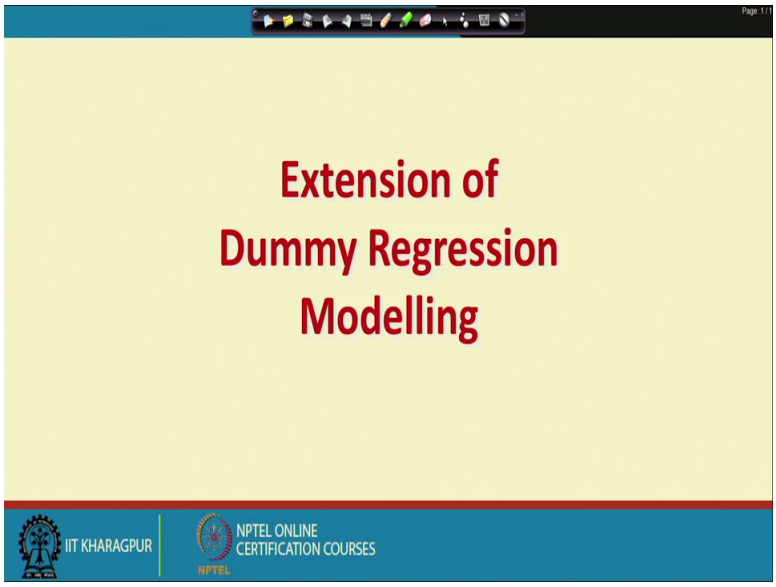
Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 37
Extension of Dummy Regression Modelling

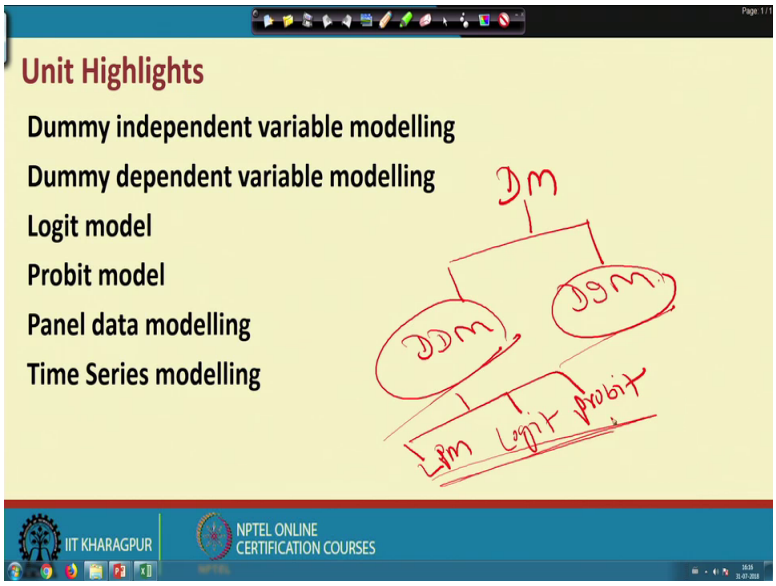
Hello everybody this is Rudra Pradhan here, welcome to Engineering Econometrics. Today we will continue with non-linear Regression Modeling and that to the Extension of Dummy modelling. So, let us see here what are the contents we like to discuss you know in this lecture.

(Refer Slide Time: 00:42)



So, basically we have discussed the concept of dummy modelling, that too while discussing the you know interactive effect and the kind of you know kind of you know structural problems where there is a need of you know introducing dummy variable and in fact, the utility of this modelling is very high and most of the engineering problems it has high relevance. So, in this lectures so, we briefly you know discuss this and then we can extend to the little bit you know higher version. Whatever we have discussed that is the basic approach about the dummy modeling and the idea is that while doing the regression modelling, the requirement is you must have a dependent variable and you must have independent variable.

(Refer Slide Time: 01:42)



If not dummy modelling, then the dependent variable and independent variable both will be numeric in natures; that means the variables information which you call as you know data, which must be numeric in nature. However, in real life scenario couple of engineering problems, where some of the variables cannot be quantified properly; that means, we may not have a numeric representation; that means, the data or the information to the particular variable is not actually quantitative. So that means, technically there is a you know structure or there is information about qualitative variable so.

So, qualitative variable means, the information is the you know qualitative in nature. For instance what we have discussed earlier is the kind of you know gender impart or the guy you know married unmarried or the kind of you know religion impart or the kind of you know stability instability impart. So, like there are you know various ways you can actually use the dummy variables, where the components or the kind of you know information will be qualitative in nature. So, now, technically if you go to dummy modelling. So, let us say this is a you know the framework about the dummy modeling. The dummy modelling in other words it is called as a qualitative response regression modelling. So, it can have you know 2 different you know structure, one is the dummy dependent modelling and dummy you know independent modelling. So, which this part we have slightly discussed earlier.

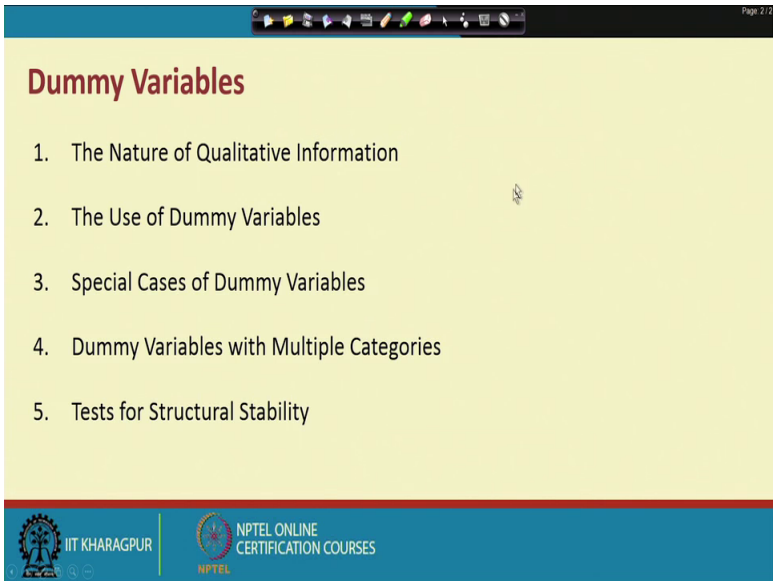
So, what will you do? We just briefly touch upon this particular component, then we will discuss this part which is not actually discussed yet. In the dummy independent modeling you know here you know one or you know few of the independent variables will be qualitative in nature while dependent variable is quantitative in nature. So that means, the information con information content is actually numeric in nature in the context of you know dependent variable. However, the information make you know numeric or maybe qualitative in the context of you know independent variable. Since it is a dummy modelling and that to dummy independent modelling at least one independent variable must be qualitative in nature that is the requirement.

So, as a result this kind of you know molding is a very special character or you know special type of you know models, and it cannot be applied each and every scenario like simple you know regression modelling. You can apply you know any kind of you know engineering problem provided it must have a dependent variable, you must have a independent variable and theoretically you may expect that there is a kind of you know relationship or logically you can establish that there is a kind of you know relationship.

Now, what in you know the kind of you know dummy modelling will tell you that you know. So, you have to specify or you have to apply a particular you know situation where at least one dependent variable will be qualitative in nature, and that will be one of the important variable which can predict the dependent variables that is what the kind of you know you know the structure which we can have in the analysis. And so, in the other sides we have a dummy dependent modeling, which we have a 3 different forms all together and that is the kind of you know say linear probability model, then we have a logit model and we have probit model.

