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

Lecture - 40
Extension of Dummy Regression Modelling- Dummy Independent Variable Modelling

Hello everybody. This is Rudra Pradhan here. Welcome to Engineering Econometrics. Today we will continue with non-linear regression modelling that too qualitative response regression modelling, where we are actually discussing the forms of model like linear probability model, logit model and probit model. In the last two lectures, we have discussed the structure of linear probability model, logit model and probit model. In fact, we have analyzed this you know the first two models that two linear probability model and logit model with a kind of you know examples and you know the kind of you know models are such a way the information will be transported into probability forms, either it will be 0 or 1 or in between.

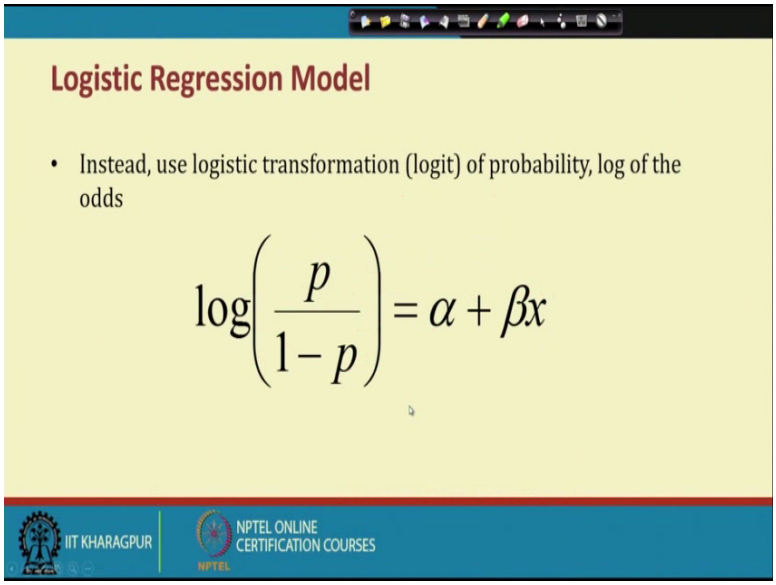
When the particular transformation will be 0 1 only, then you can use simply linear probability model. Then, when the transformation will be in between 0 to 1, then you can use either logit model or probit model which we have already discussed. So, in the last lecture what we have discussed that you know the variable in the independent variables side is the salary and the dependent variable side is the people's own house and the counter part is the rented house.

So, that means the problem which you have discussed earlier is that it is the salary which can decide whether people can stay in the rented house or staying in the you know own house. Theoretically the expectation is that you know if people have high salary or high income, they can have their own house and people having low salary, they can stay in the rented house. That is what the perception. If people have high salary and they can stay their own house, then the coefficients should be positive in nature. If not, then the coefficient should be negative in nature.

So, also in the case of you know first models that is linear probability model, we just asked the individuals about their salary and whether they are staying in a rented house or own house. So, we can generate the information and simply run the model and then, test

the model as per the theoretical expectation. Again same problems we have investigated through logit model where the forms of the model will be like this.

(Refer Slide Time: 03:02)



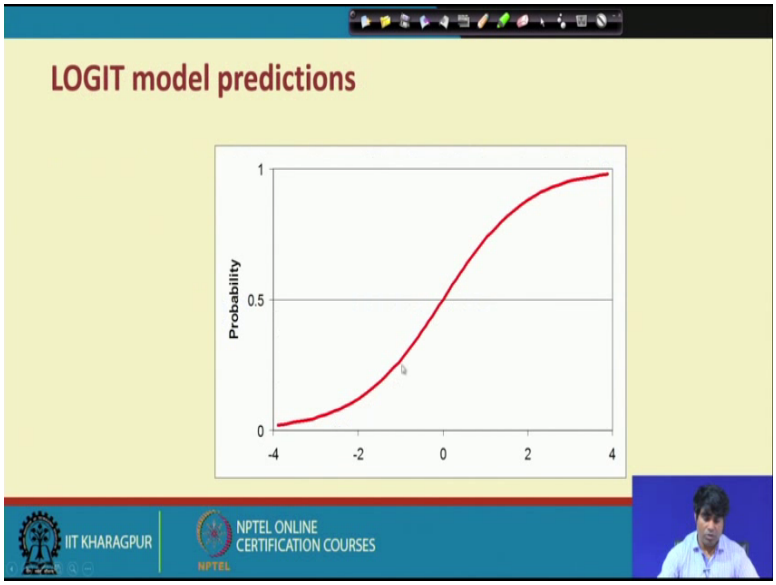
We need actually probability p like there if people having own house, then it will be 1 and then, p becomes 1 and if people have no house, then p becomes 0. So, only yes no 0 1 that is the kind of you know things and that is the requirement of linear probability model, but in this case we need actually p in between 0 to 1. Then, corresponding to p we need 1 minus p and then, we need a ratio p by 1 minus p and then, we go for log transformation of this odd ratio and then, integrate with the independent variable. That is what the case is all about and for that we have already discussed and having information about you know yes, no this model will not work and for that we have used group sampling.

So, that means instead of you know 20 individuals, we can you know have 20 organizations, then every organization you should have a couple of samples maybe 20-30-40 like this and with the expectations or the structure is that you know out of 40, few will be you know having own house. At least few will be having own house you know, so that you know we can have a probability value. So, the probability value will be you know total sample will be the kind of you know indicator and out of which how many are actually having own house. So, then the people will be having own house divided by total other you know total sampling which you have actually interacted.

So, then you will get a probability value. That is what the p value and by default 1 minus p will be the difference and then, you can just have the ratio will the quality odd ratio and then, after the transformations we can have the actual dependent variable which can be integrated with the independent variable that too in the form of you know logistic structure. So, this was the logistic structure.

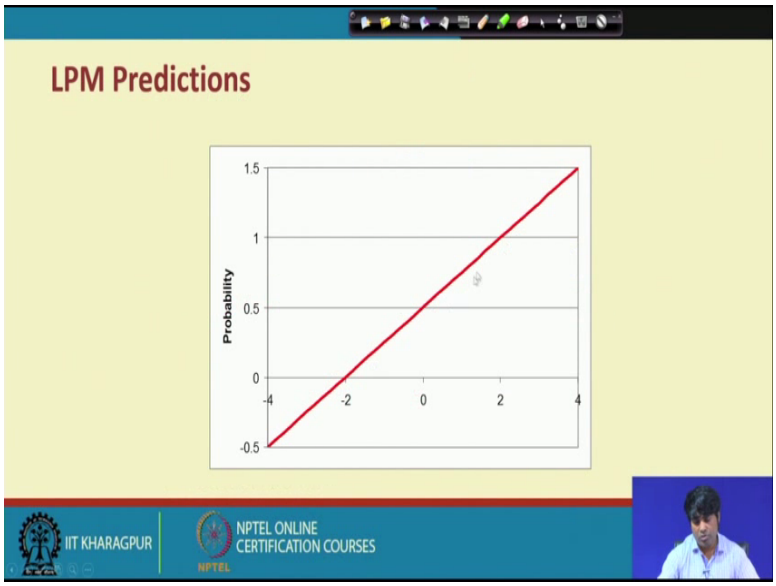
So, that means actually if you transfer into exponential form, it is nothing, but you know p by 1 minus p equal to e to the power alpha plus beta x. So, just you know and that means technically the particular form is a logistic format only. So, it is a linear format and that the linear probability model is the linear format and here non-linear format that too the application of you know logistic functions.

