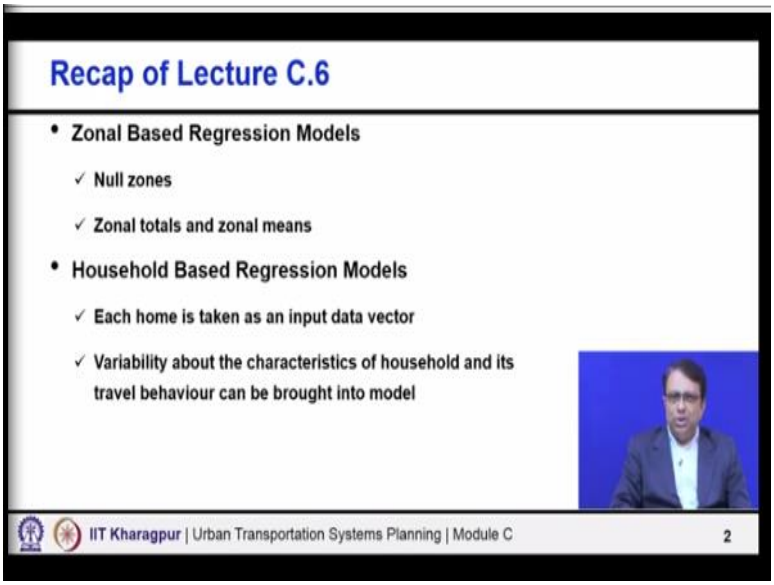
Urban Transportation Systems Planning
Prof. Bhargab Maitra
Department of Civil Engineering
Indian Institute of Technology-Kharagpur

Lecture-17
Cross Classification Analysis: Model Structure and Calibration

Module C lecture 7.

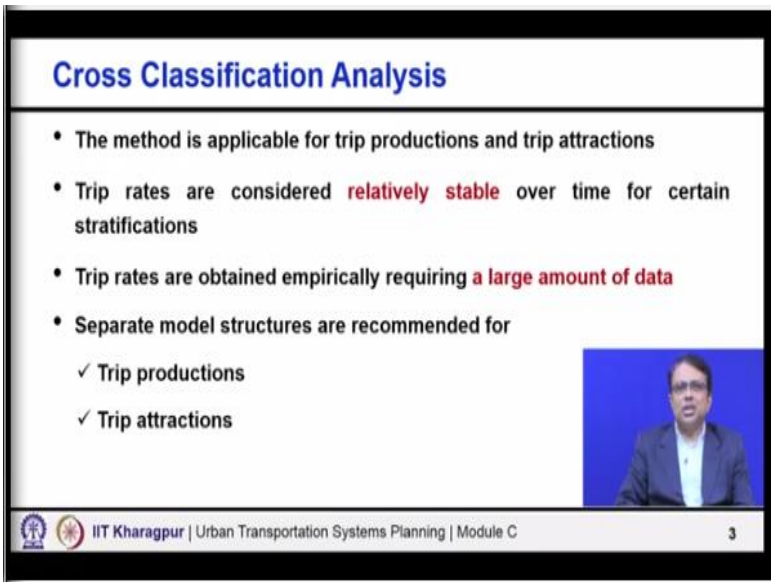
(Refer Slide Time: 00:20)



Up to lecture 6, we discussed about regression models, how to develop regression models? How to select variables? Then how to do the model specifications? Then calibration, then application, then also we discussed what are the characteristics of zonal based regression models and indicated that how the zone size influence?

And how the variable selection, also can influence model zonal total versus zonal means based on which zonal size, zone size of the zone inter-zonal variation, intra-zonal variations. And then indicating the possibility for developing household based regression models. Now with all these we generally completed our discussion about the regression based model. Today, we shall talk about cross classification or category analysis.

(Refer Slide Time: 01:29)



So, our topic for discussion in today's class is cross classification analysis. This method is applicable for trip productions and trip attractions both, like as we have seen that we can develop regression based models for trip productions and trip attractions both. Similarly this cross classification analysis or the category analysis can also be carried out separately for trip production and trip attractions.