So that means, technically in the dummy dependent modelling we have a 3 different you know models which means which are very beautifuls and applied in a situation where dependent variable is actually qualitative in natures. So, your independent variables may not you know qualitative, but the minimum requirement is that dependent variable must be qualitative in nature. So, first we touch upon the dummy independent modelling; that means, which you have already discussed, but we will just once again you know target, then we will come upon the kind of you know dummy dependent modelling. So, let us see how is this particular you know again. So, what will it do here? I will just you know the briefing the idea about the dummy modelling.

(Refer Slide Time: 06:40)



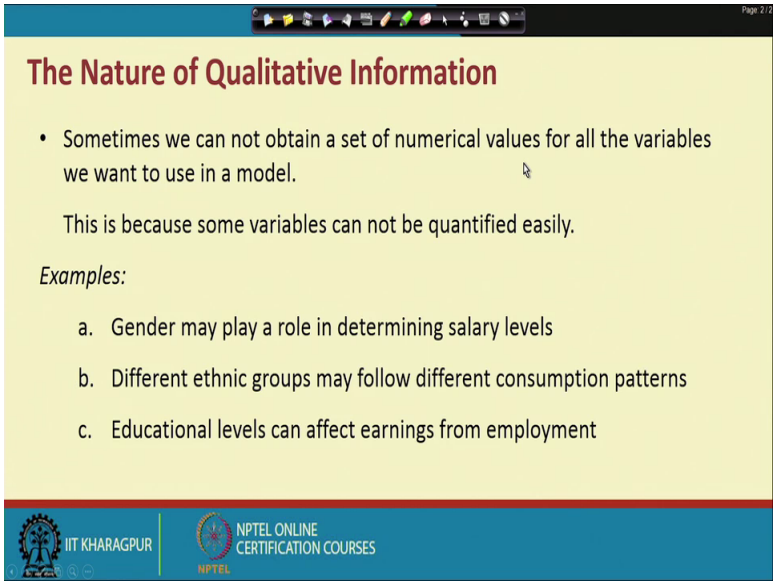
So, dummy modelling means, it is a nature of the qualitative information and then we will discuss how dummy variable can be used, then there are you know a specialty in the dummy modelling and then dummy modelling with you know multiple categories. The problem which you have discussed till now is related to only yes not kind of you know situation like male females, then married, unmarried; divorced, not divorced this means only you know in a bivariate framework only. But there are instances where dummy can be used in a multiple categories.

So, same variables can have a different kind of you know as scenarios for instance cited example may be religion. So, where 3 4 types of you know religion may be there or sometimes you are trying to study the seasonality, there may be for the 4 types you know decisions you can take or you can try to impart monthly impact on the relationship between dependent and independent variable.

So, in that case you may have a different kind of you know dummy representation so; that means, if monthly impart then you know. So, at least 12 dummies can be introduced in the system, and for that your data size should be very you know substantially very high. If we you are increasing more number of dummies and by the way this the size of the data is very small, then you may not in a position to you know address the problem because you know most of the variables may not be statistically significant which may

affect the reliability and as a result you cannot use this model for the addiction and maybe forecasting, then some of the instances you can use for the structural stability.

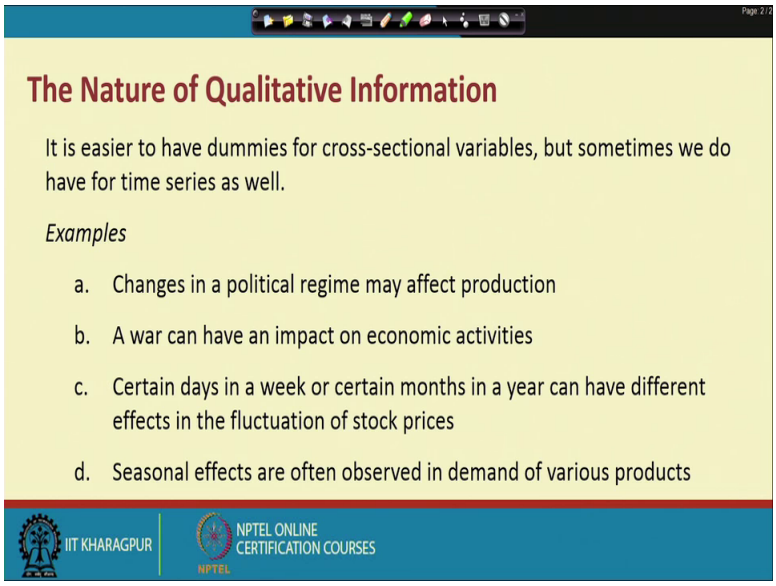
(Refer Slide Time: 08:39)



So, the nature of qualitative information which you have already discussed the cited example means base added example is the gender impact on the salary structure, then you know religion impact on the consumption patterns, educational levels or in earnings or the kind of you know employment structure.

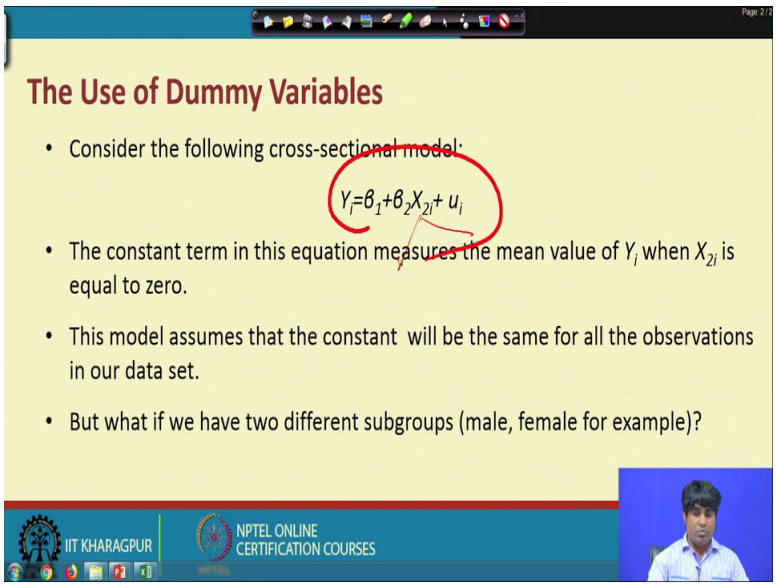
I mean these are all you know various types you know examples or you know, various scenarios where you know the use of you know dummy modelling you know very frequent in fact.

(Refer Slide Time: 09:13)



And then we will have here some of the you know more such kind of you know scenarios, you know changes in political regime may affect the production process and the a war impact on economic activities. So, he war means the war happening in a particular year and not happening. So, this is the kind of you know situation you can study, then you know crisis impart then the kind of you know seasonal team part against. So, these are all you know various types of you know kind of you know problems where dummy can be frequently huge.

