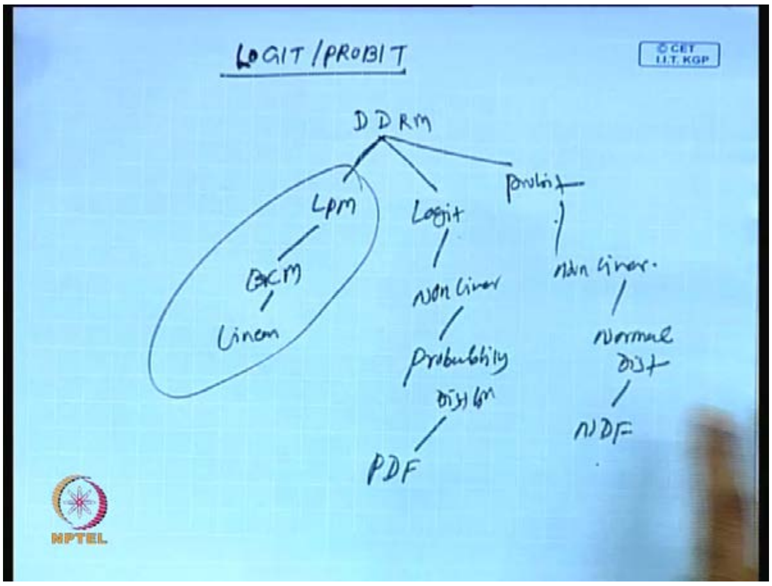
Econometric Modeling
Prof. Rudra P. Pradhan
Department of Management
Indian Institute of Technology, Kharagpur

Module No. # 01
Lecture No. # 29
LOGIT and PROBIT Model (Contd.)

Good evening. This is Dr. Pradhan here. Welcome to NPTEL project on econometric modeling. So today we will continue the logit and probit models. In the last lectures we have just entered to the a structure of regression models where the dependent variable is dummy in nature and one or several independent variables. So if that is the particular format then obviously there are three different forms of the models. One is called as a linear probability model, another is a logit model another is a probit model.

(Refer Slide Time: 00:59)



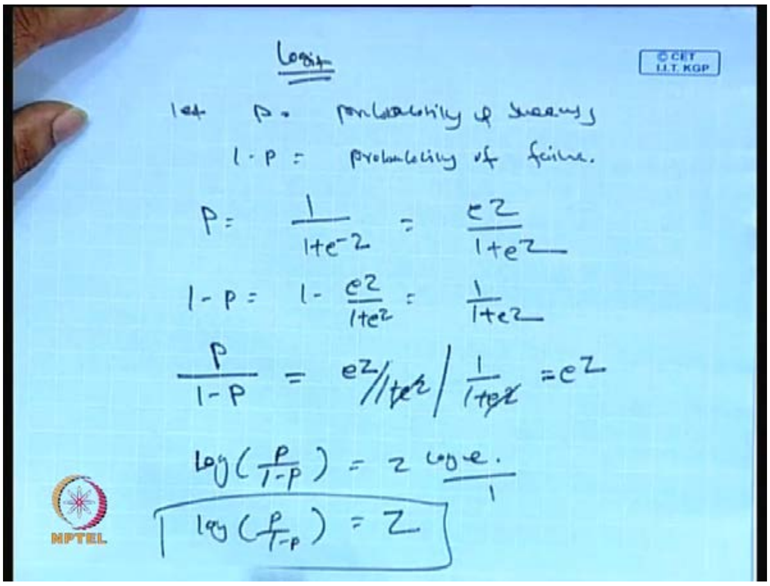
So that means so the dummy dependent regression modeling is divided into you know three typical format linear probability models then logit models then probit models.

So this is called as a linear binary choice models, it is linear in nature and this is non-linear this is non-linear this is non-linear. We have discuss this details about the linear probability models where we have formed lots of you know limitations so that limitation can be taken care of by logit model and probit models.

This is logit model is a you know pure use of the probability distributions probability distributions and this is typically use of normal distributions normal distributions. That means this is basically deals with the probability density function and this deals with normal density function PDF and NDF. So I will not again go into this direction of linear probability models. I will be quickly jump in to the logit model which we have deeply highlighted in the last class. How we have transferred this a or how we have to build the entire structure of logistic models.

So we will quickly highlight that particular issue then we will come to that particular application because you will get the job is very interesting or the model is very interesting. If will we apply directly to a particular problems otherwise it is just like mathematical derivations nothing else.

(Refer Slide Time: 02:44)



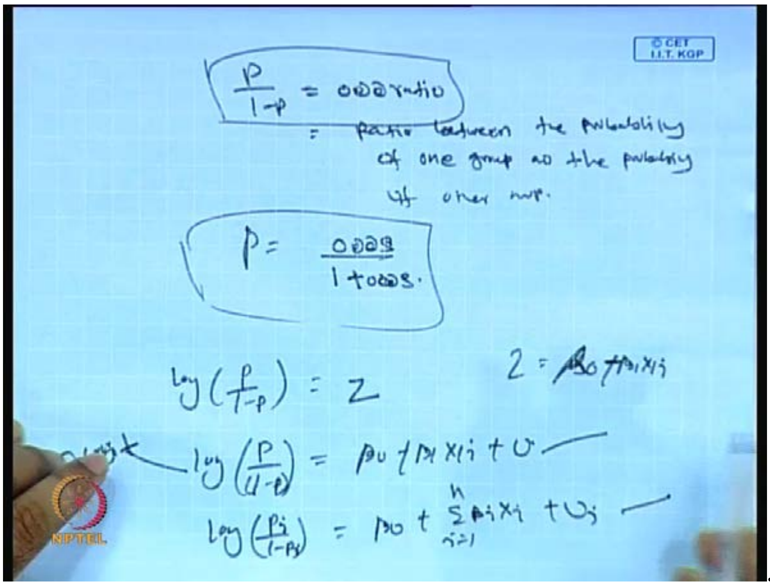
So what you have to do so for logit models specifically, we will take let we will assume that P is a probability of success then 1 minus P is obviously probability of failures.

