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

Module No. # 01

Lecture No. # 27

Dummy Modelling (Contd.)

Good morning doctor Pradhan here, welcome to NPTEL project on Econometric Modelling, so today we will discuss the same component dummy variable econometric modeling. So, in the last class we have discussed, the basic framework of dummy variable modeling, so I briefly highlight first, that dummy modeling, then we will proceed for its applications.

(Refer Slide Time: 00:49)
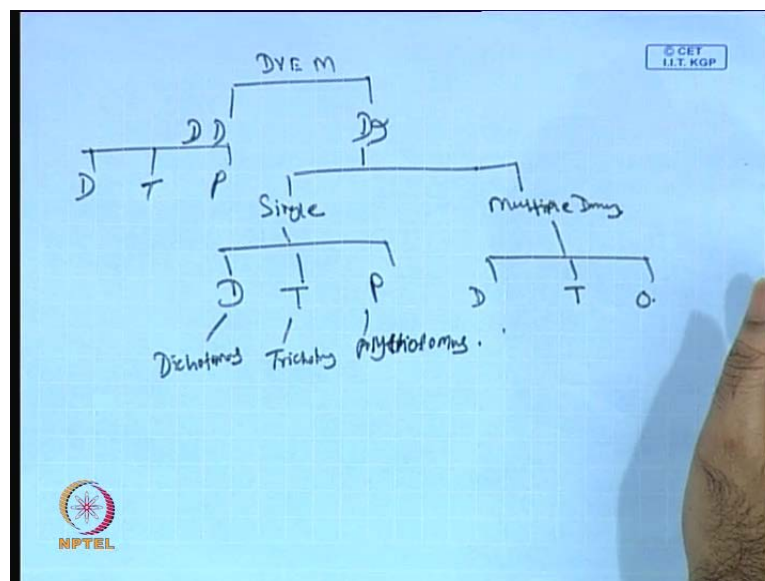
So, basically dummy variable econometric modelling can be divided into two parts, one is called as a dummy dependent and dummy independent, so in the case of dummy independent, it may be single dummy single dummy or it may be multiple dummy, it may be single dummy, it may be a you know dichotomous, it may trichotomous and polychotomous, this is how D T P dichotomous, trichotomous, polychotomous, polychotomous.

So, this how dummy can be classified, so here also dummy dependent, it may be in D T P formula so that means, by basic framework, the dummy variable structure can be

divided into two parts, dummy dependent, dummy independent. And under dummy dependent there may be single dummy, there may be multiple dummy, if it is single dummy, then within this a single dummy, the category may be a two category, may be three category, may be more than three.

So, this is how it is divided into three groups D T P order, so the in the case of D it is binary in nature, in T it is thrice three different category, then in the P it is multiple category. So, multiple category means it is not infinite, it is finite. Let us say, there are five different structure for instance, I will take a variables say religion, so religion may be may be Christian's may be Muslim may be Sikh like this way, so according to you see first you see the problem setup, so if you say religion is variable and you have a response from you know have 500 sample power. So, despondence so that means, 500 genders religions are there means, 500 religion of 500 peoples structure is that, so you have to see how many religions are all total in the 500 list.

So, accordingly, so religion has to be ordered into, that particular format so that means, if there are only four religion in the that, particular response see it then; obviously, so you have to put D equal to either 0 1 2 3 4 or you have to start with 1 2 3 4 5. So, this is how you have to structure it, so that is in the P structures or if it is thrice say, let us says you can say seasonal effect, so some are winter, you can say autumn, so obviously so this is how T structure of you can say dummy, dummy independent.

So, similarly D structure of independent is the best sited example, which is general issue, so this is how you have to categorically mention the detail about you know dummy variable econometric modeling.
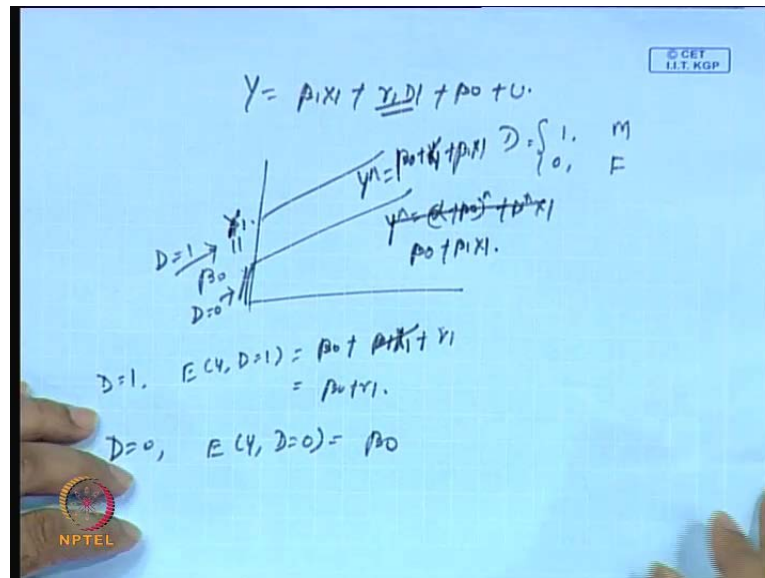
The simplest form which we have discussed in the last is that, so Y equal to beta 1 X 1 plus gamma 1 D 1 plus beta 0 plus U, so now, where D is categorically represented as a 1 form here as male and 0 is for females, so accordingly the setup will be like this, so let us say this is beta 0 and this is beta 1, this is beta 1 or you can say means, when will you put here D equal to 1 then; obviously, E of 1 Y D equal to 1 is equal to simply beta 0 plus beta 1 beta 1 X 1 beta 1 X 1 plus you can say uh gamma 1.