The basic fundamental or a most vital consideration in cross classification analysis is mentioned in the second point. Trip rates are considered relatively stable over time for certain stratification, what I mean by stratification? Stratification means, you can divide households based on certain characteristics say income, say car ownership, say number of persons in the household. Considering all these we can create many groups of household, several groups, these groups also are called category, this category are groups.

So, every group is defined whatever factors we are considering to define those groups, say we are considering household income and car ownership. So, then based on household income and based on car ownership we have stratified households. So, each household category we are taking and then we are trying to analyze the trip rates, what we are saying here that the trip rates are relatively stable over time for certain stratifications.

That means, each category of households that trip rates are generally found to be stable over time. There are certain interesting conclusions or inferences what we can draw also from this one. So, over period, let us say over period a household may have higher income, may have car ownership, maybe now zero car is there, or no car private vehicle is owned by the household. But over time the income may be more, the car ownership maybe there, maybe 1 car or 1 car household to 2 car household.

But then these households will start behaving like that category, so if your household gets shifted from one category to another category over time, then they start behaving like people of that category. In society also this is very interesting probably different categories of household starts behaving in very different manner but their behaviour remains stable once you think of a particular category of household.

So, that is what is the basic and fundamental and very important assumption, the trip rates are considered relatively stable over time for certain stratification of households. Now we obtain the trip rates empirically which require obviously large amount of data. That means if you have say consider 3 household income range and 3 household car ownership range maybe 0, 1, 2 and more, and then income suitably classified into 3 groups or 3 ranges.

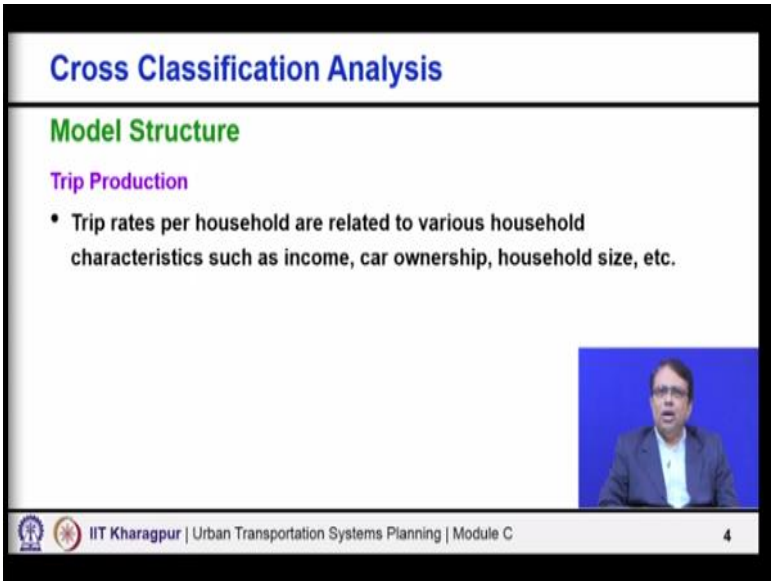
So, you have actually created is 3 by 3, 9 cells. So, this each of the cells is a category, so I have nine category I have created. Now for each of this category, we have to then calculate the trip rate, how do we can calculate? We can simply collect lot of data from different households, and we have to make each sale representative. That means, certain minimum number of households we have to actually get in that category.

So, that statistically the data or the average what your calculate is acceptable, and then from the given data empirically we are calculating the rate, so the calculation part is very simple. So, we have the rates and the second bullet point what I again would like to repeat. The trip rates are considered relatively stable over set time for certain stratifications. Rich people always behave in one manner, poor people their travel behaviour is different, middle class middle income range, they always behave in the similar manner.

So, tomorrow if some middle income household becomes high income household, they will start behaving the way the high income households behaved. So, back short is the kind of assumptions, this is very, very important, then what do we want? We want separate model structures for trip productions and trip attractions, why? You have seen even in for regression based models, that the factors affecting trip productions and trip attractions are very different.

