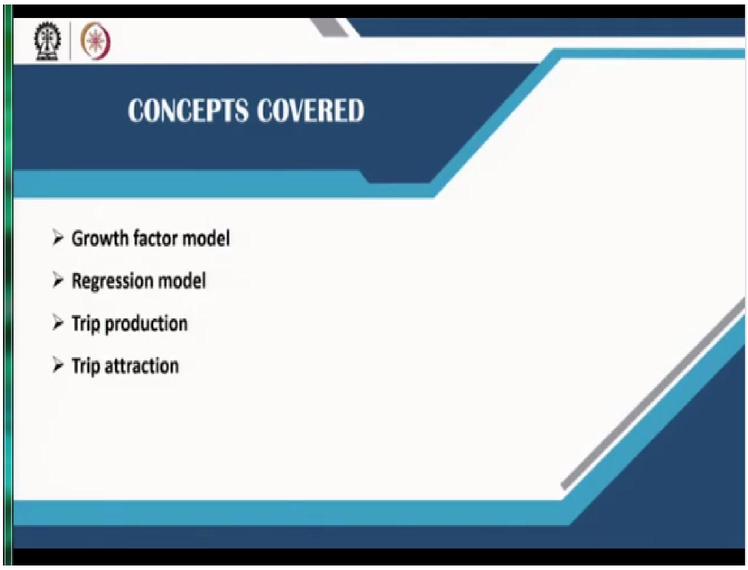
Urban Land use and Transportation Planning
Prof. Debapratim Pandit
Agricultural and Regional Planning Department
Indian Institute of Technology-Kharagpur

Lecture-33
Trip Production and Attraction 1

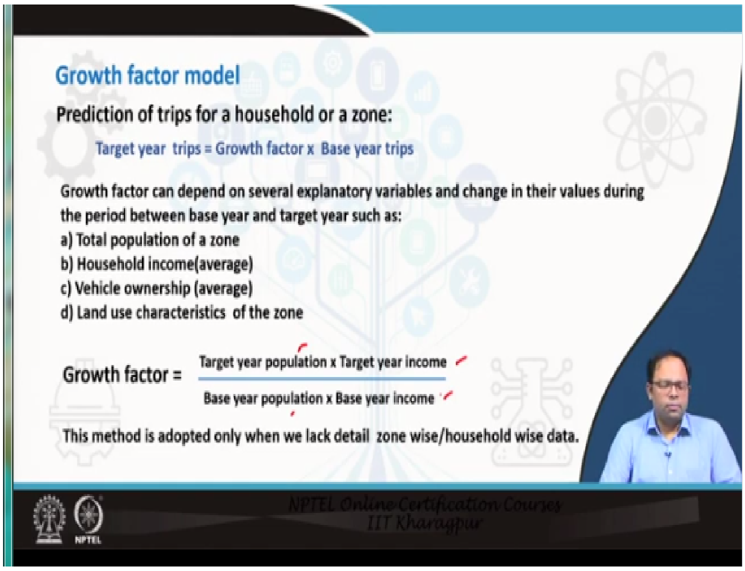
This section is the first of the two parts of trip production and attraction.

(Refer Slide Time: 00:31)



The different concepts that would be covered are; growth factor model, regression models, and trip production and trip attraction using regression.

(Refer Slide Time: 00:42)



Growth factor model:

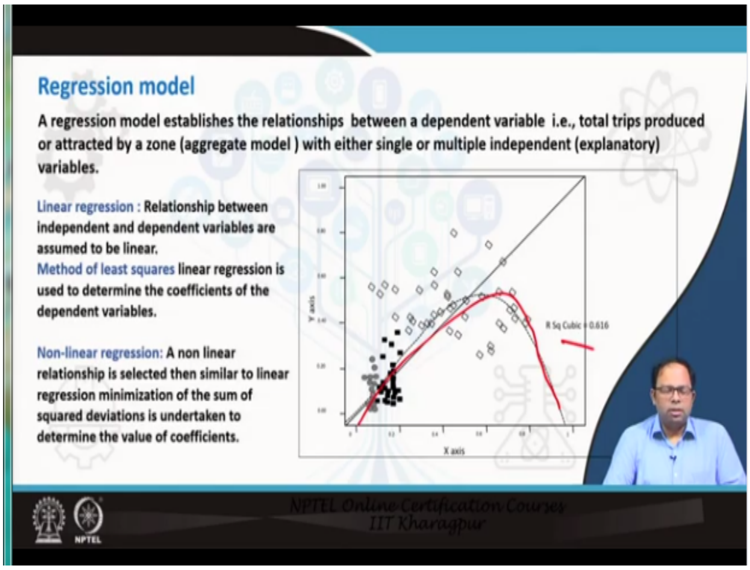
Growth factor model is a very simple way of predicting trips for a household or a zone for a given target year. Prediction is done by just multiplying the number of trips produced and attracted by a particular zone in base year with a growth factor. If predictions are to be made for a particular household, the trips produced by a household is multiplied by a growth factor (as per household category).

Growth factors can depend on several explanatory variables and it changes in values during the period between base year and target year. Variables like, total population, household income, vehicle ownership, land use characteristics etc. pertaining to zones, are some of the many explanatory variables. That means if the number of people are more, automatically the trips will grow. If the income of that particular zone or that household group or the socio economic group increases, then automatically the total number of trips will also grow or we can say, total trip generated would also increase. Thus, in order to determine the growth factor, the characteristics of the target year in respect to the characteristics of base year is determined.

$$Growth\ factor = \frac{Target\ year\ population \times Target\ year\ income}{Base\ year\ population \times Base\ year\ income}$$

As shown in the above example, growth factor is calculated as product of target of year population and income, divided by the product of base year population and income. So, instead of one variable income or population a product of both the variables has been taken. More variables can be added to both the numerator and denominator, whichever seems to be appropriate with respect to prediction. There might even be separate formulas for the prediction of production and attraction. So, only the variables which may play a role in influencing the factor subjected to prediction are chosen to determine the growth factor. Growth factor method is used when there is a lack of zone wise details and household wise data. So, this is an easy way of determining future growth of a particular area.

(Refer Slide Time: 03:10)

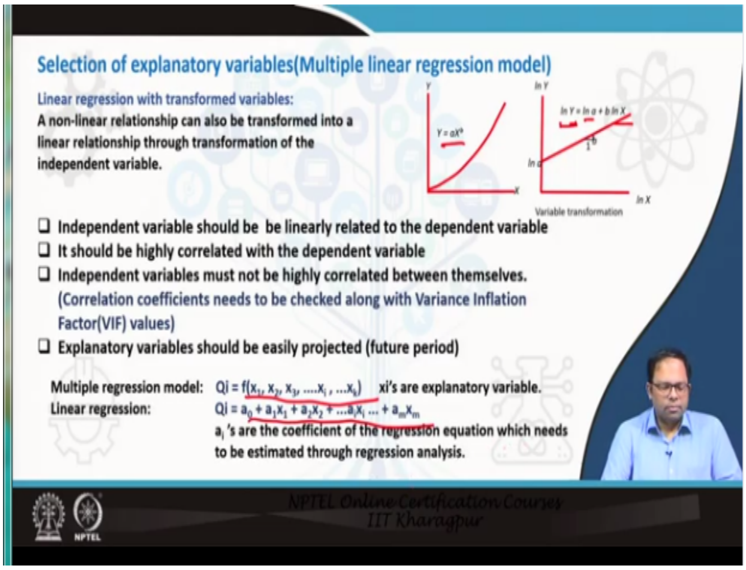


Regression model:

Regression model establishes a linear relationship between a dependent variable and an independent variable or many independent variables. It is a more popular approach as compared to growth factor model. In trip generation, the dependent variable can be total number of trips generated or attracted to a zone, for an aggregated approach. Whereas in a disaggregate approach, the dependent variable can be the number of trips produced by a particular housing category. The number of trips attracted per employee can also be worked out based on the number of employees in a particular zone.

The method of least squares linear regression is used to determine the coefficients of the independent variable(s). Apart from linear curve fitting or linear regression, non-linear curve fitting or non-linear regression can also be done. This kind of regression also estimates coefficients by minimizing the sum of squared errors.

