# COMPARISON OF REGRESSION MODEL AND CATEGORY ANALYSIS (A CASE STUDY)


| S.V.C.SEKHAR | S. ANAND | Mohamed Rehan KARIM |
|---|---|---|
| Lecturer, | Lecturer, | Associate Professor, |
| Dept. of Civil Engineering, | Dept. of Civil Engineering, | Dept. of Civil Engineering, |
| University of Malaya, | University of Malaya, | University of Malaya, |
| 50603 Kuala Lumpur, | 50603 Kuala Lumpur, | 50603 Kuala Lumpur, |
| Malaysia | Malaysia | Malaysia |
| Phone : 603-7595302, | Phone : 603-7595399, | Phone : 603-7595231, |
| Fax: 603-7595318 | Fax: 603-7595318 | Fax: 603-7595318 |
| E mail : svc@fk.um.edu.my | E mail : anand@fk.um.edu.my | E-mail:rehank@fk.um.edu.my |


abstract: Accurate forecasting of transportation demand is a challenging task. Difficulties arise because individual behavior and inner workings of economic system, both of which create the demand for transportation services, are not clear-cut. In trip generation analysis, relationships are established between the number of trips produced by, and attracted to, a given zone. The most common methods used in trip generation models are: i) Multiple Linear Regression (MLR) models and ii) Category or Cross classification technique or Category analysis. As a part of World Bank aided Comprehensive Traffic and Transportation Study for the Madras city in India, detailed household survey consisting of 22,250 households was conducted in the year 1993. Using this large database, trip end models were developed using both the methods and results are published in this paper.

## 1. INTRODUCTION

Trip end forecasting is based on the principle that 'land use generate traffic' . The level of traffic generation and in particular the mode of travel depends upon a large number of factors which can be conveniently divided into following groups : (i) Socio- economic variables (ii) Location variable (iii) public transport accessibility variables .

The goal of trip generation is to predict the number of trips that are generated by and attracted to each zone in a study area. In trip generation, methods are applied to predict productions and attractions or origins and destinations. The zones that contain the home end of home based trips or the origin end of non home based trips are considered to have produced the trip while the destination zone where an out of home activity will be undertaken is considered to have attracted the trip. Trip making is highly varied reflecting the diverse activities pursued by people in their work and non work activities. For the purposes of analysis, however, trips are typically grouped in terms of categories or purposes ; while this may disguise the variety of activities pursued, it greatly simplifies model development. The number of purposes that should be used depends fundamentally on the analytic purpose at hand and the data available. In determining the trip purposes to model, all of the steps in the modeling process, not just trip production, need to be considered. However many purposes are utilised for modeling and it is sound practice to treat work trips as a separate from other forms of trip making . The journey to work is typically the most important trip to model correctly due to the large amount of travel accounted for by this purpose and the fact that work trips most commonly occur during the congested , peak travel periods. Adding peak period capacity through construction is typically the most expensive transportation investment and the type of project that warrants the most detailed analysis. The most common inputs for trip production models are household and personal characteristics either at disaggregate level, or aggregated to the zonal level. However the models for trip attraction cannot be done disaggregated at household level. The model has to be used for arriving at projected trip matrix for horizon year and use it for developing traffic flow pattern, identify projects to meet the demand and evaluate them. Hence it is desirable to develop models purposewise, modewise and separately for peak and off peak hours. This paper covers the salient features of the approach used as part of Comprehensive Traffic and Transportation Study in Madras - India.