So; that means, it this will be removed, so it will be beta 0 plus gamma 1 only. So, beta 0 plus gamma 1 then that means, it is a gamma 1 it is gamma 1, so now, D equal to 0 means E of 1 Y D equal to 0 is beta 0 it is simply beta 0, because this will be this will be equal to 0, so it will be simply beta 0, so this is you know, where D equal to 0 and where it is D equal to 1, so this is how D equal to 0, so this is an indication this is an indication.

So, now in that case your model will be Y hat equal to alpha alpha plus beta 0 this is hat plus beta 1 hat X 1 in that case sorry. So, in fact, this 1 will be coming like this way. So, it is beta 0 plus plus gamma 1 plus beta 1 X 1 and this would be simply a simply beta beta 0 plus beta 1 X 1 beta 1 X 1, so this is how the structure is all about so, this is the simplest form of, you can say dummy dummy independent variable, let me let me highlight in a broad way.

(Refer Slide Time: 06:55)
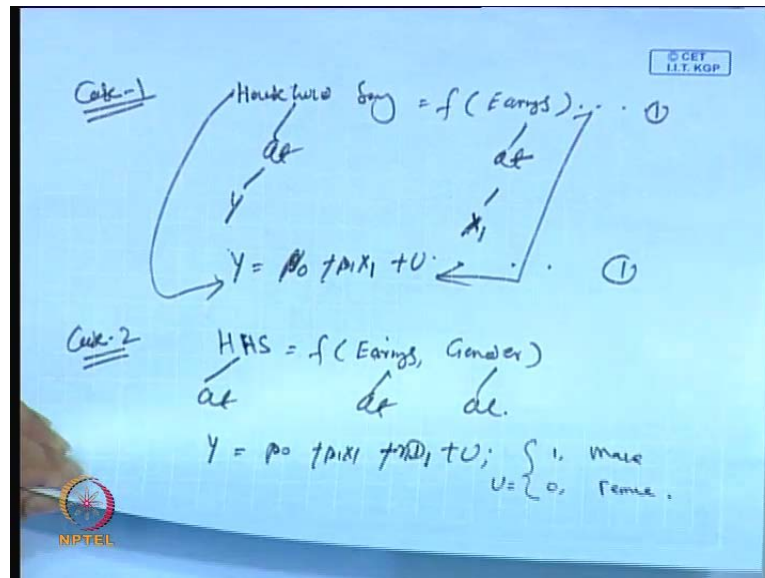
So, in that last class we have discussed a problem called as a ==household saving== household saving as a function of earnings, this is the standard one ==you know== case 1 I will give, I will show you various cases, how dummy variable can be used.

So, case 1 household saving is the earnings, so household saving is a quantitative in nature, earning is quantitative in nature, so I will call it this is Y and I will call it this is X, so then the model will be Y equal to beta 0 plus beta 1, I will call it lets better X 1 then this is beta 1 X 1 plus U, so this is the simplest model. So, now, I will this is the case 1 where household saving is equal to earnings, but you know in this particular structures ==in this particular structure==, here there is no question of dummy, so now, we will have case 2, then I will put household saving ==household saving== as a function of earnings ==earnings== and genders, if I will put earnings and genders then, this is quantitative, this is quantitative, this is qualitative.

So, now I will put like this ==this== way Y equal to beta 0 plus beta 1 X 1 plus ==plus== D gamma 1 D 1 plus U provided, so condition is that, U equal to 1 for male and 0 for female, so now, this is the case 2.

(Refer Slide Time: 08:50)
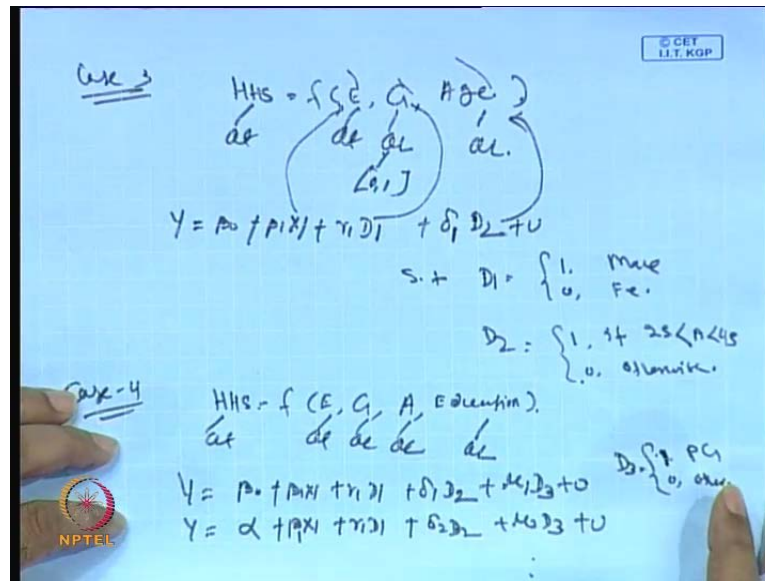
So, now, I will prepare another case, case 3, I will put age compositions, age composition is another factor, so household <mark>household</mark> saving as a function of earnings, then genders, then age <mark>then age</mark> so that means, this is quantitative, this is quantitative, this is qualitative, so that two indication of 0 and 1, so age composition, so this is also qualitative. So, accordingly I will write the model here, Y equal to beta 0 plus beta 1 X 1, X 1 is for earnings plus gamma 1 D 1, D 1 is for gender, then delta 1, <mark>delta 1</mark> then D 2, so D 2 is for age.

