# Econometric Modelling

**Prof. Rudra P. Pradhan**
**Department of Management**
**Indian Institute of Technology, Kharagpur**

**Module No. # 01**
**Lecture No. # 28**
**LOGIT and PROBIT Model**

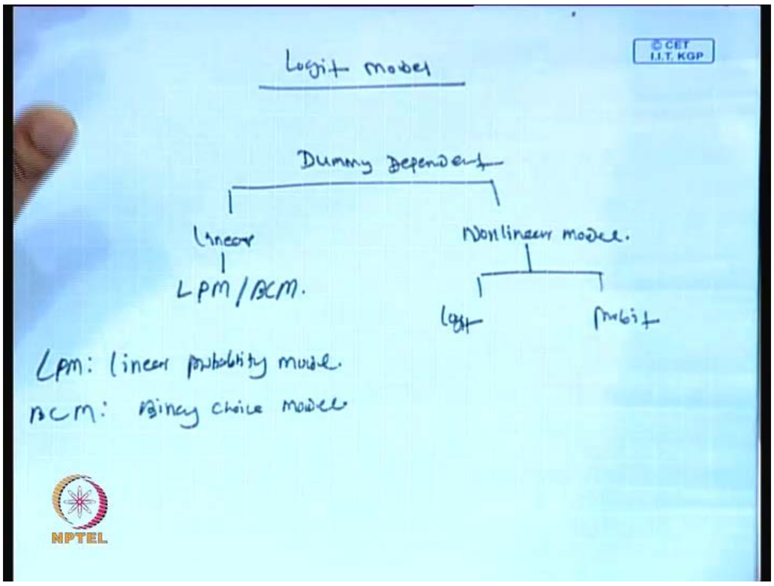
Good afternoon, this is doctor Pradhan here, welcome to NPTEL project on Econometric Modeling. So, today we will discuss Logit and Probit model. In the last lecture, we have discussed the dummy variable modeling. So, that too dummy independent variables, so today, we will specifically highlight dummy dependent variable modeling.

So, in other words, these particular structure is called as a qualitative response econometric modeling, this is very interesting topic. In fact, we have highlighted details about the independent side of the pictures where some variables are quantitative in nature and some variables are qualitative in nature.

And today, we will specifically highlight, if the dependent variables are dummy in nature means binary or categorical in nature, then how is the structure of econometric modeling, that is how we have to discuss.

So, now basically in the, before you go into discuss logit model and probit models, we briefly highlight what is the exact structure of qualitative response econometric modeling for dummy dependent.

(Refer Slide Time: 01: 35)




So, now for dummy dependent for dummy dependent, the model basically divided into two parts, it is called as a linear model and it is called as a non-linear models, linear model and non-linear model. Basically, you know, we will use simple simple models like straight line equations, you know in another way, we will represent the non-linear models like you know basically, we use logistic functions, particularly probability distribution function or normal distribution function. Because in contemporary logit models, there is another model called as a probit models.

So, we will discuss means, we will discuss this probit model after the discussion of logit models. So, now in the linear linear format, you know in linear format, the dummy dependent econometric modeling is called as a a linear probability models, linear probability models or BCM, binary choice models; that means, LPM stands for linear probability models linear probability models and this is BCM. So, otherwise it is called as binary choice models binary choice models, you know dummy dummy is a variable.

The variable basically means, which we deal in the regression analysis is that, it may be quantitative in nature, it may be qualitative in nature, it may be it may be both, but here, here we are considering both the aspects, means the model which involves both quantitative variables and qualitative variables, qualitative variables where data is not in a proper step. So, we will bring into proper step, the way we will bring the proper step, it is called as a dummy variable technique.

So, now the linear format the, in the linear format of dummy dependent econometric modeling is otherwise called as a linear probability model or binary choice models. In the other case, there is there is a model called as a non-linear model, that too dummy dependent dummy dependent econometric model where non-linear function is you know given importance.

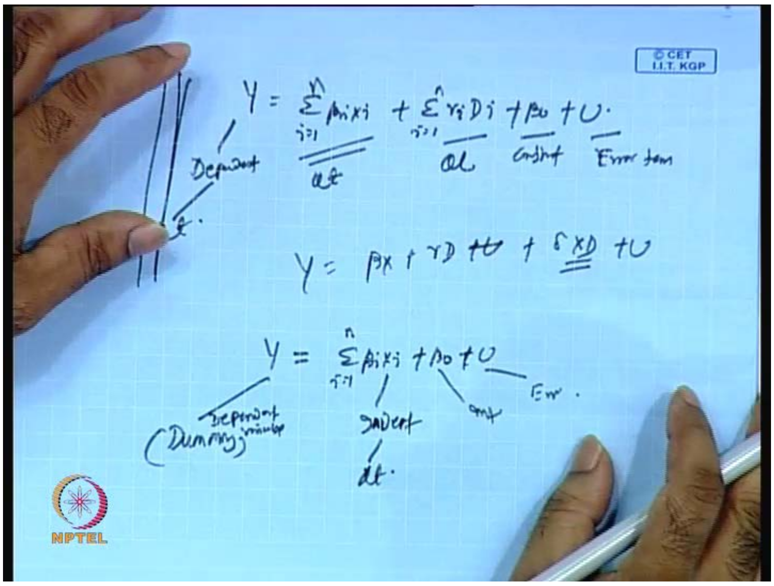
So, non-linear model is otherwise divided into two parts, one is called as a logit models and another is called as a probit models. It is otherwise, it is called as a nomit models, Tobin models like this, so there is lots of classification under this group.

So, basically we will not discuss all these classification in in details. So, what we will do, we typically we will discuss three aspects of you know dummy dependent econometric modeling, that is binary choice models, otherwise called as a linear probability model, LPM, then logit model, then probit model.

So, we start with first binary model, because it will give you signal to, entry signal to this logit model and probit models. In fact, some of the problems, it is nicely, it can be describes in the case of binary probability model. But the same problems can be discussed under logit model also, because logit model is little bit advanced in the binary choice model, because binary choice model has a certain limitation which can be taken care in the case of you know logit model and probit models, like you know it is just, it is one one way of movement for you know higher version and you can say very interesting and complex problem, like you know covariance to correlation and regression. Similarly, this you know structure of dummy variable modeling is that, you start with binary choice model, then logit model, then probit model.

So, we are just moving in the higher directions where the model accuracy will be more and more accurate. So, first of all, what is this binary choice model and linear probability model? So, before I highlight this binary choice model and linear probability model, so I like to highlight certain discussion which we have made in the last class.