One is basically trip production, means it is all produced at mostly at household end and attractions is at the other end I mean in fact the household is not there. So, depends on the commercial, retail, employment all these factors will govern. And obviously when we are creating these stratifications or the categories we are creating on the basis of what. Basis of those variables which are known to influence that trip productions and trip attractions respectively those are the variables which are used as a basis for creating the stratifications or creating this categories.

(Refer Slide Time: 09:07)



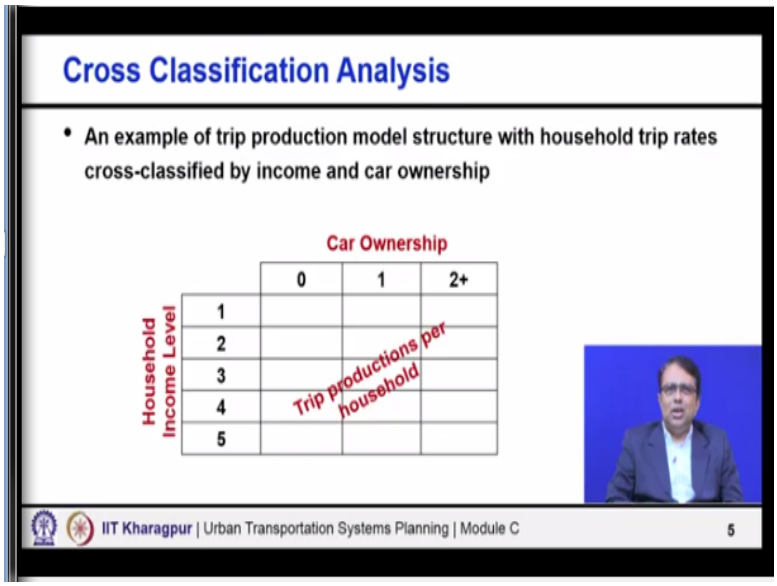
So, next let us discuss the model structure. Now as I said the model structure we need separate model structure for trip productions and trip attractions. So, first let us explain the trip production model structure. Here, trip rates per household are related to various household characteristics such as income, car ownership, household size, and all these variables which are known from research from past studies, which are known to influence the household trip rates.

So, even if I would have developed a theory regression model for modeling the household trip production whatever variables I would have considered on the basis of some or all of those variables only we shall use the or create the stratification. So, the variables are very, very similar, because we are assuming that say if we are considering income. So, that means our basic assumption is that household trip rate is likely to vary based on income.

That is why we are stratifying it classifying household based on income, same thing is household trip rate is likely to be influenced by the car ownership. So, we are actually taking car ownership and classifying households based on car ownership. So, the factors what you have studied earlier in the initial lectures of the trip generation analysis, same factors which you know are likely to influence the trip production by households.

Based on those factors, we classify the households and create several groups. If you are taking say 2 variables and classify, then you have maybe a matrix a two dimensional matrix, if you add a third dimension, you may add also 4th dimension. So, it maybe multiple dimensions may increase depending on how many variables you are considering. Let us take an example here.

(Refer Slide Time: 11:53)



We are giving you an example here which is a simple one. This is an example of production model structures with household trip rates, cross classified by income and car ownership. The 2

well established well known variables which are known to be known to influence the trip production rates by households. And in this one, we have considered 5 household income levels, how typically we do it?

Say for Indian condition, I can do it 0 to 20,000 20 to 40, 40 to 60, 60 to 80 and maybe 80,000 plus monthly income**.** These are only examples you can consider even 0 to 25,000 25 to 50, 55 to 75, 50 to 75, 75 to 100 and then more than a lakh per month, maybe those categories you can consider. But based on income and certain range of income, we are creating this class. Similarly the car ownership as per prevailing urban situation in Indian context, we can classify maybe car ownership zero, that means they do not have private vehicles in the household.