So, now D 1 is a function means where the condition is D 1, such that D 1 is 1 for male and 0 for female, so similarly D 2 classification is that age composition 1, if the guy is 25 less than less than 45 <mark>less than 45</mark>, 0 otherwise <mark>0 otherwise</mark>, so I like to know what is means here age composition, means here our fundamental objective is to know, what is the household saving and what are the factors which can influence, this factors may be quantitative means, what are the factor means, what are the variables, these variables may be quantitative in nature may be qualitative in nature.

But, <mark>but</mark> in most of the cases or <mark>you know</mark> you go to any regions or any country in the world, so most of the variables, which can influence the <mark>the you know</mark> component say household saving is that, so the variables which can <mark>which can</mark> highlight the household saving may be <mark>may be</mark> you can say earnings gender age and etcetera <mark>etcetera</mark>. So, this is

how we have to we have to fit the models, so household saving as a function of earnings, gender then age issue.

So, now this is quantitative, this is qualitative, this is qualitative so if this is the case then, this will be the models, so then I will take another case, case 4 case 4 again I will take household saving as a function of earnings, genders, age then, I will take educations education education is another variable, so this in that case, so this is quantitative, this is quantitative, this is qualitative, this is qualitative, this is also qualitative.

So, now in that case, so I will fit the model like this Y equal to beta 0 plus beta 1 X 1 plus gamma 1 D 1 plus delta 1 D sorry delta 1 delta 1 D 2 then mu, mu 1 you know D 3 mu 1 D 3 or simply you can write like this Y equal to its better put here, alpha plus beta X 1 beta X 1 plus gamma 1 D 1 plus delta 2 D 2 plus mu mu 3 D 3 then it should be beta 1 then plus u; obviously, there is U here, so plus U.

So, now here D 3 classification is that, this is education for education, so equal to 1 if it is p g respondent is p g; p g qualification and 0 otherwise, this is how education can influence particular component.

(Refer Slide Time: 13:09)

I will take another problem case 5, so I will take household saving as a function of as a function of earnings genders age educations then I will put another component called as a profession, profession I will take another component called as a profession.

So, accordingly I will put this is quantitative, this is quantitative, this is qualitative, this is qualitative, this is qualitative, this is qualitative alight, so now, the model will be Y equal to Y equal to alpha plus beta X plus gamma gamma D gamma 1 D 1 plus delta delta 2 D 2 plus plus mu 3 D 3 plus plus you can say zeta zeta 4 D 4 pus U zeta 4 D 4 plus U.

So, now here D 1 is defined, D 2 defined, D 3 defined, D 4 defined sorry D 4 is not defined, so D 4 is here is equal to 1 if the despondence despondence are say doctor 0 otherwise, means why I mentioned doctor for instance you see here, professor is a professor itself is a qualitative variable, so within the profession there is lots of quality for instance our aim is here to describe what is the profession of a particular responded on household saving.

So, now once you have 500 response data pictures then obviously, with respect to profession we have some category, so let us say out of 500 despondence there are 300 or more than 300 are doctor profession then obviously, you will set the dummy like this way 1 for doctor and the rest is for you can say other profession. So, that we can have better explanation, but you know for instance if out of such profession if we are 50 50 are there you know 500 500 divided by 50 means 10 professions are there, so you will put professions you know D equal to 1 for doctor, D equal to 2 for you can say manager, D equal to 3 for engineer, like this way.

So, then you will categorize, but you remember one thing when you will introduce dummy one after another then your variable setup will start increasing, so that is how the just like the problem of multicolinearity, so the moment we will introduce one after another then, obviously, model reliability model means, model accuracy will be very high, but in the same times it will get affected the model reliability part, because the moment will introduce one after another variable r square will start increasing then f will start increasing, because it is just like independent variable, but the thing is that, so the moment you will one after another, dummy in the system of course, you can do it, but researcher has to decide the situation, how many within a particular setup, how many dummy he can fix and how can he can structure it properly.

For instance this particular case I, I am putting for this particular variables say profession, so I am categorically divide into D into 1 1 0 only, so it can be multiple in natures, so that in that multiple case, so you have to restrict our self depending upon the

situation or scenario of the situation, scenario will be permit then you do that one, so if situation scenario will not permit, then you have to go in other way around, so I will tell you what is that other part of this particular study.