So, then, we will define P equal to 1 by 1 plus e to the power minus Z so which is nothing but, e to the power Z by 1 plus e to the power Z then 1 minus P is observed with respect to 1 minus e to the power Z by 1 plus e to the power Z which is nothing but, 1 by

1 plus e to the power Z. Then we divide we get the we make the division P by 1 minus P which is equal to e to the power Z by 1 plus e to the power Z all divide by 1 by 1 plus e to the power Z. So this is cancel so which is simply equal to e to the power z.

So now what we have done we put logarithmic means you have to take use apply logarithmic both the sides. Then we will get P log P by 1 minus P is equal to Z log e Z log e. So log e is equal to 1 so obviously the model will be log P by 1 minus P equal to Z. so this is how the logistic model is all about now.

(Refer Slide Time: 04:15)

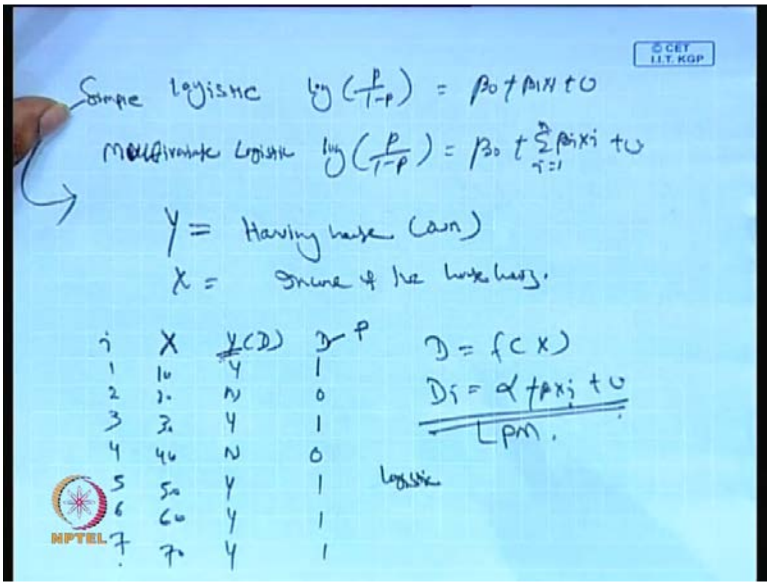


Now what happens so this P by 1 minus P is called as here odd ratio this is otherwise represented as the ratio between the probability of one group ratio and the probability of other group.

It can be also other way around so that means if P by 1 minus P is odd ratio then obviously this P equal to odd by odds means odd is the ratio between this odd by 1 plus odd this is 1 by 1 plus odds. So this is how it can be calculated. So these are the two. Our target is to find out the odd ratio. What how will you proceed for that so we will directly because here the model will be specifically log P by 1 minus P equal to Z. So Z is here equal to say beta 0 plus beta 1 X 1i.

So obviously the model will be complete model will be log P by 1 minus P equal to beta 0 plus beta 1 X 1i plus U. This is simple logistic model simple logit model. So we will go for you know means multivariate logistic model then obviously a function will be like this log P by 1 minus P is equal to beta 0 plus summation beta i into Xi. i equal to 1 to n plus U i. So this is obviously i is here is i is here. So this is the multivariate logistic model this is simple logistic model.

(Refer Slide Time: 06:31)



So that means you just summarize simple logistic model. This is simple logistic model is a log P by 1 minus P equal to beta 0 plus beta 1 X 1 plus U. Then multivariate logistic is log P by 1 minus P is equal to beta 0 plus summation beta i X i, i equal to 1 to n plus U. So you see here is so we start with a simple logistic models what you have to do so we will we will take the same problems.

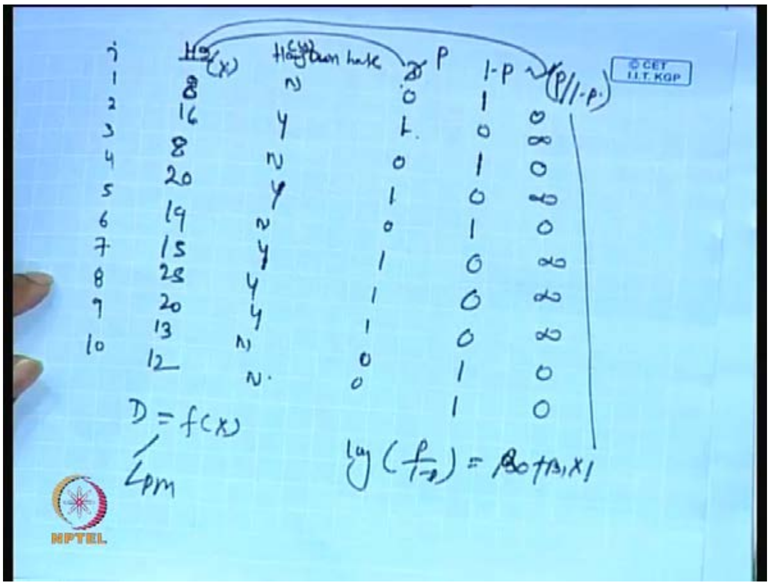
So let Y which is a you know starting point is Y equal to you can say having house means having own house and X is equal to income of the household. So this is how you have to represent the structures. So what you have to do here is we like to investigate whether before having means whether there is any connections between income levels and people having own house. So that means whether income has an impact on having house in a particular city.

So now the way we have discussed the same problem in the case of binary choice model it has a very simples or very easy to understand but, her there is a lots of complexity for instance what we have done in the case of binary choice model so we have a we have i sample observation 1 2 3 4 5 6 7 like this or a we have income level say 10 20 30 40 50

60 70 these are the income levels then we have you know Y which is nothing but, dummy which is having yes no situation. Yes, no, yes, yes, no, yes, yes, yes. So this is how the structure is all about. Now what we have to do this is Y structures dependent variables so we will transfer this D into 1 0 1 0 1 1 1. So this is how the transformation is all about.

So then finally, we will integrate D upon function of X that means D i equal to alpha plus beta X i plus U. So this is how the structure is all about in the case of this is linear probability models. But, now if we will apply logistic models for instance, for the logistic case so this is obviously represented as a P, P is probability of success then obviously we will determine you know 1 minus P probability of failures. So let me put it other way.

(Refer Slide Time: 09:44)



So here is i. i equal to 1 2 3 4 5 6 7 8 9 10. So we will take original figures here so what we have discussed in the case of binary choice model.