## 2. ABOUT THE STUDY AREA

Madras city , is the fourth largest metropolitan city in India and the largest in the southern region. This city has grown over the past 3 centuries starting with a nucleus settlement of 67 sq.km . Presently, the Madras Metropolitan Area covers an extent of 1157 sq.km of which 172 sq.km forms the city corporation area and a total of 531 sq.km forms the urban agglomeration. Madras metro area has a population of 5.36 million. The average family size is 4.81 for the city area and 6.3 in outer area as against 5.21 in 1971. The average household has 1.45 workers. The per capita trip rate has gone up to 1.28 from 0.90 in 1971. The cycle and pedestrian proportion of trips have not undergone much change in 20 years but the private vehicle trips rate has gone up from 0.39 in 1971 to 0.73 in 1992. This has resulted in the decrease in percentage of public transport trips i.e. from 53% to 44%. The vehicle ownership level has been growing rapidly and is 0.04 cars per household and 0.2 motorised two wheelers per household. This explains to a large extent the increase in private vehicle trips indicating that vehicle ownership has an influence on trip making propensity. The mean trip length has also shown a marked increase from 7.2 to 10.1 kms in 21 years. Longer trips are however performed by train (14 km) and bus (8.4 km). Major proportion of education trips is found to be made by cycle and Intermediate Public Transport (mostly cycle rickshaws).

## 3. ZONE DELINEATION

Zones have been defined such that their future population by 2011 will not exceed 50,000 , to avoid problems arising from forecasts of intra zonal trips. Wherever the density of population is high , but the area is too small to be delineated based on population a higher population index has been adopted so as to provide a reasonably acceptable area of zone. In respect of CBD ( Central business District) to the maximum extent possible earlier zoning system has been retained. In other areas earlier zoning system has been either retained or subdivided based on above mentioned criteria of population and area and network density. Considering these criteria, city corporation area has been divided into 121 ( which includes 9 CBD zones ) and rest of metropolitan area, has been divided into 68 zones (RITES, 1995). The zones have been coded adopting a three digit decimal coding system . The study area was subdivided into nine divisions on the basis of catchment / influence area for major radial corridors. The first digit represents the division under which the zone is situated. The second digit refers to the isochronal rings of 3 to 4 minutes from the central area. The subdivision within the consecutive isochronal rings is represented by the third digit. The zone within the CBD area is represented by starting digit of '0'. For example zone number 242 represent a zone, in division 2, which is lying in the isochronal band number - 4, with subdivision number equal to 2.

## 4. DATA USED

The data used has been collected as part of Comprehensive Traffic and Transportation Study (1993) for this city through a home interview survey. The home interview survey is concerned with the collection of facts relating to the socio-economic characteristics of the population and trip movements that made on the typical day within the study area. A sample size of 2 percent which amounts to 22,250 households is chosen. The frame from which the sample of households is selected from the electoral rolls. The electoral roll is updated annually and includes a list of the names and addresses of those persons who are qualified voters. Once the selection of the household sample is completed, the collection of data by home interview was commenced by specially trained field investigators and trained enumerators .Data collected from the sample household was expanded to represent the whole population. The expansion factor for each zone was obtained by the total number of households in the survey area for each traffic zone divided by the total number of successful interviews in the respective traffic zones.

## 5. FACTORS AFFECTING TRIP ENDS

*Socio - economic variables:* Socio- economic variables largely measure the desire and potential of house-holds to make trips. The variables normally considered include household size, household income, vehicle ownership, number of workers and students per household etc.

*Location variables:* Location variables reflect the surrounding environment and should ideally measure the spatial separation of households from each of the amenities which desire e.g. schools, shops, work places

*Public transport accessibility variables:* Public transport accessibility variables measure the level of public transport service available to the trip maker in the given area.

The number of variables which are thought to exert a causal effect on trip generation may be very large. Some of these variables may be interrelated and hence measure largely the same effect; other variables may exhibit only minor influences. The process of trip end modeling is: firstly, to identify those variables which have significant and separate effects on trip generation; and secondly , to develop a mathematical tool whereby the effects of selected variables acting together can be measured. The model must not only provide a good statistical fit to present -day data, but also be a logical and meaningful form. The explanatory variables incorporated must themselves be meaningful determinants of trip making behaviour, such that significant changes in their levels will produce (in the model ) realistic changes in trip making levels. In addition, the explanatory variables lend themselves to future estimation at the chosen level of aggregation. Thus it will be seen that any model developed to determine trip ends will have to take into these as determinants or variables. On the other hand larger the number of variables, more complicated it is to arrive at a relationship. It would be difficult, expensive and time consuming even to assess and quantify them in the base year also. Many planners therefore try to minimise the determinants to be considered by selecting more predominant influences and those which can be easily measured in required level of disagregation.