(Refer Slide Time: 05:44)



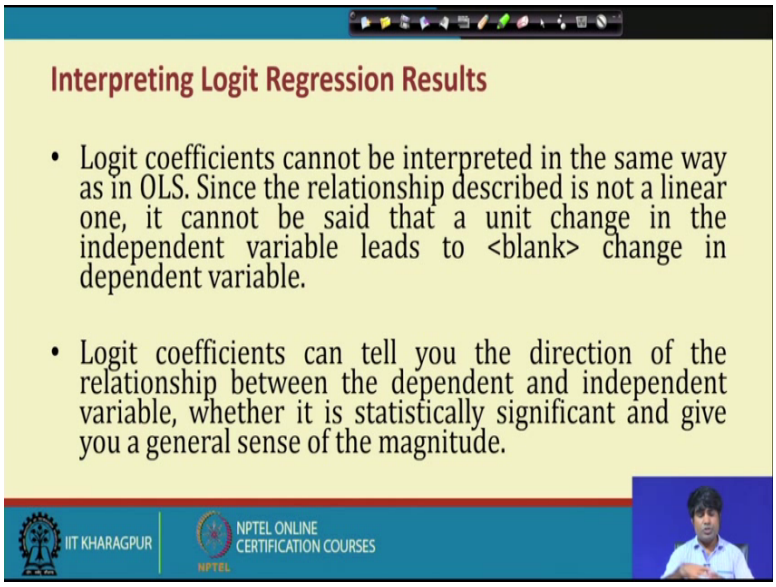
So, likewise we have already discussed, so the form of the logistic function would be like this. It is clearly indicated that you know I had seen non-linear in character compared to the previous you know Linear Probability Model

(Refer Slide Time: 05:52)



The structure will be straight line and here the structure will be simple you know like you know s types, you know structure and we use group sampling to analyze this particular case.

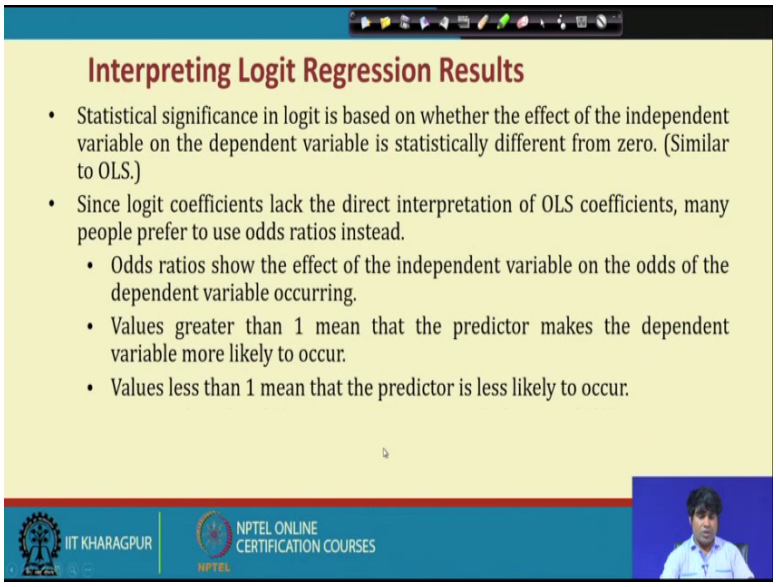
(Refer Slide Time: 06:10)



So, the interpretation will be also you know like that if odd ratio actually are high, then it is a positive impact if water is having low and this is actually low apart or you know negative impart likewise.

We can actually connect with the various other problems and then, you know analyze whether you know logit model is a very effective component to address, some of the engineering problems as per the particular you know requirement and then, come to the decision making process through which you can actually predict the dependent variable through this independent variable, where dependent variable information is the completely you know qualitative in structures.

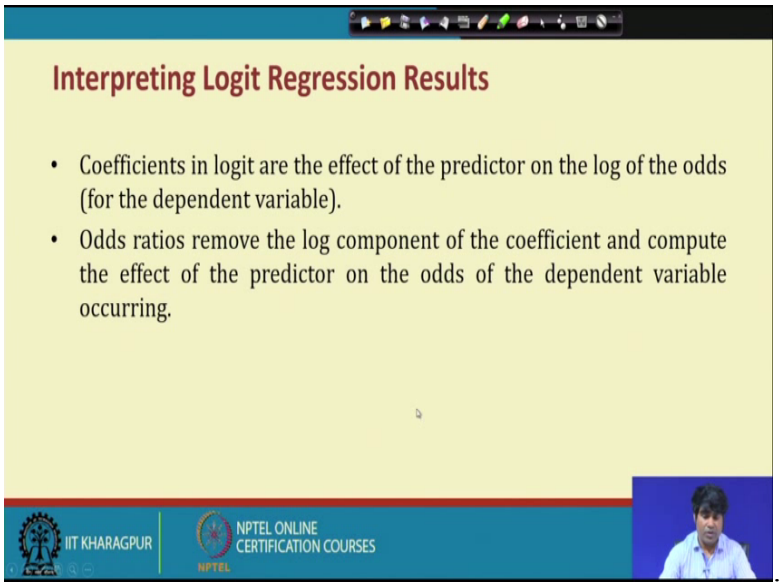
(Refer Slide Time: 06:55)



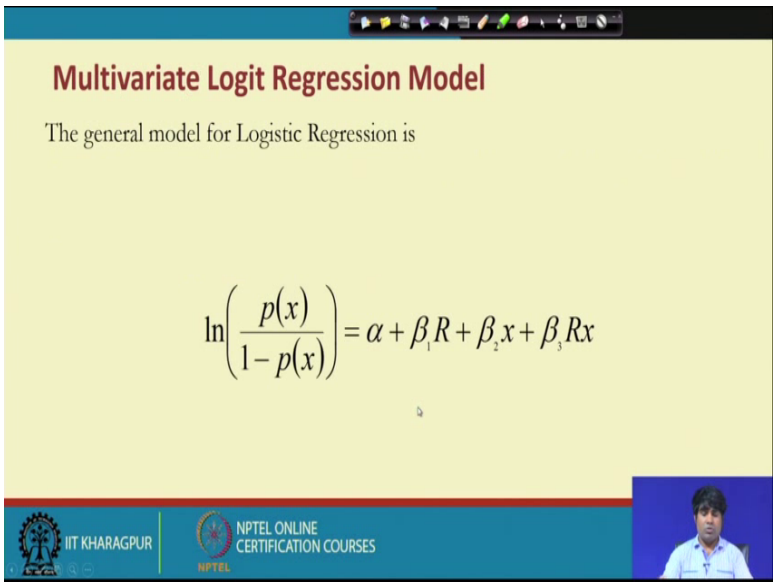
So, this is, it is another kind of you know bank examples where you know you can actually have a dependent variables like you know financial inclusions and the measurement will be people having bank account and not having bank account and we are interested to know what are the factors, which can allow the people to have bank and if not what are the factors, so that you know people having the bank. So, that means technically we like to know what are the factor responsible for the financial inclusion.