So this is family household income. So for sample size one so we will take 8 then this is 16, then it is 8, then 20, then 19, then 15, then 25, then 20, then 13, then 12. So this is the sample size now having own house. So that means it is no, yes, no, yes, no, yes, no, yes, then no, yes then yes, yes then no, no. so this is how the structure is all about now what we have to do we will transfer in to D. D is 0 1 0 1 0 1 1 1 0 0.

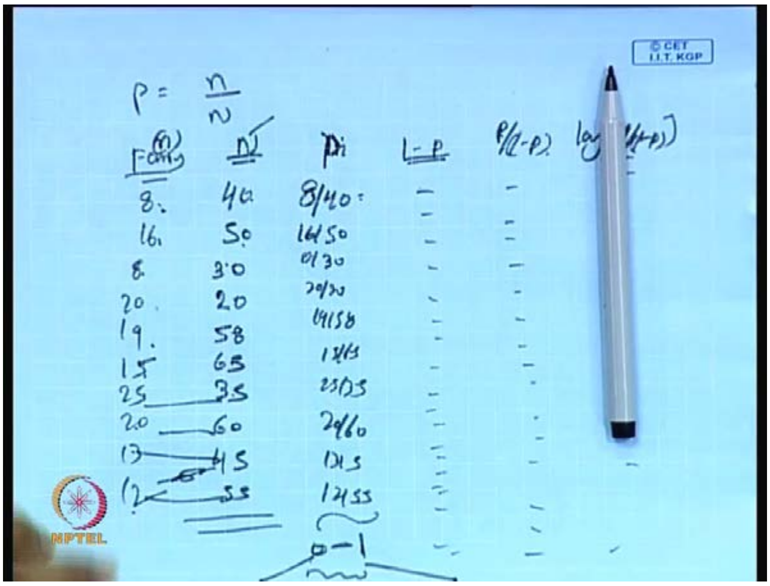
So now we are integrating these and these. This is X and this is Y. Now you are regressing D upon function of X this is what linear probability model. Now in the case of resisting models so instead of D we will call it P. So then we have to define 1 minus P so 1 minus P means this is 0. This is 1 then obviously 1 minus 1 it will be 0. Then this is 1 minus 0 means 1. Then this is 0, this is 1, this is 0, this is 0, this is 0, this is 1, this is 1. So that means this is just opposite. Now P is this much, 1 minus P is this much. So what you have to do?

So we need to have a P 1 minus P. So 1 minus P means 0 by 0 by 1, so it will be 0. Then 1 by 0 this is infinity. Then 0 by 1 0. Then 1 by 0 infinity. Then 0 by 1 0, 1 by 0 infinity, infinity, infinity, 0, 0, 0. So if it is coming infinity infinity then obviously you cannot you know estimate the model. So that means if we we need models P by 1 minus P, this is how P by 1 minus P equal to alpha beta 0 plus beta 1 X 1. This is that means we like to integrate this with this. This is P minus 1 minus P.

So now this is the system is totally inconsistent because we cannot estimate ones the figures are infinite in nature. So that means the model has a limitation. So that means whatever problems we have discuss in the binary choice models the same problem cannot be discuss in the case of logistic model. But, still we can solve this particular problem in that in the form of logistic function. So we have to change the structures. What you have to do so we use here is the original figures.

So we apply the real probability here. So what is the concept of probability? Probability means the simple the general probability is decided by n by n, small n by capital N. Means it is the individual observation by total observation. So what you have to do in this particular structures, we have taken only N case. Now what you have to do we will take N and then you will define the small n ultimately solution can be obtained.

(Refer Slide Time: 13:41)



For instance, what you have to do so forget about you know individual size. So what you have to do here is we will take a family income here. So family income is 8 16 8 20 19 5 19 15 25 20 13 12. So these are the pictures are there. So what you have to do with respect to family income, you have a samples so take N number of households, say 40. Then means we are taking 40 households having income level of 8. Then we will take 50 household having income level of 16. Then I will take 30 household income level of 8. Then 20 household income level of 20. Then 58 household income level of 19.

Then similarly, 65 household income is 15, then we will take a 35 household income level of 25, then 60 household income level of 30. Then you will say we will find 45 household income level of 12. So 1 2 3 4 5 6 7 8 9 10, 1 2 3 4 5 6 7 8 9. So then another 55 this is a like this 55. So this is how the household so what we are doing so that means these are all our family income structure. So it is 8 dollar per day, 16 dollar per day, 8 dollar per day, 20 dollar per day, 19 dollar per day, 15 dollar per day, 25 dollar per day, 20 dollar per day, then 13 dollar per day, finally, 12 dollar per day. So that means we have taken 10 different samples.

So in the case of linear probability model so we transfer with respect to the income level. Then we ask the respondents whether they have house or not if they have a house then we will put yes Y. Then if they have not, then we will put N. Then accordingly we will transfer yes to one then no to 0. Then you will we go for this as usual ordinary process of estimation. But, you know in this case so we are getting the odd ratio in infinite means undefined. So it cannot be possible to possible to go for you can say estimation.

So what you have you do instead of collecting data from a particular respondent so we will collect data in a particular city. Say for instance we will choose a homogeneous group so where income is say 8 dollars per day. Then you will find out that 40 persons are there. So what you have to do we will classify let us say what you have to do we add all together. So the moment we will add all together then obviously so you get to know what is the total sample size here. So out of total sample the moment you will get all these respondents information, let us say 500 then you will say what are the total income variations.

For instance, out of 500 respondents you are income different income structure is like this a it will vary from 8 to say 25. So either you put it in ascending order or you put in descending order or as usual no hard and fast rule. So your family income is this much with respect to this. Say sample observation then within this particular family income so you investigate number of households those who income level is similar lines.

For instance, first case we take 40 different households whose income is 8 so that means we what we will do so we will find out a ratio here called as a n i small n i. So that means it is nothing but, 8 by fourteens. Otherwise we call it P i. P i is a n by fourteen. This is small n you say and this is capital N. So then forty also out there so. 8 by fourteens so this is how you have to find out 8 by 40. Then you know 16 by 50, then 8 by 13, then 20 by 20, then 19 by 58, 15 by 65, then 25 35, then 20 by 60, then 13 by 45, then 12 by 55. So this is how you have to transfer this data.