## 6. MODELLING APPROACHES

The two modelling approaches used in trip end forecast are use of Multiple Linear Regression (MLR) and Category analysis. The MLR approach is the one used from earliest time of transportation planning and still the most widely used in India for both for trip generation and trip attraction. The planners differ only in their choice of the variables used. The numbers of variables used vary from one ( Population) to as large as six for trip generation models. For trip attraction models, fewer and land use dependent variables have been used. For studies conducted in different cities the variables used vary so much (as indicated in Table-1) that no specific rule can be established for this.

Table - 1: Variables used in Different City Studies

| Study Area | Generation | Attraction |
|---|---|---|
| London | Number of employed residents<br>Cars owned<br>Income | Retail Employment<br>Manufacturing Employment less service employment<br>Wholesale employment less service employment<br>Other employment less service employment<br>Service employment |
| Northompton | Number of employed residents | Total employment<br>Employment in public building<br>Employment in Health services |
| Cardiff | Population<br>Number of employers<br>Cars owned | -------- |
| Toronto | Population<br>Dwelling units<br>Cars owned | Population ,Manufacturing employment<br>Retail and service employment<br>Other employment |
| Thurrok | Income<br>Household | Total Employment<br>Retail Employment |

Source: Lane et al. (1971)

It has been observed that the most predominant variables used for trip generation models are: the number of workers (for work trip), number of students (for education trips), population (for other trips), number of vehicles in zone and mean income. For trip attraction, the predominant variables are work places (for work), education places (for education), population (for other trips) and commercial area.

The main advantage of MLR model lies in its simplicity, ability to develop it even with small aggregated data and possibility of checking its robustness by statistical tests like $R^2$ value, F-statistic and standard error of estimation. The disadvantage is that since data aggregated at zonal level is used, it may not correctly represent the distribution of population in each zone. For example two zones may have same mean income but the percentage of high, middle and lower income households may vary substantially and hence the trip making characteristics of the these groups would also vary widely.

The other method which has been in use since is the category analysis or cross classification method. This was first developed by Cardiff study and is the most used one in UK. This model aims to arrive at trip generation rate per household. The households are divided into a number of categories based on the major determinants of trip making. The major determinants used are household composition (number of employees for work trips, number of students for education trips and total number for other trips); vehicle ownership; household income grouped under four or five ranges. The basic assumptions underlying this method are:

i) The household as an independent unit, from which most journeys begin or end depending on requirements of its members is the basic unit of trip generation;
ii) The households with one set of characteristics produce an average trip generation unique to it i.e. quite different from the average trip generation produced by a household with another set of characteristics.
iii) The trip generation rates per household for a particular category remain same or stable over time - so long as factors external to the household are same as when the rates were first estimated; and
iv) Location specific category analysis when done also assumes that the trips or journeys generated by the household depends on the characteristics of that household and its location in relation to the facilities required like work places, schools, shops, accessibility to mode etc.

This methodology overcomes the main criticism against the MLR models but involves some other problems. For example, there are large number of categories depending on the level at which trip rates are required. If trip rates are required for 3 types or purposes, for 4 modes (car, two wheeler, train, bus), for 4 income groups and 4 levels of vehicle ownership, the number of categories involved would be 192 each for work trips, education trips and 256 for other trips, which amounts to total 640 categories.