(Refer Slide Time: 06:10)




So, that is nothing but we will put Y equal to Y equal to, what is that summation? beta i X i plus summation gamma i D i; i equal to 1 to n and i equal to 1 to n here, plus beta 0 plus U. So, this is this is quantitative variables, this is qualitative variables, this is constant and this is error terms, this is dummy, sorry this is simple dependent variables and it is also quantitative in nature.

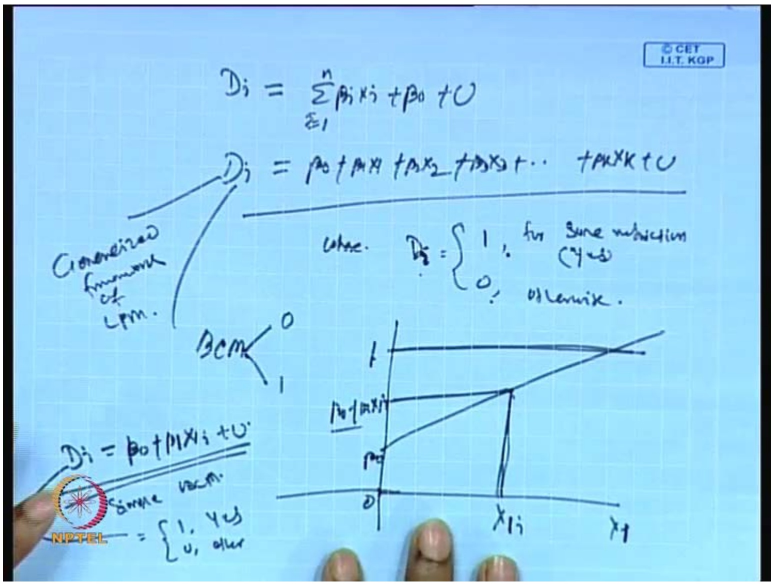
So, in that case, this is in the last last class discussion where you know variables are, some variables are quantitative in nature and some variables are qualitative nature. That means, in the simple regression dummy modeling, so we we may use one dummy variable or we may use multiple dummy variables. So, accordingly we have discussed the individual effect, and we have also discussed the interactive effect. For instance, if there is model like this Y equal to beta X plus you know gamma D plus U, then obviously, we will create another variables say delta into X D.

So, then we will call it plus U. So, this this is called as an interactive effect, which we have already discussed in our last lectures. So, here we will not discuss all these details, so we will directly proceed to linear probability models.

So, now in contemporary to this is you know, in contemporary to last class discussion. So, what we will do, for linear probability models, we like to write like this. So, Y equal to Y equal to summation beta i X i plus beta 0 plus U. So, I equal to 1 to n, so that means, here this is dummy dependent this is dummy dependent means first, it is dependent, dependent variables dependent variables and it is purely dummy in natures.

So, this is independent variables independent variables then it is it is independent variable, then it is otherwise called as a means, otherwise it is represented as quantitative quantitative variables. So, this is intercept and this is error terms. So, this is how the linear probability model is all about, so I will get it highlight in other way.

(Refer Slide Time: 08:44)




So, we can call it like this delta, instead of Y i we can put it also D i, D i is equal to summation beta i X i plus beta 0 plus U, so i equal to 1 to n. So, it it can be otherwise written as D i equal to beta 0 plus beta 1 X 1 plus beta 2 X 2 plus beta 3 X 3 plus continue plus beta k X k plus U. So, this is the generalized generalize framework of generalize framework of linear probability model, this is the generalized framework of linear probability models.

So, now we have dependent variable in the left side, and all these independent variables in the right side, where the dependent variable is use as a proxy variable, that is dummy variables and other variables are you know as usual quantitative variables, means all these variables can be measured in a quantitative way, so it is not a problem. So, here problem is the dependent variable is the dummy in nature, that means it is it may it can be binary, it can be categorical. So, many ways it can be response variables, so we will we will represent in many ways.

So, now here, where you can say, what we will the condition we will obtain here is that D i equal to 1 for some you know for some restriction for some restriction or you can say for yes and 0 is otherwise 0 is otherwise. So, that means in this this binary choice model, this is otherwise called as a binary choice model.

So, that means once we will call it binary choice model, then there are two way representation, two way representation is that, so it may be value lies between, means value, its value lies between 0 to 1. So, this this is binary choice model, so its value lies between 0 to 1. So, if yes, then it will come to 1, and if it will no, then it will come to 0.

So, that means, that is how it is called as a binary choice model. So, the. So, the. So, the basic formats of you know dummy dependent variable is that, it is the model is called as a linear probability model. So, otherwise called as a binary choice model where the dependent variable is categorical or binary in nature. So, it will it will move from 0 to 1 only and you know in other case, you know other case independent variables are completely independent and you know they are very much quantitative in natures.

You remember one thing, so after you know proper specification, the model estimation is more or less same, like which we have discussed couple of lectures back. So, like you know bivariate, trivariate or multivariate, so usual format is as usual same. But only thing is here means, only extra thing we are adding here is that, sometimes dummy variable means, some variable are qualitative in nature in the right side and some of the variables are qualitative in nature in the left side left sides.

So, that means till now, we have not touched the structural equation modeling where you know, there are many dependent variable and you know many independent variables. But here, till now we are discussing, there is one dependent variable and several independent variables that means, one dependent variable with one independent variable or one dependent variable with multiple dependent, independent variables.

So, now in the second case, so we are discussing the dummy variable technique where dependent variable is one with several independent variables, some are quantitative in nature, some are qualitative in natures. So, that we have discussed in the last class, but to in our today's discussion, so we are taking keeping all variables in the right side constant, multiple in natures, other sides we are taking single dependent variables which is purely qualitative in nature and its lies its value lies between its value value lies between 0 to 0 to 1 0 to 1.

So, if it is 1, then it is yes situation; if it is 0, it is no situation. So, means the way the the way we will define the problem, so accordingly we will categorize it or you can say you will call the, or ensure the binary code all right. So, this is how the structure, you check it here, so how I will represent this particular you know binary choice model. So, this is this is how the structure is all about, so this side I will measure X i. That means, let me highlight, we will put it in a simple format; this is in fact, generalized formula, means generalize models.

