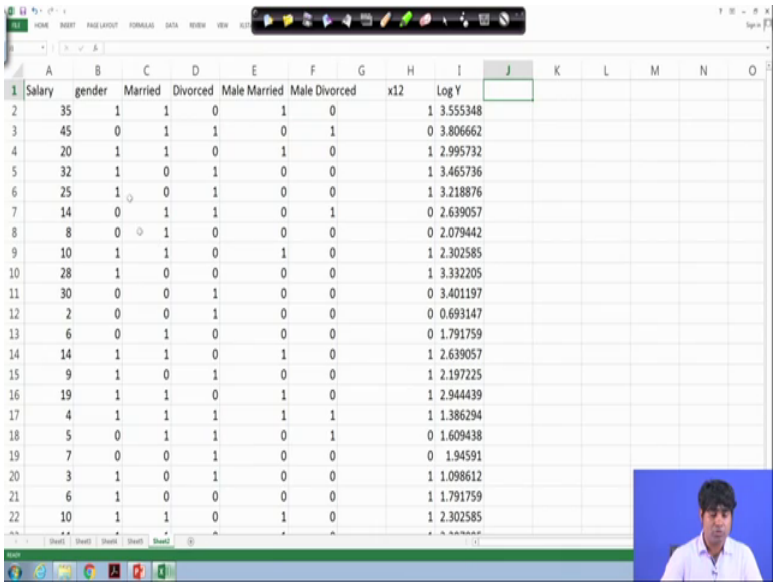
So, we like to check how male impact on salary you know colors you know salary average salary, then merits impact on this average salary, then what is the divorce impact on you know salary then again the kind of you know interactive effect for that we have actually created here the kind of you know examples and for that.

(Refer Slide Time: 19:24)



So, this is what the kind of you know understanding, the detailed description about this model and then let us see we will work out here o, how it can be operated in real life situation see here. We have 3 technical variables a you know x ones then X 2 and X 3 and then what will you do will create a interactive effect male and you know merits.

(Refer Slide Time: 19:48)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Salary	gender	Married	Divorced	Male Married	Male Divorced		x12	Log Y						
2	35	1	1	0	1	0		1	3.555348						
3	45	0	1	1	0	1		0	3.806662						
4	20	1	1	0	1	0		1	2.995732						
5	32	1	0	1	0	0		1	3.465736						
6	25	1	0	1	0	0		1	3.218876						
7	14	0	1	1	0	1		0	2.639057						
8	8	0	1	0	0	0		0	2.079442						
9	10	1	1	0	1	0		1	2.302585						
10	28	1	0	0	0	0		1	3.332205						
11	30	0	0	1	0	0		0	3.401197						
12	2	0	0	1	0	0		0	0.693147						
13	6	0	1	0	0	0		0	1.791759						
14	14	1	1	0	1	0		1	2.639057						
15	9	1	0	1	0	0		1	2.197225						
16	19	1	1	0	1	0		1	2.944439						
17	4	1	1	1	1	1		1	1.386294						
18	5	0	1	1	0	1		0	1.609438						
19	7	0	0	1	0	0		0	1.94591						
20	3	1	0	1	0	0		1	1.098612						
21	6	1	0	0	0	0		1	1.791759						
22	10	1	1	0	1	0		1	2.302585						

The just you put you know equal to signs and this is what actually male and you know females that transformation is 0 1. Just you multiply and have this particular variables so that means, this is the integration between this and this.