If location specific rates are also to be derived to satisfy assumption (iv), for a city like Madras, three or four such sets will be needed e.g., for core area; for intermediate area; for area outside city; for zones within the influence area of rail corridors. In order to satisfactorily derive trips rates for 4 x 640 categories, a very large set of data needs to be collected. The major difficulty would arise for the planner in providing data for forecasting since one will have to estimate likely number of households of each category which would lie in each zone at the future date. In order to overcome this difficulty, in some studies, the number of categories has been brought down to a level at which such estimates (providing future planning data) are possible. For example, since vehicle ownership and household income are inter-related, only one of these two is accounted . Since choice of mode by different member of household is dependent upon distance of travel, cost of travel, time of travel, and accessibility, etc., apart from vehicle ownership, this (modal split) could be done at zonal aggregate level using another sub-model. Thus the number of categories can be reduced to trip purpose (3) household composition (3,3,4 for work, education and other respectively) and 4 or 5 groups based on income or vehicle ownership. The number of categories would then be 120 or 150 as the case may be.

For purpose of comparison, models have been developed for whole day trip and for peak period trips. The peak period trips are those trips which are generated from the zones or residence, between 8.00 hours and 9.30 hours. This period has been chosen after comparing the traffic flows across various corridors in the metro.

## 7. MULTIPLE LINEAR REGRESSION (MLR) MODELS:

The concept of regression can be applied to trip end modeling in different ways. The most commonly used method has been that of least-squares regression in which a response surface is fitted to observed data utilizing the principle of least squares. For computational purposes the response surface generally been assumed to be linear. In this study zonal least squares regression has been used. In this application each traffic zone is treated as one observation. The dependent variable is some measure of the number of zonal trip ends and the independent variables are typically such measures as number of households, number of workers and number of vehicles, etc. Alternatively the variables are expressed as mean values for zones. In using the model as a forecasting tool it is assumed that the derived relationships will remain stable over time and those future values of the independent variables can be estimated outside the model. Forecasting, often involve extrapolation beyond the range of observed data, although this is a problem inherent in all trip end forecasting techniques.

From the Development Authority, Household interview survey, and from the basic records, the data which would be reliably obtained are population, residential area, number of workers, number of students, number of school places, work places, vehicle ownership, household income etc. at the zonal level. A software MLINREG in basic language has been used. This programme has an in-built facility of carrying out stepwise regression analysis.

The first requirement is choice of variable for the model. In many earlier studies, population as a single variable was used. The earlier Study conducted for Madras in 1971 had developed purposewise, modewise models (totaling 22 numbers) and the variables used in different models were 3 or more. The list included population, number of cars and motor cycles, residential areas, public and semi public areas, open space, commercial area, social and recreation areas and agricultural areas in different zones. The $R^2$ value varied from 0.0428 to 0.9238. All these models considered full day trips.

The latest thinking in respect of evaluation is that level of benefits that accrue in peak period is distinct from those in off peak period. In order to correctly estimate this it is desirable to work out separately peak period trips and off-peak period trips, build separate matrices assign them separately. Hence it was decided to develop trip end models separately for these two periods. In view of different values of time savings attached to trips made for different purposes, models had to be developed purpose wise (work, education and others). The models have been developed only for motorised vehicle trips and cycle trips have been excluded in rigorous modelling.

Thus it will be seen a total of 3 x 2 = 6 models each for trip generation and trip attraction were developed. This paper describes the detailed approaches in case of one set i.e. peak period work trips. Finally evolved models for all purposes are listed at the end of this section.

## 8. CHOICE OF VARIABLES

In order to select the most appropriate variables to be used in the model, a correlation exercise was carried out to see how far each is correlated with the dependent variable. The value of coefficient of determination equal to or over 0.75, with dependent variable has been taken as the lower threshold value for first choice. This showed that undermentioned variables have influence on respective trips generation.