So the moment you will transfer all this data then now you will find some figures. Then you will find out 1 minus P i. So you can able to find out because all are in ratio format 0 point something or 1. So that means this particular items will be vary from 0 to 1. This said item particularly 0 to items but, in the case of binary choice model, we are taking this extreme and this extreme now in this particular case so we are taking altogether, so different structure altogether. So then we will find out 1 minus P. Then you fill up this gap. Once you fill up this particular gap then obviously what you have to do then you have to find out the ratio.

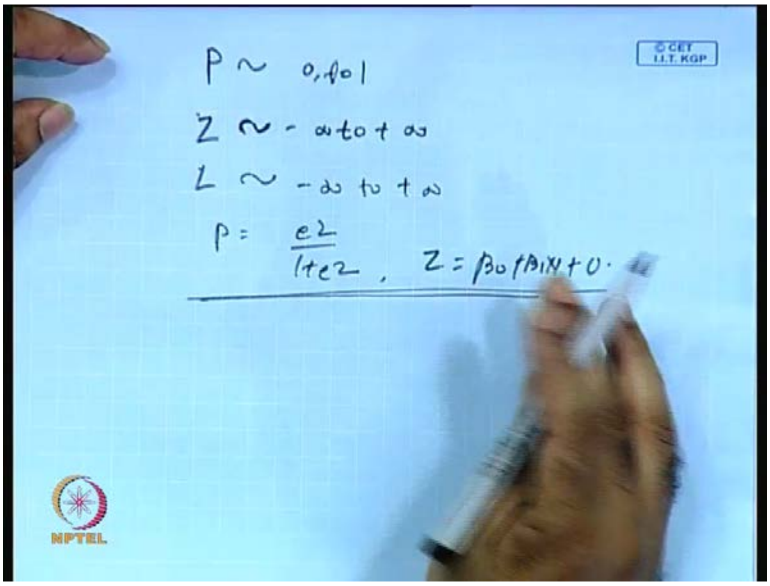
So P by 1 minus P. Then obviously the moment you will find out P by 1 minus P then you have to find out its log. So log upon P by 1 minus P so this is how you have to observe now. log of 1 by 1 minus P has to be obtained. So this is also P by 1 minus P has to be obtained sequentially. so that means now so this is you know this is your family income and this is your log transformations. So what you have to do now you have to integrate the log transformation with that a you know family income. So that will give you the indications.

So that means a it is means it is question of yes no situation but, it is in a different format. So that means here the structure is we ask the 40 respondents, the moment you will ask the 40 respondents then 8 persons family response is that they have house. That means rest 32 families having not any house. So that means in a particular group homogeneous groups, we are investigating or you are taking a response from 40 household. Then out of 40, 8 are having you can say their family means having their house.

Then we investigate 50 peoples then out of which 16 having house. Similarly, we investigate 30 and 8 having house. Similarly, we investigate 20 peoples so then all 20 people have their own house. Similarly, we investigate 50 and 19 people have their no house. Similarly, we have 65 we investigate 65 group then 15 people have their you know house. Like this we have to classify the entire structure.

That means instead in the binary choice models you are specifically targeting one particular section of the people. So then you are going for it but, here what you have to do in that particular logit models. So what you have to do you take this particulars one sample, this value so that means you will take a this structure is altogether very broad in nature. Means instead of going a particular specific area or specific small cluster you have to go to big clusters. So where within that structure there is a some kind of homogeneous setup. So that homogeneous setup will give you signal for the logistic functions.

(Refer Slide Time: 22:05)



So now with this particular structure so we have to calculate the moment of this logistic functions. Now this you know what you have to do so P generally lies between here 0 to 1.Then obviously the Z here is will be lies between minus infinity to plus infinity or a you know logistic functions it will be moved from minus infinity to plus infinity. So this is 0 to 1.

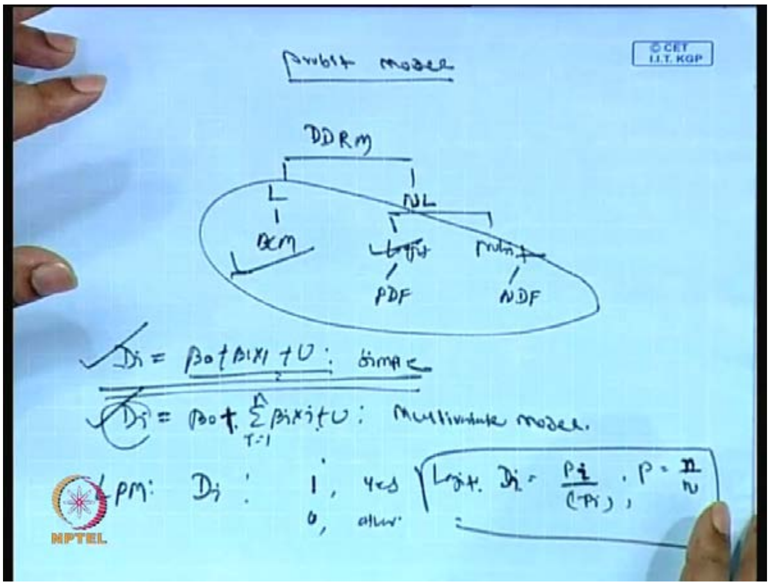
Now so what you have to do, so it is normally the structure is that, so logistic format is that, so P equal to e to the power Z by 1 plus e to the power Z, where our Z can be written as a beta 0 plus beta 1 X 1 plus U. This is for simple models so which we have already discussed, because here the target is so we like to know what is the means family income and they whether they have house or not. Means our hypothesis is that a person having higher income may have own house and person having less income they have not house this is our hypothesis.

So accordingly we gather the information and you apply the particular models. That is whether binary choice model or logistic model. Then we come to conclusion that what whether it has a specific implications or you can say means whether there is any linkage between this income and having household or not. Now we find means what we have received here is so there is a you know different methodological difference and you know there is also sample as different structures.

So means all together this structure is very complicated in the case of logistics. So to simplify their particular structure so you have to redesign or restructure the sample size. So that it can be possible to prepare the or it can be possible to use this logistic models. So logistic model is typically more advanced than the binary choice models where the format is more or less you know in the case of binary choice model the format is 0 to ones but, in the case of logistic model so we are taking the various ranges of that is from 0 to 1 then we are integrating with the you know having their income level.