So, now in that case, if we will introduce one after another dummy then; obviously, there is multicolinearity problem, so r square will obviously, high f will obviously, high but when you will introduce one after another, then whatever variables already in the system they should be significant in the same times, the variable which is introduced in the new variable, which is introduced in the system, that should be also significant if introduction of that particular dummy will affect the dummy, model accuracy you can say other part of the model its better you have to drop that, particular variable or that particular variable qualitative variables, so that is the profession component.

Let us see, let me highlight another way, how the profession can be categorically divided into various groups, so for instance you see here, let us let us have a look here, household saving then obviously, earnings is another variables. So, earnings is this is Y this is X 1 then then forget about in the mean times.

Let us say we are remained silent about this three factors, first we start with the profession obviously, profession is more important variable than this one this is so, now I will take it a profession here, professions I will call it say, you can say D 1 in this case it is D 4, so we will introduce another other qualitative variables after that, so let me equal to me highlight here, within a particular variable, how many dummy you can create further, so now, that is our aim here, so Y X 1 and D 4.

So, these are the data points you have, so this is 500 data points and this is; so earnings point quantitative, quantitative, qualitative, quantitative like this this is quantitative, this is quantitative, this is quantitative, this is quantitative, so you will go ahead, so now, this is qualitative, this is qualitative, this is qualitative, this is qualitative, so now, this is question mark, so you cannot directly regret, so you have to transfer it.

So, now once you have this particular format and this particular column, so forget about other columns now, first you decide about this particular column, then you have to look for other columns, so now, here the professions figures 500 professions are there 500 peoples professions are there, so now, we have to see, how many professions are there in this particular 500 observations.

So, now accordingly D can be categorized for instance there are, if there are 4 professions, so let us say in that 500 samples there are 4 professions say doctors, then manager, then you can say you know chattered accountant, then you can say company secretary, where these are the peoples are in this particular group; so, so accordingly so now, you will 4 dummy's, so now you shall take it, how I will do this one.

(Refer Slide Time: 20:16)

So, now, you see here HHS earnings, then professions 1 2 3 4 5 6 7 8 9 10 up to you can say 500, so this is figures are already there in quantitative format, this is quantitative format figures are there, so now, here in the profession side, so let us say this is doctor this is a this is a manager, so this is doctor, this is chattered, this is manager this is company secretary, this is company secretary, this is chattered, this is chattered accountant, this is doctor, this is doctor, this is medical, I mean managers, so this is how it is get filled up.

So, now you have to transfer into quantitative information, let let you know we have 4 professions, doctors, and then manager, then chattered accountant, then company secretary. So, this is 1, this is 2, this is 3, this is 4. So now, we will transfer accordingly 1 2 then this is doctors, then 1 chattered accountant, 3 manager manager is 2, then company secretary 4, this is 4 then chattered accountant 3, then doctor 1, doctor 1, manager 3, so these are the problem.

So, now the sample size is of course, let us assume that, this is up to you can say, you know 10 12 vector, this 12 samples are there, let us start with this 12 sample structure, so now, this 12 sample structures the moment you have 12 sample structure, so then obviously, what you have to do, you can here there are two specific objective, means another side objective you have to create here means, here many objective is to study or to integrate what is the impact of profession and household saving so that means, our hypothesis that, if your profession is good say doctor or you can say chattered accountant or you know company secretary, then your household saving will be very high, if your profession is manager, then your household saving is very low like this way, so this is our observations.

So; that means, which profession is more appropriate to household saving or total profession is is there an impact on household savings, so first you to target total profession, then you will go which particular profession has a impact so that means, when you will go for total professional impact on this household saving then, obviously, So, dummy may not be serious problem, because you are using only 1 dummy and if we will use 1 dummy then obviously, expand uh overall fitness of the model will be high and most of the chances most of the chances more chance is here, that you may have significant of all these variables.

But you know if your objective is secondary in nature for instance you you like in addition to know overall impact of profession on household saving you you are also interested to know what is the influence of doctors, doctor profession on this household saving and chattered accountant profession on household saving, then company secretary appropriation on you can say household saving, then manager profession on household, if you like to have this type of secondary objective then; obviously, you have to create another 3 dummy's.

So that means, in the professions, so you will the model will be like this, so Y equal to simply alpha plus beta X plus forget about other dummy, so now, here I am just regressing household saving is function of earnings earnings plus profession only profession skipping other things remain constant. So, now D 1 this is beta 1 X 1 this is put it beta 0 then D D 1 D 1 put here D 2 X 2 plus D 3 X 3 plus D 4 X 4 plus D 5 plus D 5 X 5, so D 1 so means where where the condition is that D 1 equal to 1 for 1 for doctor and 0 for others, so similarly D 2 1 for manager and 0 for others, so D 3 1 for chattered

accountant 0, for others similarly D 4 equal to 1 for c s company secretary at 0 or other this is how you have to classify. So, now in that case, what you have to do <mark>what you have to do</mark> you see here, so now, let us say 10 data points.

(Refer Slide Time: 25:22)

So, 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17, let us say 17 data points are there, so households having then earnings <mark>earnings</mark> 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17, so all are quantitative information are readily available, so then earnings 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17.

