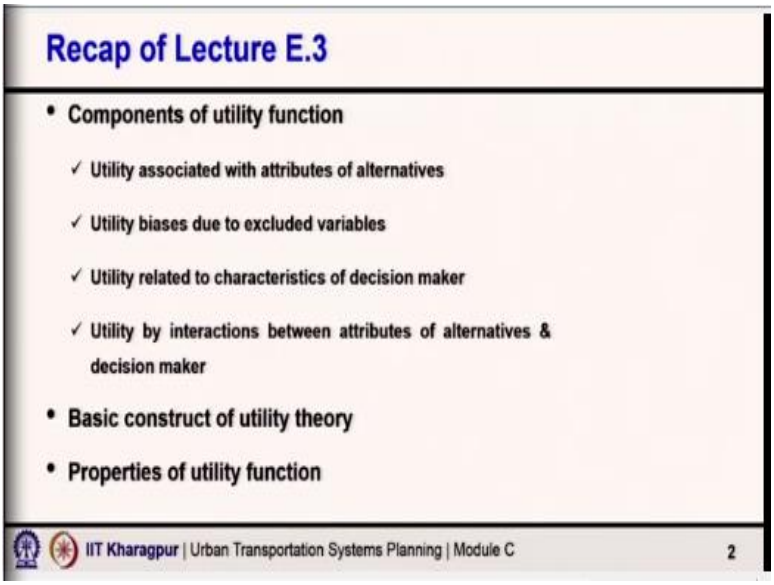
Urban Transportation Systems Planning
Prof. Bhargab Maitra
Department of Civil Engineering
Indian Institute of Technology-Kharagpur

Lecture-34
Disaggregate Mode Choice Models-III

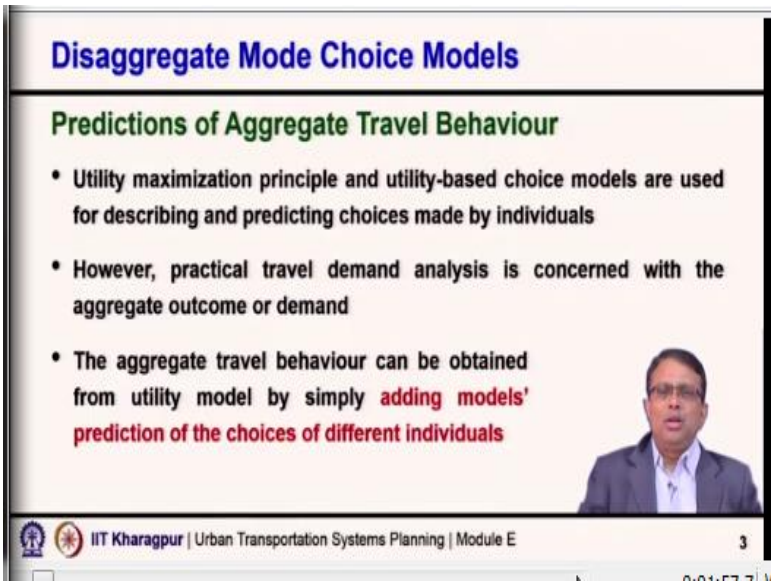
Welcome to module E, lecture 4. In this lecture we shall continue our discussion about disaggregate more choice models.

(Refer Slide Time: 00:23)



In lecture 3, we discussed about various components of utility function namely one part we said that one part would be attributes of alternatives. Then one part would be also the attributes of the decision maker, then the interaction part attributes of alternatives and the socioeconomic characteristics of the decision maker. And then one more component also we mentioned that was the effect of excluded variable or what we said as use of alternate specific constant that the variables which we are not included how to consider their effect in the overall utility calculation. Then we discussed about the basic construct of utility theory and said to you the important properties of utility function.

(Refer Slide Time: 01:32)



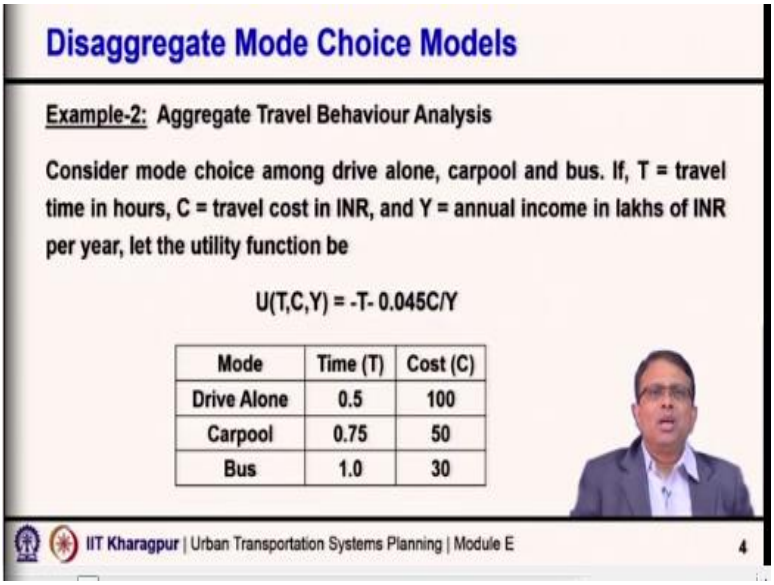
So, with this today we shall discuss first about the prediction of aggregate behaviour using disaggregate mode choice models. Why we are saying this? Because utility maximization principle and utility based choice models are used for describing and predicting choices made by individual. So, these are disaggregate model, so we are actually predicting the choice of an individual, given the characteristics of an individual, characteristics of the alternatives which are available.