So if this a particular item is significant then you can justify that you know family income has a substantially impact on the peoples having their house or not. So this is how you have to observe. So in addition to logit models there is another model is called as a probit models.

(Refer Slide Time: 25:25)



So what you have to do that in the case of probit models so let us see here is so what is the probit model means. Typically probit model is the one second different structure of non-linear response models.

For instance, we have already highlighted that you know you are dummy dependent regression modeling is basically divided in to linear format and non-linear format. Under non-linear format, this is linear format means binary choice model this is logit model and this is probit models. So this is typically based on the probability density function and this is based typically based on normal density functions. Now otherwise the structure is more or less same.

So what you have to do ultimately in if it integrate all these three then the setup is like that way. The D i equal to beta 0 plus beta 1 X 1 plus U. So this is the simple models and D i equal to beta 0 plus summation beta i X i. i equal to 1 to n plus U. This is for multivariate models. Now whatever we have discuss in the binary choice model or logit model so we use simple models.

So let me first highlight this you know probit model in the case of a say simple structure. Then we will discuss another problem which is a based on a multivariate models where the response variable is very much means dependent variable is very much categoricals or otherwise it is a dummy in nature, where other variables are completely independent in nature and they are somewhat quantitative iN nature.

So we are not mixing two together but, mean mixing means you see in the last lectures we have discuss the setup where a dependent variable is quantitative in natures and independent variable is qualitative in nature.So we discuss one dependent with one independent and one dummy variable then obviously one dependent with the several independent variables which we are quantitative in nature and several dummy independent variables which we are totally qualitative in nature.

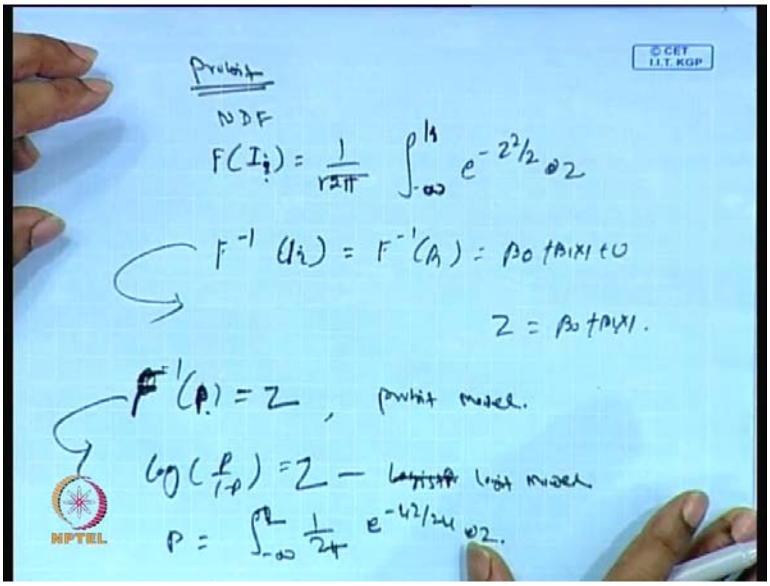
But, in this particular format so our independent variables may be one may be multiple in nature but, dependent variable we are discussing only one level only because we are not touching upon this structural equation modeling. So what we have done so like whatever we have discussed you know like bivariate trivariate and multivariate where Y is always you know only dependent variable, so here also same things.

So every case your dummy variable is a single ones so far as a this particular dummy dependent regression modeling is concerned. That means one dependent variable is there which is a purely categorical in nature or sometimes binary in natures. In other sides so we have series of or you can say one independent variable that means every times the right hand side will be dummy then dummy with one independent variables. Then dummy with a several independent variable that is multipolarity models and this is called as a simple model.

So now there are three different structures. Binary choice model can be also integrated with the this case or it can be integrated with this case. Then logit model can be also discuss under this head. Only difference is that this particular concept even if in the case of probit model these two are also same. But, only difference is with respect to D i representation. That means in the case of linear probability model D i such that so it will be one for yes situations and 0 for otherwise. So in the case of logit models so this D i. D i is represented as a P i by 1 minus P i.

So this is a P i by 1 minus P i where P is represented as a n by n. This is how the constant we have to add in the case of logit model because if will we apply 1 0 1 structure in this logit model then it will give you inconsistent setup. So as a result so we will take a large samples. Then within large samples you have to ask 10 respondents then you will see out of 10 respondents how is their response. So the probability level will be find out. So with this basis of probability levels then you have to find out the odd ratio and that too with the with the help of odd ratio we like to investigate the entire problem.

(Refer Slide Time: 30:30)

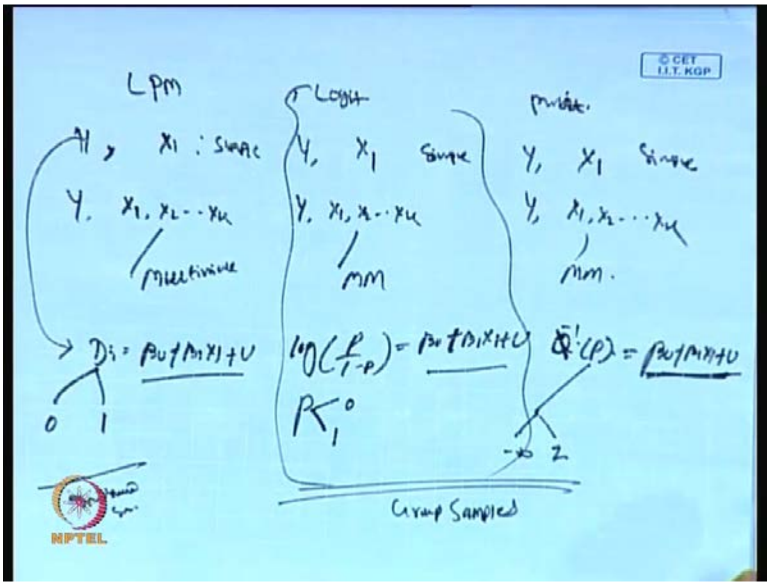


So now the way we have discuss this linear probability model then logit model. Probit model also can be discuss in the similar cases. Now for probit models we use you know called as a normal density function. What is this normal density function? Normal density function is a represented as a normal distribution for that is nothing but, 1 by 2 pi square root into minus infinity to i.It will start from minus infinity to e to the power minus Z square by 2 D Z.

