Journal in Urban and Regional Planning

UP School of Urban and Regional Planning
http://journals.upd.edu.ph/

Trip Generation Modeling in the Residential Areas of Downtown Cagayan de Oro City, Misamis Oriental, Philippines

Chris Lawrence O. Edulan
Department of Civil Engineering, Xavier University – Ateneo de Cagayan, Corrales Avenue,
9000 Cagayan de Oro City, Misamis Oriental, Philippines
ched.5882@gmail.com

Leasusana C. Ty
Department of Civil Engineering, Xavier University – Ateneo de Cagayan, Corrales Avenue,
9000 Cagayan de Oro City, Misamis Oriental, Philippines
leasusanaty@gmail.com

Anabel A. Abuzo
Department of Civil Engineering, Xavier University – Ateneo de Cagayan, Corrales Avenue,
9000 Cagayan de Oro City, Misamis Oriental, Philippines
aaabuzo@xu.edu.ph

Abstract

Rapid urbanization will adversely affect the community and its environment without careful planning and intervention. This study was conducted to investigate the trips generated by the people residing in residential communities in the downtown area of Cagayan de Oro City, Misamis Oriental, Philippines. It also developed a mathematical model to estimate the number of trips produced in the study area. A multiple regression analysis was employed to achieve these objectives. It was identified that the number of trips produced per household corresponded with the increase in the number of family members, the number of vehicles owned, the household's income level, and the dwelling type. The developed model will serve as a reference for regional and urban planners in forecasting future trips in the area and in the formulation of policies and recommendations that will advance the land use and transportation setting of the city.

Keywords: travel demand, trip generation, multiple linear regression, modeling, sustainable transportation planning

1. Introduction

For the past few years, the city of Cagayan de Oro has been subjected to a rapid increase in economic, infrastructure, and land use development, alongside the increasing demand for shelter, facilities, and access as the growing population continues to change its urban landscape. This state of urbanization and economic progression resulted in an increase in demand for mobilization and transport (Muromachi et al., 2015). Aside from the positive effects of urbanization, it also has negative implications to the people and society as it contributes to the rise in various transport issues, namely traffic congestion, road accidents, air pollution, climate change, energy use, and land take (Lohani, 2005; Puomanyvong et al., 2012; Fillone, 2015). Moreover, with an insufficient understanding of the trip patterns of the city and the population's travel behavior, these consequences are worsened, putting the city's sustainability at risk (Daniel and Ituen, 2013; Aloc and Amar, 2013).

One way to alleviate the worsening condition is to develop an effective and sustainable transport system that primarily contributes to conserving the environmental, social, and economic sustainability of a city (Lohani, 2005). For this to be achieved, empirical studies should be conducted to provide sufficient knowledge pertaining to the trip patterns of the city (Daniel and Ituen, 2013; Aloc and Amar, 2013). Therefore, there is a good reason to conduct local travel demand studies to serve as basis for planning and management of local traffic problems (Soltani and Ivaki, 2011).

Travel demand models are used to determine the number of vehicles that would occupy a certain road or any transportation element. The models give transport planners an idea of what problems or complications would be expected in the future and the possible management strategies and solutions to carry out in the process (Padmini et al., 2013). Travel Demand Modelling includes four steps:

Trip Generation, Trip Distribution, Modal Split, and Route/Traffic Assignment.

This study aimed to conduct a trip generation study in the residential areas of downtown Cagayan de Oro. Specifically, the research intended to establish a baseline data collection procedure on trip generation to identify the various factors affecting trip generation and formulate a general trip generation model of the study area.

Trip generation is the first step in the conventional four-step transportation forecasting process, which is used to forecast travel demands of a certain area. It is concerned with determining the number of trips that will begin or end in each traffic analysis zone within a study area (Garber and Hoel, 2009; Huntsinger et al., 2013; Sanghai, 2014). The quantification of trip generation essentially involves the development of models based on several factors affecting the travel patterns in the study area. The prime consideration for the formulation of trip generation models is the compilation of baseline data acquired from surveys and existing local data. Individual preferences drive the socioeconomic data, and trip characteristics and circumstances are derived from the survey data. Furthermore, Aloc and Amar (2013) specified in their study, Trip Generation of Lipa City, that the most significant parameters affecting the production and attraction of trips in a particular zone are the zone's population, number of households, number of students, and household head monthly income. Additionally, vehicle ownership and the type of dwelling have considerable effects on trip generation (Gadepalli et al., 2013; Padmini et al., 2013).