There attributes, the levels of that attributes at the current or prevailing condition, then what would be the choice in terms of or considering the utility maximization principle. But remember that practical travel demand analysis is concerned with the aggregate outcome or the demand. We want to know finally end of the day from a zone whatever total trips are or total trips will be produced how many of those trips would use car? How many of those trips will use bus, metro, rail and other available modes of transport?

Because zonal level or aggregate level finally we are interested to know how the split will happen. We came to disaggregate level because of certain limitations of the aggregate models and because of certain advantages of the disaggregate models over the aggregate models. But end of the day we need the assessment or the prediction again at the aggregate level. So, the aggregate travel behaviour can be obtained from the utility model that means this disaggregate model by simply adding model predictions of the choices of different individuals.

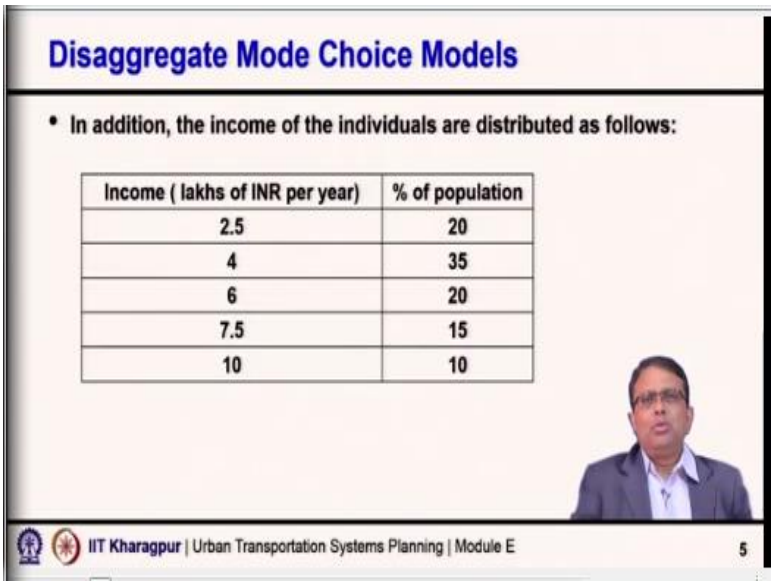
That means I am saying that if there are say 5 categories of people in a given context. Then take each category, so every in one individual representing each category. And for that individual belonging to a particular category what will be the choice. Similarly for each of these groups we can take an individual from each group, and then try to see then what will be the choice for the mode. And then that way we can again see that at aggregate level what will be then the prediction.

(Refer Slide Time: 04:28)



Let us take this example, consider mode choice among drive alone as I said that using private vehicle alone. Secondly, the carpool where a particular vehicle is shared by 2, 3 or 4 people, known people and among known people we are sharing it. And then the third option is use of bus and the utility is expressed in terms of as I have shown -T, -0.045 C by Y, here T is the travel time by each mode, C is the cost or the fare of using that particular mode, and Y is the annual income in lakhs in Indian rupees. So, here suppose the values are given by drive alone by carpool and by bus the time and costs are given.