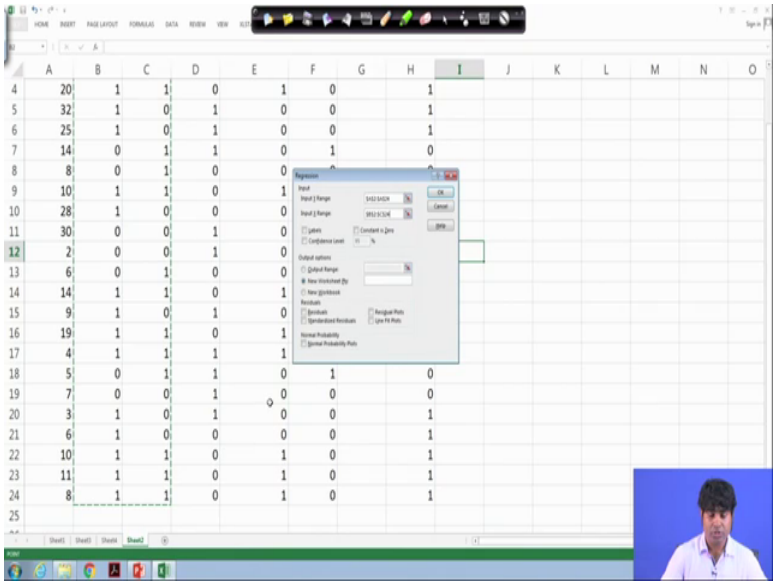
So, you just you know put. So, then you just you know generate this series. So, this will be the first interactive effect. So, the data set will be created and then second variable same inference will be like this you know this is a this is what actually a and. So, this is what actually male married and then in this case will be put here male divorce this is what male divorce so; that means, this is another variable which we will create. So, equal to; that means, this end and the end the kind of you know and this. So, this is usually this will be created again.

So, now, we can generate this series. So, now, the data set is ready. Is I am doing this because the data analysis in excel software is you know directly cannot take. So, you can actually manually process it then allow this software to run the model. So, now, what is happening here, we have actually here the dependent variable, then first independent variable second independent variable third independent variable and the first interactive effect and the second interactive effect.

Well the structure of the model is here and then we use this you know models and then go for the estimation process. So, ultimately you can go to the data analysis package again and you know highlight the regression, then it can actually reset because the earlier

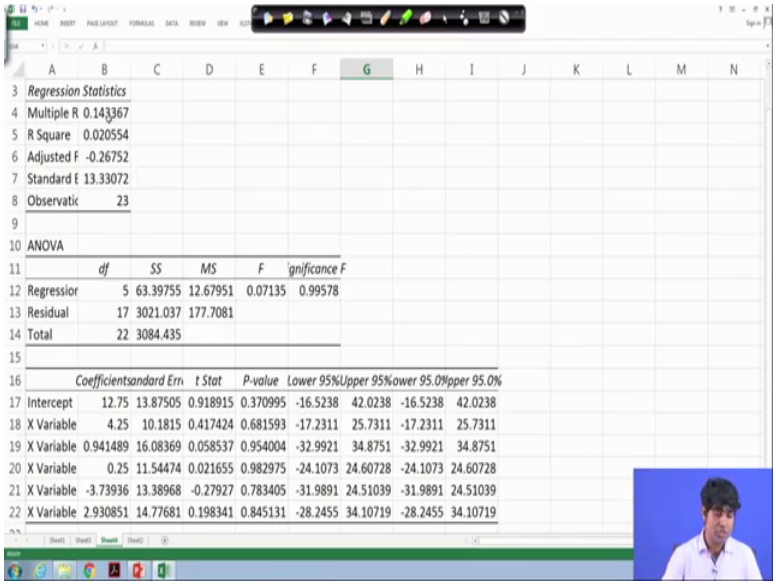
set was the you know kind of you know the structure about the a dummy representation. So, here so the same dependent variables. So, the dependent variable will be like this.

(Refer Slide Time: 21:56)



And then the kind of you know independent variables. So, what will you do? So, we will take all the independent variables simultaneously for this dependent variables, it includes both you know independent variables and the kind of you know Interactive effects.

(Refer Slide Time: 22:11)



So, here 2 any 3 independent variables and 2 interactively effect and now here is the output.

So, this is what the output. So, now, yes of course, this there is little bit of improvement corresponding to the previously you know model outcomes, and of course, it is not so, good because F is not coming statistically a significant and R is technically very low and again adjusted R square is also very low. So, what is happening here? So, what is more important here the coefficients the coefficients and here this is the intercept and this is the first independent variable estimate second variable independent variable estimate.