Sometimes it can be like this F inverse. F inverse i is equal to its better let us put i here is because sample observation is i so i is equal to here is F inverse P i so which is equal to beta 0 plus beta 1 X 1 plus U. Here again Z is consider as a beta 0 plus beta 1 X 1so that means the particular format can be written like this way. So that means in other words we can put it even if in the you will put in the probability set so it will like this q inverse P or F inverse P is to equal to Z so F inverse P equal to z.

So the way we will calculate you know log of P by 1 minus P is equal to Z, so this is case of logistic models. It is derived from logistic function in fact. So this is F inverse P equal to Z. So this is derived for probit models. Now so what is P here is so this P is mentioned as a you know minus infinity to you know i or Z. Its better you since we are using Z here so then it will be a Z here so 1 by 2 pi e to the power minus you can say U square by 2 U D Z or 2 Z D Z. So this is how you have to calculate.

(Refer Slide Time: 33:00)



So that means you just make a comparisons here is so we have three different formats linear probability models then logit models then probit models. So in a what is this similarity is that so every times in the you know Y and Y is a function of X so here Y is also function of X here Y is function of X. Now this is simple models, this is simple one, then this is also simple one. Now Y as a function of X 1 up to X 2 up to say X k so this is

Y upon X 1 up to X 2 up to X k then this is Y upon X 1 F 2 X 2 up to X k so its better put X 1 here X 1 here X 1 here is.

So now in the 1st simple setup means this what is this similarity here in both the all the three cases, these three different models are typically use in the case of dummy dependent econometric modeling, linear probability model, logit model and probit model. So similarity is that it is applied in a particular situation where the dependent variable is a dummy in nature or dependent variable is categorical or binary in nature. So simple model is Y upon X 1 Y upon X 1 Y upon X 1 though the functional form is completely different.

Now this is simple model and this is multivariate models, this is also multivariate model, this is also multivariate model. So what you have to do in the case in this particular structures so we will write D i equal to beta 0 plus beta 1 X 1. So here D categorical divided into 0 1 limit and in this case, this particular logistic case we will write log of P by 1 minus P is equal to beta 0 plus beta 1 X 1 beta 1 X 1. Obviously it is U here.

So then this you know here P lies between 0 1 this also binary in number but, the range of the samples will be in between 0 to 1. In the case of probit model so we use q means psi inverse P. Psi inverse P is equal to beta 0 plus beta 1 X 1 plus U. So this you know this particular length will be from you know it will start from minus infinity to minus infinity to Z. That is how the range is all about means there is alpha plus beta X. So this is how the minus infinity to Z. So this is the total step.

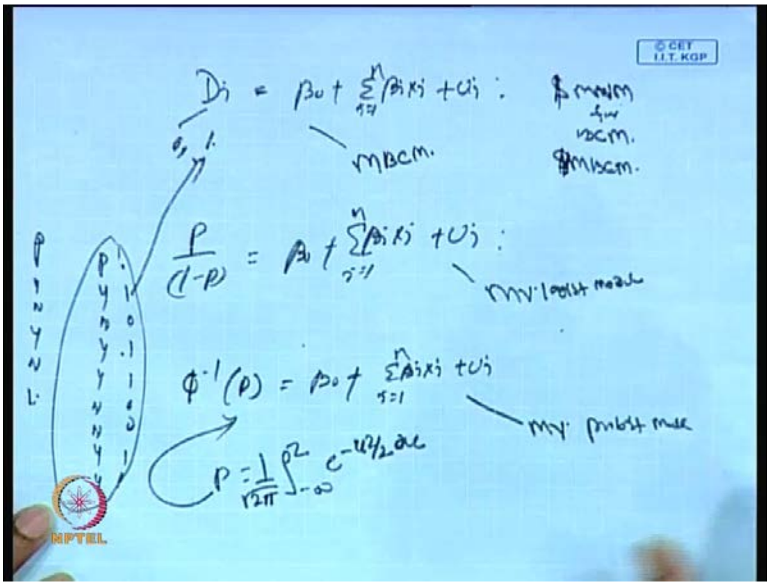
So the entire structure is like this way. So what we have observes here is the right hand side all the sides are very much equal. Only the difference is a the left hand structures. So that means in the case of binary choice models it is the simple structures we will put simply D i then transfer all the sample points to 0 1 binary format. In this particular structures so we may transfer in to 0 in to 1 format but, it is going against the means the structure is going towards the inconsistent level. So what you have to do instead of you know individual sample we will take group samples here.

So the only difference in this two cases you will go for group samples rather than individual samples. This is called as a individual sample case. So it is case of individual sample case now what you have to do we have to find out what is the specialty of this three models. So that means it is little bit advanced the reason is that here, it is little bit more complicated and advanced than this one and this is more complicated than this one. It is more advanced you know by statistical angles this particular function is more reliable.

Because, it follows the normal distribution step if anything attach with normal distribution then the accuracy of the model will be very high. So that is why a probit model is better choice than the logit model. Similarly, logit model is better than the linear probability model but, out of all these three, most typical you know this way so most of the instances we like to use this logit models because the outcome of the logit model and the outcome of probit model is more or less same. Only difference is it is a functional format.

One case you are transferring the group [means transferring the variables in to logistic function and another case you are transferring the information through you can say normal distribution function that is the only difference between a logit model and probit model. Otherwise the setup is more or less same.

(Refer Slide Time: 38:40)



So this is the simple setup which we have discussed in the case of you know multivariate models. So D i equal to simply a beta 0 plus summation beta i X i. i equal to 1 to n plus U i. So this is called as a simple multivariate models for binary choice model. Otherwise called as a simple multivariate binary choice, simple binary choice models or you can say that simple multiple binary choice model multivariate binary choice models.