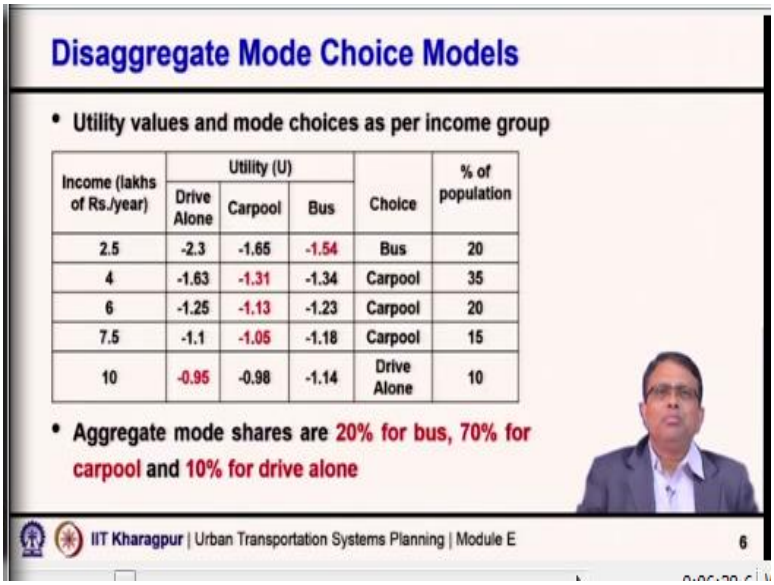
(Refer Slide Time: 05:23)



Now we are considering let us say about a zone and where the income distribution is something as given in the table say 2.5 lakh per year INR is the income. Let us say for 20% of people, then 4 lakh 35% people, 6 lakhs 20% people, 7 and half lakhs 15% people, and 10 lakhs for 10% people. So, it is a heterogeneous society, so not everybody will have same income. But we know what is the income, average income or what is the income of different groups?

And overall populations, how that is distributed to different group in terms of their share? So, given this how to predict then the aggregate behaviour?

(Refer Slide Time: 06:27)



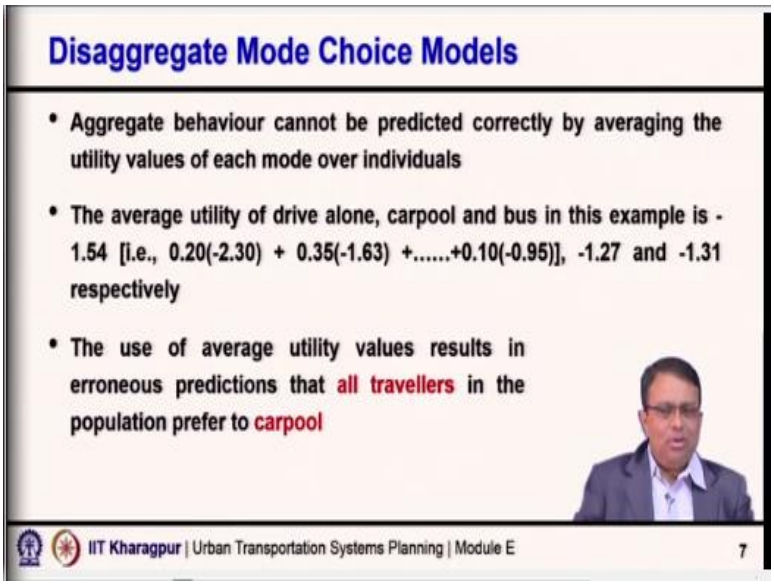
So, it is again simple, we take each group consider an individual in that group, and then taking the travel time, travel cost and the income in that particular group we calculate the utility. For example 2.5 lakhs per annum INR income, for that group, for by person in that group the utility of drive alone is -2.3, utility of carpool is -1.65 and utility of using bus is -1.54. So, we calculate the utility and as per utility maximization although in this context all are disutility.

So, as per the utility maximization principle the least negative value is for bus or we can say utility is maximized by use of bus. So, for this person or for people belonging to this particular income group, choice is going to be the bus. And we know that 20% of the overall population who are belonging to this particular income group will then use bus. Similarly each income category we can take, and we can take one person from that income category.

Try to calculate the utility, and try to say what will be the choice of mode for that particular group. Similarly you can 3, 4 lakh income group the choice is likely to be carpool, for 6 lakh per annum income group the choice is again going to be carpool, 7 and half lakhs again going to be carpool. But for 10 lakhs the utility maximization principle indicate that, that the choice will be drive alone.

So, each taking each group, choice and the representation of in that group in the overall population. We can then easily say that whatever is the overall population in that particular zone 20% of them would use bus 35 + 20 + 15, so 70% will use carpool, and the remaining 10% will use drive alone as their choice of mode. So, we are able to actually still give the aggregate choice ok, how the choice will be distributed among different modes that we are able to predict. So, this way although we are using a disaggregate model but we are able to predict the aggregate behaviour.

(Refer Slide Time: 09:29)



But interestingly you may note that the average aggregate behaviour cannot be predicted correctly by averaging the utility value of each mode over individuals. That means, I know say for example that consider the mode drive alone. The utility for drive alone is -2.3, for income group 2.5 and the representation is 20%. So, taking -2.3 into 0.2+ for the next income group utility for drive alone is 1.63 so -1.63 into 0.35.