So, these are the things we like to actually check here.

(Refer Slide Time: 07:30)



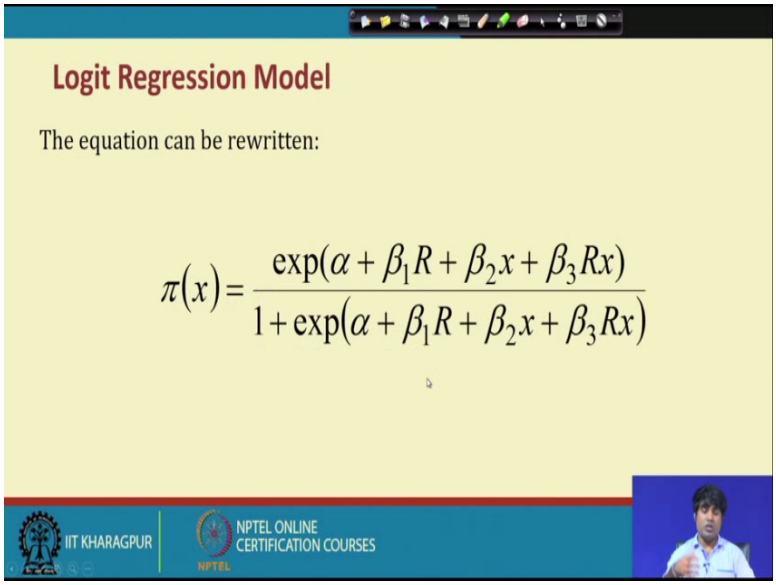
(Refer Slide Time: 07:33)



So, for that you know this can be the one of the you know multiple regression model. That is the extended version of the previous ones like you know people having house and the salary. So, now here bank account and then, the factors affecting the bank account like you know let us say income is a factor, then the kind of you know family members can be actually factors. So, like that you know a kind of you know structure which you can have to analyze the situation and in between you can also go for interactive effect and then, you check whether interactive effect can have the impact on the you know

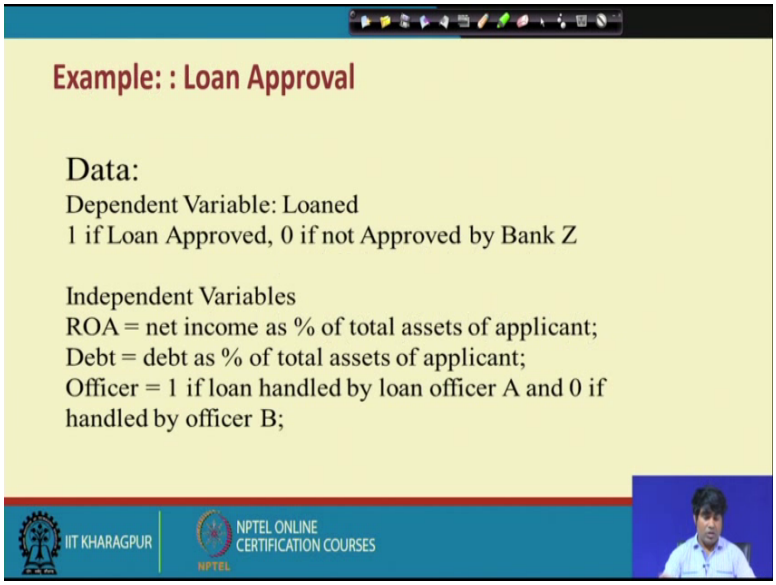
dependent variable whether we like to study the determinants of you know financial inclusions.

(Refer Slide Time: 08:23)



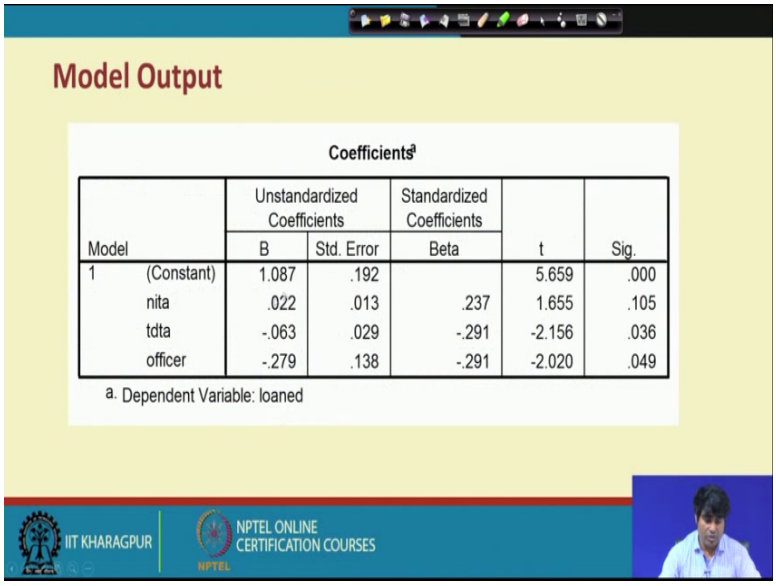
So, now it can be extended to any forms like that means I told you earlier it is a logistic function. So, ultimately you know the dependent variable will be you know a ratio between exponential of you know alpha plus beta 1 R plus beta 2 x plus beta 3 Rx. That is the independent variable structure and 1 plus e to the power x again that with respect to R x and you know R x that is the interactive effect.