(Refer Slide Time: 09:51)

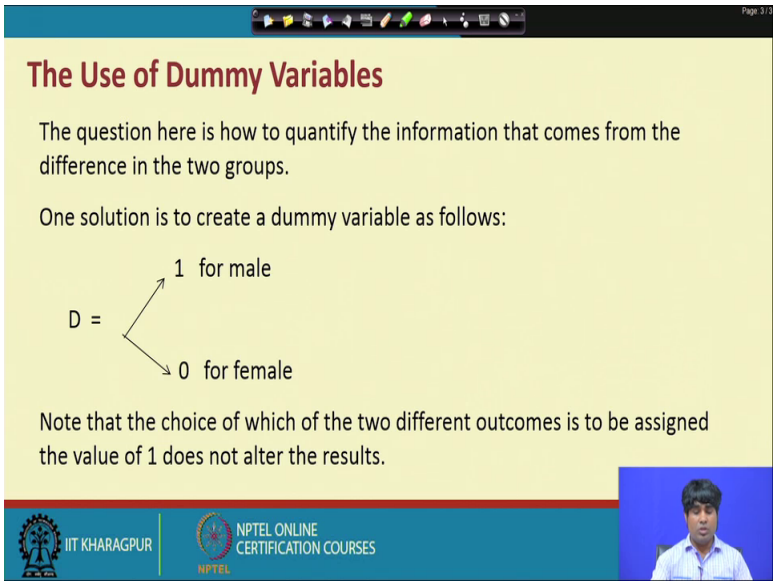


So; that means, once you once you know the kind of you know structures. So, you can specifically use this dummy molding and it is as usual actually regression modeling only thing is you know the variable understanding will be little bit you know different, and that to it should be you know classified in 2 different groups or 3 different groups something like that and to understand more in a much better way. So, we may have a equation like this ok. So, this is the equation where you know we have X 2 and this X 2 you know maybe you know I mean say it is only 1 variable.

So, by default it would be dummy if it is a dummy modelling concept, then let us say it is a gender impart. So, it will be male female and then the transformation will be 0 1 and y is the salary structure, which you have actually discussed in our last couple of lectures so; that means, that is very easy to you know understand the dummy and with the you can actually easily you can study the impact. That means, technically there are various factors, which can affect the you know salary structure, but in even there is no independent variables then still you know gender can be created and we can study the gender impact on the in a salary. So that means, that is a very we you know beautiful component through which you can you know model can extend as per the particular you know requirement.

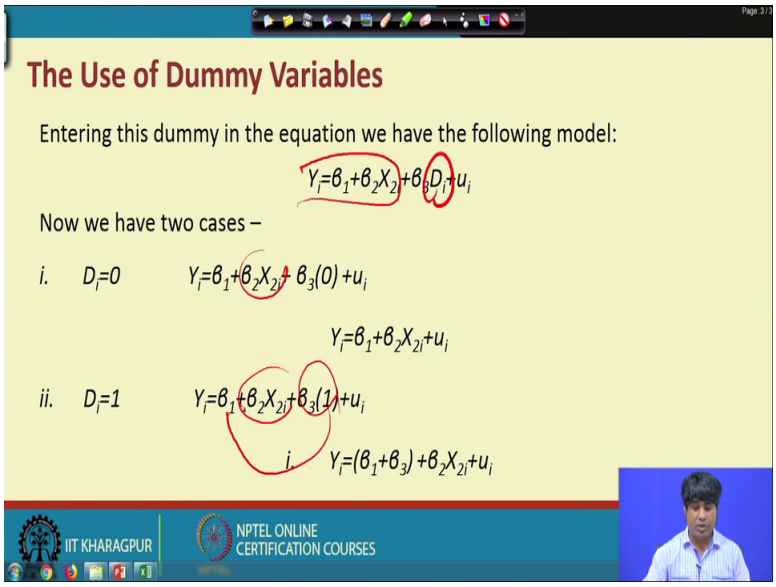
So, it brings some kind of you know consistency or you know stability to you know analyze the problem. Then you will have actually different kind of you know structure all together that is what the description.

(Refer Slide Time: 11:29)



So, these generally it varies from 1 to 0. So, 1 for male and 0 for female so, as a result once you get the estimated output. So, in 1 case the y equal to alpha only in under case y equal to alpha beta alpha plus beta, where you know if the representation is the male and when beta equal to 0, then the representation will be female and in that case y estimate will be only alpha. So, and if you add beta then that will go to the you know male side.

(Refer Slide Time: 12:04)

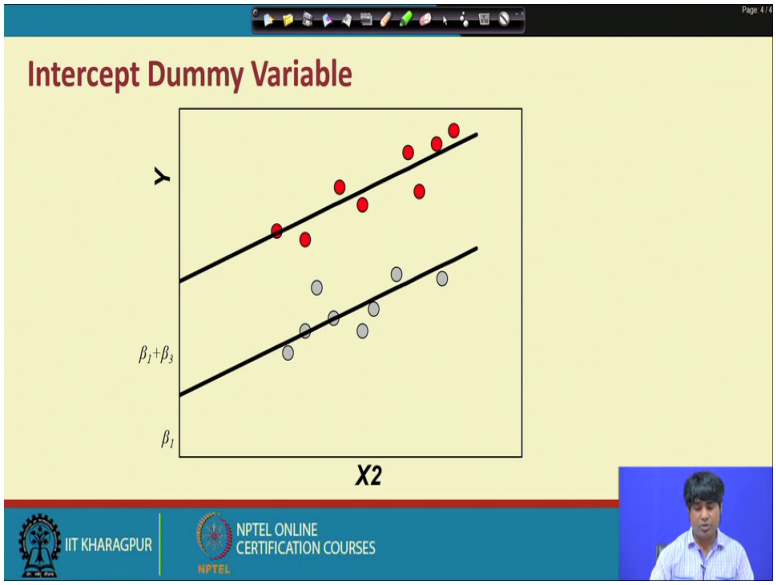


So, that is the kind of you know difference so, you see here. So, the model will be like this. So, in this case we have one dependent one independent variables and 1 dummy

variables both are independent, but this is numeric in nature. So, in this case so, this may be numeric in nature and this side this is a dummy in natures.

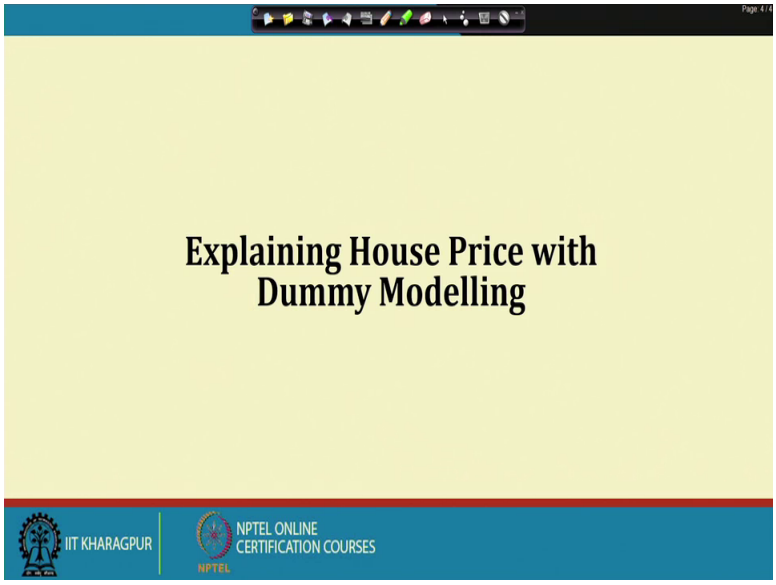
So, now if you put D equal to 0 then the model will restrict to this much and when D equal to 1 then the model will be restricted to beta 1 plus beta 3 plus beta 2 x 2. So, beta 2 x 2 will be common both the sides only thing is the extra component will be beta will be added that will be discriminate into 2 different groups that is the kind of you know examples which you can have in that case of you know dummy modelling.

(Refer Slide Time: 12:54)



So, this is how the various you know you know you know clusters where we can differentiate the 2 different categories.