(Refer Slide Time: 05:46)



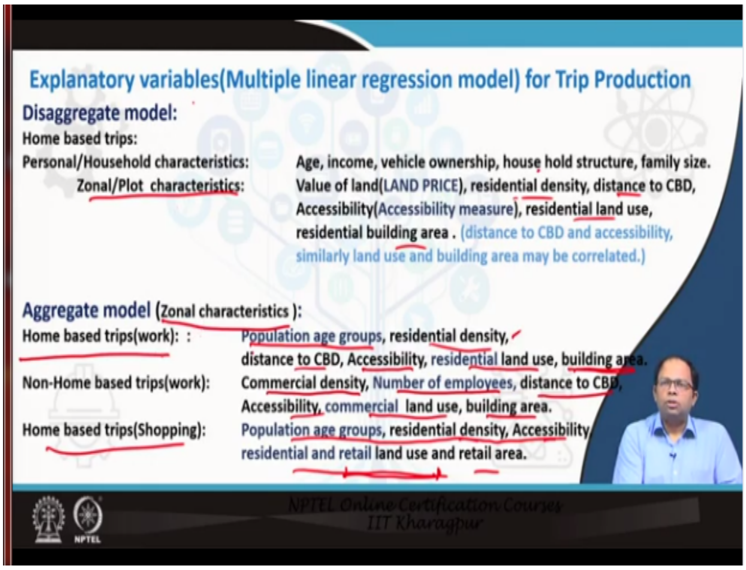
Selection of variables:

The selection of variables for a regression model is based on the assumptions of MLR that has been discussed separately. MLR essentially is a linear function of multiple explanatory variables. The variables which exhibit a linear relationship with the dependent variable are selected. There might be cases when the independent variables might have a nonlinear relationship with the dependent variable. In such cases, a transformation (for example. Logarithmic transformation) can be used to make the relationship linear. For example, the nonlinear equation $Y = aX^b$ can be transformed into a linear relationship $\ln Y = \ln a + b \ln X$ using a logarithmic transformation which can be seen in the graphs shown in the figure.

The independent variables should be highly correlated with the dependent variable, and low to no correlation should be present between the independent variables. Variance inflation factor (VIF) and bivariate Pearson correlation can be used to detect multicollinearity of independent variables.

Another thing that should be kept in mind is that the variables that are included in the model, should be such that they can be projected in future period. If not, then such a model is of little use from policy point of view.

(Refer Slide Time: 07:43)



Explanatory variables for Trip Production:

Regression model for trip production can be done using an aggregate approach and a disaggregate approach. In aggregate models, the trip productions of a particular zone is determined for different trip purposes. Trips are classified as home based work trips, non-home based work trips, home based shopping trips, etc. Different type of explanatory variables are used to determine the number of trips produced due to a purpose. For example, home base work trips can be determined by population age groups living in a particular zone, the residential density of a zone, the distance to CBD, accessibility of that zone, residential land use, building area etc.

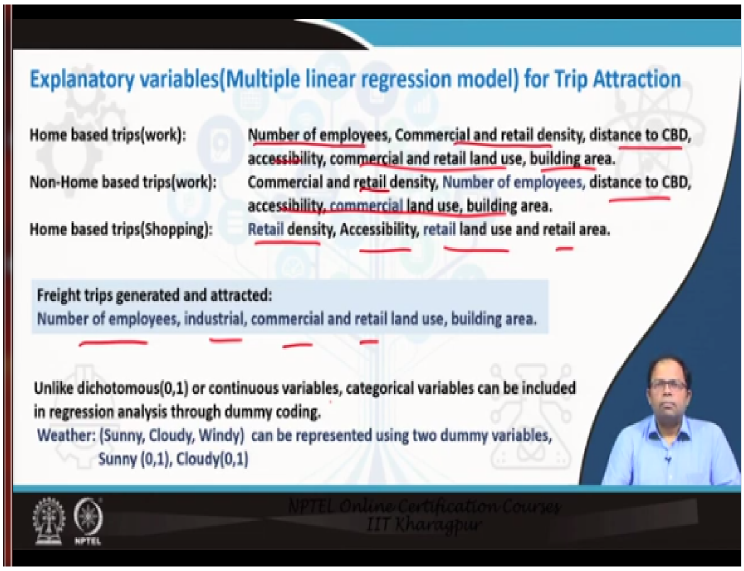
Certain things may obviously be correlated. For example, the total number of people living in an area and the total building area may be correlated. People from different age groups usually perform similar kind of activities throughout the day. So, dividing the population into the age groups probably gives us more subgroups of the population and each subgroup could have different coefficients in predicting the total number of home based work trips. Residential density may play a role and so can distance to CBD or accessibility.