(Refer Slide Time: 08:53)



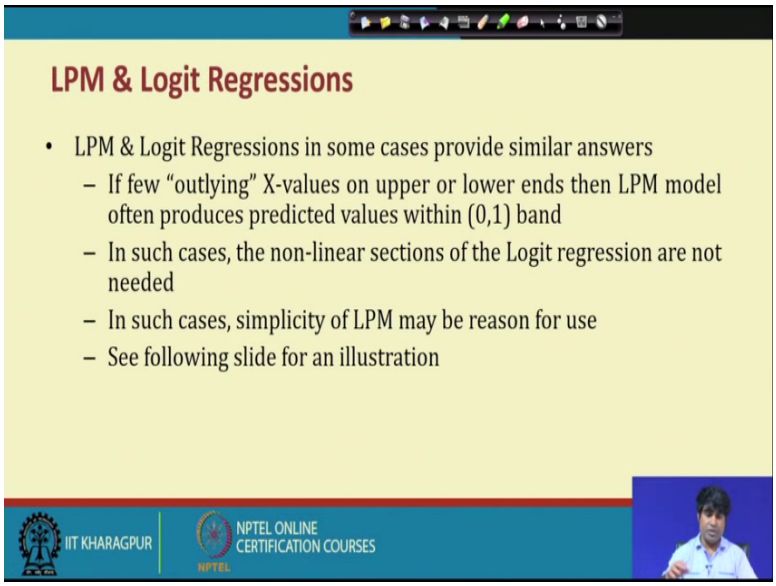
Here the sample examples you know another kind of you know example is you know where that loan will be approved and not approved. So, again 0 1 kind of you know structure and what are the factors responsible for that written on assets debt and officer in charge. So, these are the things which can affect the dependent variables. So, ultimately the dependent variable case is the, you know loan approved which is actually 1 eps and 0. If not, again the same problem can be analyzed through linear probability models and can be applied to logit model and again in the case of logit models, you need group sampling rather than individual sampling, otherwise it is more or less same like the previous problem which you have discussed.

(Refer Slide Time: 09:43)



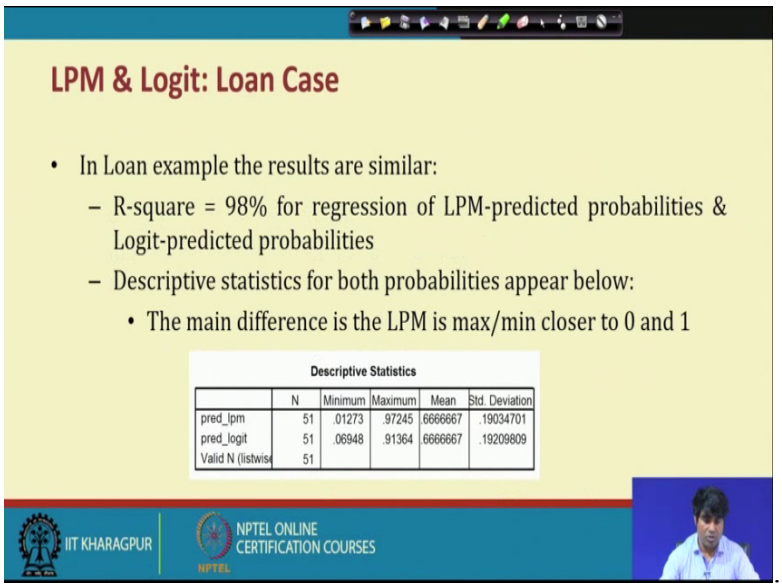
These are you know results and we find actually it statistically significant.

(Refer Slide Time: 09:52)



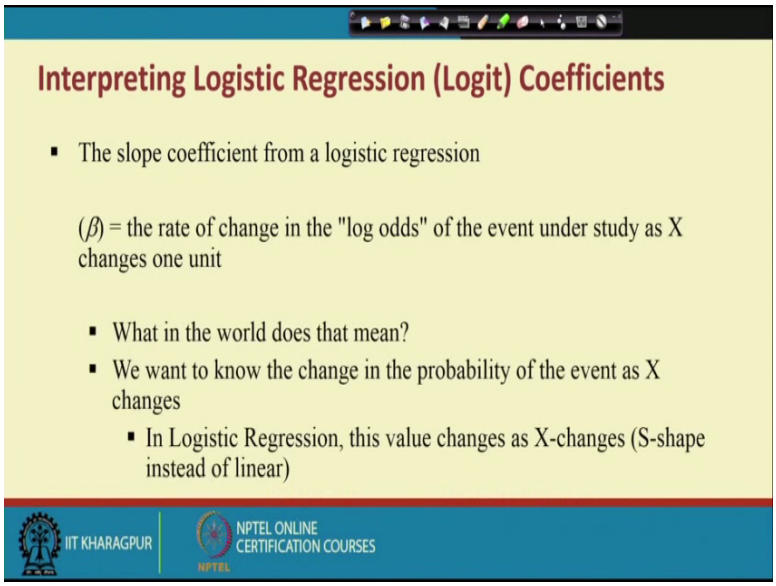
So, that means this model is very effective again you know to eh predict the dependent variable with respect to the independent variables, where the dependent variable is the qualitative structures that too either in between 0 to 1 or 0 you know 0 and 1 only. So, this is how the you know structure through which you can actually analyze you know problems.

(Refer Slide Time: 10:19)



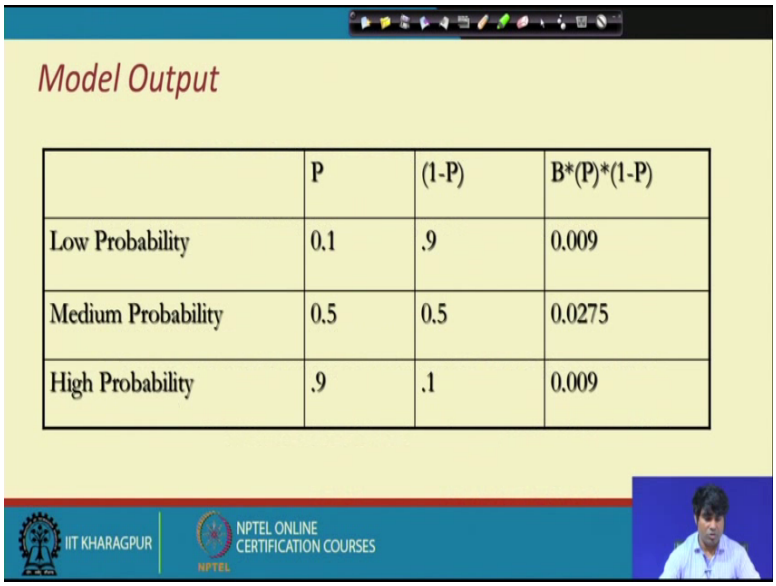
Then, come with the kind of energy.

(Refer Slide Time: 10:22)



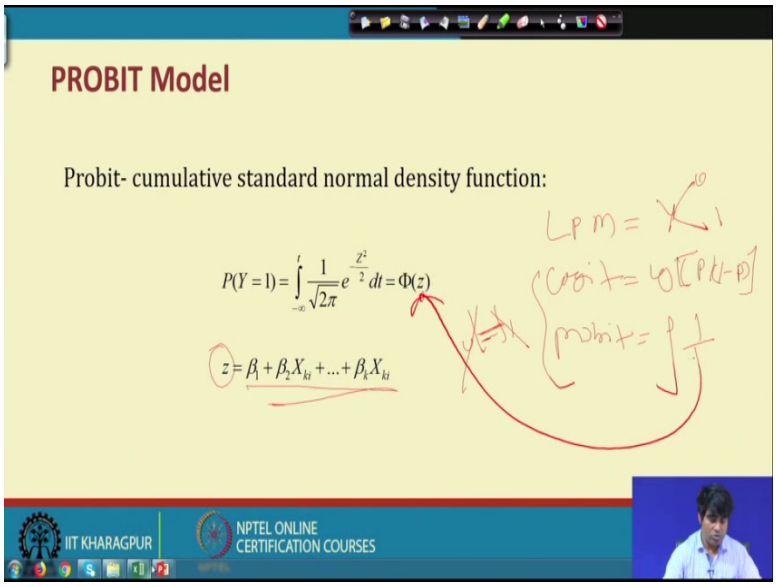
Of course if you compare it to a linear probability model and logit model, it will find drastic difference compared to the last problems which you have analyzed with respect to people saving house and then, the salary structure. The logit model gives better result. So, now you know all these 3 forms of the models are similar in character. Only thing is you know the functional forms which you can apply to estimate the structure is different. In the first case, we simply use straight line equation. Second one is the logistic logistic equation and the third one is the normal density you know functions which we will discuss just now. That is what is called as you know probit models. So, let us come back to the probit model structure.