Work trips : Number of workers, population, number of households and total vehicles.
Education trips: Number of students, population, households and total vehicles.
Other trips : Population and households

Variables selected which have lowest inter correlation among them and are highly correlated to dependent variable. However care has been taken to select the most logical variable. Stepwise MLR runs were be done and the set which gives best $R^2$ and F value was selected. Public transport accessibility variable was not selected as it did not show any logical results. The stepwise run made for peak period work trips showed the following results:( refer Table - 2)

Table - 2 : Trial runs for trip generation models

| Variables - in order of | $R^2$ | F- test |
|---|---|---|
| Workers | 0.8176 | 33.59 |
| Total vehicles | 0.8433 | 30.41 |
| Residential area | 0.8535 | 12.77 |
| Industrial area | 0.8588 | 6.90 |
| Population | 0.8617 | 3.76 |
| Income | 0.8653 | 4.86 |
| Households | 0.8660 | 0.96 |
| Commercial area | 0.8663 | 0.41 |

For the selecting the best variables for the model two criteria is adopted to: one is that it should give good statistical fit and next criterion is that the variable should be such that it can forecast accurately for future years. Considering these criteria the variables selected for work trips are: number of workers in zone and number of motorised (personal) vehicles in zone. By similar exercise it has been found that the best results are obtained by using number of students and number of vehicles for education trips and population and number of vehicles for other trips. The t- statistics were also worked out and the results were within permissible limits at 10 percent level of significance.

The trip generation models developed for peak and off peak period for entire metropolitan area are given below: ( Refer Table 3)

Table -3 Regression Equations for Trip Generation

| Purpose | Peak Period Model | | Off Peak Period Model | |
|---|---|---|---|---|
| Work Trips | 109.093 + 0.446 * V + 0.3748 * W | $R^2$ =0.832 | 77.98 + 0.0458 * V + 0.046 * W | $R^2$ =0.26 |
| Education | 95.38 + 0.1645 * V + 0.2033 * S | $R^2$ = 0.617 | 7.005 + 0.0309 * V + 0.005* S | $R^2$ = 0.18 |
| Other | 56.38 + 0.01608 * P + 0.048 * V | $R^2$ = 0.360 | 127.37 + 0.00581 * P + 0.157 *V | $R^2$ =0.39 |

Where
V = Number of motorised vehicles , W = Number of workers, S = Number of students
P = Population and $R^2$ = The Coefficient of determination

## 9. CATEGORY ANALYSIS

In many developed countries, category analysis or cross-classification technique is more popular because of better results obtained using this method. Category analysis can deal easily with quantitative variables and does not make any assumptions about the shape of the response surface. The technique is based on determining the average value or average response of dependent variable for defined categories of the independent variable. The categories are defined by a multi-dimensional matrix where each dimension represents an independent variable and where independent variables are themselves stratified into categories. In applications to trip end modelling each observed analysis is assigned to a particular category or cell of the matrix depending on its values of the independent variables and the average trip rates are subsequently determined for each cell.

Category analysis which was tried first for grouping of number of workers, students, with second grouping of household income. The income grouping range which is compatible with the census study is adopted for this analysis. The results are given in Table - 4. Though it was easier to group the households under different range of income, results did not indicate any consistent trend.

Table - 4: Category Analysis for whole Day Trips

| Household composition | Income Range < Rs. 1101 | Income Range Rs. 1101-1800 | Income Range Rs.1801-3600 | Income Range Rs.3601-5700 | Income Range >Rs.5700 |
|---|---|---|---|---|---|
| | Work trips | | | | |
| No worker | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| One worker | 0.75 | 0.77 | 0.75 | 0.78 | 0.76 |
| Two workers | 1.34 | 1.31 | 1.33 | 1.23 | 1.33 |
| Three or more | 1.58 | 1.97 | 2.06 | 1.98 | 1.84 |
| | Education trips | | | | |
| No Student | 0.02 | 0.04 | 0.03 | 0.03 | 0.02 |
| One Student | 0.71 | 0.69 | 0.61 | 0.61 | 0.58 |
| Two Students | 1.34 | 1.26 | 1.24 | 1.09 | 1.10 |
| Three or more | 2.01 | 1.90 | 2.05 | 1.53 | 1.66 |
| | Other trips | | | | |
| One member | 0.28 | 0.21 | 0.28 | 0.23 | 0.60 |
| Two members | 0.30 | 0.39 | 0.29 | 0.31 | 0.46 |
| Three members | 0.31 | 0.45 | 0.35 | 0.41 | 0.40 |
| Four members | 0.32 | 0.38 | 0.40 | 0.38 | 0.51 |
| Five or more | 0.36 | 0.36 | 0.36 | 0.46 | 0.64 |