Then only one car, 2 and more car because we will not really have many households where you will have 3 cars, 4 cars. So, if you create separate groups, you may not be able to get good number of households under that category, so that you can calculate a trip rate which is statistically acceptable. So, you may keep it like 2+, remember that interestingly these groups we create always the last group is kind of open ended one lakh plus income or 2 and more cars.

Even if suppose I would have considered that family members in the household, I would have taken maybe up to 2 then 2 to 4. And then maybe take greater than 4 example only or if you want more classification maybe 4 categories, then you would have considered something in that range only, maybe up to 2%, then up to 4, maybe up to 6 and then mean saying greater than 6.

So, generally the last group what we create is we keep it open ended. And how many groups you should create? There is no formula, it is your logic and your understanding about the city context and the given context. As I said that we want to create group, remember that creating group is not the end of the work, it is the beginning of the work. After this you have to collect data for each category like each box represents a category.

So, you have to collect data for each category and you should get certain minimum acceptable sample size. So that you can say, I have taken at least minimum this many sample, say, I am just

saying for example, that at least maybe I need some 50 households. I cannot get 2 households and calculate a trip rate, that will not be acceptable.

Because a third household gets in you will probably find the rate is very different. So, we want a rate which is really representative and which is almost like stable, even if you collect little bit more data from more number of household, still you will find the average is remaining there only more or less remaining there. So, that is what that is the way we create the group, how many again gets controlled by that?

That you have to get representative data and we should cover the whole range. So, probably we will find more than one lakh, you may not find or more than 100,000 rupees per month income, that number may not be very high. Some maybe getting also income 500,000 rupees, 200,000 rupees, 300,000 rupees even more but they also need to be represented. So, you cannot exclude anybody I mean any household today I should be able to place it here.

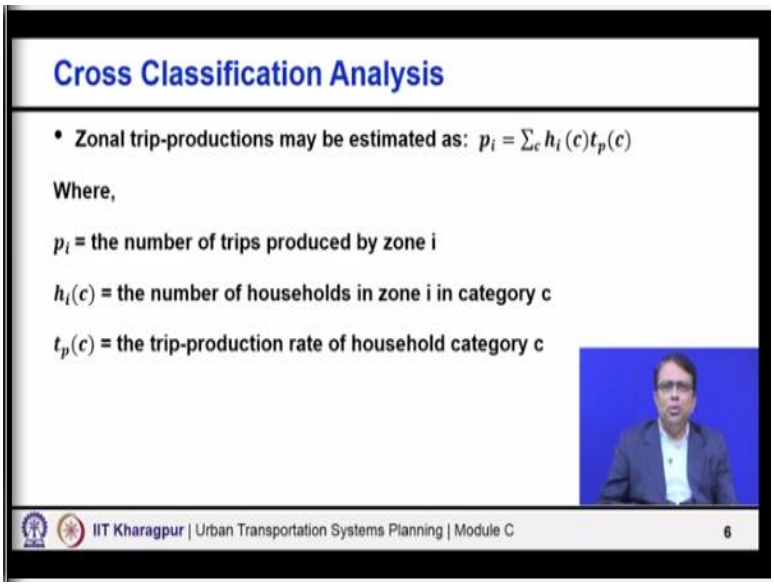
So, that is the reason why we try to keep the last group open ended. But whether we should create further grouping, yes you can create further groups. If you feel that you can get adequate number of households from your study area and you can calculate the trip rates in a proper way an acceptable manner. Also there are other consequences, how many households are categories you are considering that will depend on how much data you have to collect.

So, here you have 5 income categories and 3 car ownership, so 15 categories, so each of these 15 sales I should collect data. Now if you consider one more characteristics, maybe let us say household size, 3 characteristics, then your number of cells will increase drastically. So, for that many cells you have to get the data. So, how many characteristics like household variables to consider? How many stratification you make?