(Refer Slide Time: 11:07)



That is what the case says.

(Refer Slide Time: 11:11)

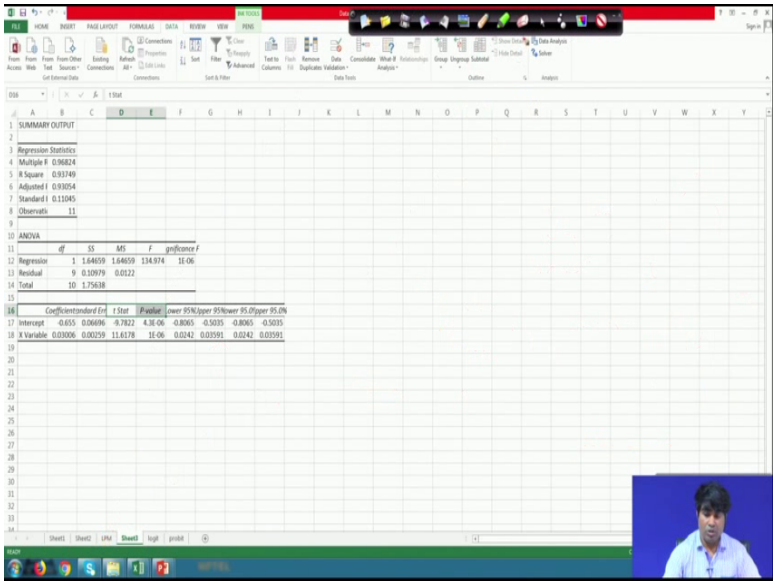


This is what the probit model and in the probit model and like the previous case, you know log p upon 1 minus p a b with you know alpha plus beta x 1 beta x 2 and so on. So, here is the same structure. So, that means technically there is no difference again. So, this side is more or less sense only. Z is the kind of you know transformation. So, that means we have a linear probability models, we have a logit and we have actually probit, ok. So, in the logit case, it is the log P by 1 minus p, ok. Log p upon 1 minus P and in case of

linear probability model, simply Y which is actually in between 0 to 1 and in the case of you know probit, it is the kind of you know you know normal density concerned as to that is here actually it is the case here and in both the cases, the sample information will be in between 0 to 1 only. So, no 0 no 1.

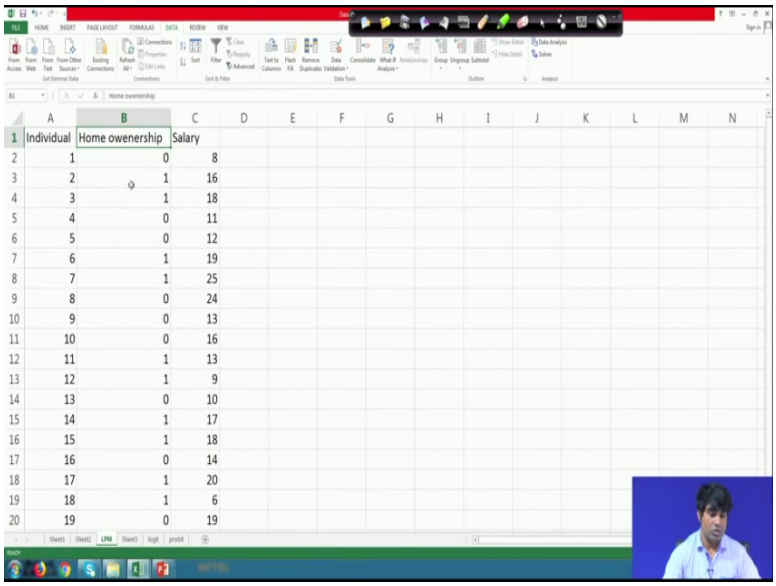
So, it is in between that is the game which you can like to apply here in the case of you know probit model. So, what we can do actually. So, let us take the examples so; that means, what we have discussed earlier.

(Refer Slide Time: 12:30)



The similar examples I am just taking. So, that means if you solve this you know problems like you are not is the linear probability case.

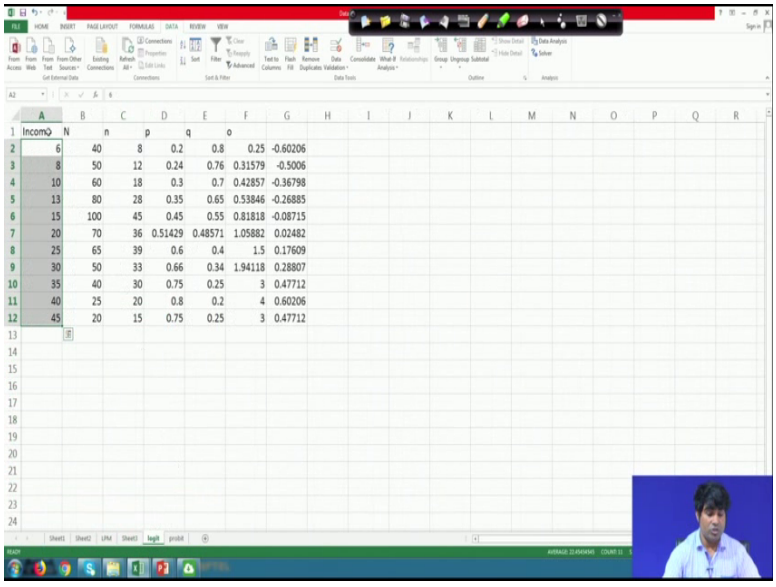
(Refer Slide Time: 12:41)



Individual	Home ownership	Salary
1	0	8
2	1	16
3	1	18
4	0	11
5	0	12
6	1	19
7	1	25
8	0	24
9	0	13
10	0	16
11	1	13
12	1	9
13	0	10
14	1	17
15	1	18
16	0	14
17	1	20
18	1	6
19	0	19

You know people having house and the salary structure that too through linear probability model, these were a final sampling which you can use for the estimation and again if you go to the actual logit model, this is the data structure which you like to follow finally to estimate the models.