So, now this is also quantity information then, we will call it D 1 D 2 D 3 D 4, D 1 D 2 D 3 D 4, so we mentioned here or I will call it here simply D, so you see here this is for professionality this is for professionality, so now, let us say doctor case is like this way. So, 1 1 1 then others will be 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 then 0 0 0 0 0 1 1 1 0 0 0 0 1 0 0 0, so similarly D 3 profession, so this is 0 0 1 1 1 this could be 0, this would be 0, this would be 0, this would be 0, this would be 0 and this could be 0, this could be 0, this could be 0, this could be 0, this is 0, this is 0, this is 0, this is 0.

So, D for it would be obviously, 0 0 0 then 0 0 0 0 0 0, then this should be 1, this would be 1, this would be 1, this would be 1, then this would be 0, this would be 1, this would be 1, this would be 1, this would be1, so this is how the complete sample structure is all about like this.

So, now, now if I will if I will like to know what is the total professional impact on household saving then what we have to do I will put like this 1 1 1 then 2 2, no it is better it is 3 3, then this should be 2 2 2 2 for 2 this is this is 4 4 4 4 4 and this should be 2 this should be 4 4 4 4, so this is how

So, now once you have professional impact on household saving then obviously. So, now, either you use this particular dummy or you use all dummy for this you know with respect to 1 dummy so; that means, this is depend upon your objective specification if your objective is just to uh you know if your objective is like this way, if your objective is what is the total professional impact on household saving, then you have to follow this particular structures, you have to integrate this 1 with this 1 with a earning salary.

So, now if your secondary objective, what is the impact of doctor professional in this household saving, what is the manager manager professional household saving or what is the C S professional on household saving or what is the you can say chattered accountant professional on household saving, then in that case, if this just to be removed, then you have to create for dummy's you see, now the total setup if we will highlight the total setup here.

So, now how many dummy is here, so this is with respect to gender, this is with respect to age, this is with respect to education, now you have to create 4 dummy's. So, D 4 D 5 D 6 and D 7 so that means, it is a purely multivariate model, so as a result r square will be very high f will be very, but I have a doubt about you know with respect to significance of individual parameter, that is how we call a specification test.

So, now when we will go for you know specification test, that times if your sample observation is not exclusively very high, then it will create problem, but here with the respect to 500 sample points it may not be serious problem, later it will give you you know excellent you know result, so that is why you have to go for this you know categorization.

So, when you have a dummy variable modelling, so you have to see, because dummy you have to categorize it, you have to decide, we have to represent this structure, so because it is response variable, so once you get the response, so accordingly with the response you have to design the dummy variable; so whether it is unique in nature, multiple in nature, you have to decide provided your objective must be like that way if

your objective is <mark>you know</mark> secondary, then you have to go for this one, if your objective is primary, then you have provided it <mark>it</mark> must be supported with database.

For instance if your objective is secondary then, obviously, you have to create 4 dummy's for 1 is for doctor, 1 is for manager, 1 is for chattered accountant and 1 is for company secretary, but in the meantime suppose you have a sample observations say only 50, so 50 is I am not, I am very sure with 50 sample observation it will go up to you know 9 dummy's then, obviously, it may give indication, that the model may not be perfectly sample forecasting our position; so it is better when you <mark>you</mark> are creating secondary objective and you are creating additional dummy variable then accordingly you have to increase you sample size.

So, sample size is absolutely very high then; obviously, it will give you better <mark>better</mark> estimation, better forecasting or you can say it can be; obviously, useful for policy uses, so, this is how that, this is one of the interesting example, which you can sight in the case of dummy variable econometric modeling.

So, now I will take another example, so let us let us take a case of I will take same things you know there is another problem called as a savings and income.

(Refer Slide Time: 32:14)

So, I will put another problems called as a savings and incomes; so which we have called just now we called it is in the form of earnings.

So, now the way we have discussed the previous problem is that, it is a cross sectional problem, so now, I will introduce another type of problem where dummy can be used, but that to time times series till dummy, so what is what is a the structure of that particular component, so, you see here, in that case saving is quantitative variables and this is quantitative variable, this is quantitative variable.

So, I like to know what is the impact is of and you see here, before I highlighting objective let us say have a sample size here. So, this is time factor. So, let us assume that this is in the case of say Indian economic in the case of India, so we are taking 1970, 1971 upto you can say 2009. So, my objective is whether whether the globalization of 90 globalization the globalization is 90 has an effect on the saving income in equality saving income in nexus this is my objective.

Because, ultimately dummy variables use a dummy variables representation depends upon the objective specifications, if my objective is to know what is the impact of income on saving, then the problem is very simple one, but if with this two variables I will create additional dummy, because dummy is a most of the cases is a artificial creations, so you have to create a artificially for the model reliability, model simplicity and model feasibility.

So, now if I will add another objective whether whether globalization has an effect on saving income nexus or you can say saving in income equation, so in that case, so you have to introduce a dummy, so it is a globalization impact you know in India globalization 1990 is the most important year for globalization; so what I i have to do instead with the total sample observation 1970 to 2009, so I will take here 1990 is benchmark, 1991 is the globalization error, so it is the benchmark.