All this will be guided by several factors, how many sales we are getting? Whether we can get adequate data then how much effort I have to put to get collect the data, more number of sales I create, then more will be my effort for in terms of data collection, I must get representative

sample size in order to calculate the rate. Every household must be able to fit somewhere in this group, all such considerations will govern.

(Refer Slide Time: 19:10)

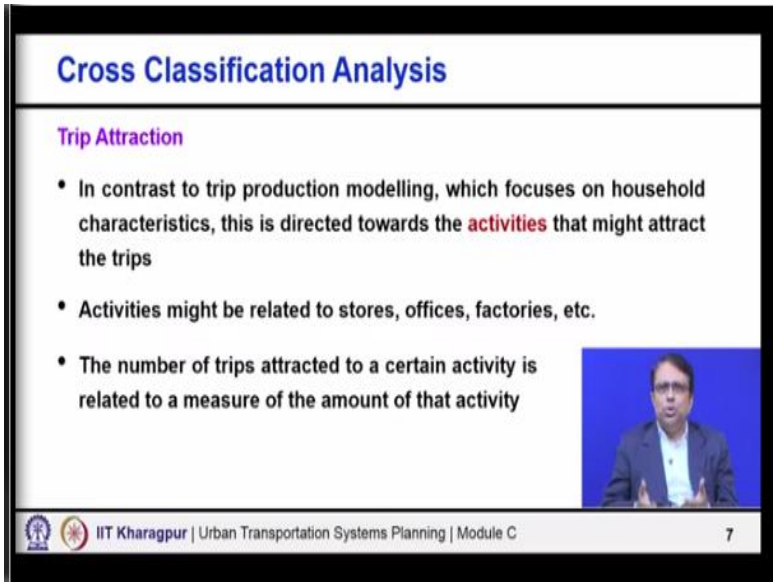


Then what we do? Suppose each is one category, each box represents one category and in this case, then there are 15 categories of households, so how I will get the total from each zone? These rates are like trip production, so t p c represents that trip production rate for household category c. So, you have rates you have established and then you know what are the trip rates? So, for trip rate for category c the trip rate is t p c.

You also know how many households are there in a zone i with in that category? category c. So, simply what I will do then? I will multiply a h i c number of households in category c, in zone i multiplied by the trip production rate of household category c. That keeps me the total trip productions by household category c, summit over all category which are there and that gives you the production from that zone.

So, that way we can actually estimate the zonal trip productions from a zone. Using these categories, for each category the trip rate the corresponding number of households in a zone and then sum over all the categories that will give you the total productions.

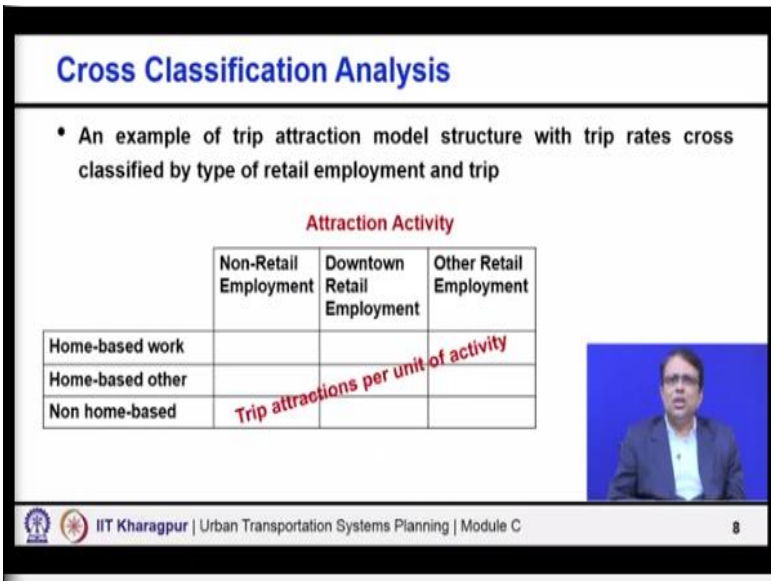
(Refer Slide Time: 21:07)



Now if I go to trip attraction, then in contrast to trip production modeling which focuses on household characteristics, in this case it is directed towards the activities that might attractive. So, what kind of activities will attractive? Maybe stores, offices, factories. So, considering these we have to make the category. So, number of trips attracted to certain activity is related to the measure of the amount of that activity.