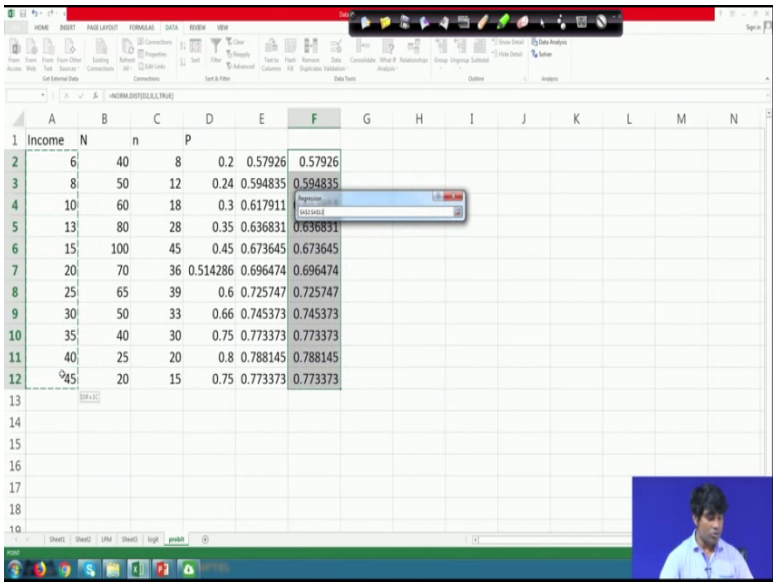
(Refer Slide Time: 12:55)



Incom	N	n	p	q	o		
6	40	8	0.2	0.8	0.25	-0.60206	
8	50	12	0.24	0.76	0.31579	-0.5006	
10	60	18	0.3	0.7	0.42857	-0.36798	
13	80	28	0.35	0.65	0.53846	-0.26885	
15	100	45	0.45	0.55	0.81818	-0.08715	
20	70	36	0.51429	0.48571	1.05882	0.02482	
25	65	39	0.6	0.4	1.5	0.17609	
30	50	33	0.66	0.34	1.94118	0.28807	
35	40	30	0.75	0.25	3	0.47712	
40	25	20	0.8	0.2	4	0.60206	
45	20	15	0.75	0.25	3	0.47712	

That means this is the final transformation data and that with independent variables and now, we come to the third you know third one that is the probit model.

(Refer Slide Time: 13:11)



Here the probit model structure is it seems like you know logit model and we need actually group sampling because the first hand information should be in between 0 to 1. So, is also best way to represent the group sampling rather than individual sampling.

So, what we can actually do here is, we have actually same structure of the data. So, income then you know we have actually a you know organizational choice or a household choice, locational choice and then, you know you try to find out how many are actually very effective. Same questions you know whether they have the own house or they are staying in the rented house.

So, that means first organization we are you know targeting 40 and out of which 80 are actually having own house. So, by default probability will be 8 by 40. Similarly second organization 50 will be the target. 50 will be the final target out of which 12 is the case of you know having own house. So, by default well by 50 is the probability values likewise 60 80 28 100 75.

So, we have actually the probability that is simply the ratio between small n to capital N like the previous one which you have discussed after getting the probability value. So, we will go for actually you know normal density functions in the case of logit. So, we need actually 1 minus P that is Q, then you find P by 1 minus P and then, you will be go for you know log of this particularly you know odd ratio.

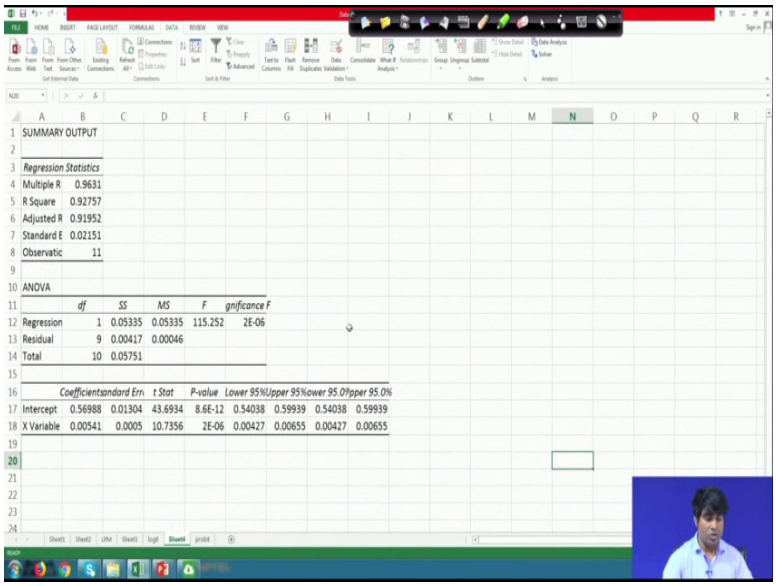
So, here what will you do after getting the P; you can directly go to the normal density function. So, that means what here we have done here is, so this is the normal density function. So, we just apply the transformation and then, you know scroll. It will get you know the entire you know observations, right. So, this is what actually the kind of you know case through we can get the data and what you can do here you just scroll it, ok. Then, you know we can get to know this value, let us come. So, this is what is actual transformation.

So, what we are doing actually just you put equal to sign here. I am just showing how it appears, then you ask for you know normal distribution transformation, ok. So, this is you know normal distribution transformation and click here. So, the moment you click, you will find here the first hand requirement dust and that is actually the kind of you know P value, then the requirement is mean and standard deviations which you can put actually let us say 0 and 1 and then, the cumulative cumulative structure, ok. So, then all right all right, ok. So, this is what actually be kind of you know I just add it, all right. This is coming, ok.

So, there is a slight error here all right, ormal distributions. See yes you can click this once and then, you find out here the value and again then 0, then 1 and then, yes now it is clear. You just put this one's and you know the close you will get the value, that is exactly coming, ok. So, now you can scroll this once. Yes the symbol is appearing here. So, now this is what the actual requirement of the dependent variables and that too is actually NDF transformations, Normal Distribution Constant Transformation.

Like logit models, we have locked P upon 1 minus P here. Simple actually in the kind of you know transformation, we have to apply for you know probit model, then again you go to the data analysis and I can put you know, and reset the structure and here this will be the you know information which will be for probit model and as usual this is the income structure and then, around the models which will appear like this.

(Refer Slide Time: 17:58)



Again this is an interesting model again and we will find you know very you know higher square and F is the statistically significant and then, the coefficient is also when say the x variable coefficient is also statistically significant and the kind of you know prediction is very good here and that means, technically if you compare let us say here 0.96, then R square is 0.93 115

So, if you compare with logit model output, the logit model output will be 1, F is 134.97 and here if we say probit model, then it is 115 5, ok. So, that means 134 and 15. So, here I think you know if you compare these two models, then ultimately logit model gives you a better kind of you know you know kind of you know indications while predicting the independent variable to dependent variable that too oh you know salary with the you know, ok. So, you know other people having own house or rented house.