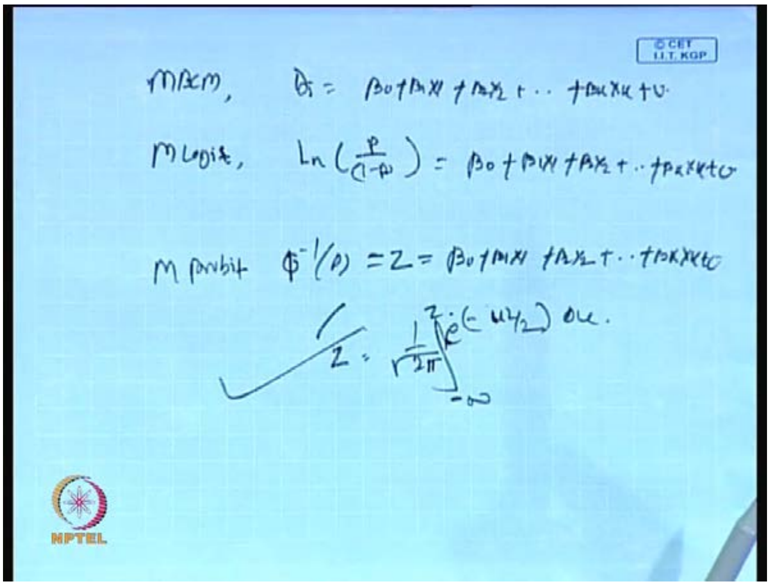
So this is where these lies between 0 and 1 limit. So similarly, in the other case we will write P by 1 minus P equal to beta 0 plus beta i X i summation, i equal to 1 to n plus U i. So this is called as a multivariate binary choice models. So this is called as a multivariate logistic model logit models.

Similarly what you will do you will puts this particular format Z equal to beta 0 plus summation beta i X i, i equal to 1 to n plus U i. It is called as a multivariate probit model where P equal to minus infinity to Z e to the power minus U square by 2 d U. Obviously square root of 2 pi is here. This is the normal distribution function.

So what you have to do so that means you see here is we start with simply you know P component probability yes no situation yes no situation yes no situation sorry probability yes yes no situation yes yes yes no no yes yes yes yes like this. So this is how the structure is starting point is this one so we will transfer in to 1 0 1 1 0 0 1 1. So this is how the transformation. Now this direct transformation can be applied to this one see if will we directly handle this particular data with this particular setup then it is called as a multivariate binary choice models.

So now if will you transfer P in to 1 minus P but, this particular structure is an insignificance so, what you have to do in that case you have to go by group samples find out total sample N capital N and small sample small n. Then obviously you will find the probability value n small n by capital N. So with that particular value it will be come in the ratio format it will not exactly 0 and 1. Most of the items will be in between 0 to 1 so obviously P by 1 minus P cannot be undefined. So in that case you can go for estimation that means the model will be more accurate or more authentic.

(Refer Slide Time: 42:38)

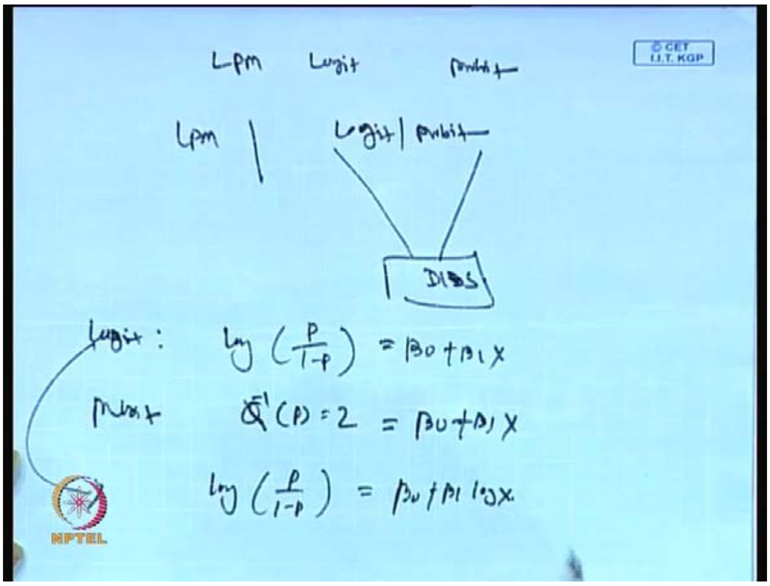


Similarly in the case of probit model you will transfer the entire data in to this normal distribution form shape. Then once it will be transferred then you have to regress with this particular models. So this is how it is called as a probit setup. So similarly, we can write like this way also, the simple structure is a in the case of binary choice model multivariate binary choice model so we will write like this way D i equal to beta 0 plus beta 1 beta 1 X 1 plus beta 2 X 2 continue plus beta k X k plus U.

And similarly, in the case of multivariate logistic model that is you know Ln in fact this is I have made wrong here. It should be log of 1. This 1 it is to be taken log here is so that means in the logit case it is log of P by 1 minus P so it is equal to beta 0 plus beta 1 X 1 plus beta 2 X 2 continue up to beta k X k plus U. So similarly, in the case of multivariate probit models so we will have an inverse P equal to or it will call it a Z. Then Z equal to beta 0 plus beta 1 X 1 plus beta 2 X 2 plus beta k X k plus U. So this is how the model is all about.

Now here the Z is followed by 1 by square root of 2 pi, e to the power minus U square by 2 d U. Obviously summation I will say minus infinity to Z. So this is how the probability [distribution] sorry normal distribution function is all about. All right now I will just briefly highlight the exact difference between these two that is logit model and probit models.

(Refer Slide Time: 45:06)

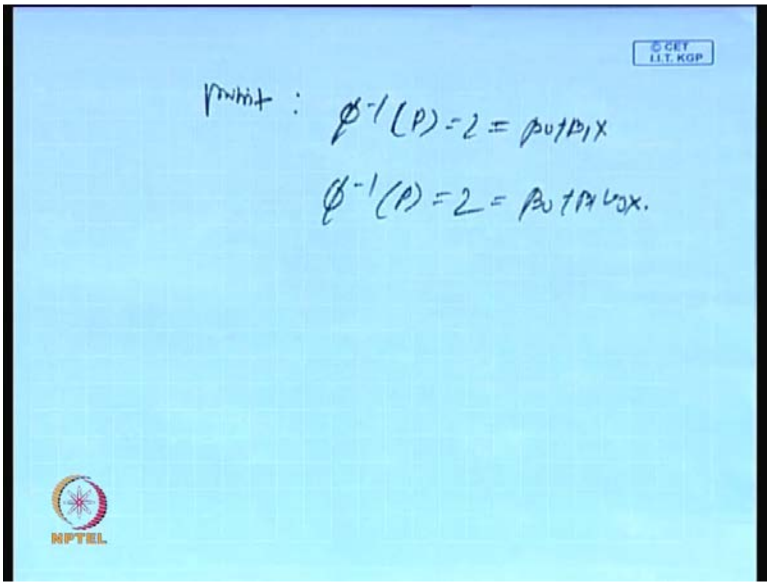


So what we have to do so out of the three models linear probability model, logit models and probit model, linear probability model is one step and logit models and probit model is another set. Now we will see what is the similarity and dissimilarity between these two. So linear probability models is not so much headache because it is a simple structures where only the transformation is to binary form 0 1 situation that is yes no situation but, in the case of logit model and probit models so it is not a question of yes no situation or binary numbers.