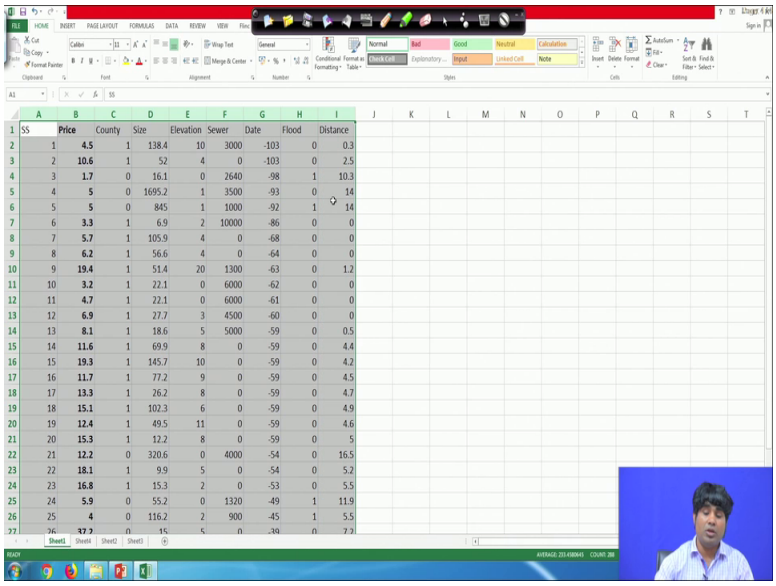
(Refer Slide Time: 13:00)



So, this is a kind of you know example which you like to discuss this is a case study and it is a civil engineering problems just to real state problems here usually you know the for the builders or the kind of you know consumers, they are very keen to know what should be the housing price. And what are the factors responsible for that and by using this case you get you can get to know how dummy involving can you know give some kind of you know better structure to predict the housing price. Because the usually or theoretically a house you know plot price will vary from various factors with respect to number of rooms and with respect to distance, with respect to facilities with respect to quality.

So many items are there which can affect and dummy modelling is a very beautiful you know concept where we can actually use and very easily capture all these factors and then predict the housing price as per the builders requirements and as per the customers requirement. So, before we start this particular case so, let me give you as you know examples yes you know data set how does it look actually in the dummy modelling case. So, let us say so, this is a kind of you know example we will be just you know discussed some case differently.

(Refer Slide Time: 14:27)



SS	Price	County	Size	Elevation	Sewer	Date	Flood	Distance
1	4.5	1	138.4	10	3000	-103	0	0.3
2	10.6	1	52	4	0	-103	0	2.5
3	1.7	0	16.1	0	2640	-98	1	10.3
4	5	0	1695.2	1	3500	-93	0	14
5	5	0	845	1	1000	-92	1	14
6	3.3	1	6.9	2	10000	-86	0	0
7	5.7	1	105.9	4	0	-68	0	0
8	6.2	1	56.6	4	0	-64	0	0
9	19.4	1	51.4	20	1300	-63	0	1.2
10	3.2	1	22.1	0	6000	-62	0	0
11	4.7	1	22.1	0	6000	-61	0	0
12	6.9	1	27.7	3	4500	-60	0	0
13	8.1	1	18.6	5	5000	-59	0	0.5
14	11.6	1	69.9	8	0	-59	0	4.4
15	19.3	1	145.7	10	0	-59	0	4.2
16	11.7	1	77.2	9	0	-59	0	4.5
17	13.3	1	26.2	8	0	-59	0	4.7
18	15.1	1	102.3	6	0	-59	0	4.9
19	12.4	1	49.5	11	0	-59	0	4.6
20	15.3	1	12.2	8	0	-59	0	5
21	12.2	0	320.6	0	4000	-54	0	16.5
22	18.1	1	9.9	5	0	-54	0	5.2
23	16.8	1	15.3	2	0	-53	0	5.5
24	5.9	0	55.2	0	1320	-49	1	11.9
25	4	0	116.2	2	900	-45	1	5.5

So, now, in this case so, we have actually 31 samples. So, this is the 31 sample and then we have the housing price. So, let me let you know maybe this is these are all in lakhs, and then it is a different you know you know factors which can responsible for that size of the house, then elevation, sewers, date, flood, distance what I have already mentioned. You see here so, this price this price is the dependent variable and then these variables are actually independent variables. So, now, in this problem so, we have a dependent variables like this, then dependent variables like this and then the you know the dependent variable like this.

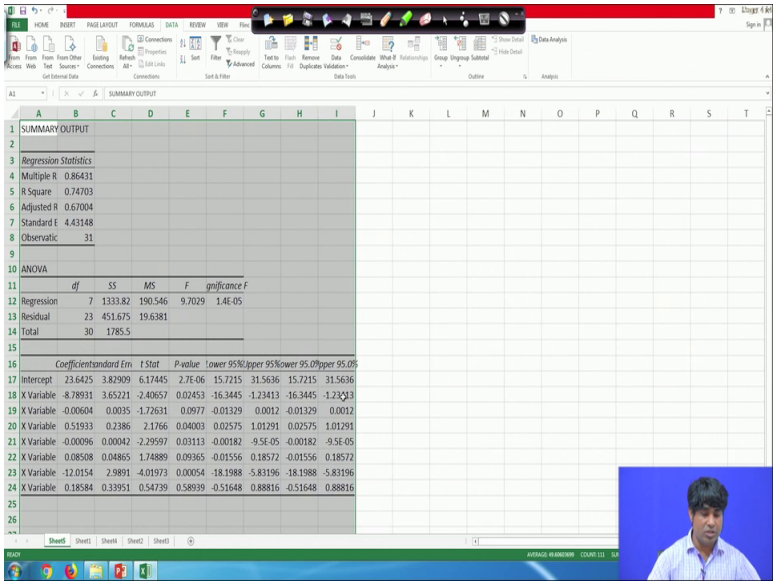
So; that means, technically so, the entire set is actually dependent variables, where we have altogether 7 independent variables and out of which the first independent variable and sixth independent variables they are you know dummy in character. So; that means, this is actually county yes no situation. So, where yes it is one and where no 0 similarly whether flood can affect the housing price.

So, in that case so, if flood is happen the during that locality then it will be 1 if not it will be 0 so; that means, it is a full of you know data from different locations and then we are trying to predict the actually the housing price. After you know this these are all called as you know sample you know structuring. So, once you do the sample structuring then the estimation process will be as usual you know very same for instance after having this spreadsheet, but that; obviously, you have to understand these variables you have to also

understand this variable, how the impart can go to the housing price and ultimately you go to the data analysis and as usual allow this subtest to run the regression.

So, this is what the regression tool and now just indicate the housing price here's up to this 31 samples, and then you have independent variables and you include all the independent variables simultaneously starting with county to distance and where we have couple of you know dummy variables and couple of you know quantitative independent variables.