So, that means actually once you identify the problem and you know structure, the problems you know to know the dependent variable structure, independent variable structure, then you know just you know apply the estimation by linear probability models, then logit model and probit model. Of course, the sampling structure and the dependent variable structuring is a little bit you know different, but the way we have actually done the transformation, it is not so difficult one and in fact, some of the softwares you need not require to do the transformation. You must have the faster net output and by default software will transfer into the logit format and probit format, then

we will go for the kind of you know estimation. Ultimately what is happening that you know we have 3 different you know kind of you know alternatives and same problems you can investigate with these 3 alternatives.

In one case, the things are in a linear in character, then the other two cases it is in non-linear in characters where 1-1 case we use logit functions and another case there is actually normal density function since we are using different functional for linear and then, non-linear with logit and non-linear with normal density function, then there is a high chance you know you know the model results will not be uniform. Of course, what is the unique of this you know 3 models whatever we have tested here that you know every time the coefficients are coming positive.

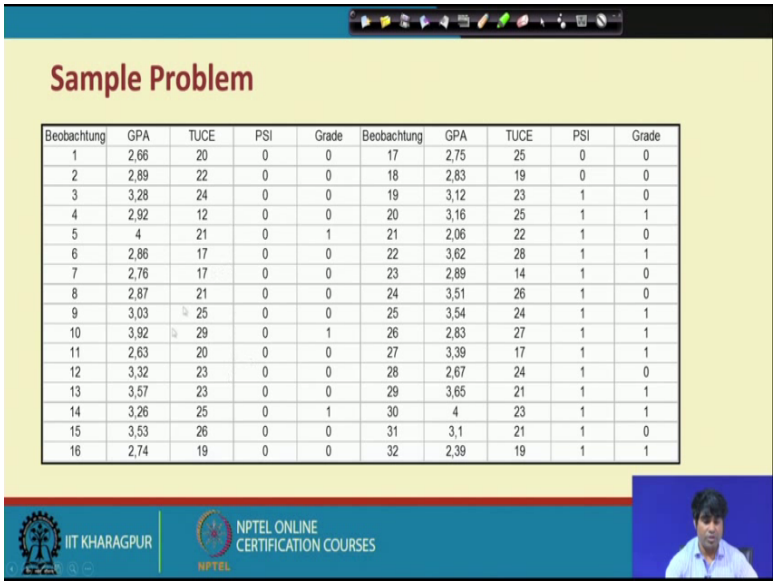
So, that means it is going as per the theory, but so far as reliability is concerned and you know the logit model is much better compared to linear probability model and probit model, of course not necessarily every time by the logit model will be better choice than the probit model or the linear probability model. In some instances, linear probability model may be good to predict the dependent variables, which is actually qualitative in nature corresponding to independent variable and the other case you know again budget model may be good to predict the dependent variable with respect to independent variable.

So, technically you know we have no idea which model is actually very effective for this case. So, we can just test it, then we can get to know which one is the you know good models and then, final choice for the prediction will be on that model only. So, the other models by default will be discarded. In fact, it is a kind of you know robustness check process since you know same problems we are investing you know with respect to 3 different formats.

So, then finally we have to choose the particular format which is very effective and that will be on the basis of reliability check and the kind of you know model requirements and more or less in nature of the kind of you know problems will not you know you know different, but you know the reliability part will be specifically you know different which we have already seen here in these samples while you know collecting salary with you know people having the own house or the rented house. So, that is the kind of you

know situations where you can use dinner probability models, logit models and probit models.

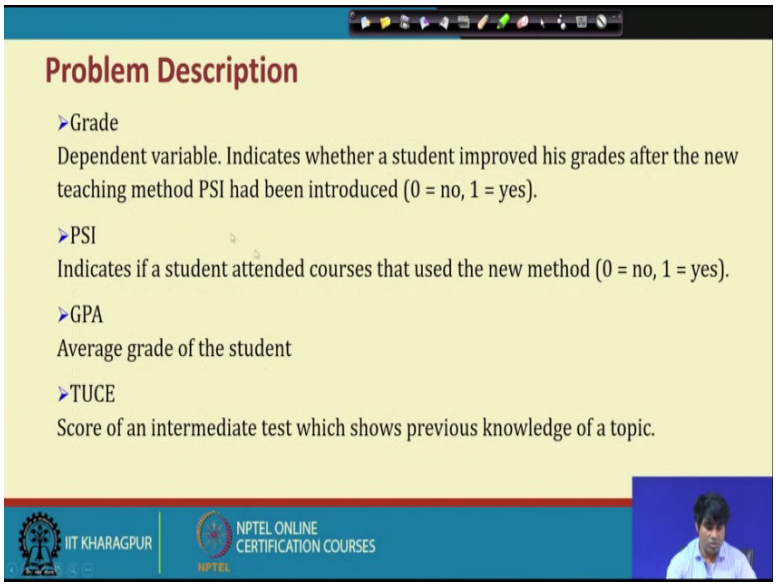
(Refer Slide Time: 22:40)



Beobachtung	GPA	TUCE	PSI	Grade	Beobachtung	GPA	TUCE	PSI	Grade
1	2,66	20	0	0	17	2,75	25	0	0
2	2,89	22	0	0	18	2,83	19	0	0
3	3,28	24	0	0	19	3,12	23	1	0
4	2,92	12	0	0	20	3,16	25	1	1
5	4	21	0	1	21	2,06	22	1	0
6	2,86	17	0	0	22	3,62	28	1	1
7	2,76	17	0	0	23	2,89	14	1	0
8	2,87	21	0	0	24	3,51	26	1	0
9	3,03	25	0	0	25	3,54	24	1	1
10	3,92	29	0	1	26	2,83	27	1	1
11	2,63	20	0	0	27	3,39	17	1	1
12	3,32	23	0	0	28	2,67	24	1	0
13	3,57	23	0	0	29	3,65	21	1	1
14	3,26	25	0	1	30	4	23	1	1
15	3,53	26	0	0	31	3,1	21	1	0
16	2,74	19	0	0	32	2,39	19	1	1

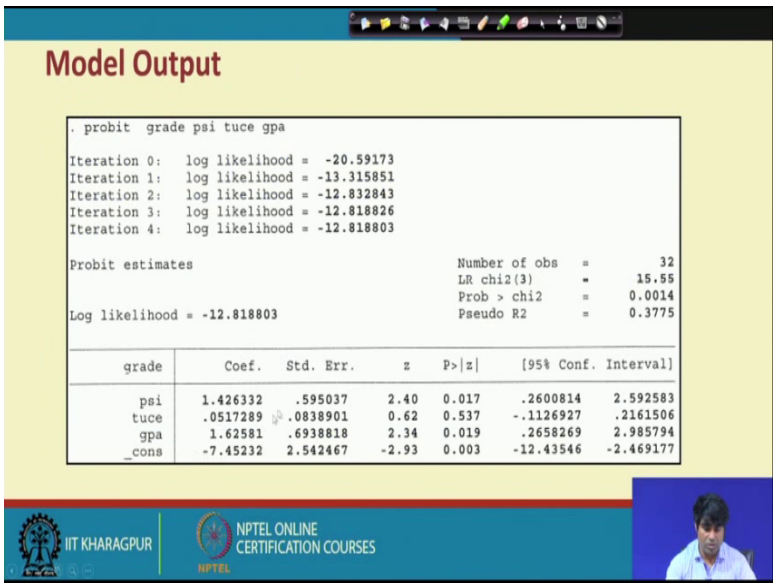
So, now you know this is how the kind of you know structure and similarly there are you know various you know other forms of models you know problems where you know we can use either you know linear probability models or logit model or probit model, but ultimately the transformations should be actually you know either in 0 1 format that with dependent variables or in between 0 0 to 1 and that too again dependent variable. So, no restriction or no specific indication about the independent variables group, but dependent variable structure should be like that to apply linear probability model, logit model and probit model.