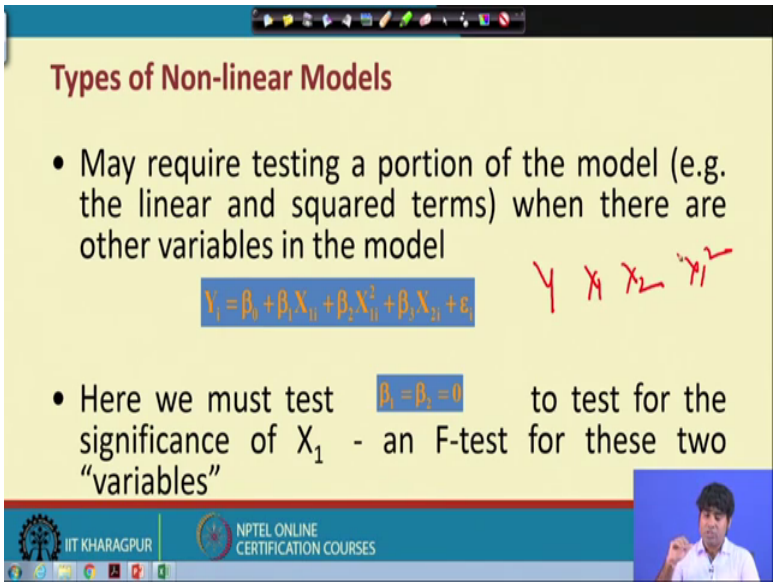
And third a third you know independent variable estimate then first interactive effect and second interactive effect. So that means, see the ultimate after getting this you know estimated output you can again come back to the kind of you know you know discussion. So, here so, just you know put this coefficient in place of you know beta, because we have all these beta results including alpha and then what is the you know difference yes.

So, every time it should be 0 1 kind of you know structure so; that means, you know when you put actually 0 and ultimately it will go to the a female structures. So, when it is actually male it will be 1. So, ultimately so, that in the effect will be here alpha plus beta ones or you know a simply alpha then if you know I say yes no so, again. So, if the guys married then its 1 and not then it will be 0. So, then again so, every times you can actually a put you know allow 0 and allow 1.

So, then you will find you know big difference then finally, you can have the outcomes and then I check the interactive effect whether it is a statistically significant or not. Of course, the theoretically this model is the it you know there is the possibility of you know adding interactive effect, ultimately the you know best requirement or the effectiveness of model depends upon the you know statistical reliability or econometrics reliability. That means, technically it will pass through all the reliability test and the diagnostic test then after that you can interpret how is the individual variables impart and how is the interactive effects impart.

So, it is like you know more or less same like you know as usual multiple regression. So, after you know creating interactive effect, which is logically sound theoretically then there is no issue to go ahead with the estimation process, get the estimated output interpret the model and go for the decision making you know things as per the particular you know engineering requirement. So, this is a special correct type of you know models.

(Refer Slide Time: 25:11)



Then after that we have a various types of you know other non-linear models starting with you know quadratic form. So, here is the kind of you know structure is same why is the dependent variables and X is a kind of you know independent variables, and X 2 is a kind of you know independent variables and we are allowing X 1 in a kind of you know or non-linear format so; that means, technically. So, we have here is; we have here actually Y X 1 and X 2 and then we will create a X 1 square you see its not actually a big deal. So, we have a dependent variables and we have X 1 and we have X 2.

Like we have created an interactive effect you know for instance male you know male married and male divorce. So, here also you can create a you know X 1 square that is again kind of you know interactive effect. So, there X 1 X 2 or X 1 X 3 it is a kind of you know cross product interaction so, X 1 square is a kind of you know dot product interaction. So, you can multiply a variable with X 1 with X 1 and X 2 with X 2.