Non-home based trips mostly start from workplaces like, meeting a person, going for a sales call and so on. These kind of trips mostly go from some commercial area to another commercial area. The variables chosen for determining these kind of trips are commercial density, number of employees, distance to CBD, accessibility, commercial land use, building area. Again, commercial land maybe correlated with building area or both of them are found to be correlated with the number of employees.

Home based shopping trips can also be there which can be related with population age groups, residential density, accessibility, residential and retail land use and retail area. Even though these are home based shopping trips, people may choose to travel both near and far for shopping. So, both population or the zonal characteristics as well as residential land use and retail land use data is required.

These are the different likely parameters based on zonal characteristics, which could be included while developing a regression model for number of trips generated from a particular zone. Similarly, we can also have a disaggregate model, where prediction of the total number of trips is based on different household types. So, for home based trips, household characteristics and zonal plot characteristics can play a role. Household characteristics could be age (of household head, household members), income, vehicle ownership, household structure, family size. Variables like value of land, land price, residential density, distance to CBD, accessibility residential land use, residential building area can also be used to determine total number of trips produced by a particular socio-economic group in an area.

(Refer Slide Time: 12:34)



Explanatory variables for Trip Production:

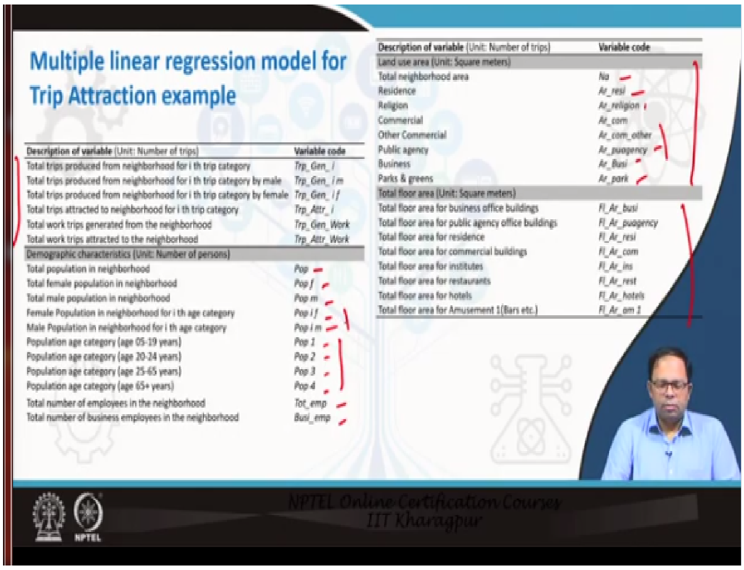
Similar to trip production, trip attraction can be explained using different explanatory variables using a multiple linear regression model. For example, number of home based work trips attracted to a particular area can be determined using the number of employees, commercial and retail density, distance to CBD, accessibility, commercial and retail land use, building area etc. for the area. Although attraction is mainly a function of the number of employees, but in case of commercial area, the commercial and retail density also plays a role in determining the total number of employment in respective sectors. Other people employed in ancillary services like cleaning, catering, etc. could be determined by landuse and hence details of other landuse is also important.

Non-home based work trips attracted depends on the commercial and retail density, number of employees, distance to CBD, accessibility, commercial landuse, building area, which is similar to the home based work trips attracted**.** Number of home based shopping trips that are attracted depends on the total retail density, accessibility, retail land use and retail area. The more is the retail area, more number of people would be attracted to go to that particular zone for shopping. So these are the different variables which play a role in explaining trip production and attraction.