So, if we will put it simple deviation models, then delta equal to simply beta 0 beta 1 X 1. Obviously, i is there because I represent this sample observation plus U, so this is the simple simple binary choice model simple binary choice model. So, that means here, D i equal to 1 for yes and 0 for otherwise 0 for otherwise. That means, so the value lies between like this way. So, this is 1 now corresponding this 1, you will draw the line here beta 0. Now, you know this this this will continue. So, obviously some point of time, the mean has to be calculated. So, this is you can say you know X X 1 i, so this should be X X 1, X 1 i then this side beta 0 plus beta 1 X 1 i.

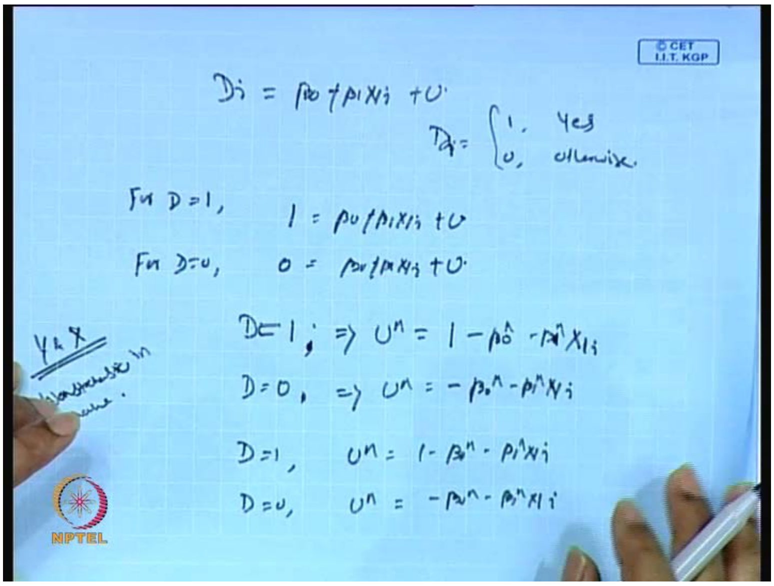
So, this is the this is how the structure, so that means, we are considering this particular area, so this maximum limit is 1, so it will lies between 0 to 1. So, that means whatever information we have, we will transfer into 0 to 1 format. So, that means the maximum limit is 1, then minimum limit is 0. So, that means in this particular in these particular structures, there are in fact two possibilities only. So, in one possibility is say, suppose this situation is yes, no, then one possibility, yes for 1 and no for 0 or vice versa.

So, that means, so there is no other way round. So, you have to design the question in such a way, the responds only answer yes or no, or you have to develop the question accordingly in such a way, so it will always yes and no situation. That means, either you can some of the you know dummy variable technique, you have to artificially create the particular structure or you can say feasibility.

So, accordingly you have to proceed with that structure and feasibility. So, because it is not a natural process, the other process may be little bit natural, but this is purely artificial and you have to design very very very perfectly, we for you know going for the estimations.

So, now D 0 is equal to beta 0 plus beta 1 X 1 i plus U. So, the maximum limit will 1, so 0 to 1, so minimum will meet 0. Now, if you will put here, suppose if I will put here 1, here if I will put 1 here, then beta 0 plus beta 1 X 1 i will be simply equal to 1, so that means, you see here.

(Refer Slide Time: 16:19)



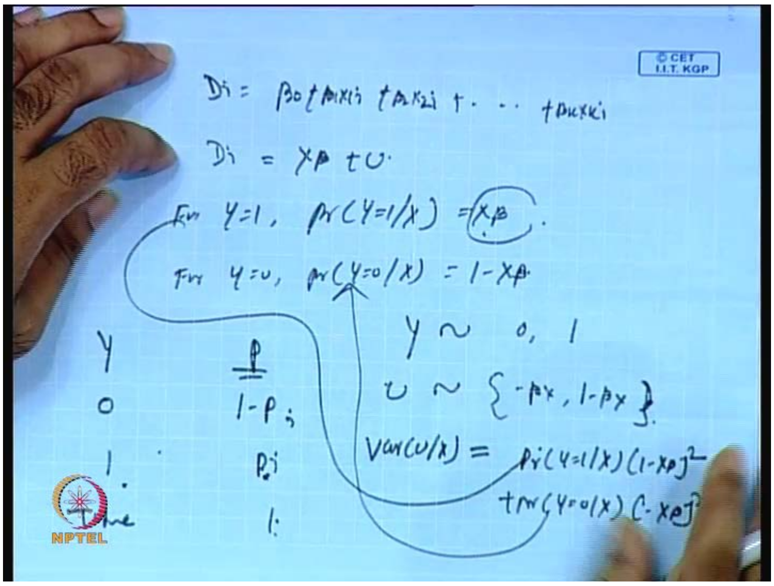

So, I will put it now other way, so for the simply simplest model is the D i equal to beta 0 plus beta 1 X 1 i plus U. So, D D equal to D i equal to 1 for yes and 0 for otherwise 0 for otherwise. all right So, now, so for for D equal to 1, for D equal to 1, then for D equal to 1, then the equation will be 1 equal to beta 0 plus beta 1 X 1 i plus U, for D equal to 0 for D equal to 0, then 0 equal to beta 0 plus beta 1 X 1 i plus U, so this is the simplest formula.

So, now the, if we will go for estimations, so if go for estimation, then you know here the Y and X relationship are non-stochastic. So, these are non-stochastic in nature non-stochastic in nature, so that means so when we will go for estimations. So, when D equal to 1, then obviously, we will estimate the model. So, obviously we will get error term Y hat equal to Y minus Y bar, so Y is equal to 1 here.

So, that means 1 minus beta 0 beta 0 hat minus beta 1 hat X. So, similarly, when D equal to 0, so then, that implies U hat equal to U hat equal to minus beta 0 hat minus beta 1 hat X 1 i X 1 i.

So, this is beta 0 hat beta 1 X, this is X 1 i and this is X 1 i, this is X 1 i. Now, there is no such for D equal to 1 that when there is D equal to 1, then that means, other words D equal to 1 means, so U hat equal to 1 minus beta 0 hat minus beta 1 hat X 1 i and when D equal to 0, so then U hat equal to minus beta 0 hat minus beta 1 hat X 1 i. So, this is how the model can be represented, but you know so when we will go for generalize model generalize models.

(Refer Slide Time: 18:45)



So, for instance if I will put here D i equal to D i equal to simply beta 0 plus beta 1 X 1 i plus beta 2 X 2 i plus beta k X k i. So, now if I will put like this way, so D i equal to simply X beta plus U. Now, for Y equal to 1 now for Y equal to 1, so p r Y equal to 1 for all X equal to X beta, now for Y equal to 0, then p r upon Y equal to 0 slash X is equal to 1 minus X B, because this structure is like this way Y then this probability levels.