Trip characteristics refer to trip purpose, trip origin and destination, mode of transport, and travel time. Trip purpose is defined as the traveler's motivation to travel, directed to different types of land use (Ewing, et al., 2011). In this study, four home-based trip purposes were used in profiling trip characteristics in the residential areas of downtown Cagayan de Oro City: (1) education trips, (2) work trips, (3) market trips, and (4) social and recreational trips. The trip origin and destination primarily account for the direction of travel. The trip origin serves as the starting point of trips mostly generated by households in traffic analysis zones. On the other hand, the trip destination serves as the endpoint of trips, as these generate attraction to the travelers (Garrick, 2008). The transport mode is the means of transportation the traveler is willing to avail (e.g., jeepneys, motorela, bicycle, and private cars); whereas the travel time is the particular time of the day and the days in a week the traveler is driven to travel (Garber and Hoel, 2009). Trip origin and destination, mode of transport, and travel time are the key information to represent the dependent variables in the trip generation models. For various trip purposes, separate models must be developed, considering that travel behavior essentially depends on trip purpose.

Identifying the independent variables permits the formulation of trip generation models from which the number of trips that will begin or end in each traffic analysis zone is quantified (Garber and Hoel, 2009).

Since the study employed several explanatory variables (i.e., household size, number of employed persons, number of students, number of persons with a driver's license, and the number of vehicles owned by the household), a multivariable linear regression analysis was utilized in the development of the general trip generation model. The most basic linear regression model between an explanatory variable X and the mean response variable Y takes the form $E(Y) = \alpha + BX$. This model, which contains a single predictor variable, is referred to as a bivariate model since it contains only two variables (i.e., Y and X). However, multivariable regression problems start with a collection of potential predictors.

A set of terms are created from the pool of potential predictors, which appear in the multivariable regression model as the X variables (Weisberg, 2005). In a multivariable study, the explanatory variables are denoted $X_1, X_2, X_3, \ldots X_k$, where k denotes the number of predictors. The bivariate model then generalizes to:

$$E(Y) = \alpha + B_1X_1 + B_2X_2 + B_3X_3 + \cdots + B_kX_k \quad (1)$$

2. Methodology

2.1 Survey Instrumentation

This study made use of the household interview survey as the mode of gathering the necessary primary data for this research. A household interview survey is concerned with collecting facts related to the population's socioeconomic characteristics and trip movements made on a typical day within the study area (Sekhar et al., 1997). It is a crucial tool in identifying the explanatory variables as it could improve the method in the analysis of trip generation (Rhee, 2003).

Additionally, existing local data statistics, which includes the quantification of demographic, economic, as well as previous and current transport data, were obtained from different government agencies, such as the Philippine Statistics Authority (PSA), City Planning and Development Office (CPDO), Cagayan de Oro City Roads and Traffic Administration (RTA), and Land Transportation Office (LTO).

2.2 Unit of Analysis

Variable identification is highly dependent on the chosen unit of analysis in the process of trip production modeling. There are two types of units of analysis to consider: the household, and the individual. However, between the two, the household unit is widely used and highly preferred by most studies (Padmini et al., 2013) since households are considered the main trip generator contributing to more than 80% of generated trips in an urban area. It is in consideration of the family as the basic unit of society to which all needs of the individual are usually met, and the household as a place where nearly all trips start and end. Also, when the household is considered the base unit, the individual income of each working member in the family is consolidated into the household's total income. Likewise, vehicles owned by specific family members are included as part of the total number of vehicles owned by the household. It unifies all complex elements into a single unit of measure. Allocating the degree of contribution and ownership of the mentioned variables becomes more difficult to analyze when considering individuals as a base unit, or a different set of qualities would have to be considered (Dodeen, 2014).

2.3 Sample Size and Respondents per Barangay

In conducting a survey, selecting the appropriate sample size to represent the entire population is essential. A formula that considers the margin of errors and confidence levels must be utilized. For this study, the researchers utilized Slovin's formula since the equation is based on the estimation of the population proportion with a specified value for the margin of error.

Using the 2010 PSA Census of Population and setting the confidence interval to 95%, the total number of households included in the survey is 400. Upon determining the sample size, the calculated value was further stratified to get the required number of respondents per *barangay*. The sample size distribution was calculated based on the number of household units of each of the 40 *barangays*.

The study surveyed 419 household units, which is 18.59% of the total household population of downtown Cagayan de Oro. Furthermore, 80% of the gathered data was used for model development, while the remaining 20% was used for model validation.