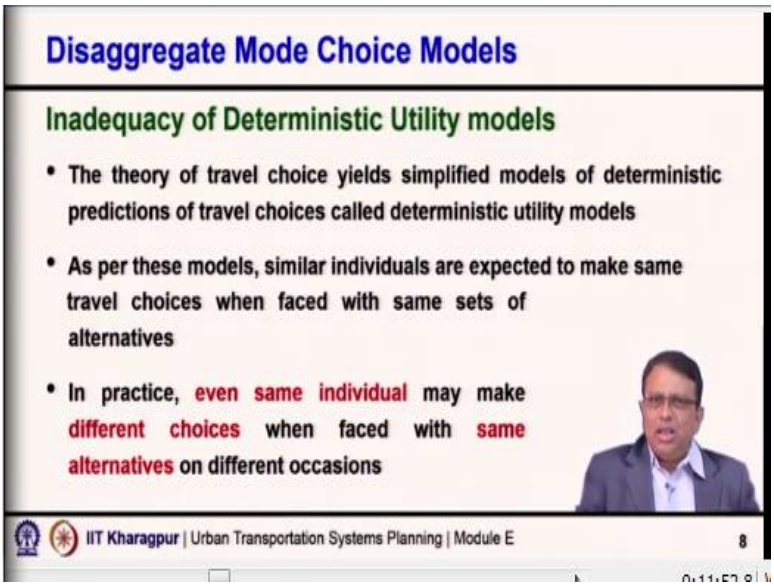
Like that if I calculate the utility of drive alone considering all segments what is the average utility? I will find the value is maybe -1.54 similarly the average utility of carpool across all segments will be -1.27 and the bus it will be -1.31 respectively. Then as per this average value, then we would say everybody would use carpool because according to utility maximization principle.

That is the value which will be, that is the mode that will be selected, but then that will be incorrect. So, we cannot do that, we cannot really calculate the average utility value of a particular mode or a particular option across different segments, that we cannot do. So, you have to consider each segment their choice and the how much percentage of overall population is included in that segment.

So, for that segment this is the choice, take another segment this is the choice, take a third segment this is the choice. So, like that you have to take as I have shown in this particular slide.

So, we cannot really take simply the average of utility of a particular option across different segments.

(Refer Slide Time: 11:46)



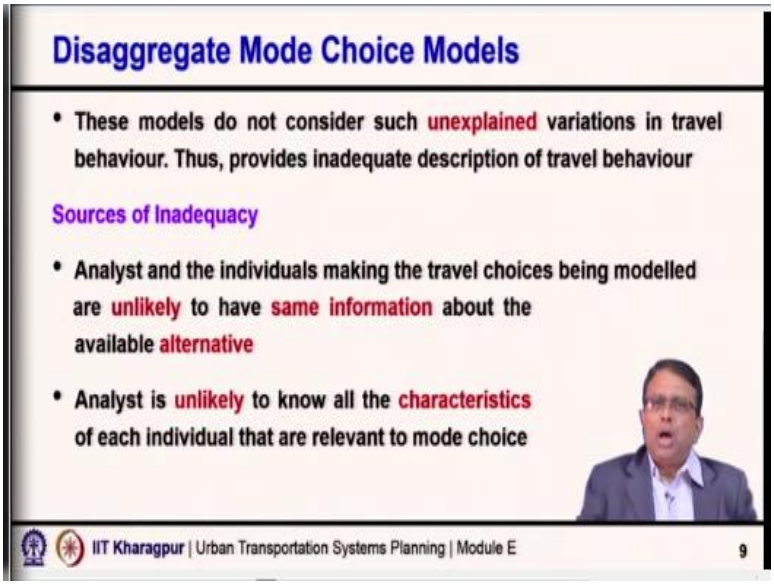
Now there are theoretically it is sound, we have clearly said you how the utility maximization principle works. But these are this kind of models we say deterministic utility model, why it is deterministic? Because we said that the maximum utility that option would be selected, that means that option will be selected and other options are not going to be selected. There is no uncertainty, no other error; nothing else is there; so it is deterministic.

So, that means everybody who have that income and who are facing with the similar choices will make the same preference or same choice, the outcome will be same. So, that is what we said the theory of travel choice yields simplified models of deterministic prediction of travel choices called deterministic utility model, deterministic because there is no uncertainty. We are saying the one with the maximum disutility will be selected that is all, so it is deterministic.

So, as per this model similar individuals are expected to make the same travel choice when faced with same sets of alternatives. Because let us take the example what we discuss now, all what we are taking is the income of a person. So, anybody, any individual facing the same 3 choices and having same income, the choices should not be different they are all then should select the same option.

But in practice even same individuals may take different choices when faced with same alternatives and different occasions. So, different occasions they may say different thing, even same occasion face different individuals with the same income level may also choose different mode or make different choices. Now that is what is found and that is called the inadequacy of the deterministic utility model.