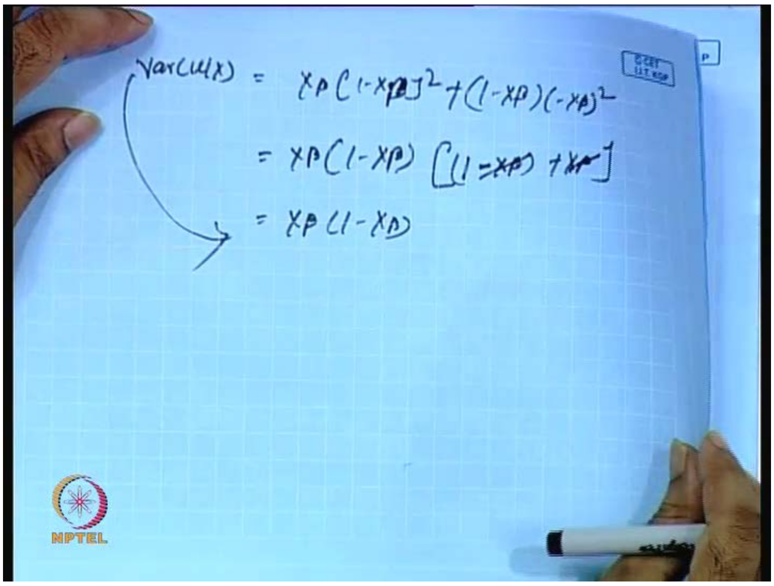
So, when Y equal to 0, then probability equal to 1 minus P i. So, when it is 1, then it is equal to P i. So, that means the total probability is always equal to 1. So, 1 minus p 1 minus P minus plus P i is exactly equal to 1.

So, now similarly, so when Y equal to 1, then you know probability of probability of, probability is total probability is equal to 1, so one one probability of Y equal to 1 for X equal to simply X beta.

Similarly, when Y equal to 0, then probability of Y upon 0 our X will be 1 minus X beta. So, that means the structure is if Y stands to 0 1, then you know this residuals this is this is this is in fact is residuals, this is in fact residuals. So, now so, the residuals U will stands for U will be stands for minus beta X when 0 it will be 1 minus, 1 minus X beta 1 minus beta X, then it will be 1 minus beta X 1 minus beta X. So, this is how the structure is all about.

So, now when Y stands for 0 1, so U stands for minus beta X into 1 minus beta X. So, similarly we we can calculate the variance of U, variance of U U upon, so means our X will be probability probability of Y equal to 1 over X into 1 minus X beta X beta whole squares plus p r into Y equal to 0 upon X into into minus X beta whole squares minus X beta whole squares.

(Refer Slide Time: 21: 42)



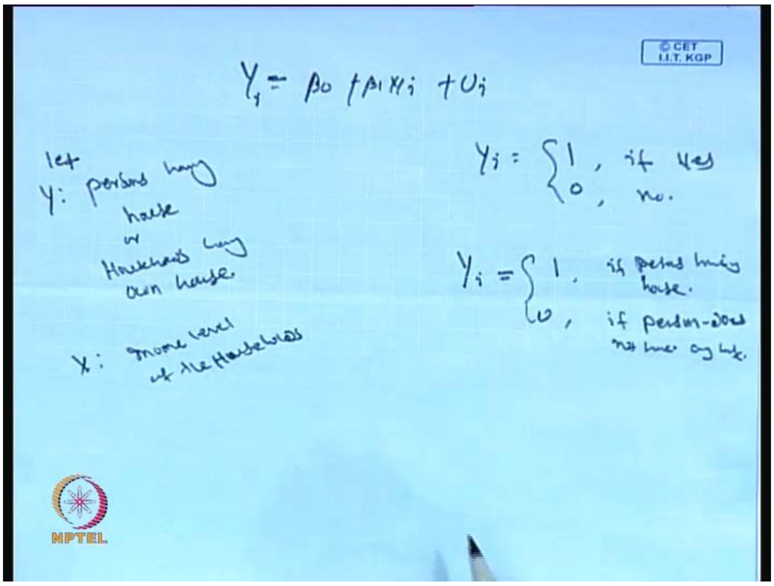

So, that means if we will if we will insert if we will insert this particular structure, and if we will insert this particular structure, then obviously, what we will what we will get. So, variance of U variance of U over X will be equal to X beta X beta into 1 minus X beta whole square plus 1 minus X beta 1 minus X beta into minus X beta whole squares. So, that implies if X beta X beta into 1 minus, this is X beta only, this is X beta 1 minus X beta will come on. So, then it will be 1 1 1 minus X beta 1 minus X beta plus X beta plus X beta. So, that means X beta into 1 minus X beta because this and this will be cancelled, so this is the variance of variance of U.

So, now what we have received in the linear probability model, so there are two different limits. So, one limit is Y equal to 1 and another limit equal to limit is Y equal to 0. So, that means the the value, the variable which is qualitative in nature. So, we will transfer all these information into two two different formats. So, one some of the items will be in the form of 1 and some of the item in the form of 0, depending upon the information availability.

So, for instance you know, I am targeting the issue of you know what is the person having household and the income. When if you will integrate these two, then obviously, then I will ask the respondent according to their income levels, whether you have you have a house or not. So, obviously they will say yes or no situations, so that means, I have to specifically I have to ask two questions, one is do you have any house, means answer is yes or no, then I like you what is the income levels. So, we we like to the corresponding income levels, so that is X component and then Y component is the persons having house or not.

So, that means like this, let me take it this with a practical example. So, what we will do, so instead of you know elaborating in a bigger size, so will take a small problem, then we will highlight the exact issue of linear probability models all right.

(Refer Slide Time: 24: 10)




So, what you will do? So, we will take Y equal to beta 0 plus beta 1 X 1 i plus U i. So, this is U i, so where Y i equal to 1 1 and 0, if some yes and if it is no.

So, let me give a example here. So, let Y equal to say, persons having having household, having house or or you can put households, having households, having one house, this is how the question where you have to design. So, that means here the the specific problem is that, we like to know what is the impact of income on you know having household. Basically, when you discuss about to binary model, binary choice model, probit model and logit model, so it will be applied on the basis of their problem setup, if the problem is like this, then you have to apply the binary choice model, but the same problem can be analyzed also in the case of logit and probability, probit model, but it is in a different, with different setup.