(Refer Slide Time: 17:04)

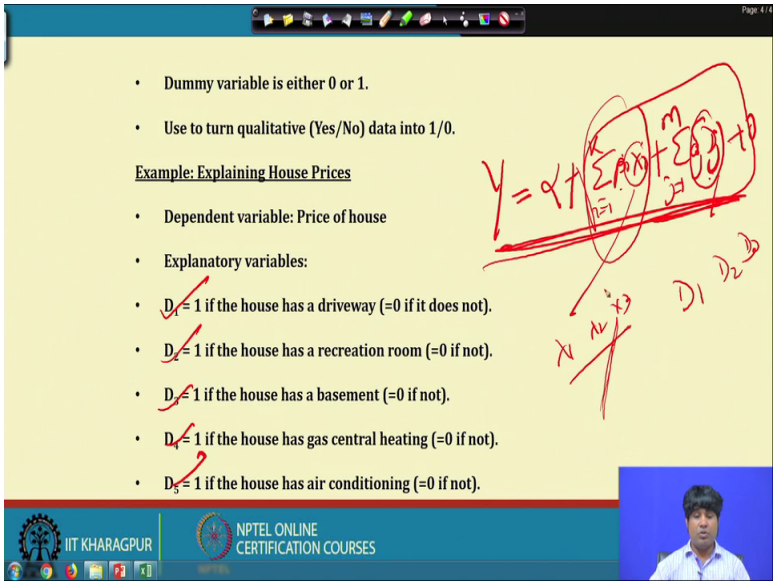


Then you just put you will find you know this results. You see this very interesting results and R square is very high and register R square is also very high, F is statistically significant and these are all the variables coefficients starting with the county to distance and then these are the coefficients some are having negative related some are positive related. And what is more important that you know these are all t statistic beyond these variables, and most of the instances the variables are statistically significant a variables are statistically significant; that means, there is a impart. So, and that to all these variables are you know significantly influencing the housing price.

So, now after knowing this so, now, what is it more important is, how you actually plan or you know structure the models for the customer attractions? Because usually most of the instances a customer will attract your you know product, here the product is the a house and if you have a different kind of you know variety. Of course, you know price

will be high depending upon the variety, but the beauty of the dummy modelling that you know it by default will give you lots of flexibility or diversity for the customers attractions and that you know itself will you know brings big business to the builders and the a kind of you know real state problem. In this contest I will just discuss this particular you know case.

(Refer Slide Time: 18:39)



And here you see here what will you do? We have 2 sets up you know variables so; that means, technically this case you know is like this where Y is the dependent variables, heres we have let us say alpha is the constant for summations beta i x i i equal to 1 to k then plus summation delta j D j delta j D j plus u. So, j equal to let us say 1 to m so; that means, this is x. So, this may be x 1 x 2 x 3 and so, on and this D j it may be D 1 D 2 and D 3 so; that means, all our independent variables and this particular cluster is the dummy independent sorry independent variables which is numeric in nature, but in this dummy is actually this variable is a dummy independent.

So; that means, this variable represents qualitative information's and this variable represents quantitative information. So, as a result so, we like to check how these independent variables affect the dependent variable, and how these you know dummy variables effect the dependent variables. So, now this is the model which you have applied in this particular case to understand the price variations and that to how beauty the dummy modelling is and to represent the housing price with you know different

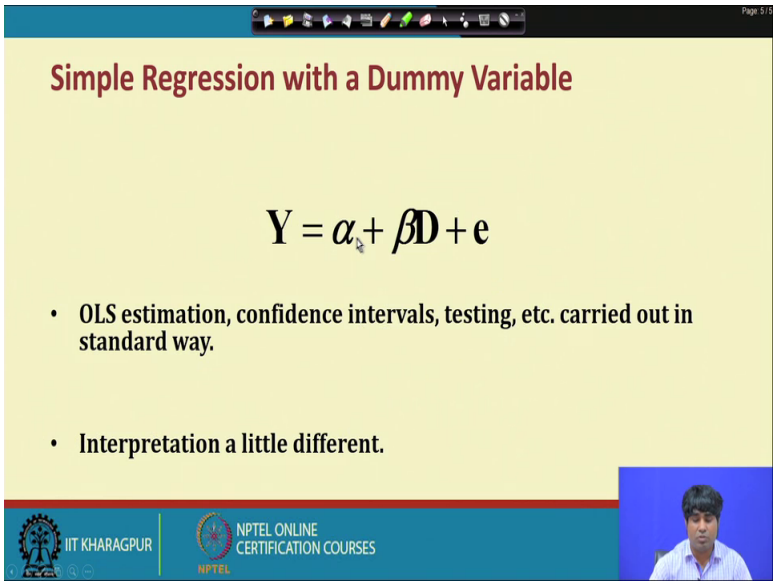
features and different character. So, now, all together so, here j equal to 1 to m means number of dummies. So, now, we have taken for dummies here D 1, D 2, D 3, D 4 and D 5 the expert will incorporate you know after this.

So, now, first one is the D 1 that is the first dummy, and which represents you know 1 and 0 if one then represents house as a driveway and if D 1 equal to 0. So, the house has not driveway so; that means, technically any future is yes means it will be add to the housing price. So, that by default housing price will be out that is basically add ons all these add ons will be yes no kind of you know situations and by default it will be add value to the housing price. So, if you reduce the add ons then by default housing price only a reduce and D 2 is equal to 1 if they are house has recreation you know facility, and 0 if that is not there.

Similarly a D 3 is a dummimal, dummy variables which represents 1 if the house has a basement and 0 if not similarly D 4 is another dummy variable which represents house is a central heating facility and D 4 equal to 0 if the central heating facility is not there then final D 5 if the house has ac facility then the D 5 equal to 1 and once D 5 equal to 0. That means, there is no ac facility so; that means, technically if all the facilities are there so; that means, D 1 equal to 1, D 2 equal to 1, D 3 equal to 1, D 4 equal to 1 and D 5 equal to 1 and that means, 5 different impart will go to the intercept and as a result the housing price will go up.

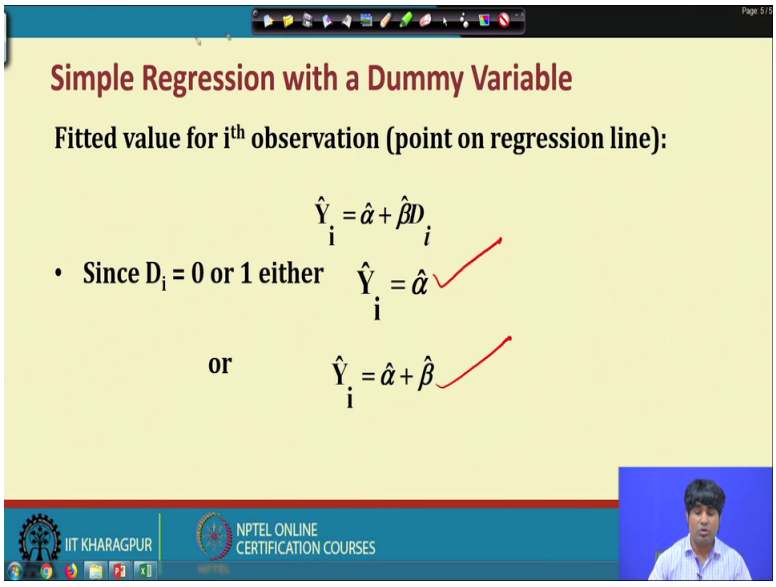
So, that is the kind of you know things which you can have to predict the housing price with respect to different features. So, let me explain so, how we can go ahead with the interpretation and for that you know we have already estimated this output.

(Refer Slide Time: 22:38)



Let us start with you know as dummy and where the model starts with you Y equal to alpha plus beta D and that is the error terms, then we will go ahead with the estimation process and then finally, we have estimated model.