Freight trips produced and attracted can also be similarly predicted. This can be done based on number of employees, industrial commercial and retail land use, and also building area in that particular zone. These variables actually determine the number of freight vehicles that would be both produced as well as attracted to that particular zone.

Categorical variables, apart from dichotomous (0/1) or continuous variables, can also be included in regression as already discussed. To see the effect of these type of variables, they need to be dummy coded. For example, weather (sunny, cloudy, windy) can be represented using 2 dummy variables. A cloudy day can be represented as sunny (0) and cloudy (1) to include these kinds of categorical or qualitative variables in regression equation.

(Refer Slide Time: 15:14)



Example of MLR for Trip Attraction:

The example shown is an MLR for trip attraction. A long list of all the variables considered for modelling is shown. Corresponding to each variable are the variable code (in italics) that should be designed to have resemblance to the actual variable name so that a person can understand which variable this code belongs to. Apart from that, these codes should use the syntax of nomenclature of variables supported by statistical packages in order to avoid any inconsistency while analysis.

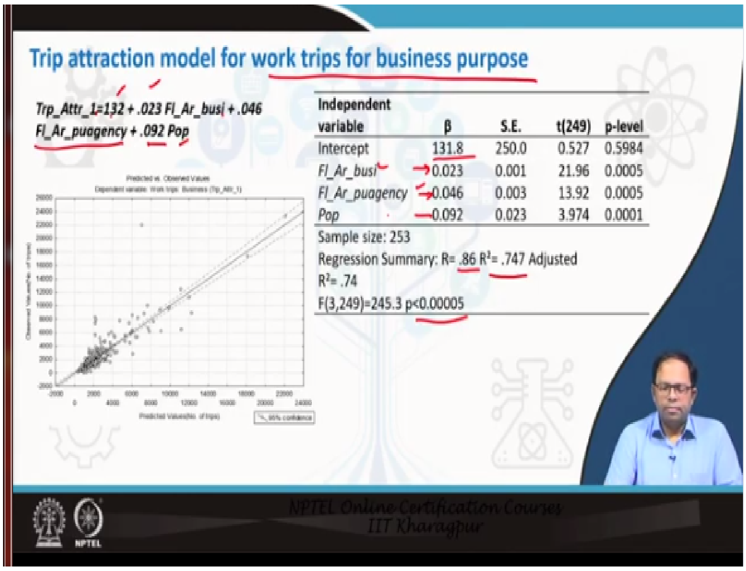
The trip related variables that has been used are, total trips produced from neighbourhood for i^{th} trip category (*Trp_Gen_i*); total trips produced from neighbourhood for i^{th} trip category by males (*Trp_Gen_im*); total trips produced from neighbourhood for i^{th} trip category by females (*Trp_Gen_if*); total trips attracted to neighbourhood for i^{th} trip category (*Trp_Attr_i*); Total work trips generated from neighbourhood (*Trp_Gen_Wrk*); Total work trips attracted to neighbourhood (*Trp_Attr_Wrk*) etc.

The demographic variables that has been used in this example are, total population in the zone; total female population; male population; total female population; male population; population based on different age groups; total number of employees in our neighborhood and total number of business employees in the neighborhood.

Land use related variables are, total neighborhood area, which is a continuous variable; area of residences; area of parks and greens; area of business; area of public agency (*Ar_puagency*); all other commercial area; religion landuse area and so on. Total floor area also is considered for different kinds of buildings, office buildings, public agencies, residences, commercial institutions and so on.

Many of the variables considered like building area and land use maybe correlated. Similarly, variables like including population for different age groups and gender wise categorization of variables may seem to be redundant. This is done because at the beginning of the regression it is not evident which representation of a particular variable will give the best possible result to predict the total number of trips attracted to a particular neighborhood or a particular zone. So, all possible combinations of variables are tested to check which derives the best possible result. After checking for all the assumptions and other checks, the most appropriate regression which is both theoretically and statistically consistent, is developed.