So, now what I have to do. So, I will put after 1991 or greater than equal to 1991, I will put 1 1 1 rest item I will put 0, So, then I will go for specifications, so like this it is the dummy I will create the time dummy. So, I will create time dummy then; obviously, this should be 0 0 0 0 0 0 0 0 0 0, then this way continue this is 1 1 1 1 1 1 1, so this is 1, so this is how you can go for classification, but you know again if I will put third objective what is the impact of global financial crisis.

What is the impact of global financial crisis it is introducing the year say 2008 made of 2008 so; obviously, if it is annual annual then obviously, there may be some problem,

but you have say 2008, but if it is more 3 data then; obviously, you may not have such problem, so now 2008, so instead of 2008 you will put 2 2 2 2 2 2, So, now, the entire data has a classified into 3 parts.

So, 1 part is having 0 figure, another part sample structure as a 1 figure, another sample as a 2 figures, so now, if your objective is to have you know structural brakes on this particular saving and income then obviously, you can get to know by this particular dummy if your objective is very much clear about the issue, you can say global financial crisis or you can say globalization impact, then in that case what you have to do. So, you will create 3 dummy, so D 1 D 2 D 3, so here you will put in this case uh these are all in that case you will put 1 1 1 1 1 1 for 1972 to 90 then others will be 0 0 0 0 0.

Second case you put all 0 only from 1991 to you can say 2007, we will put 1 1 1 then, rest will be 0 0 0 then in the D 3 case you will put to 0 0 0 all cases and after nineteen after 2008, so you will put 1 1 1, so now, you can if it is supported by alpha 1 it is supported by alpha 2 it is supported by alpha 3 if alpha 1 is significant alpha 2 is significant and alpha 3 is significant so that means, we can conclude that uh there is impact 1972 to 1990, so this income has a substantial impact on saving.

So, then if this is the alpha 2 is significant, then you'll say that globalization has a significant impact on income saving nexus, again if this is also significant global financial crisis figure we will say that global financial crisis also contribute impact, that affect the saving income nexus, this is how the dummy variable can be properly setup.

So, now various ways you can say classify the dummy structures you know particularly dummy dependent structures, dummy independent structure; where there is 1 dummy 1 dependent variable, where which is purely quantitative in nature and in addition to that, there is some uh independent variables which is which are quantitative in nature and some are qualitative in nature.

So, I will put there you know there are many application like you know dummy variable modeling, so we have already highlighted 2, 3, I will put before I conclude in this particular session and we like to move to other particular item, that is a dummy dependent variable econometric modelling, so I I like to highlight 1 particular means I I like another case under this dummy independent technique; so that is how it is called as regional effect, so to just to determine the regional effect.
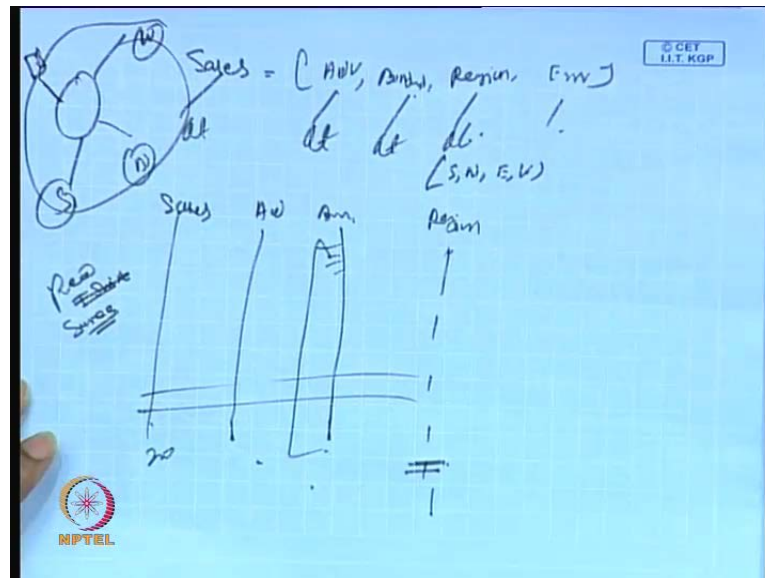
So, I have briefly highlighted little bit earliest, the sales impact on say advertising expenditures, then bonus region and of course, error terms this is how the this this is how the model is all about, so now, sales is quantitative, advertising is quantitative, bonus is quantitative, region is qualitative, error forget about it yes the way we have discussed here you remember, I have discussed all including the this one, I have discussed for different problems, under dummy variable modelling, but you know I I am just giving you the estimation process of this dummy variable modelling.

So, how you have to build the setup and how you have to estimate the setup, after having estimated equation, then you have to go for lots of you know authentic check for instance we have to go for specification test, reliability check, then autocorrelation problem, heteroscedasticity problem, multicolinearity problem, etcetera etcetera. So, these are all you know possible uh you can discuss all these in details, but you know we are not highlighting here, all these discussions means we are not testing, we are not going to we are not going to discuss anything about the specification test, reliability test or multicolinearity problem, heteroscedasticity problem or autocorrelation problem.