(Refer Slide Time: 23:24)



So, this is another example and by the way these are all similar kind of you know testing you can do the same you will find.

(Refer Slide Time: 23:27)



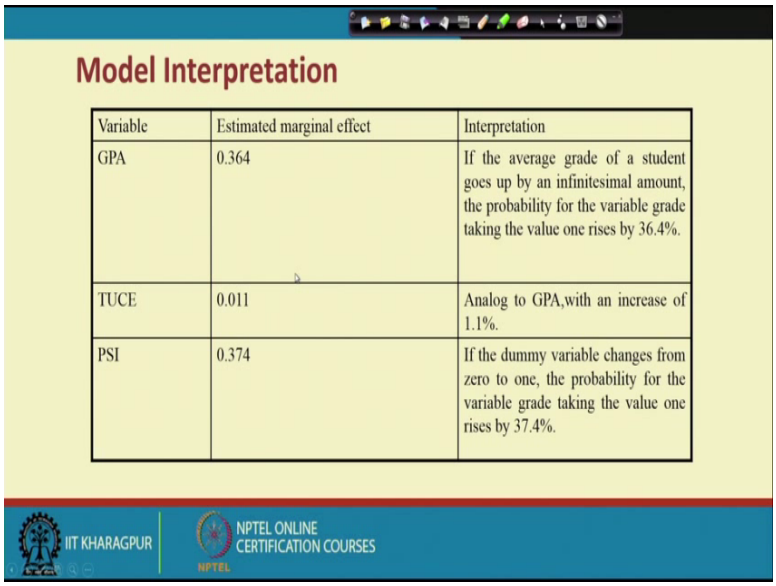
Then, you have to check you know the kind of in fact log likelihood ratio also can be an indicator which can be actually a project whether the model is equal to it or you know benefit. So, you know sometimes you know in that case of you know logit and probit model we use actually a shadow R square instead of you know simple R square in adjusted R square, but ultimately these are all different you know dynasty parameter

through which you can actually justify the reliability of the model, but you know whether R square or F that you know or you know some kind of you know shadow R square or you know chi square, all these indicators will be you know very useful, but ultimately if the model is good, everything will be in a right shape. So, the reliability will be also very good.

So, then final choice will be on the basis of reliability and then, it can actually pick up a particular you know models to analyze the situations where the dependent variable will be qualitative in nature and that with respect to independent variables. So, that means we have here lots of you know flexibility to predict a particular you know engineering problems where you know again dependent variable will be qualitative in character that too in a kind of no probability structure and then, we remain linked with you know or we will link with you know the independent variables, where you know all are numeric in nature or you know few are actually numeric and few are qualitative or maybe few are also means all are maybe you know qualitative in nature.

So, that means, technically dependent variable is here restrictions. So, no restriction to the independent variable it may be any point, but a dependent variable should be in a specific form, ok. So, either in 0 1 format or in between 0 1 format so, that is why it is a specialized kind of you know modeling and as a result it cannot be applied each and every engineering problems. So, you have to apply this problem in a specific kind of you know situation where there is a need and you know there is a kind of you know requirement.

(Refer Slide Time: 25:50)



So, these are all various examples and how you have to interpret exactly. So, what I have already told you know in the case of you know logit model and probit model, the kind of you know reliability and the kind of interpretation is a slightly different and through different kind of you know indication. For instance, in the case of logit, we use here odd ratio to you know give the interpretation or to analyze the kind of you know problem while you know predict independent variable with the independent variable.

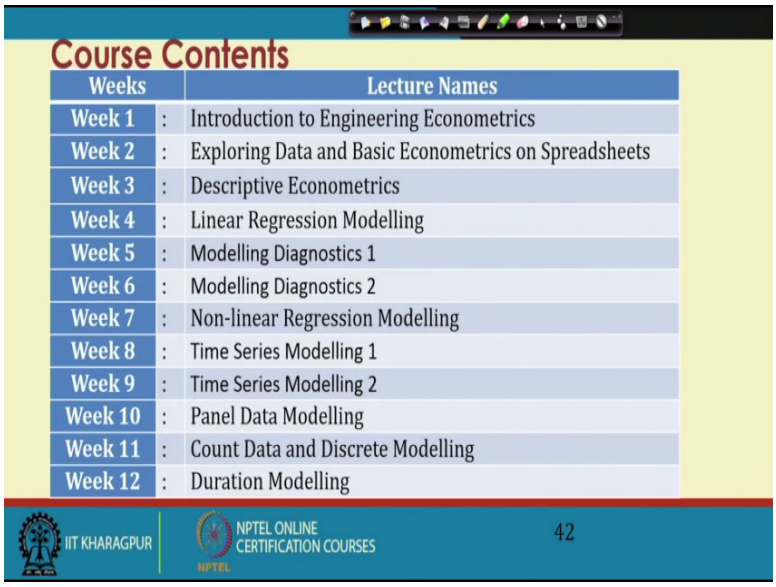
So, that means technically you know in the case of you know non-linear regression modeling, we have plenty of choice and you know to you know pick up a particular problem and to analyze the problem as per the particular engineering requirement and the kind of you know management decision requirement.

With this, we will stop here and in the next class, we start with you know time series modeling, then after knowing the time series modelling, we can go for the panel data modelling since we have discussed the dummy modelling concept and that is you know one of the big component or you know important component which you can understand the panel data modelling and since in the panel data modelling, the structure will be both time series structure and you know cross sectional structure.

So, first we discuss the time series structure and whatever problem we have discussed till now is a kind of you know cross-sectional structure and after knowing the time state structure, we can move to the panel data structure and various other types of you know in

a special kind of you know problems which I have already highlighted you know like the case here you know or like what we can call you know Count data model discrete you know modelling and these kind of duration modelling.

(Refer Slide Time: 27:41)



So, also this is you know flow if you like to maintain and you know before you go to this you know specialized is kind of you know modeling. So, let us you know see how is the time series modeling and with this you know we can stop here.

Thank you very much.