(Refer Slide Time: 14:22)



So, why it happens? Because these models do not consider such unexplained variations, because we cannot explain then, why? Because utility is same, utility is maximized through that choices then why others or some others are using some other options, choices are why going to be different? These kinds of models are unable to consider such unexplained variation in travel behaviour thus provide inadequate description of travel behaviour in that sense.

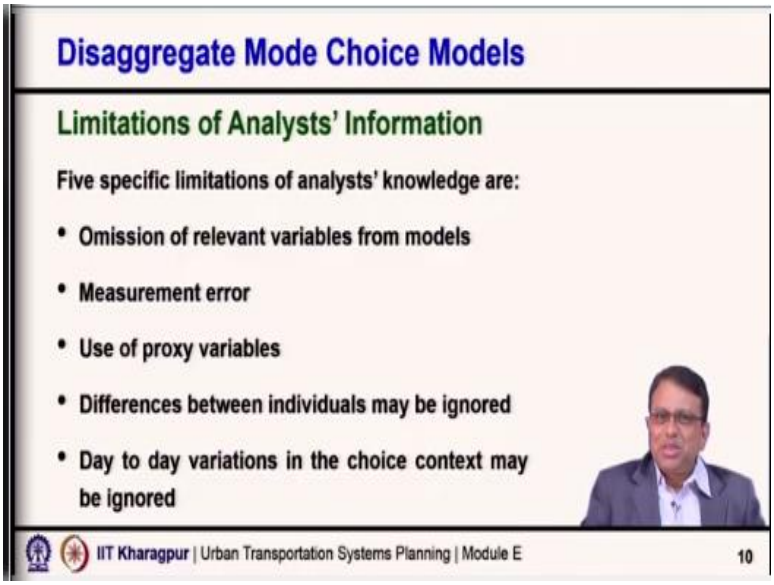
Now what are the sources of inadequacy, why it happens? Analysts and individuals making travel choices being modeled are unlikely to have same information about the available alternatives. The information available to all the travelers let us see if you consider in a city context, the information available to all traveler is not really same. So, information itself the level of information available maybe different that is one.

The second is analyst is unlikely to know all the characteristics of it individual that means the choice of mode, in this particular case it is mode, choice of mode. So, in the choice of mode may depend not only on income but maybe so many other variables as well, so many other attributes also. Human being is not so simple that just I take 1 or 2 characteristics income or age or gender and then everybody having the same income and same age will behave in the similar manner that looks very strange.

If you consider human being they cannot be represented in a so simple manner**.** So, analysts are unlikely to know all the characteristics of each individual that are relevant to mode choice. So, one is it could be limitations from the point of view of analyst or from the perspective of the analyst also all the individuals being modeled are unlikely to have same information about the available alternatives, both might be true.

And that is the reason why the actual choice may be different than what we are predicting from using this deterministic utility model. It does not contradict the utility maximization principle but it is because not everybody has the same information and nor we are considering all the variables which are relevant. So, actually whatever we are saying as deterministic utility, the true utility may be somewhat different from that, that is the reason.

(Refer Slide Time: 17:49)



Now let us then discuss in details, because this is an very, very important area. This can make the deterministic utility model the prediction from the deterministic utility model wrong, so we must know. Then what are the limitations of analyst information? I said 2 things, one is the perception or the information same information may not be there to all the traveler who are making the choice.

And second essential I say it is the limitation of the analyst, I am developing a model I may not have all the information that or all the relevant variables I may not be able to consider which are likely to influence the choice behaviour. So, here I would like to highlight five specific limitations of analyst's knowledge which cause all these problems first omission of relevant variables from the model.

For some reason or other you did not consider a variable in your model but you know that, that is a relevant variable. There could be so many reasons, maybe I want to keep my model simple, that is again a point. Because later on in not in this course but probably when you understand how the models are developed you know that we really cannot probably so easily consider every variable we want also we may not have the data also.

Some cases considering those variables even may not be relevant also for me. So, there could be so many reasons, why? A relevant variables may be dropped or may be omitted, so naturally that may bring this error. There could be measurement error, because say zonal level we are saying zone centroid to zone centroid what is the travel time? Now you just imagine you stay in a zone, for every person from every house is the travel time going to be same, no, that is cannot be, so there is an approximation.