So, once you have the estimated model, then you have to follow that particular principle the moment the you know the importance of dummy variable is that, a fix of dummy variable and it is boundary, that is more important once you fix, that introduce the

number of dummy in that boundary, then the problem is as usual the econometric models original general econometric models, so after that the procedure is almost all same.

So, if any any problem in between for instance, let us say you have a problem, say summation uh beta Y equal to beta 0 plus summation beta I X I equal to 2 and summation gamma I D I where I equal to 2, so that means, it is 4 variable case. So, now, in the 4 variable case, 4 variable must be significant if, so I have to first I have to first give the structures, so what is Y, what is X 1, what is X 2 and what is D 1 and what is D 2, so I will give you the complete information in a quantity format; of course, I have downloads of homework here, before coming to the contrary information.

Once the quantitative excel sheet is ready, then I will hand over to analysis to do it your job, so he has to do it and he will continue the specification means you will go with estimation process, then you will come out with the estimated results, after that he used to go for specification test, reliably test, then you know all these problems like more (( )) heteroscedasticity, etcetera, etcetera, but if we found there is any fault, then he go back to original pictures and you will find where is that possibility; whether there is need of transformation or dropping of any dummy etcetera, etcetera.

So, once you have the estimation result and having the quantitative information data, so then no problem you know to go through all these procedure or you can say structures, so it is very easy after that; so in that particular context you know I will like to highlight another different problem with respect to this dummy independent variable.

So, here a region, I like to know the regional impact on this sales advertising in bonus for instance, the structure will be sales here, advertising here, bonus here, so now, this is region. So, now, I will I will let us say the region is classified into four groups, here south, north then east, west so that means, I have to collect data figures, means this is a company here, company is distributing for different; I mean sales is going to east, sales is going to west, sales is going to south, sales is going to north; so so I will collect all these all these you know for sales volumes, then I have to represent here.

Let's say there are 200 sample per sales, so I will put a total advertising expenditure of the company, total bonus given, total amount of bonus given to the employees and where are the regions, where are the regions I have to highlight, so if the if the objective is only regional impact on sales then; obviously, your problem is little bit simple, but in the

simple that means, you are putting in south, west, north sense out, just you are putting 1 2 3 4 the way, we have highlighted last problems in the case of profession, similarly here either you apply 1 2 3 4 and then you go for estimation and check the status or else if you are very keen on you know what is the Hindu peoples impact on sales, whether Muslim people impact on sales, etcetera, because sometimes it depends upon what is the problem, what is the product and what is the service you are delivering.

For instance, let us say you are you are delivering a product sales means, this sales may be a product with respect to say red red t shirt, so generally these are, you can say these structure will be well designed or feasibility is very high, if you have sound theoretical knowledge without having a theoretical knowledge to apply or to use dummy can be very problematic, because dummy is designed with respect to theoretical knowledge I just I am just highlighting this particular issue.

So, let us say the sales with respect to red t shirt red t shirt or you can say a red saree, saree is mean for female, so now, you are respondence first of all your respondence will be females, for instance initially you if you are not, if you are using red saree, then if you are influencing regions, so it is not a problem for if it is male response respondence or female respondence until and unless you use the term gender; since gender is not there. So, you just calculate what are the sales, red saree is the a product.

So, now the company is producing only red saree or any different saree, but we are taking only volume of the red saree, so now, this is the sale picture of red red saree and this is the advertising expenditure for this red saree and this is the bonus we are giving to employees, so whether there, first of all whether they are using the this for saree or not; so if obviously, if here there may be you know within the bonus it is a quantity picture, but within the quantity picture we can create a dummy also, for instance bonus is extra amount, generally company used to provide to the, to them to the employees.

So, now, the issue is here with what is the bonus impact on sales volume so that means, if there getting bonus and there are not purchasing any sales of that particular saree, then there is no such impact so that means, a bonus is mean, we are giving bonus or the company is giving bonus to increase the or to change the structure of red saree only, if that guy is not purchasing, then forget about it, so in that case you can apply dummy to know, whether bonus has a having impact on saree or not.

For instance, whether they are purchasing the saree with the bonus amount for instance I would ask the response, whether you are using your bonus amount to purchase saree, then you put 1 if no then you will put 0, so answer is, obviously, you can get to know whether bonus has impact on sales saree, so if that 1 is significant then; obviously, bonus has a impact otherwise it has no impact.

Similarly, region, so now, once you know bonus means you see here out of I founded only 2 hundred cases the bonus is used for purchasing rate saree then; that means, suppose a regional impact is concerned if we will integrate all these things it is regional impact has to be modified accordingly instead of taking 500 sample it is better to take 200 samples, where all these items will be very consistent, then within a region, then again you will see which region is more effective if it is only just to know the regional impact on the particular issue then; obviously, you all go ahead with original this particular say if you like to know what is the means Muslim impact on this 1 or Hindu impact on this 1 or Christian impact on this 1 then; obviously, you have to create additional dummy D 1 for Hindu D 2 for Muslim D 3 for you can say any others like this way then you will see the impact of each each religion on that particular sales. So, this is another type of dummy variable econometric modelling which we have to highlight in that particular structures alright.