So you have to take a group sample then you transfer these all items in to some functional form with the help of some functional form that is through probability distribution function and through you can say normal distribution function all right. So this is how you have to proceed so what is the exact difference in the case of logit. So the probability mass function I will call by simple rule so this log P by 1 minus P log P by 1 minus P is equal to beta 0 plus beta 1 X.

So this is one and sometimes what happens and in the case of a probit models this is nothing but, inverse P equal to Z then inverse Z inverse P psi inverse P P equal to Z which is nothing but, beta 0 plus beta 1 X beta 1 X.You remember here is so what you will do so this is the simple model for logit and probit but, you know we can be like this way. So the model can be further modified by like this. So log P by 1 minus P is equal to beta 0 plus beta 1 log X. So this can be possible otherwise the same model it can be also for you for probit model.

(Refer Slide Time: 47:17)



So probit model it can be written as a either like this is equal to beta 0 plus beta 1 X beta 1 X but, in the same times you can write like this way P equal to Z equal to beta 0 plus beta 1 log X.

So this can be another format through which you can discuss this. You can say logit model and probit model. So what I like to mention here is that we have three different sets of models to highlight or to discuss the dummy dependency case only so that means here the econometric model or regression model in a particular format where dependent variable is a proxy or it is a usually called as a categorical or its binary and others are quantitative in natures. So if I will summarize all this details means last triple lectures which we means last you know last couple of lectures we have discussed the dummy variable modeling.

So dummy variable modeling basically divided in to two parts one part is called as a dummy independent variables modeling and then called as a dummy dependent variable modeling. So in the case of dummy independent variable modeling then the structure is completely very simple. So within the samples there is a little bit complexity with respect to bivariate setup and multivariate setups. So when there is a dummy means the dependent variable is quantitative and independent variable is dummy. Then you know we have a different setups where in one situation there is a one variable which is a very much proxy and others are quantitative in nature that means all these are with respect to independent variable.

In other case so we have a multiple dummies and multiple quantitative variables independent quantitative in nature. So but, in all the cases so we are assuming Y is dependent Y is dependent and which is quantitative in nature. So there is no dummy in the left sides that means dummy is not attached with the a dependent variable. So, alltogether there are two situations in this first (( )). So that means here a dependent variable is very much quantitative in nature and all the cases it should be one. So one you know dependent variable with a one dummy and a one independent variable which is quantitative in nature.

In other case you know one dependent variable which is purely quantitative in nature and there is a series of independent variable which is which are quantitative in nature and series of dummy variables means other variable which is a binary in representation or categorical in representation or you can say some way of classifications. So because this particular structure otherwise you know there is lots of integration in the multivariate analysis with respect to other application like discriminate analysis conjoint analysis etc so we will discuss details when there is a such needs.

So in another setup we have dependent variables you know dummy and independent variables are very much quantitative in natures. So what you have to do in this particular case so we have three different models one is called as a linear probability model another is called as a logit model and another is called as a probit models. So in the linear probability models so we may have a again simple one and you have again multivariate one.

Similarly in the logistic case you have a simple one you have a multivariate one and in the case of probit model we have again simple one or multivariate ones. So in the case of binary choice models when the setup is a simple then obviously it is just like D i equal to alpha plus beta X i or beta 0 plus beta 1 X 1 i. So here D i is just you know representation of 0 1 representation. So it is very simple one but, when we will go for multivariate then D i equal to beta 0 plus summation beta i X i. So we are assuming that all these independent variables are very much quantitative in natures but, in the dependent variable is qualitative in nature or you can it is purely binary in a form so that is a linear probability model.

But in the case of but, here one thing is that every time you have to use you know individual sample size but, in the case of logit model and probit model so you have to go for group samples and in that case dependent variable is a very much dummy and you have to transfer with a particular function. So, we use logistic function for this transformation and for that you have to use probability. Basically we use probability density function for logistic model and we use normal distribution function for the you can say probit models.

But the setup is almost all set the way we will discuss that particular problem with a respective household income and household having house. This can be analyzed under you know binary model logit model and probit model but, binary model structure is completely different where its setup has an integration with individual sample structure. But, here in the logit and probit it is an attachment with the you can say group samples.

But, you know ultimately the right hand side structure is more or less same for logit model and probit model. Ultimately it is in the form of function of X only so where X may be individual in nature or multiple in nature.So if it is a means whatever case it is purely quantitative in nature. Only thing is in the right hand side it is a qualitative in nature. So you have to be very careful about its categorizations you can say its transformation. So what you have to do so for logit models your transformation rules follows probability distribution function.

So whatever step is there so you have to transfer in to odd ratio so that is the ratio between you know probability of success to probability of failures. So that is how you have to find out the that particular item is very much important. Higher is the odd ratio higher is the relevance of the model, lower is the odd ratio, lower is the relevance of the model.

Similarly you will be transfer this particular information to some form of some normal distribution set in the case of probit models. But, you know logit model and probit model are more or less give this similar kind of results. So that is why since you know in classroom situation it is very difficult to handle you know normal distribution transformations. So it is better to go by you know logit models because means it is empirical. We have investigated that there is no much difference it is a value has there is a small difference. But, so far as a significance of the model is concerned in the case of probit model or logit model are more or less same.

So but, there is a little bit difference still you have to find out the difference. Means for research point of view it is better to have the talk to analyze that particular problem with respect to probability distribution functions and as well as with normal distribution function. That means that is with the use of logit model and with this use of probit models. So this is how this is all about this you know in the structure of dummy variable modeling.

So there are two varieties one is dummy dependent and dummy independent. Now this particular topic is very useful for panel data modeling so which we will discuss in details in the next class only. So with this we will conclude this lessons. Thank you very much have a nice day.