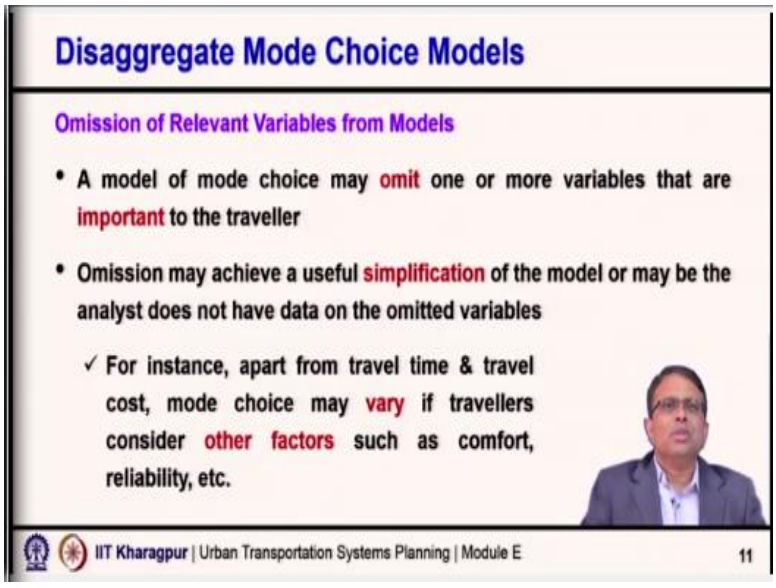
And every data you take, every measurement you take such kind of measurement, every measurement you take and specially such kind of measurements you will probably have some level of measurement error, so that may be the reason. Use of proxy variable, what is proxy variable? I want to use a particular variable but that data probably is not available. So, I use another variable which indirectly represent that variable, that is the proxy variable.

So, use of proxy variable could be also another reason, then difference between individuals may be ignored**.** Everybody the kind of even the same characteristics individuals are not same, let us accept that whatever characteristic you say the people with same height, same weight, same education, same income are they then same, still they are not same, there will be variability. Say also it could be day to day variations of the choice context.

My choice of mode may be influenced by so many factors, maybe you are carrying lot of money, so that day you do not want to use bus, you want to use probably taxi or your private vehicle. Or you travel by private vehicle but your family wants to go to a family function or for some work. So, you leave the vehicle that day and maybe you are travelling by taxi or travelling by bus. Such kind of day to day variation, or simply one day maybe somebody is not feeling well, simply may not be very something very serious, but simply not well not feeling that energetic.

So, maybe he decides ok let me travel by taxis today instead of taking the crowded train or crowded bus, such kind of day to day variations also. So, all these are reality but analyst really it is not possible for analyst to correct all these problems**.**

(Refer Slide Time: 22:22)



Let us take little bit more discussion about each of this. First coming to omission of relevant variables from the model, a model of mode choice may omit one or more variables that are important to traveler. So, I mentioned this grossly, but little bit more discussion we want to make

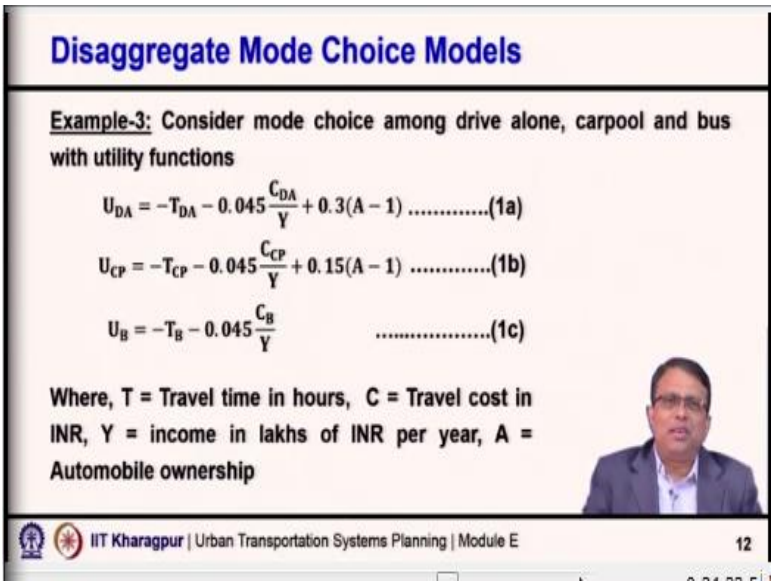
and take an example to explain it further. So, omissions may be done because maybe we want to simplify the model.

Or maybe analyst does not have the data of the omitted variable, I know that is important. But I cannot consider a variable unless I have the data. Or I can collect some data from individual but it is to be applied on the secondary data**.** So, I can ask some 1000 people and I can I probably develop a model, but how I will apply because, that data is not available for the whole population in that area.

So, applicability of the model because the model is not just for developing only for the sake of developing but ultimately the model is to be used. So, I need to only consider those variables where the data is available now, and that data will also be available for the prediction purpose. So, I gave in at example here, for instance say apart from travel time and travel cost mode choice may vary if travelers consider other factors may be comfort, convenience, reliability and so many other things.