Next refinement was done by separating peak and off-peak period trip rates on the basis of income ranges and results are shown in Table 5

Table - 5: Category Analysis : Peak - Off peak period Trips

| Household composition | Income Range < Rs. 1101 | | Income Range Rs. 1101-1800 | | Income Range Rs.1801-3600 | | Income Range Rs.3601-5700 | | Income Range >Rs.5700 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wok trips | Peak | offpeak | Peak | offpeak | Peak | offpeak | Peak | offpeak | Peak | offpeak |
| No worker | 0.027 | 0.000 | 0.016 | 0.060 | 0.020 | 0.000 | 0.000 | 0.000 | 0.00 | 0.00 |
| One worker | 0.50 | 0.092 | 0.521 | 0.105 | 0.566 | 0.101 | 0.642 | 0.121 | 0.57 | 0.08 |
| Two workers | 0.949 | 0.168 | 1.069 | 0.175 | 1.188 | 0.157 | 1.321 | 0.136 | 1.33 | 0.09 |
| Three or more | 1.403 | 0.280 | 1.605 | 0.236 | 1.773 | 0.311 | 1.877 | 0.231 | 2.00 | 0.25 |
| Education trips | Peak | offpeak | Peak | offpeak | Peak | offpeak | Peak | offpeak | Peak | offpeak |
| No Student | 0.019 | 0.004 | 0.030 | 0.009 | 0.018 | 0.005 | 0.031 | 0.000 | 0.01 | 0.00 |
| One Student | 0.817 | 0.118 | 0.811 | 0.090 | 0.783 | 0.044 | 0.768 | 0.055 | 0.75 | 0.06 |
| Two Students | 1.613 | 0.177 | 1.586 | 0.143 | 1.551 | 0.109 | 1.487 | 0.116 | 1.33 | 0.16 |
| Three or more | 2.473 | 0.309 | 2.498 | 0.223 | 2.403 | 0.206 | 2.355 | 0.067 | 2.16 | 0.22 |
| Other trips | Peak | offpeak | Peak | offpeak | Peak | offpeak | Peak | offpeak | Peak | offpeak |
| One member | 0.103 | 0.123 | 0.031 | 0.129 | 0.108 | 0.094 | 0.125 | 0.250 | 0.33 | 0.00 |
| Two members | 0.156 | 0.156 | 0.143 | 0.132 | 0.115 | 0.121 | 0.067 | 0.095 | 0.17 | 0.35 |
| Three members | 0.171 | 0.145 | 0.166 | 0.161 | 0.102 | 0.138 | 0.071 | 0.137 | 0.12 | 0.19 |
| Four members | 0.166 | 0.136 | 0.166 | 0.136 | 0.132 | 0.156 | 0.199 | 0.135 | 0.18 | 0.22 |
| Five or more | 0.198 | 0.169 | 0.178 | 0.186 | 0.191 | 0.183 | 0.227 | 0.215 | 0.18 | 0.33 |

However it was noted that the trip rates did not indicate any consistent trend of increasing with number of the particular category of persons or increase in income, both of which are logical expectations. Hence it was felt it might be suitable to try it on vehicle ownership basis. This was found to yield better results in other contemporary studies. It was also considered better to group the category of nil worker, nil student and one person household and merge it with the next level. The ideal grouping of vehicle ownership would be to consider: no vehicle + cycle only owners, two wheelers owners, one car + one two wheeler owners, and over one car + over one two wheelers. The trip rates are obtained for city area and outside city area separately. However this grouping also showed ( refer Table 6) low values especially in the off peak period.