Of course creating this kind of you know variables by the help of you know dot product, it should be again theoretical sound and logically supported. So, then after that you can help the general you know data generating process, and then you have to go with the a kind of you know you know estimate you know estimations and can have the estimated results. For instance again you can go to this you know the kind of you know excel sheet.

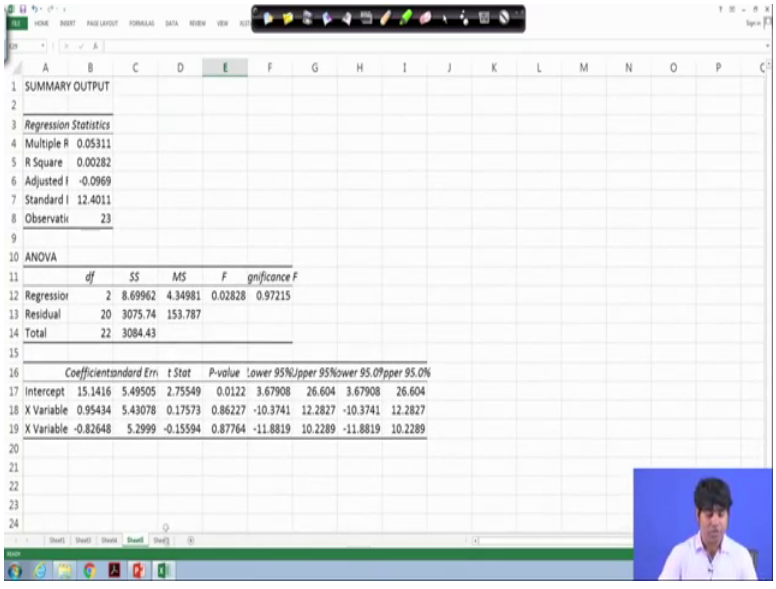
So, let us say this is actually a you know kind of you know what you can call it as original data sets, let us say this is actually Y and we will get a you know let us say

another kind of you know x you know 1 squares let us say this is the kind of you know thing. So, what will you do provided this is X 1 and. So, what will you do? So, you just you know put square equal to here and this against multiplied by this a multiplied by this, and then you created this particular variable.

In fact, this is a dummy a kind of you know representation it will not change at all it will change at all, but it can be created like this after we create after this variable creation. So, what you can do? Again you go to the data analysis package and then put the a as the you know kind of you know regression requirement. So, reset again and here the requirement is actually dependent variables. So, first you go to the dependent variable structures. So, this is what the dependent variable structure and here in the case and the independent variables.

So, you can actually a allow or you know gender and then this ones and against and the interactive effect.

(Refer Slide Time: 28:22)

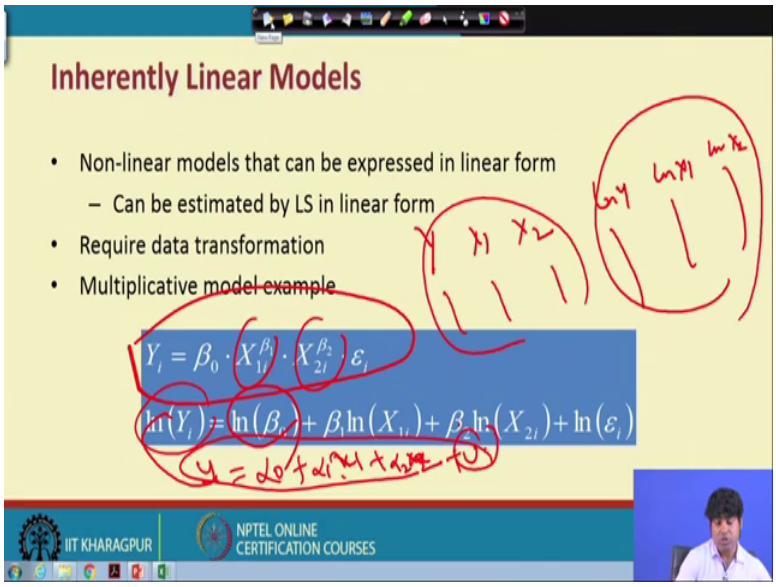


So, then you put so, this will be this will become like this so; that means, the process is a more or less actually same. So, you will have a kind of you know output and after getting this output again come to this you know interpretation so; that means, you create X 1 square is a third into you know interactive type of you know variables.