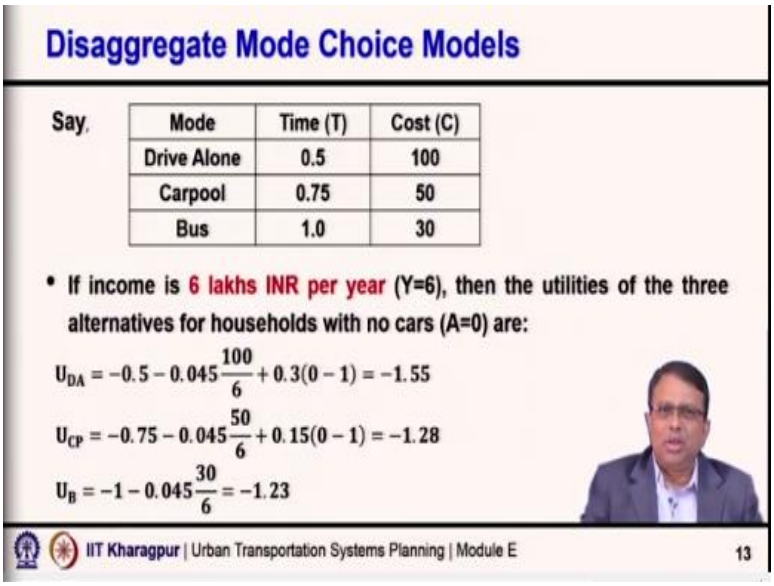
So, all these data I may not really have. So, sometimes maybe quantifying it is also a challenge, of course there are you can very well consider qualitative attributes, that is not a problem in travel behaviour model. But there are this issue is a very, very pertinent issue, omission of variables.

(Refer Slide Time: 24:32)



So, let us take an example here, I have taken an example where there are 3 again alternative options the options I did not change throughout all the lectures. Always we are talking about drive alone, carpool and bus just to keep the context very, very easy and you can compare across different example. Now let us see this 3 utility equation I have given, here you can see we are considering travel time, we are considering travel cost, we are considering income and we are also considering car ownership data, all these are considered.

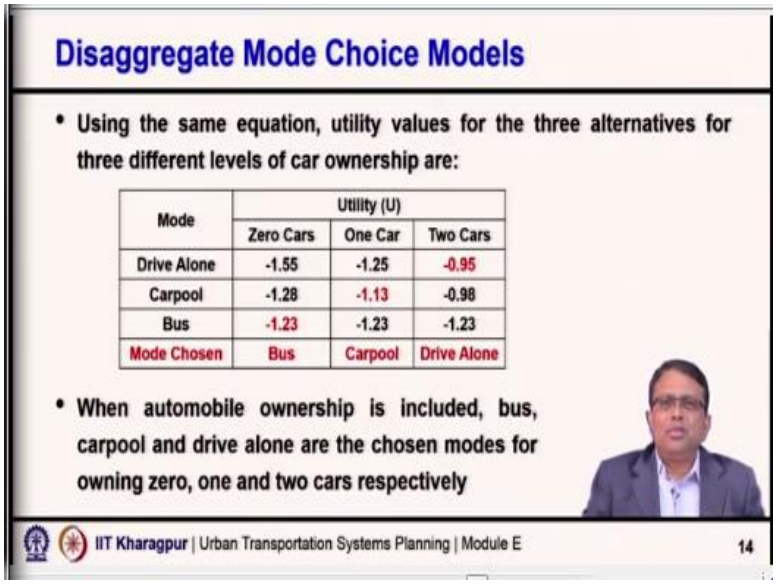
(Refer Slide Time: 25:21)



Then as usual if I know that travel time, travel cost and for this drive alone, carpool and bus what are the values? Those are also known. And let us consider a person whose income is 6 lakhs rupees INR per year, so then I can very well calculate the utility. So, utility for drive alone is

calculated as -1.55, utility of carpool is calculated as -1.28 and utility of bus is calculated as 1.23. And 3 alternatives and this is for all people where the A is 0, that means car ownership is 0, automobile availability is 0.