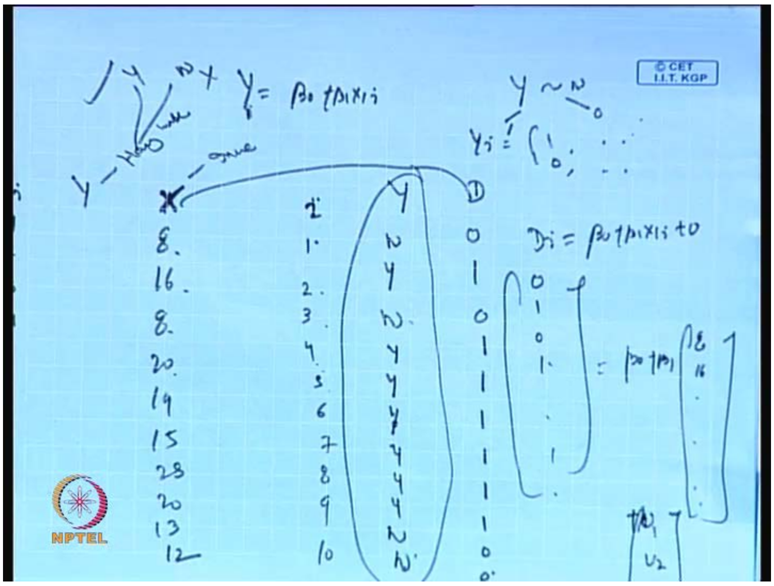
So, now in the mean times, let us assume that Y is is a variable which which is recognized as a person's having house and or or you can say household having one house, this is Y representation. Similarly, X representation is the income level of the income level income level of the households, income level of the households.

So, now you see here, so the presentation will be like this. So, that means, so here, persons having housed or household having on house means obviously, the question is yes or no situation. So, what we will do, we will categorically divided into 1, if persons having persons having you know house person having house, either you write it here or else what you will do.

So, questionnaire is already designed, so obviously, if we will put Y equal to 1, then it means the person has a house. That means, it is yes situation, then if person has not house, then it will no situations. So, that means, so it is 0, so that means, if persons does not have house or household does not have any house, person person does not have does not have any house have any house this is 1 0 situation.

So, how do we represent this particular structure, binary phase model? So, now what we will do, so so we like to know, means the interesting interesting fact in the binary phase model is that, how to bring this particular setup, so that the binary choice model can be applied. Initially, you you may not have such options, the moment you will get data, then obviously, with basis of data, you you have to apply some you know dummy variable structures. So, the moment you will transfer the available information and to dummy variable structure, then dummy variable technique can be applied. That means, if the binary choice model can be applied for estimating the parameters or the significance of particularly income to households.

(Refer Slide Time: 27: 48)




So, that means what you will do, so our basic model is Y equals to beta 0 plus beta 1 X 1 i. Here, Y i, Y i is dummy, so Y i is 2 range 1 and 0, 1 for something and 2 for something. So, that means, so with this particular problem, so we have taken Y and X.

So, this is say sample size i, so, let us take it here, so i is the sample size, so 1 2 3 4 5 6 7 8 9 10, so 10 samples size we have taken. So, income is income is you know quantitative in nature, household income is quantitative in nature. So, what we will do, I will put it here 8, then 16, then 8 20 19 15 25 20 13, then 12. Then if you want to increase, you can also increase or of you want to decrease, you can also decrease.

But this is a simple hypnotically problems, where you know income level of that particular set is like this. So, then these are all individual units like you know 1 2 3 4, this is already written here, so 1 2 3 4 5 6 7 8 9 10. So, similarly, for person 1, his income level is 8, person 2 income level is 16, person 3 income level is 8, person 4 income level is this much, then this is the first question, means the question value will be designed in a such a way that, the first question must be what is what is your income levels.

So, you have to find out the answer first, the moment you will say the, my information is this much, then obviously, you have to ask the second question, do you have any house, one house? So, obviously there will be yes or no situation, that means in the first one, we will ask the income level, then you will give some quantitative value, means say you know 5000 or 20000 or something something. So, this is how you can say income is represented.

So, similarly corresponding income level, then I will ask do you have any house, means, so his answer or her answer will be simply yes or no, So, that means Y is X is here income and Y is having having house. So, if house is there, then I will put yes, if not then I will put no. So, that means the having house, it will turn into yes situation or no situation, yes means it is there, no means it is not there. So, that means persons with the for instance take a case of 8, a household having income level is only 8 dollars.

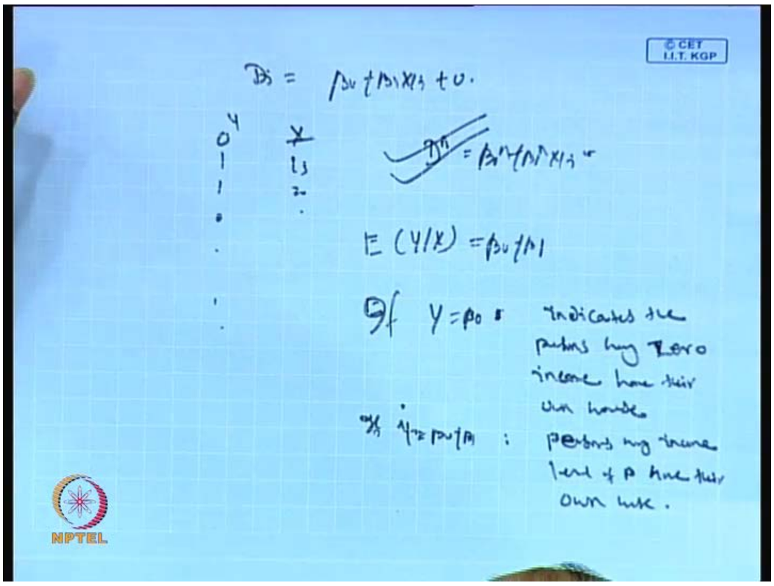
So, now with 8 dollars, so now whether he he has house in that city or not, so that means, if we will ask do you have any house, so obviously, answer is yes or no. So, if it is yes, then you put yes, if it is no then you put no. So, that means I will I will take it this sides, so Y is here. So, 8 eight 8 I will take said no 16 yes, then then 8 no, then 20 yes, then 19 yes, then 15 yes, 15 yes 15 yes then then then then then 9 20 19 yes, then 15 yes, 25 yes, then 20 yes, 13 13 13 no, then 12 no. So, this is how the structure is all about.