Then again you run the models. So, you will get you will get a kind of you know estimated outcome and then check whether the estimated outcome is a statistically significant or not for instance you know here we have 2 different kind of you know structure in 1 case we have Y equal to beta 0 beta 1 X 1 sin beta 2 X 2, and then again in betweens we are allowing and you know beta I 3 X 2 X 1 squares and then check whether you know the particular you know interactive effect will be statistically significant or not.

If that is the case you can go ahead and then you see you know, how best we can you know interpret the kind of you know requirement.

(Refer Slide Time: 29:29)



Then in thus in the same times in the same times you can have another form of the models like here and this is a exponential type you know models or sometimes in business or you know management.

So, we call as you know CD type you know of production function Cobb Douglas production functions. So, here the forms of the connection are completely non-linear. So, what we can do? The entire models can be transferred into linear format for instance you know the first model in the original model. In fact, is like this. So, this is what the original model. So, what we can do here we will go for you know transformation. So, that is actually the you know best way to you know bring into the linear format.

Then after that we can go ahead with the LS mechanism and we can estimate the process and have the estimated output. And after that the interpretation process will be more or less same. So, in this case we have Y beta 0 then X 1 to the power beta 1 X to the power beta 2. So, that is here so, what will you do here? So, we can apply log transformation both the sides and mathematical it is allowed and that is the best way to simplify the structure and go ahead as usual you know simple LS mechanism.

So, now so, we have actually originally we have Y X 1 and X 2. So, now, the transformation process will be a Lny Ln x 1's and Ln x 2. So, this is how the three variables can be created. So, this is the original and this is the transport data now once you have a transport data. So, the log y can be treated as you know y then it will be simply alpha.

Then let us say alpha 0 then this will be say alpha 1, I say a you know small x ones then plus alpha 2 small x 2 small x 2 and then the error term. So, now, the same models can be represented in this form where y is nothing, but log y then alpha 0 is nothing, but you know log of beta 0 and alpha 1 is nothing, but actually beta 1 and x 1 is nothing, but log x 1 and alpha 2 is beta 2 and then log x 2 is nothing, but you know small x 2 and then log epsilon will be the error terms.

So, this is how the final transformation and then in order to simplify this structure. So, what we can do? It is actually a model transformation is required. So, again you go to the excel, you know excel spreadsheet and this is how the original data set. So, what will you do? So, we will have here our transformation for instance you can go ahead with you know a log transformation.

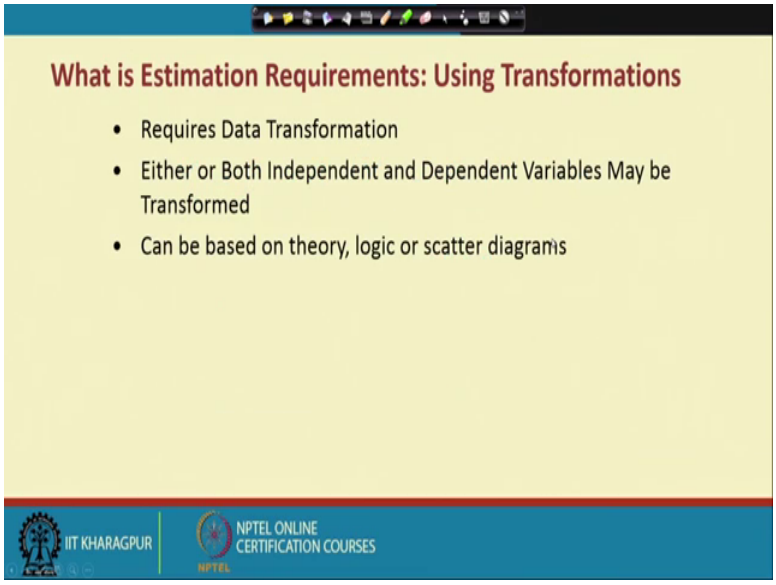
So, just you put and allow here actually a log transformations all right. So, yes this is what the log transformations and then you can just indicate the particular variable and put. So, this will be created here. So, this once you transfer then you can generate this series this is actually say log of Y and similarly we can go for you know and this variable transformation so; that means, the entire variables original setup is like this, and this is the transport system of course, these independent variables are dummy and it is in the form of 0 1 which cannot be transferred here.