That means you calculate the again the rate for certain activity type or certain attraction type, what is the rate multiplied by what is the amount of that activity because the rates will be per unit activity. So, multiplied by the amount of activity and then again sum over all such activities to get that total attraction.

(Refer Slide Time: 22:21)



Let us see the next slide what I have shown it like a table. So, this is an example of trip attraction model structure with trip rates cross classified by type of retail employment and type of trip. For example if you see the attraction activity, we have made it other retail employment, downtime retail employment and non retail employment. So, like that different employment wise I have classified 3 classifications we have made.

And row wise we are considering 3 types of trips you know the trip can be classified in so many ways. I have discussed all these different potential classification of trips, one way we say that home based work, home based other non home based trips, so here let us say we have considered that category. So, I have 3 types of trip saying that home based work, home based other and non home based.

And then different types of employment, here we have considered not just type of retail employment. But we have considered non retail employment, downtown retail employment and other retail employment. This is only an example one can that different attractions factors which again we have discussed earlier, and any of those factors you can consider.

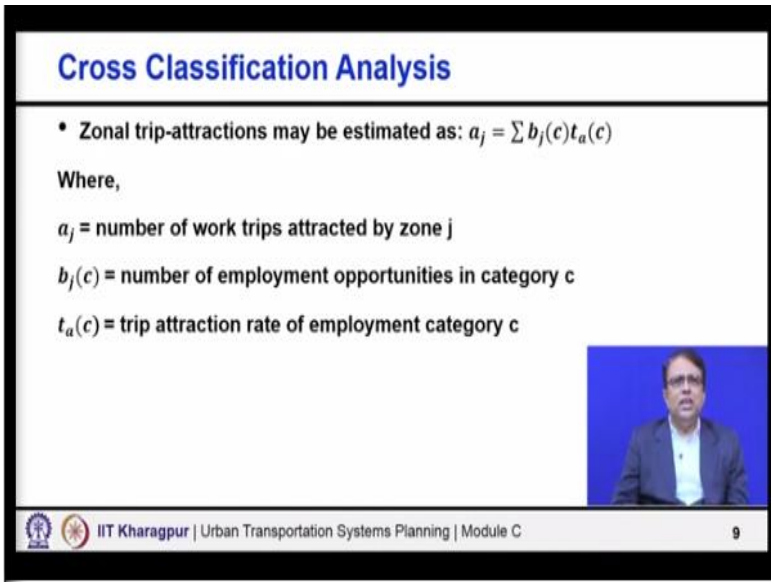
So, 3 by 3, so I have got 9 cells, so again what I will do? I have to take maybe so many establishments which are pertinent to maybe say non retail employment and home based work, non retail employment and home based other, non retail employment non home based trips.

Similarly downtown retail employment home based, downtown work, downtown retail employment, home based others, downtown retail employment which have is non home based and then we will calculate the trip rate.

So each sale with multiple observations, we shall calculate certain rates. Once these rates are calculated then we know that we can also calculate the total attractions. Remember that here also we try to represent different types of landings maybe this is only an example where not all different types of land uses are considered. We have not considered them, just we wanted to show that how we structure the model?

Saying that you take a variable make 3 groups, you take another variable make 3 groups, so you have nine elements. And then you can have one more variable consideration, you may consider one more variable and accordingly decide but the number of cells obviously will increase. Now whatever you do, you have to get adequate amount of data, so that you can calculate the trip rate and accordingly you will try to get the total attractions.