## 10. MODIFIED CATEGORY ANALYSIS

As the conventional method did not indicate logical results, different method which was earlier evolved by Stopher (1983) in one of his studies. [Stopher et al] was adopted to find the trip rates. The modification is carried out by working out global means, row means and column means and applying corrections to all cells. This is an iteration process in which case the first iteration is obtaining the trip rates for each cell representing that group in conventional way. At this stage there will be some cells with abnormal values. In the second iteration by applying corrections based on global averages again trip rates are calculated. The process is repeated till there is no large variation in the trip rates of the cells which showed normal values.

Table - 6: Comparison of Conventional and Modified Category Analysis - Peak Period

| Household composition | Method | With in Corporation Area: Vehicle Category | | | | Outside the corporation Area: Vehicle Category | | | |
|---|---|---|---|---|---|---|---|---|---|
| Peak | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Upto one worker | Conventional<br>Modified | 0.480<br>0.463 | 0.551<br>0.626 | 0.617<br>0.798 | 0.417<br>0.606 | 0.354<br>0.264 | 0.492<br>0.427 | 0.582<br>0.599 | 0.476<br>0.407 |
| Two workers | Conventional<br>Modified | 0.934<br>0.992 | 1.098<br>1.155 | 1.287<br>1.327 | 1.191<br>1.135 | 0.707<br>0.793 | 0.994<br>0.956 | 1.283<br>1.128 | 1.127<br>0.936 |
| Three or more workers | Conventional<br>Modified | 1.394<br>1.585 | 1.885<br>1.728 | 1.902<br>1.900 | 1.129<br>1.708 | 1.251<br>1.366 | 1.421<br>1.529 | 1.762<br>1.701 | 1.333<br>1.509 |
| Off Peak | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Upto one worker | Conventional<br>Modified | 0.224<br>0.221 | 0.236<br>0.252 | 0.316<br>0.399 | 0.284<br>0.353 | 0.122<br>0.074 | 0.152<br>0.106 | 0.249<br>0.252 | 0.096<br>0.206 |
| Two workers | Conventional<br>Modified | 0.381<br>0.394 | 0.382<br>0.426 | 0.583<br>0.484 | 0.469<br>0.438 | 0.209<br>0.247 | 0.281<br>0.279 | 0.433<br>0.401 | 0.625<br>0.626 |
| Three or more workers | Conventional<br>Modified | 0.562<br>0.624 | 0.704<br>0.656 | 0.817<br>0.802 | 0.738<br>0.756 | 0.376<br>0.724 | 0.427<br>0.756 | 0.766<br>0.878 | 0.000<br>0.832 |

## 11. TRIP ATTRACTION

Trip Attraction models are generally developed using MLR approach since it is difficult to quantify the parameters in a required form from data available in Indian cities for defining the categories. In order to arrive at the suitable variables to be used for MLR models of different purpose a correlation exercise was carried out first (vide table for work trips). In addition for work trips, stepwise MLR runs were made for full day trips to see how different variables influence choice and the reliability of model. The different combinations gave results as shown in Table - 7.

It will be seen that the last mentioned would be the best combination. However considering likely availability of data for horizon year, it was decided to use work places as the only variable for work trips. On the same analogy, for education trips also, only school places have been considered the variable. For other trips, it was found that population and commercial area would give the best results. These two are easily obtainable based on the approved Master Plan.