So, for that the sample structure will be changed, but why is actually numeric? So, that can be go for you know log transformation. Again so, with this you know samples you

know data set it is very much clear that you know in this form of you know non-linear regression models, the fast and transformations should not be actually 0 1 type; that means, dummy type. So, it should be absolutely numeric.

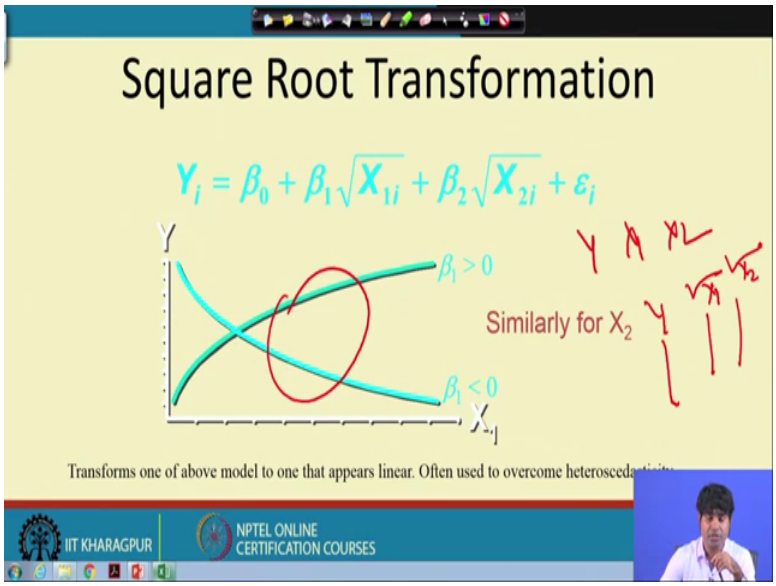
So, that you can go ahead with the transformation, proper transformation and then you can you know use this transport output for the estimations and then we can go for the decision making process. If that is not the you know case then you cannot use this form of you know model to you know go ahead with the estimation process and as per the kind of you know engineering requirement.

(Refer Slide Time: 34:08)



So, this is another form of non-linear regression models and in same way.

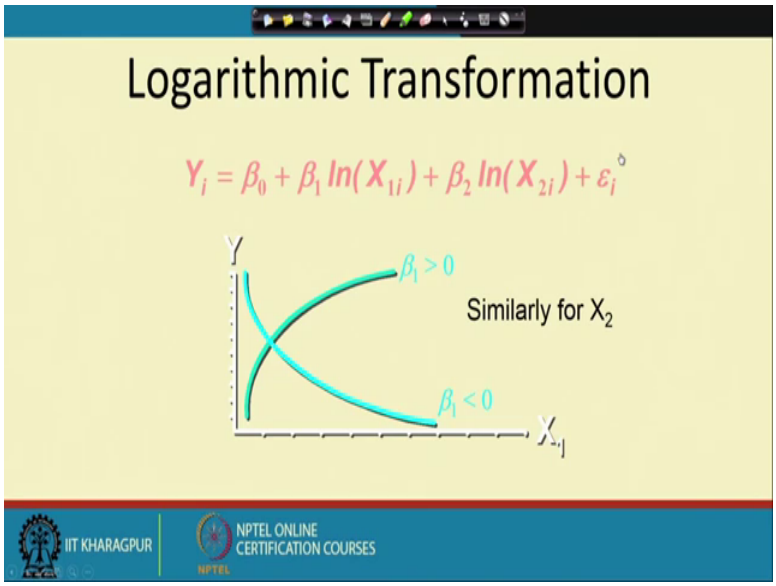
(Refer Slide Time: 34:10)