(Refer Slide Time: 26:22)



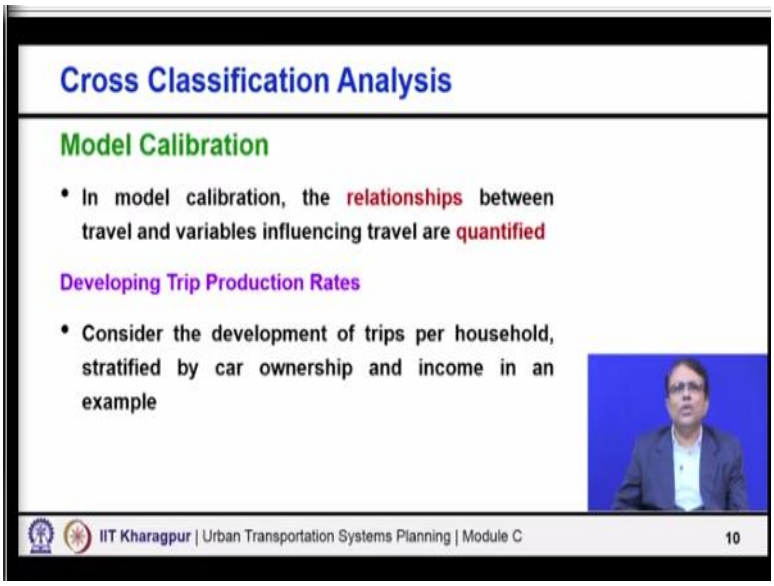
So, here also I said how you will get calculate zonal trip attractions for a zone. Let us say t a c is the trip attraction rate of employment category c. Then you know in a zone how many number of employment opportunities are there in category c? Say simply multiply that number with the

corresponding rate, so that gives you for that category how much is the attraction? Now submit over all categories, that gives you the total attractions.

So, structure more or less is similar, every case we considered the relevant variables, which are likely to influence the productions or the attractions. Then represent the observed range of that variable into certain groups, like that multiple variables we can consider. Then we will have multiple sales or category, then for each category we need good amount of data, so that we can calculate the rate, so like that the rates are established for each category.

And then we know whether it is productions, whether it is attractions in a zone for each category what is the actual number and what is the corresponding rates, what we have already obtained. And then we can get them the total productions or attractions for with respect to that category, similar things we can do it for every category. And add them together to get the zonal productions or zonal attractions, that is what we do actually.

(Refer Slide Time: 28:22)

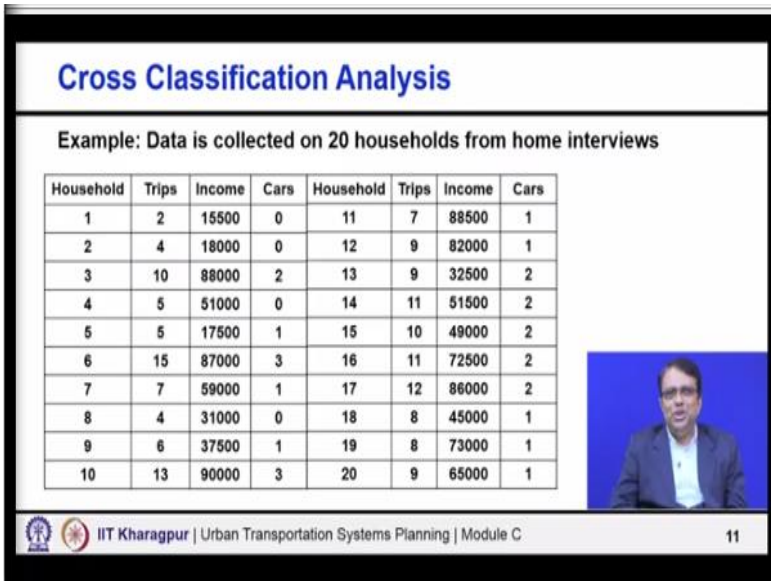


That is I said here, that in model calibration relationship between travel and variables influencing travel are quantified, so exactly the same what I said. For that we actually develop the trip production rates. Let us say consider the development of trip productions say it is stratified by car ownership and income. Let us take an example to understand further, how we develop the trip production rates?