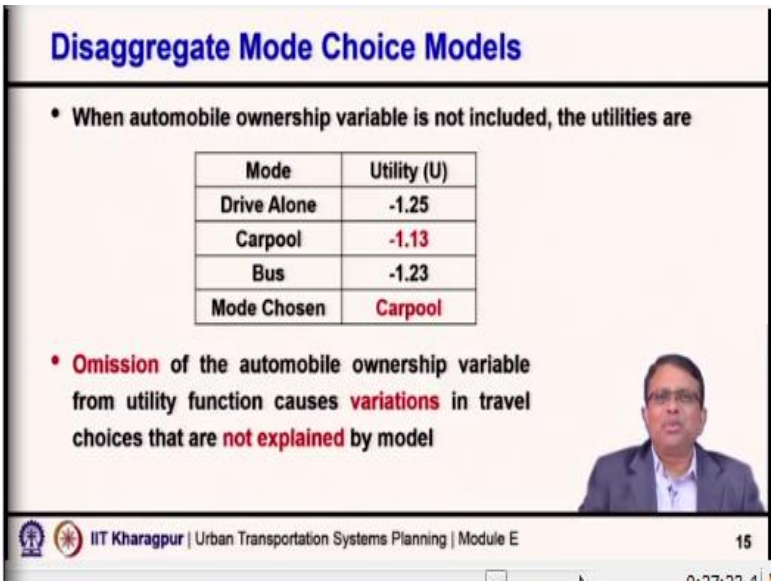
(Refer Slide Time: 26:14)



Like that similar way, I can calculate it for a person who is having one car and also for a person who is having 2 cars. Look at this table, the remaining thing is very simple in terms of calculation. So, we know for 0 car, that means so those who do not have car that kind of individual is going to choose bus. One car, the person is going to choose carpool and who is having 2 car, the person is going to choose drive alone.

So, the choice of mode actually depends in this case as you see it depends on the car ownership**.** Now just consider if I do not include this car ownership in the model, so what happens?

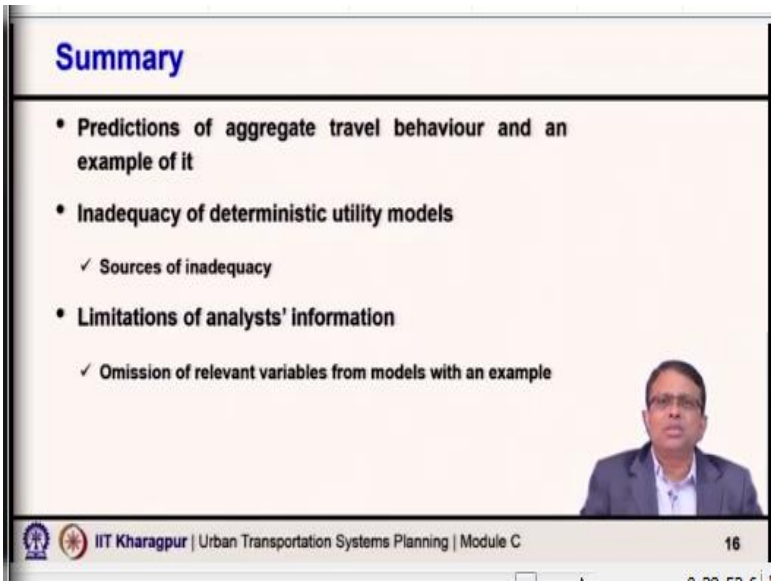
(Refer Slide Time: 27:15)



I will simply calculate my utility like this, drive alone -1.25, carpool -1.13, bus -1.23. So, I will say everybody who is having income of this and facing options, 3 options with this travel time, travel cost, all of them are going to use carpool. So, both cases we are using utility maximization principle, principle is not going to be different. But because in one case we consider car ownership we find the choice of mode for 0 car, 1 car, 2 cars are going to be different.

When we do not consider car ownership everybody will be predicted to use only carpool. So, this omission of variable may cause erroneous prediction. Or I would say if I do not consider this variable or if I omit this variable my model will predict that everybody will use carpool. But in reality some will use carpool, some will use drive alone, some will use bus as well. So, my error predictions will be wrong in that way, why this happened? This happened because of this omission of variables, relevant variables.

(Refer Slide Time: 28:44)



So, what we discussed in this lecture? We told you how using this aggregate model, you can even predict the aggregate travel behaviour, taking different segments, characteristics of each segment. take a person there. Think of a person who is representing that segment, calculate the utility and accordingly say what is going to be the choice? So, for each segment you can separately calculate the choice and you know that in the overall population how, what is the contribution of each segment and their respective choices?

So, you can say then overall 100% population how much percentage will use option a of how much percentage will use, option b and so on. So, aggregate prediction can be done. Then we said grossly what are the mature inadequacy of the deterministic utility model? We discussed several sources of inadequacy, first to broad thing that every individual making a decision may not have the same level of information that is one part.

The second is coming to the analysts like us who develop the model, we have our own limitations. Then we started discussing what are the limitations of analyst information we discussed again mention various items, various aspects. And then took one of them omission of relevant variables and explain clearly with an example, how due to the omission of relevant variable the model choice.

In one case where the variable is not excluded, and the other case where variable is omitted or excluded, how the choices could be different and therefore apparently how due to omission of variables the choice could be erroneous. So, we shall continue in the next lecture, thank you so much.