(Refer Slide Time: 18:31)



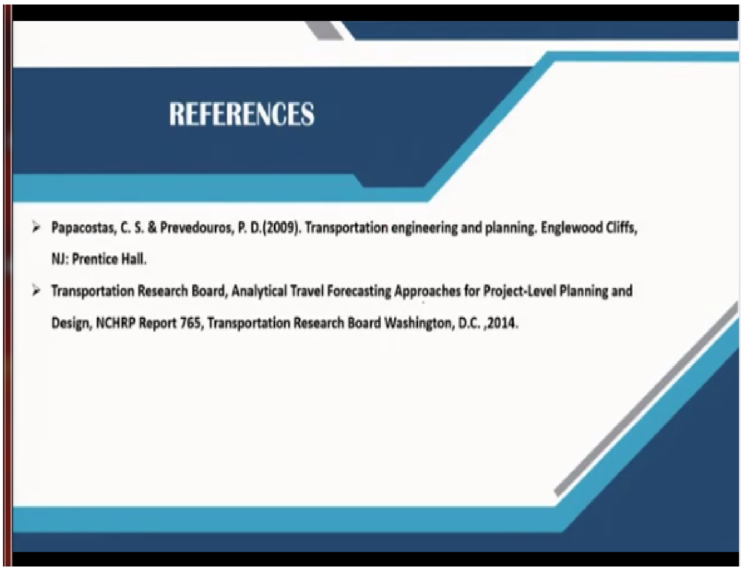
Results of a Trip Attraction model for business purpose work trips:

The table shows the final result of the model, where total trip attracted to a particular zone is predicted, particularly work trips for business purpose. The regression equations shows that total trips attracted is dependent on the intercept(131.8), floor area of businesses, floor area public agency, and population. The equation with variables and respective coefficients is as follows:

$$Trp_{Attr_1} = 131.8 + \ +0.023 \times Fl_{Ar_{busi}} + 0.046 \times Fl_{Ar_{puagency}} + 0.092 \times Pop$$

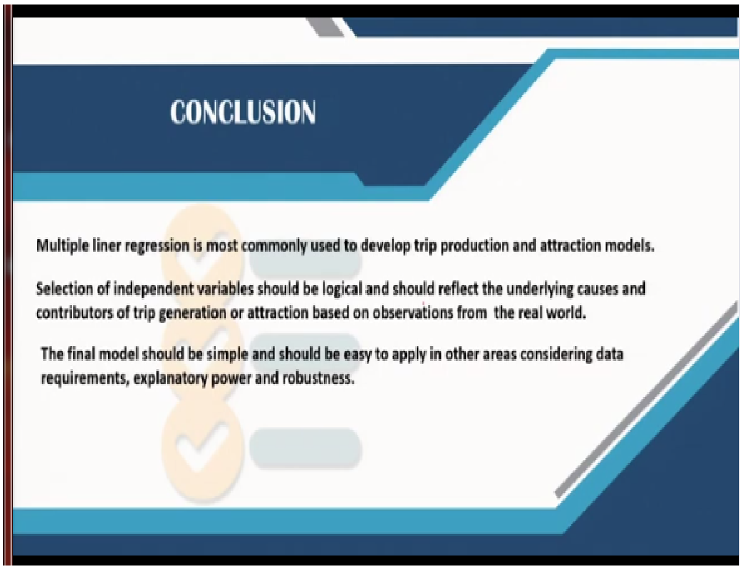
The above model returns the best possible result, given the data. The adjusted R square is 0.747, and the total F value is significant which means the model structure is proper. The model has intercept at 131.8 and as population of a zone increases, then trip attracted to that zone for different purposes also increases but at a very small magnitude. Primarily trips attracted is related with total number of buildings, i.e., floor area for businesses and public agencies which attracts people to come to work in these particular areas. These are the explanatory variables that are included in the model.

(Refer Slide Time: 20:00)



These are some of the references can be considered for further reading.

(Refer Slide Time: 20:03)



Multiple linear regression is most commonly used to develop trip production and attraction models. Selection of independent variables should be logical and should reflect the underlying causes and contributors of trip generation or attraction based on observations from the real world. The final model should be simple and should be easy to apply in other areas considering data requirements, explanatory power and robustness.