So, this comes once the structure, structure is almost like the model specifications in regression model, so we are specifying the structure here. And once you have specified, the next step is what? Calibration. So, how we calibrate the data? We have to collect a lot of data and then as we develop regression equations or regression models, and then try to taste which model is acceptable and so on.

And look at the constant, look at the sign, and magnitude and all those things, in this case what we will do? We will calculate the trip rate for each category or each sale from the given data, so this part is very, very simple. So, let us take an example as is given here.

(Refer Slide Time: 29:48)



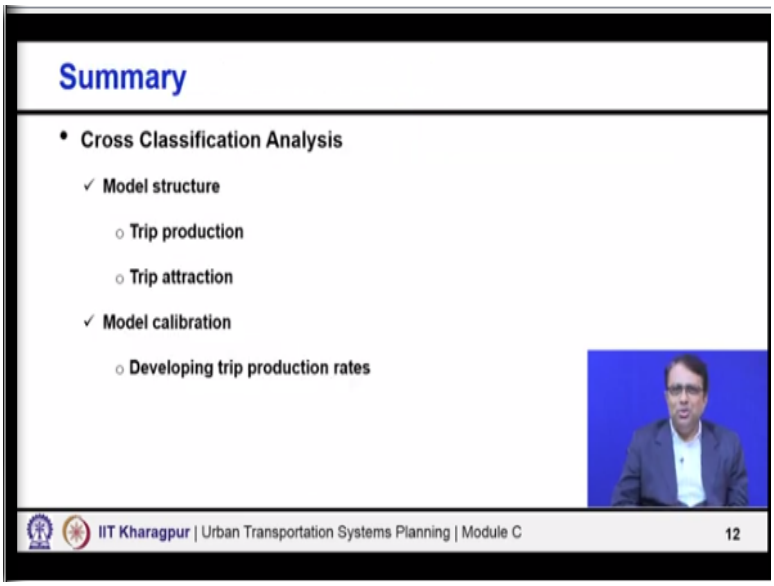
Household	Trips	Income	Cars	Household	Trips	Income	Cars
1	2	15500	0	11	7	88500	1
2	4	18000	0	12	9	82000	1
3	10	88000	2	13	9	32500	2
4	5	51000	0	14	11	51500	2
5	5	17500	1	15	10	49000	2
6	15	87000	3	16	11	72500	2
7	7	59000	1	17	12	86000	2
8	4	31000	0	18	8	45000	1
9	6	37500	1	19	8	73000	1
10	13	90000	3	20	9	65000	1

It is an example only, because no real world studies you can do with 20 households, but here I cannot take several 1000 households just to explain, in a class. So, I have considered just say let us say 20 households as an example. So, I have surveyed, I have approached say maybe 20 households, I have found how many trips they are making, and what is their monthly income? So, based on that, I have just summarized this data.

So, I have 20 household, household numbers are given 1 to 20, each household I have the data, how many trips they make or they made? And I also know that their household income. So, we

using this data, we shall see in our next class how to develop the trip tables or trip rates. So, we shall continue in the next class.

(Refer Slide Time: 31:01)



So, then what we discussed as a summary, we told you what is cross classification technique, what we really try to do in that and the fundamental assumptions, the trip rates remains stable for certain stratification of households over time, that is the fundamental. And then we say how we do it? We simply collect the data and for each of this category, we calculate the rates. So, to do that, how to proceed with the cross classification analysis?

First step is to define a model structure like specifying model in regression models, specifying the model structure. Here also we have to specify the model structure, we explained that how you specify the model structure for trip production and trip attraction taking some examples. And then the next step is actually once you have defined the structure, then you have to collect a lot of data and calculate the trip rate.

That is basically comes under model calibration. So, the calibration we just started, we took an example told you that we have 20 household data and then we shall continue in the next lecture. And calculate the trip rate and move forward, thank you so much, thank you all.