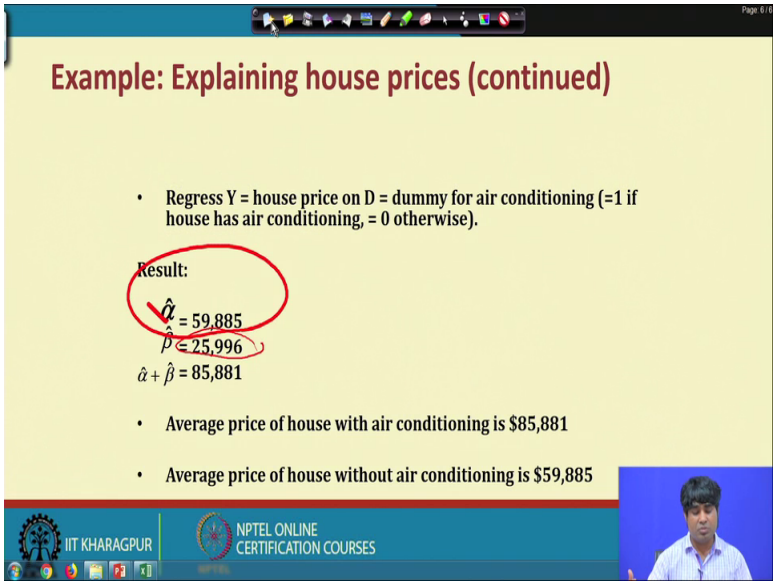
(Refer Slide Time: 22:56)



Y cap equal to alpha k plus beta hat D. So, now, we have 2 options. So, if D equal to 1 then the Y cap equal to alpha cap plus beta cap, if D equal to 0 then we simply have Y cap equal to alpha hat. So, that is the kind of you know difference. So, we have a 2 different options only. So, first option and second option so; that means, when you apply

different dummies. So, we have a plenty of you know options by default right. So, now, you know what we can do.

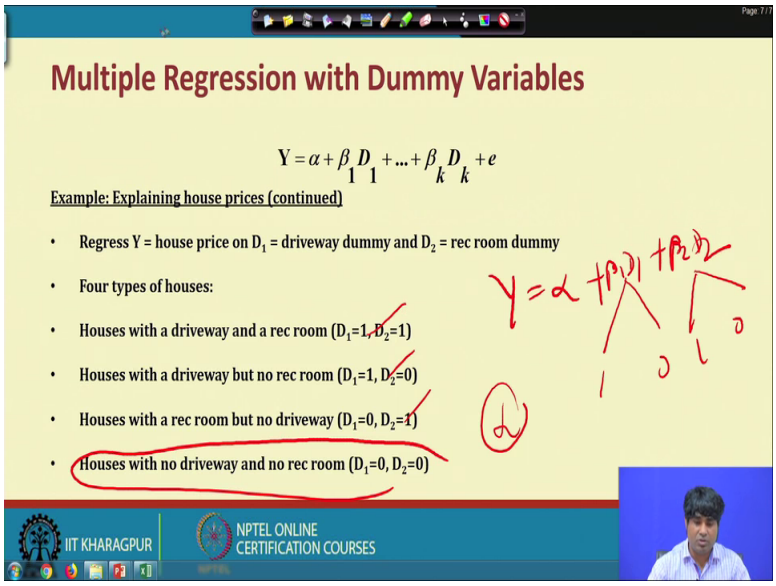
(Refer Slide Time: 23:31)



So, let us assume that alpha equal to alpha hat equal to 59000 approximately and beta hat equal 25996, then you know housing price will be a you know having let us say dummy is you know ac facility. So, now, what is the house price with you know ac facility. So, in that case. So, the value will be alpha plus beta, that is in 85881 and if there is no ac facility, then y cap is simply alpha hat in that case the housing price will be simply a 59000 you know 885 and if ac facility is there. So, 25996 will be extra.

So, that; that means, technically dumb the introducing dummy by default give 2 different package, where we have one situation ac facility another situation not ac facility. If you require ac facility then the housing price will be 85581 and if not then the housing pricing will be 59885. So, likewise you can have a another dummy also. So, now, in this case. So, what will you do? So, you allow with you now 2 dummies.

(Refer Slide Time: 24:53)

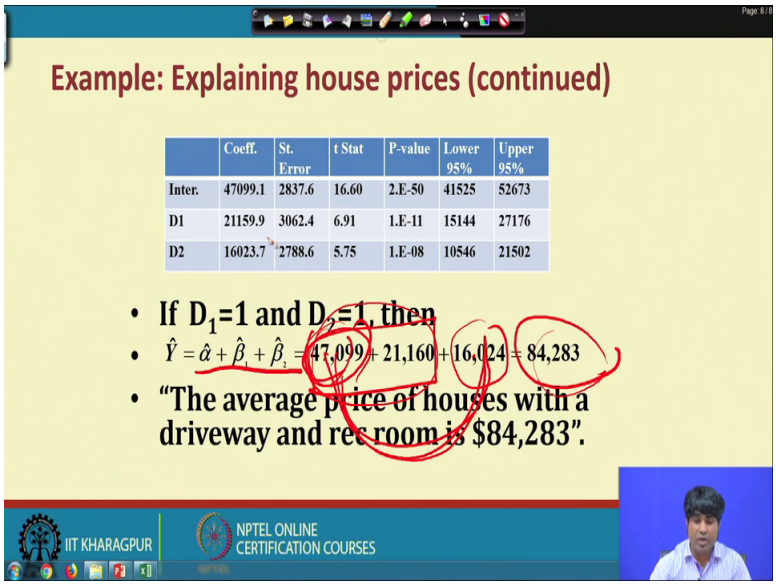


So, let us say you know b beta 1 D 1. So, in beta 2 D 2 and likewise you can extend up to you know k number of you know dummies or in this problems we have a 5 different dummies then you can you know extend one by one. So; that means, technically. So, if there are 2 dummies in a particular models let us say like this a you know see a it is a case like this. So, here what you can do. So, y equal to alpha plus beta 1 d ones and plus beta 2 D 2.

So, this will be 1 0 and this will be 1 0. So, ultimately alpha is the common, then in one case alpha plus beta 1 plus beta 2 that is the maximum housing price and then another model will be alpha plus beta 1, and third one will be alpha plus beta 2 and then the final one will be alpha plus beta 1 plus beta 2 and then fourth one is y simply alpha. So, in this case.

So, as a result we have a 3 4 different (Refer Time: 26:27) if you have a 2 different features. So, we can bring 4 different price all together so; that means, technically we have a different options altogether to represent the situation.

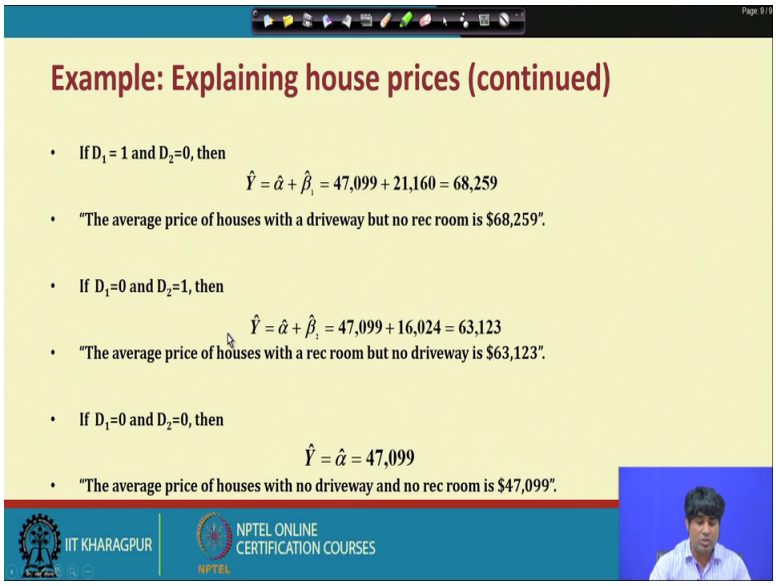
(Refer Slide Time: 26:19)



So, likewise so, let us say this is actually having D 1 and D 2. So, let us assume that this is the estimated output and by the way the variables are statistically significant. So, by default so, if both the facilities are there, then the housing price will be alpha plus beta 1 plus beta 2. So, as a result so, technically you will have this much of price and if both the facilities are not there, then the housing price will be this much only. If only one facility is there let us say this one then the housing price will be this much and if this facility job only there then the housing price will be this plus this. So, by default so we have a 4 different package.