So, now, so this is yes no situation, so you know the computer will not recognize this you know letters. So, what you have to do, you have to transfer into some binary information, so that means, it is only difference is Y and no. So, that means if we will put Y equal to 1, then n obviously 0, because total probability equal to 1. So, that means means you forget about this probability, but means yes no situation means, by default you know there are, if you will go up to 1, then obviously, there are two possibilities, 0 and 1. If it is yes, we will code 1, if it is no we will code 0.

So, then accordingly, you have to transfer this particular series. So, n transfer 0 Y transfer n 1 n transfer 0, Y transfer 1, Y transfer 1, Y transfer 1, Y transfer 1, Y transfer 1, Y transfer 1 then 0, then 0. So, this is how the table is prepared, so now, once the table is prepared. So, you will forgot now Y, so Y is the, in the mean time, there is no such you know rule, so what you have to do. So, we have to transfer Y X into D that means our standard equation is, so D i equal to beta 0 plus beta 1 X 1 i plus U. So, that means X and D is already there, now what you have to do. So, you have to regress like this, 0 1 0 1 1 like this.

So, then in the other sides, other sides this beta 0 beta 0 plus beta 1 into you can say you can say 8 16 8 like this. So, this is how the representation is there plus U, so U 1 U 2 U 3 like this. So, U 1 U 2 up to U 1, so this is how the picture is all about all right.

(Refer Slide Time: 33:23)




So, now what you have to do, so you need to have a binary choice model. So, what you have to do, that means, so D i equal to D i equal to beta 0 plus beta 1 X 1 i plus U.

So, now as usual, you have one is 0 1 1 0 0 0 like, then X this is Y information, then X information, similarly quantitative in nature. So, 15 20 like this way, so then obviously, you can go for estimation. So, you can get the D hat equal to D hat equal to beta 0 beta 0 hat plus beta 1 hat X 1 i, that is all.

So, this is this is how the estimated regression equation for binary choice model binary choice model, now you have two options, what is the two options? So, that means E upon Y X, E for Y, our X indicates the beta 0 plus beta 1.

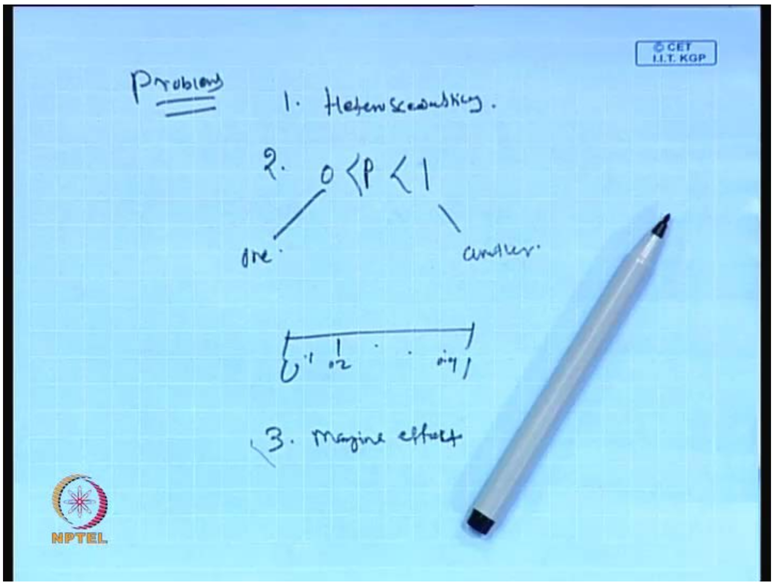
So, now E upon you know if if if if Y equal to beta 0, if Y equal to beta 0, then you know the probability that family with 0 income, that means it indicates, it indicates if Y equal to beta 0, then it indicates the persons persons having low, persons having persons having 0 income 0 income right person having 0 income persons having 0 income have their own house. Then if I will put like this, Y equal to beta 0 plus beta 1, then we will write persons persons having income having income level of beta 1 having income level of beta 1 have their own house. So, this is how the interpretation is all about.

So, that means binary choice model can be evaluated properly with respect to our problem setup. So, that means in this particular setup, so Y is very categorical or binary number, so its value is always lies between 0 to 1. So, once you will transfer this entire information qualitative information into yes no situation, that is 0 1. Then obviously, X is already quantitative, so you can go ahead with regression. So, you will finally, get the estimated models like this, so you will get the finally estimated model.

So, now the moment you will get this estimated model, then obviously, you have to go for lots of means, you have to specifically go for the validity of the model. So, as far as a validity of the model is concerned, so you have to go for specification test, Eigen test and overall fitness of the test. Then finally, finally this you know the problem of heteroscedasticity, multicolinearity, and then autocorrelation etcetera. etcetera These are things has to be, means have to be discuss again, so that means, here the thing is that, once you will transfer then everything will be in right direction or proper set, you can proceed accordingly.

But the thing is that, this binary choice model has a limited applications, the reason is that there are several limitations associated with binary choice model. So, first of all what are the problem associated with binary choice models? So, then since there is a serious problem in binary choice model, so by default, you have to solve this problems in another format, that is called as a logit model or you can say probit model. So, we we first highlight the problems of binary choice model, then we will move to logit models or probit models.

(Refer Slide Time: 37:33)




So, what are the problems? So, most important problems you will place in the binary choice model is the heteroscedasticity issue. So, there may be heteroscedasticity issue there is heteroscedasticity issue for instance, you see here, suppose its depends upon you know how how is your respondent. Suppose, you have you have taken a some respondents, 500 respondents and out of 500 respondents only 1 or 2 having higher level of income.

Say let us say, 10000 and other peoples having, say you know 5000, less than 5000. So, then obviously, there are out layers problems, so that will lead to heteroscedasticity issue. So, to minimize all these things, so either you will transfer apply the transformation rule, then you simplify or else or else what you have to do, you have to change the sample structure. So, that means you have to choose the respondent such a way so that there income level will be very more or less homogeneous, not perfectly homogeneous. That means, if we will say 10 10 10 10 10, that is one type of homogeneous or if you will put 10 9 8 like this. So, there is variations, if 10 is average then 2 3 points below and 2 3 points above, so that that in that structure it is not a problem. But if there is huge difference, then obviously, there is serious of serious issue of heteroscedasticity.

So, that means one of the most important problem we will face in the case of binary choice model is the heteroscedasticity issue. In fact, heteroscedasticity can also be visible in the case of as usual simple model or you can say, like you know bivariate model, trivariate model and multivariate model. But in this particular context, even if one variable is dummy whether is dependent side or independent side, but in this particular case like binary choice model, so we will assume that this particular case is totally binary in natures and the data variation is there.