So, this is a square root transformation same thing instead of you know log Y. So, we have here originally we have actually y X 1 and X 2. So, now, same y will be remain there. So, instead of X 1 you can go for X 1 square root of X 1 transformation and square root of X 2 transformation. So, usually the idea is you know reduce the volatility and put in a kind of you know structural form and against you re estimate the process, then the process of the estimations with the transport output it will be more or less like the original structure and then you can you know interpret.

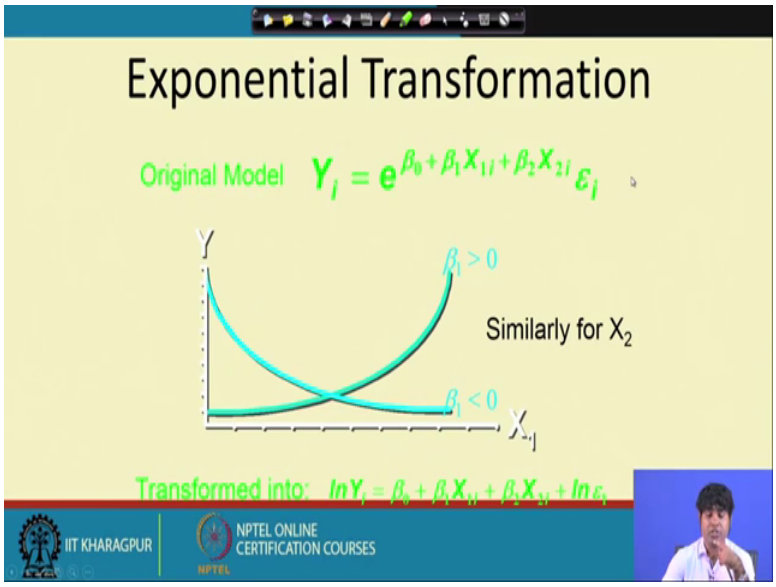
Of course the kind of you know elasticity of these outputs will be different, but more or less the processing through software is a you know very much similar. So, it is not actually big difference, just you need actually data transformation. And if your model is it being like this and for that you can actually go for you know data visualizations like here we have actually a 1 to check the actually then we can transport the you know model like that and then we can apply the transport output for the model estimations.

(Refer Slide Time: 35:22)



And the too as per your requirement this is what the transport output.

(Refer Slide Time: 35:28)



And against there is another form of you know our models, in our non-linear regression model that is exponential types of you know again. So, this is the originally it is in non-linear. So, we can apply log so, log of Y equal to beta 0 beta 1 X 1 beta 2 X 2 so; that means, here this in this case only log Y can be transferred; and this Y can be transferred to log Y then X 1 X 2 as usual remains same so; that means, compared to previous one. So, one way you can first find out you know Y and log X 1 and logs 2 X 2. So, like this

case and then you can actually keep y remain constant and X 1 will be changed to root X 1 and X 2 can be change into root X 2 and then go ahead with estimations and in the third case y can be log y and that is the only transformation you do and the remaining beta 0 beta 1 X 2 and beta 2 X 2.

So; that means, X 1 and X 2 is the original kind of you know requirement, then the over with the estimation process. So, the estimation process altogether actually same and ultimately you have to just check various forms and the same problems can be processed through various other forms. That is one kind of you know robustness check and by default you will find various alternatives same variable same you know problems and just we are changing the various you know modeling forms and various in with respect to various functional forms and structural form.

Then you check you know with estimations and the model which is really good and best as per the reliability is concerned diagnostic diagnostics are concerned, then that model will be finally, kept for the decision making process and as per the particular you know engineering problems requirement. So, with this we will stop here and.

Thank you very much have a nice day.