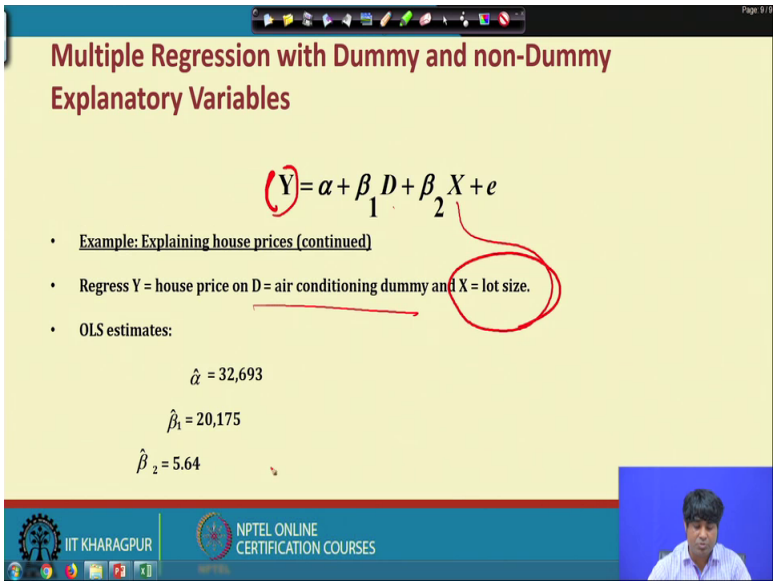
So, with all facilities this is the price with no facility this is the price with one facility this is the price and with he another features only. So, that is the price so; that means, for different price paid we will have to you know to we know attract the customers, and that too it is very easy to you know predict the such kind of you know situations by using a dummy variables.

(Refer Slide Time: 27:37)



So, this the case heres. So, four different cases which you have I highlighted.

(Refer Slide Time: 27:43)



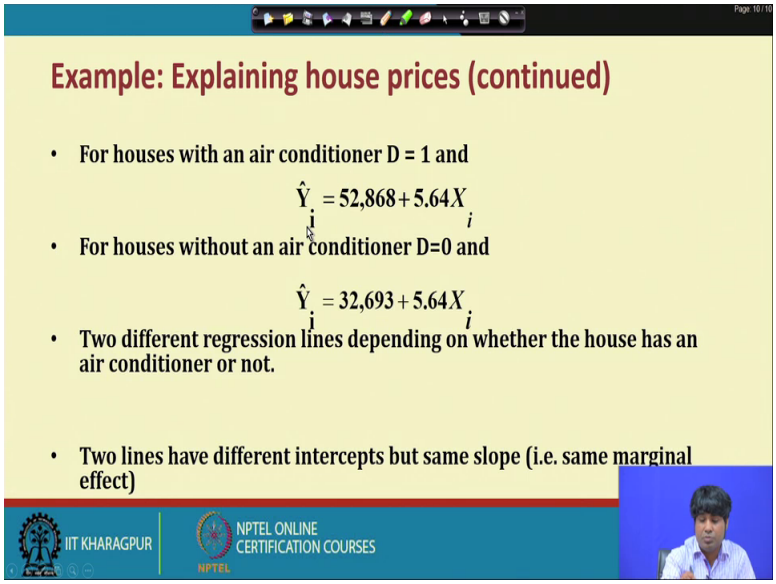
Now, likewise you connect to the extent one after another dummy. So, in this case we have actually 5 dummies. So, we can have actually 5 different you know situations, and then you can have a different price pack as per the availability of you know particular you know features, and that too where you know dummy equal to 1 and dummy equal to 0. So, now we will have a different kind of you know structure where we can go with the dummy with you know one more independent variable, which is actually numeric in

nature for instance why is it by default here we are discussing about the housing price and these the dummy variables let us the first dummy ac facility and x is the another independent variable, which is a numeric in natures; that means, quantitative stock variable.

So, here you know the X represents the lot size. So, if you book the house in a lot then you know you can have you know high and low price or some kind of you know discounting something like that. So; that means, technically you know booking the house in a lot. So, may have you know less price if you go by you know individual purchasing. So, as a result lot size can you know say effect the housing price in total. So, in this case so, we have actually 4 different package again. So, one package will be alpha plus beta 2 X after the estimations and then another one will be alpha plus beta 1 and beta 2 X that is that us the 2 different because only one dummy is there.

So, 2 different proxies. So, let us say alpha is this much beta is this much beta 1 is this much and beta 2 is 5.64. So, ultimately so, the estimation process will be like this. So, which we can analyze you know in this case. Where the dummy can be represented as a one, then this model will be restrict to this much.

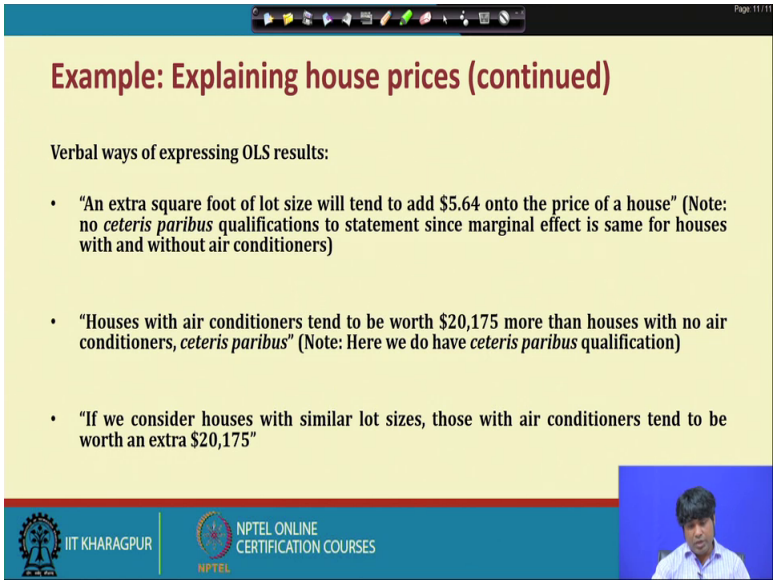
(Refer Slide Time: 30:00)



And in this case where dummy equal to 0, then you know we have the situation like this we have the situation like this. So, 2 different regression lines depending upon whether the house is in you know air conditioning or not against the lot size will also get affected.

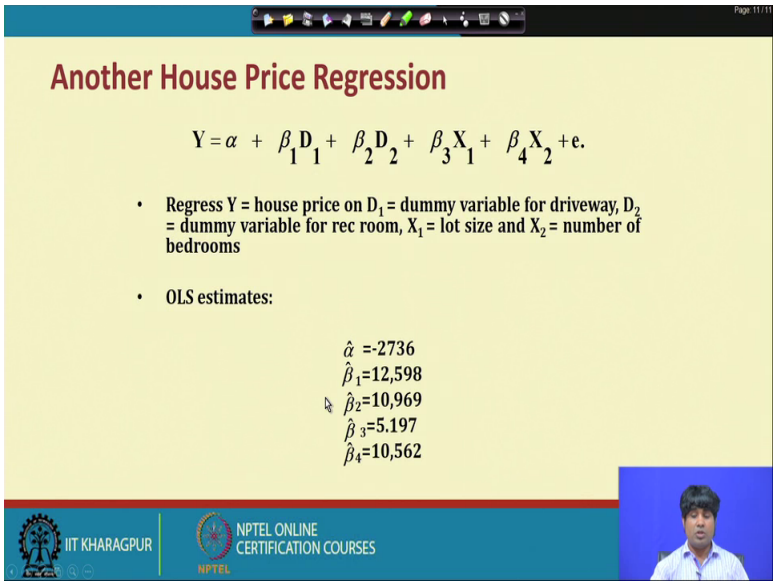
So, depending upon the lot size so, the housing price will get affected right. So, this is how the kind of you know structures, which you can have in the help of you know dummy modelling.