So, as result, so there is a heteroscedasticity problem. So, that means one of the most important limitation we will observe in the binary choice model is that, the existence of heteroscedasticity heteroscedasticity problems. So, since it is you first check it and accordingly you have to go for proper transformation or proper structure to solve this heteroscedasticity issue. In fact, this particular problem can be solved through logit and you can say probit models.

Second, so there is difference of you know difference of interpreting p. So, we you know p is always probability is always lies between 0 to 1, 0 is one extreme and 1 is another extreme another extreme. So, that means we are taking two extremes in between in between there is several points 0 1. So, it is 1.1, 0.1, 0.2 like this way it will continue 0.9.

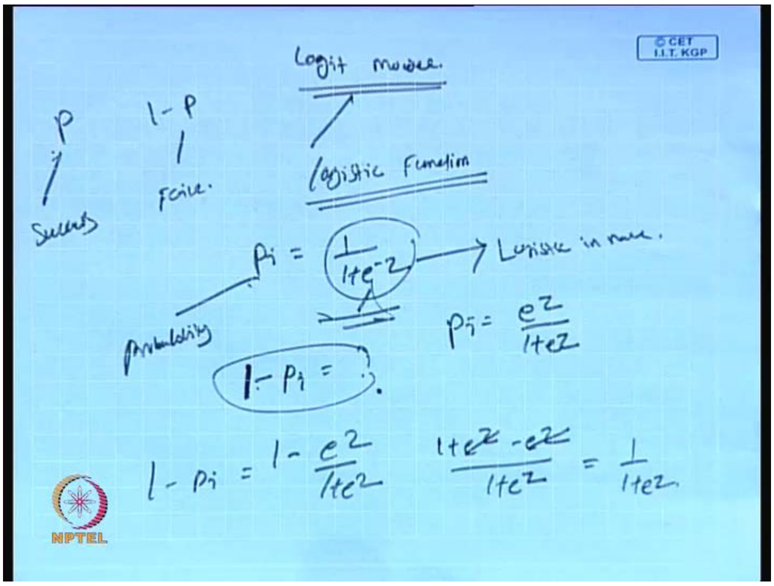
So, many points are there, but you know, we are not touching any other points. So, we are just we are just preparing the setup you can say, 0 1 interval. So, that means one extreme this side and one extreme other side. So, that is why binary choice models have a limitation when the picture will be in between 0 to 1. But here, we are creating artificial scenario, but some of the cases, artificial scenario may not be possible. In that context, binary choice model has a, you know limiting use or limiting applications.

So, this is second second problem associated with the interpreting probability having you know you know less than 1 or you can say greater than to 0, so if it is greater than 0 and less than to 1. So, in between then it it has a serious problem with respect to interpret interpretations.

Then third problem is the marginal effect, marginal effect marginal effect can be also studied here, marginal effect can be studied, so the marginal effect can be observed. That means, once you have D D i equal to alpha plus means beta 0 plus beta 1 X 1 plus beta 2 X 2 up to beta k X k plus U. So, then obviously, you will get the estimated model, then with respect to estimated model then with respect to estimated model you can depend with respect to particular parameter you know with respect to particular variables say X 1 i X 2 i like this way. So, that particular effect will give you the marginal effect, sometimes you know since there is only limit two limiting factor, and then the marginal effect will more or less you know very limiting use.

So, as a result, it is it is a serious problem in binary choice models. So, with this, we will we will we will this finish this particular binary choice model, because all these limitation will be highlight in the case of also in the in the case of logit models.

(Refer Slide Time: 42:50)




So, then next item we like to discuss is called as a logit models. So, what is this concept of logit models? So, logit model is the non-linear format, non-linear format means non-linear form of the dummy variable modeling, so particularly in the case of dummy dependent. So, it is derived, the route is derived from the logistic functions, so the route is derived from logistic function. So, the route is derived from logistic functions

So, what is a logistic function? So, that means, so here, we will consider P i equal to 1 by 1 plus e to the power minus z. So, P i stand for probability and you know e stands this particular this particular equation is exponential format and particularly the all whole function will be related to logistic in natures logistic in natures.

So, that means logit logit modeling is a special type of modeling, means it is advance advance to linear probability models, where where you know the dependent variable is proxy in natures. So, when dependent variable is proxy in nature, then we will use linear probability model. But certain cases, this you know linear probability model has a limitation, so the way we have already discussed.

So, as a result, we will prepare another type of beautiful structures through which we can observe this you know problems. So, that means the problem is where the dependent variable is proxy and independent variable is quantitative in nature.

So, now in this logistic format, we start with a particular function say P i, so which is equal to 1 by 1 plus e to the power minus z. So, what I will do, so you know probability value, probability value has a 2 limit. So, 1 minus P this is probability of success and this is probability failures, this is success side and this is failure side. So, all right

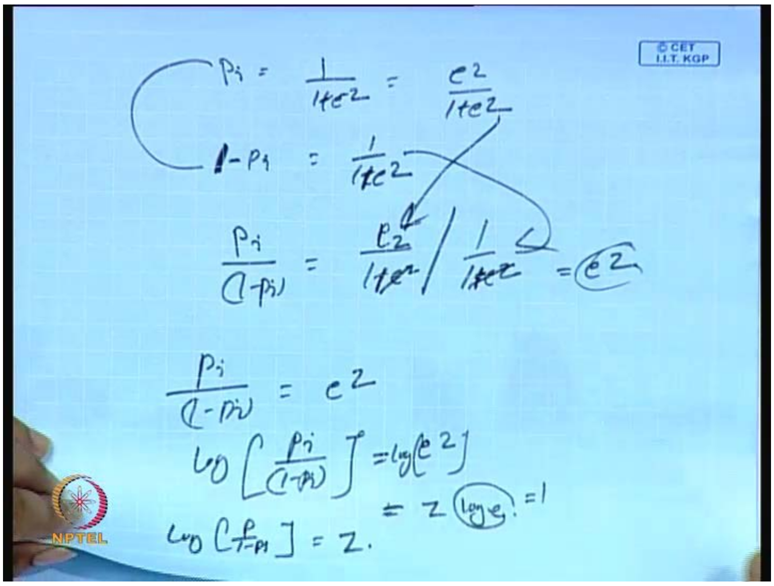
So, now we like to find out the ratio, so the ratio will give you the signal or indication of the logistic you know logistic model. So, P i is a logistic functions which is equal to 1 by 1 plus e to the power minus z, so corresponding to you know logistic functions. So, we will find out 1 minus p i, so which which we will observe just right now. Before explaining 1 minus P i, let us lets we will first highlight the particular structure P i. So, p i is here, 1 by 1 plus e to the power minus z. So, if I will simplify this particular one, then p i equal to e to the power z by 1 plus e to the power z am I right, so this is this is how the P i structure can be developed.