Table - 7: Trial runs for MLR - attraction models for work trips

| Attractions Variables | $R^2$ | Corrected-$R^2$ | F-Value |
|---|---|---|---|
| Employment places | 0.5931 | 0.5866 | 91.82 |
| Employment places and other work area | 0.6097 | 0.5871 | 48.43 |
| Employment places ,other work and commercial areas | 0.6102 | 0.591 | 31.63 |
| Employment places and total work area | 0.61 | 0.597 | 40.51 |

The final sets of trip attraction models derived for peak and off peak period are listed below: ( refer Table 8)

Table -8 : Regression Equations for Trip attraction

| Purpose | Peak Period Model | | Off Peak Period Model | |
|---|---|---|---|---|
| Work Trips | -366.06+0.4955*WP | $R^2$ =0.636 | 4.225 + 0.0652 * WP | $R^2$ =0.70 |
| Education | -136.12 + 0.309 *SE | $R^2$ = 0.509 | -15.055 + 0.01520 *SE | $R^2$ = 0.22 |
| Other | 215.50 + 19.99 * CA + 0.0060 * P | $R^2$ = 0.163 | 390.73 + 14.5* CA + 0.00091* P | $R^2$ = 0.06 |

where
CA = Commercial area Hectares, WP = Work places, SE = Student enrollment, P = Population $R^2$ = The Coefficient of determination

## 12. VALIDATION OF MODELS

In order to test the reliability of the models, the results obtained in both cases have been applied to base year data and zonewise trips synthesised. The synthesised trips have been validated against the values in the matrix obtained from observed (HHI survey) trips.

Using regression models the results obtained will be closer to the one obtained by category technique especially for peak hour data but for off peak hour data the results from regression models are not very accurate. However in order to develop a model using a cross- classification technique the data base used should be quite large.

## 13. CONCLUSION

The household data base was comprehensive which was collected as a part of detailed Traffic and Transportation Study for the city Madras- India, and so an attempt is made to compare and appreciate the two methods in Trip Generation process namely; Multiple Linear Regression Analysis and Category or Cross Classification Analysis. In most of the developed countries Category Analysis approach has replaced the Multiple Linear Regression Method as later approach gave more accurate results. The main problem of Category analysis approach is that ; when multiple grouping is done some cells go underrepresented and tend to have abnormal values. For example in this case study, even though the data base is large i.e. about 22,250 households, when data was categorised, some cells in the off peak period show zero values. Hence the prediction for the future years becomes major problem . To overcome this problem, different approach as suggested by Stopher in one of his studies was adopted and it showed better results. Another problem encountered was the cells where grouping were done according to the income range and they did not show any logical trend and to over come this it was decided to group based on the vehicle ownership basis and it showed better results.

This paper has brought out that vehicle ownership taken as determinant in developing trip generation models, the results obtained using MLR and Category analysis do not differ much, spscially for peak hour data. This paper has also clearly brought out the high significance of growing vehicle ownership on trip making propensity in an Indian city.

## ACKNOWLEDGMENTS

The authors are grateful to Mr. Imtiyaz Ahmed - Chief Project Manager, RITES, Madras and Dr. S. Ponnuswamy, Retired A.G.M, Indian Railways for their valuable guidance during the study. The authors also thank the Madras Metropolitan Development Authority for the opportunity given to them in studying the problems in depth and for use of data collected for CTTS Study in the preparation of this paper. The authors are also thankful to the management of RITES - A Govt of India Undertaking and University of Malaya for allowing and encouraging the authors to prepare this paper.

## REFERENCES

M/S RITES, M/S KCL and M/S PTCS (1995) **Comprehensive Traffic and Transportation Study for Madras Metropolitan Area(CTTS)** - Study Report Madras Metropolitan Development Authority, Madras

Stopher, Peter R.., and Mc Donald, Kathig (1983) " Trip Generation by Cross Classification - An Alternative Methodology " **- Transportation Research Record- 844 TP 84/91 ,** Transportation Research Board, Washington.

Lane R., Dowell T.T. and Smith P.P. (1971) **Analytical Transport Planning** - Gerald Duckworth and Company, London.

Edward K.Morlok, **Introduction to transportation engineering and planning**. The McGraw-Hill Companies, New York

Sheldon M.Ross (1996) **Introductory statistics**.The McGraw-Hill Companies, New York