(Refer Slide Time: 30:35)



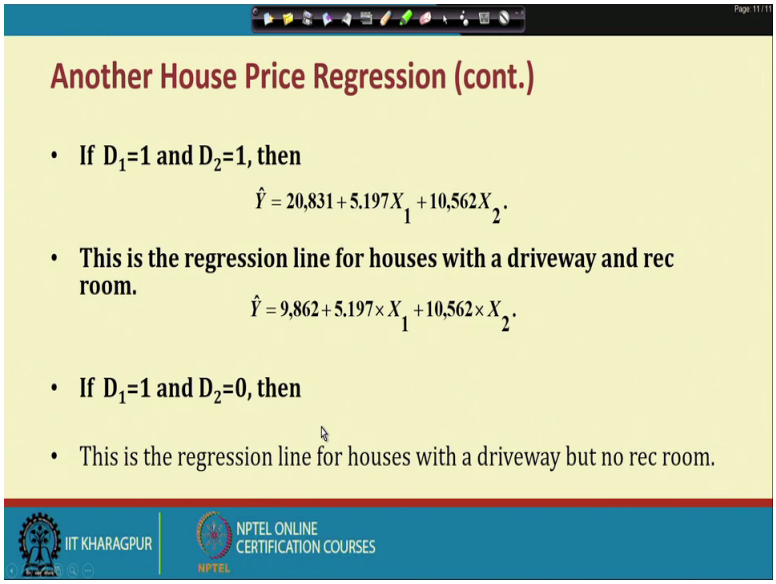
So; that means, a you know while you know addressing the you know or predicting the housing price. So, what is more important is that you know, you have to find out to various variables which can affect the housing price.

(Refer Slide Time: 30:49)



Both you know both you can say you know quantitative variables and qualitative variables, and then you can you know predict as per the particular requirement. So, in this case we have a little bit extension. So, 2 dummies, one is dummy for you know driveway another dummy is having a recreation facility; and as a result. So, if all dummies will be yes then the model alpha beta 1 plus beta 2 then beta 3 X 1 beta 4 X 2 where X 1 is the lost size and X 2 is the number of bedrooms so; that means, when you have more number of bedrooms the price spec will be high, then less number of bedrooms the price spec will be low. So, likewise we have a different price break which can you know represent the housing price.

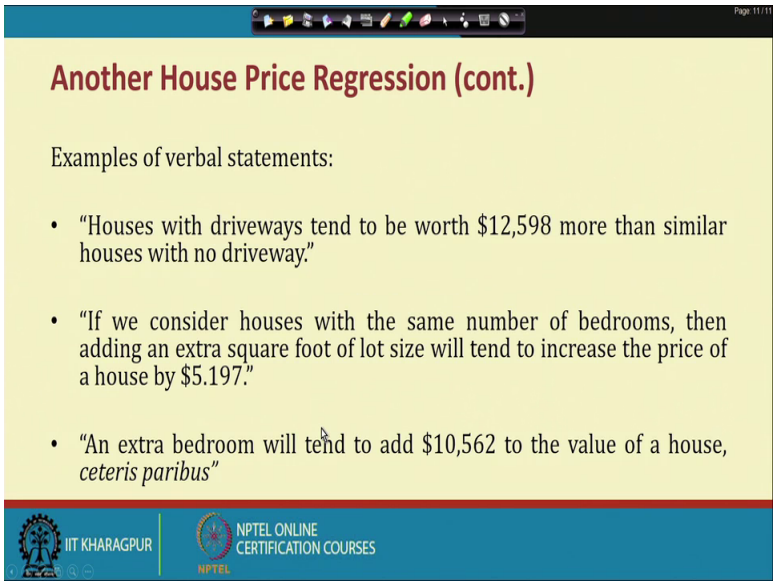
(Refer Slide Time: 31:42)



So, likewise so, if it is d equal to D 1 equal to 1 and D 2 equal to 1.

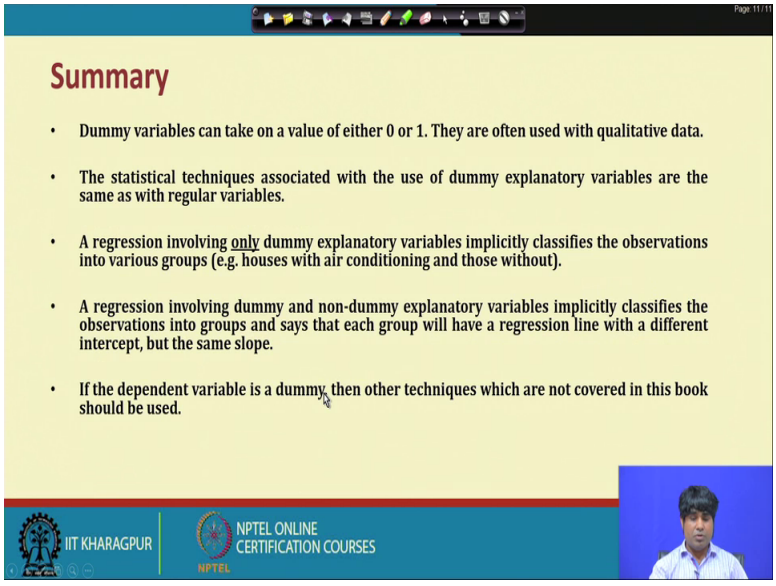
So, this is the price specs and if D 1 equal to 0 and D 2 equal to 1, then this another price spec and D 1 equal to 1 D 2 equal to 0 then another price specs and both will be 0 and there will be another price spec.

(Refer Slide Time: 32:02)



So; that means, by using different dummies. So, we have a different price spec, and that is how the beauty of this particular you know technique to give little bit you know flexibility or even a kind of in diversity while you know credit in the housing price and that too for attracting the builders or the kind of you know customers. So, that means at the end you can get to know that you know the dummy modelling is a one kind of you know.

(Refer Slide Time: 32:34)



You know beautiful structure through which different kind of you know engineering problems can be analyzed you know mostly you know very useful for this you know civil engineering problems and similarly other engineering problems can be used. For instance a product with you know different chemical features and then you have a different colors, then you can use a different kind of you know dummy, and can fix the price as per the particular you know requirement.

So, likewise you know, you have to bring a particular problem, and then you check whether dummy variable can be used to analyze this process this situations and a as per the kind of you know engineering requirement and that too as per the management you know addition is concerned.

So, also or in this lectures, we have specifically highlighted the you know dummy molding that to the structure of you know dummy independent molding and by the way we have already discussed couple of problems related to the main dependent modeling. And now we like to see the other side of the games where the dependent variable is the dummy types in that, case we have a 2 different, 3 different models altogether linear probability models logit models and probit model which we discuss in detail in the next class this will be stopped here.

Thank you very much.