So, now this is 1 minus 1 minus 1 minus P i, then once you have P i, then you can find out the 1 minus P i. So, that means 1 minus e to the power z by 1 plus e to the power 0, so this is this is how 1 minus P i is all about.

So, now if we will simplify, then it is nothing but 1 plus e to the power z minus e to the power z divided by divide by 1 plus e to the power z, now this is this is cancelled. So, it is simply equal to 1 by 1 plus e to the power e to the power z, 1 by 1 minus P i minus P i equal to 1 by 1 plus e to the power z.

So, what what you will do, if you will simplify, so that means, we start with P i. So, P i equal to 1 by 1 plus e to the power minus z, which is nothing but e to the power z by 1 plus e to the power z then 1 minus P i is equal to 1 by 1 plus e to the power 1 by 1 plus e to the power z.

(Refer Slide Time: 46:45)




So, we use P and 1 1 minus P i, P i and 1 minus P i, then you you will develop a ratio, that is called a word ratio which is difference which is the difference between probability of success and probability of failures all right. So, that means we will define a functions called as a P i by 1 minus P i 1 minus P i which is nothing but 1 by 1 plus e to the power 1 by 1 means e to the power z is e to the power z by 1 1 plus e to the power z divided by 1 by 1 plus e to the power z.

So, this is just I have put the value here, so this is P i value. So, I am putting P i value here is and I am putting here 1 minus P i value. Now, if we will simplify, then this and this cancels, so that means, which is is simply equal to e to the power z. So, that means P i by 1 minus P i P i by 1 minus P i is equal to simply e to the power z.

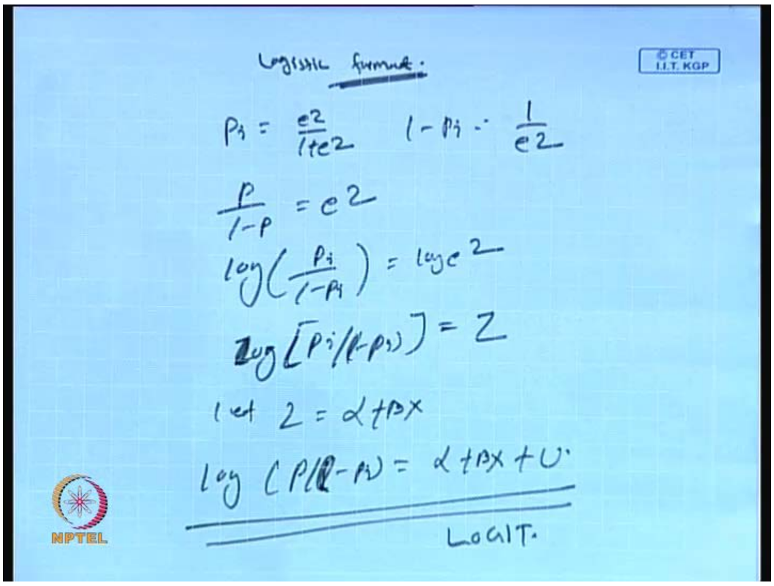
So, now this is still leads a non-linear format, still non-linear format. So, what we will do, we will transfer this into linear format, because the basic basic point of starting means, basic format of this econometric modeling we start with OLS techniques.

So, the OLS technique one of the restriction is that, the functional form must be linear in natures, means the parameter which you are going to estimate must be linear in nature. So, accordingly you have to transfer this function into linear format. So, what you have to do, so you apply log in both the sides.

So, the moment will put log in both the sides, so the format will be like this, so beta log P i minus 1 minus P i. So, this will be come to e to the power z, so if we will apply log this side then this should be log other side also, then log P i by 1 minus P i equal to 1 log e to the power z.

So, if we will simplify further, then it will be z log e z log e, but log e always equal to 1 log e always equal to 1. So, obviously the transformation will be log P by 1 minus P i log 1 minus P i is simple equal to z only. So, z equal to, so z is a z is a function actually so z is a function actually. So, what we will do, so you see the the basic logistic format is like this.

(Refer Slide Time: 49:45)




So, we will find out P i, P i equal to e to the power z by 1 plus e to the power z 1 minus P i equal to 1 by e to the power z. So, as a result, P by 1 minus P is equal to e to the power z, then log of P by 1 minus P i; p i is equal to log of e to the power z, then obviously, z equal z means log of P by 1 minus P i 1 minus p i is equal to simply z.

But let z equal to alpha plus beta X, so that means, we have to find out log P by 1 minus P i is equal to, is equal to alpha plus beta X, then by default we will put error term. So, this is the general format format of logit models, this is the general format of logit models.

So, now what we have discussed in this particular class is that, so we are very much you know highlight highlighted the entire structure of binary choice models, then the introduction of logit models. So, that means typically we have discussed the situation where the dependent variable is dummy in nature or categorically in nature, then when the situation is dummy dependent, then there are three difference ways we can discuss or we can analyze, one is binary choice model, another is logit model, another is probit model.

So, binary choice model, the you know the advantage is that it has 0 1 limit only. So, it is very simple and very easy to understand and easy to estimate, but in other case, logit model and probit model, it is a somewhat little bit complex, because one case we will use probability distribution and another case its normal distribution, the way we have discussed which is purely probability functions, that is you know that is how we use probability here and that too logistics functions.

So, now with the help of logistic function, we develop that particular model. So, that is model is called as a logistic models. So, logistic model is that log P by 1 minus P i equal to alpha plus beta X plus U, this P by 1 minus p i is the odd ratio.

So, now with respect to the same problem, which we have highlighted, just now in the case of binary choice model, so where you know family family income is integrated with family having house or not having house. So, in that problem, so we can also sight here, but there is interesting problem here, because in that binary choice model, we are limiting the size 0 to 1.

So, if we will limit the size 0 to 1 here, then obviously, the problem will be insignificant, means the model itself will be inconsistence. So, we will highlight the, what is the actually inconsistence here so that next class, we will discuss in details about that inconsistence part of the logistic model. So, to make the system or consistent, then what sort of things required, so we will highlight in detail in the next class. Thank you very much. Have a nice day.