So, with this say with this we we can somewhat, means we can close this particular structures like you know quantitative sorry qualitative response, dummy independent variables, before I close this particular chapter and move to the other part qualitative response dependent econometric modelling.
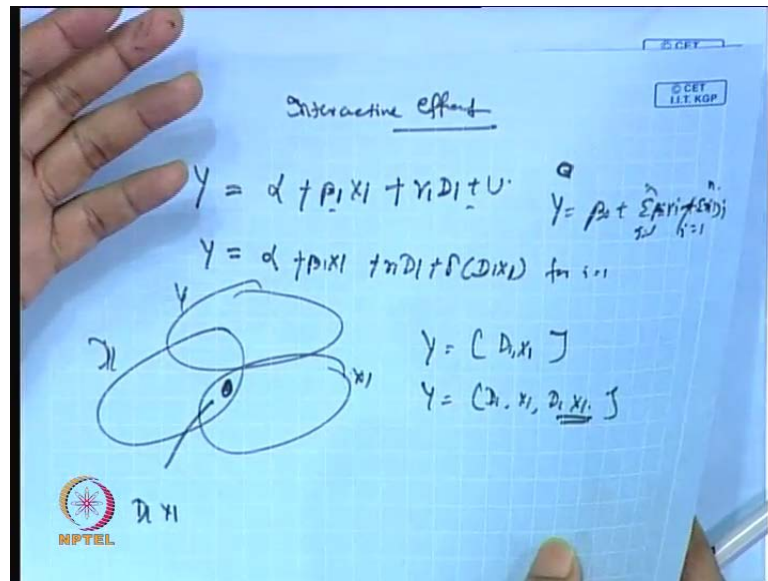
So, I will highlight one thing here the component called as a interactive effect, for instance this simple quantitative means interactive effect this term is called interactive effect.

What is this, so let us start with a simple uh you know qualitative response regression modeling, so Y equal to alpha plus plus beta 1 X 1 and plus gamma 1 D 1 gamma 1 D 1 then U right, so far as interactive effect is concerned we are just starting the individual impact. So, now, if there is such here what is the general formula, means generalized structure, here Y I equal to beta 0 plus summation beta I X I plus summation gamma I D I gamma I D I, I equal to 1 to n I equal to 1 to 1 to n, this is how this structure is all about, so now now for for I equal to 1, then, obviously, this is the setup.

So now, suppose interactive effect is concerned, so I will take like this, alpha plus beta 1 X 1 plus gamma 1 D 1 plus delta into D 1 X 1, so this is called as a interactive effect so; that means, you see here, so this is Y and this is you know this is you know X 1 and D 1 this is y. So, I like to know this this this is this variable is known as D 1 X one. So, what I have to do I will create; now initially we start with a function of D 1 X 1 alright.

So, now we have a function have a D 1 X 1 then D 1 X 1, so this is called as a indirect effect, so this this delta is significant, then this interactive effect has a significant impact on y, that means, the combination of D 1 and X 1 has a has a impact direct it is individually D 1 may not be effective individual X 1 may not be effective, but the join D

1 X 1 may be very effective for instance <mark>uh</mark> you if you take any good case, the goods may be complementary and substitutes.

For instance you see here I like to take here coffee, then I like to glass with a what is the coffee volumes with respect to milk and you can say milk and <mark>you know</mark> sugars milk and sugars are complementary in natures, so now, if I if my analysis is in that case then; obviously, the milk interaction on sugar will have a effect on you can say coffee. So, that is how the interactive effective effect can, you can say studied it can it can be studied so; that means, it is a joint impact on joint impact of dummy and independent variables to depend variables. Similarly, the complexity will be start more, when we will add another variable in the system.

(Refer Slide Time: 52:30)

So, let us say Y equal to alpha plus beta 1 X 1 plus beta 2 X 2 plus <mark>D 1</mark> D 1 gamma 1 gamma 1 D 1 plus gamma 2 D 2 plus. So, this sigma the simple models, but when we will go for interactive, then Y will be alpha plus beta 1 X 1 plus beta 2 X 2 plus gamma 1 D 1 plus gamma 2 D 2 plus delta into D 1 X 1 delta 1 into D 1 X 1 plus delta 2 into D 2 X 2 plus you can say U.

So, this is how you can go for it again we can add also another component, then delta 3 <mark>uh delta 3</mark> you can say D X 1 square and <mark>X</mark> X 1 square X 2 in the meantime, this can be another interactive effect. So, this interactive effect <mark>you know</mark> we like to know what is the joint impact on this one even if you know D 1 X 1 you can have a another variable D

1 X 1 uh and X 2 so; that means, is D 1 X 1 D 2 X 2 so; that means, D 3 X 1 square D 4 D 4 X 2 squares and D 5 D 5 X 1 and X 2, so this is how the interactive effective can study.

This is simple interactive effect and there can be also complex interactive effect, so when you will add one after another dummy the system with you know increase of independent variable then; obviously, the interactive effect will be more complex and it is also more interesting sometimes it is very essential and sometimes it is a it is it is very much objective specification, if your objective is not with respect to interactive effect then you go ahead with the simple model, if your objective is with respect to interactive effect, then you have to be very careful about this interactive issue with this we will conclude this particular session here; so in the next class we will discuss the qualitative response dummy dependent variables.

Thank you very much have a nice day.