2.4 Variables Used in the Model Formulation

In the data analysis, multiple linear regression (MLR) was used, where the main goal was to develop an equation that would predict the number of trips generated per household (dependent variable) of the study area with the use of the acquired data (independent variables). Table 1 shows the list of the independent variables used in the modeling process. The age variables are divided into four groups. The people who fall under the first age group (between six to 16 years old) are assumed to be elementary and high school students. The people included in the second age group (between 17 to 30 years old) are considered to be either continuing tertiary level of education or already part of the workforce. The third age group (between 31 to 45 years old) and fourth age group (between 46 to 65 years old) encompass household heads and people who are part of the workforce or have retired.

Table 1. Independent variables setting

Variable	Variable assignment
X_1	Household size
X_2	Number of employed persons in the household
X_3	Number of students in the household
X_4	Number of persons with ages between 6 years old to 16 years old
X_5	Number of persons with ages between 17 years old to 30 years old
X_6	Number of persons with ages between 31 years old to 45 years old
X_7	Number of persons with ages between 46 years old to 65 years old.
X_8	Number of persons with driver's license
X_9	Number of vehicles owned by the household

Table 2 shows the dummy variables considered in the development of the trip generation model. Dummy variables are numerical variables extracted from independent variables with numerous categories or subgroups. In this study, the variables which were subdivided into individual categories were gender distribution of the household members (S1 and S2), household average monthly income (I1, I2, and I3), the type of dwelling (D1, D2, and D3), and house ownership (O1).

Table 2. Dummy variables setting

Variable	Dummy Variable	Binary Coding
S_1	Gender of the head of household	0 if male and 1 if female
S_2	Gender of household members	0 if male and 1 if female
I_1	Average Income of the household	monthly income of PhP10,001.00 to PhP 15,000.00 (0 if No and 1 if Yes)
I_2		monthly income of PhP15,001.00 to PhP 20,000.00 (0 if No and 1 if Yes)
I_3		monthly income exceeding PhP20,001.00 (0 if No and 1 if Yes)
D_1	Dwelling Ownership	Multi-storey type (0 if No and 1 if Yes)
D_2		Apartment type (0 if No and 1 if Yes)
D_3		Boarding type (0 if No and 1 if Yes)
O_1	House Ownership	0 if Rented and 1 if Owned

Note: PhP10,000.00 is approximately equivalent to 201.96 USD, where 1 USD = PhP49.51

2.5 Statistical Treatment of the Data

In the study, descriptive and inferential statistics were used. Descriptive statistics allowed the use of correlation and collinearity diagnostics. On the other hand, inferential statistics were used to develop a trip generation model and conduct other multivariate methods such as t-test, analysis of variance (ANOVA), regression analysis, and factor analysis.

2.5.1 Multiple linear regression analysis

One of the most widely used statistical tools in model development is the MLR. This approach was utilized in the study to model the linear relationship between a dependent variable (Y or the number of daily trips) and one or more independent variables (i.e., household size, number of employed persons in the household, and number of students in the household).

This study used the variable "Y" to represent the dependent variable which is the trips produced or attracted to a specific zone. Variables "$x_1, x_2, x_3, \ldots x_n$" were used for the independent variables that would predict the value of the dependent variable. "$\beta_1, \beta_2, \beta_3, \ldots \beta_n$" represented the regression coefficients, which were obtained through the regression analysis, and "ε" as the random error term. In general, the model appears in the form of:

$$Y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \pm \varepsilon \qquad (2)$$

2.5.2 Modeling Process

In the regression analysis, 80% of the data was used for the model development, while the remaining 20% was utilized to validate the resulting model. To produce the regression model (Dooden, 2013; Raagas and Sinco, 2013), the following procedure should be followed:

1. Inspect the relationship between the dependent variable (Y or the number of daily trips) and each independent variable (i.e., household size, number of employed persons in the household, number of students in the household) to check for noncollinearity. When non-collinearities are detected, linearization of the relationship must be executed by transforming the dependent variable, the independent variable, or both.
2. A correlation test involving all explanatory, as well as collinearity diagnostics, must be conducted. Examine the result to check whether a multicollinearity problem exists. If the issue is present, either of the two solutions can be done: submit the variables to factor analysis or eliminate one of the highly correlated variables.
3. Upon selecting the correlated independent variables, the regression models can be formulated.
4. Every formulated model should include several statistical tests such as coefficient of determination (R^2), t-statistics for the variables of the models, and the overall goodness of fit of the models through the F-test.
5. When all the models are assessed based on the test results, the outcome gathered from the fourth step is summarized and assessed to select the best model.

2.5.3 Variable Selection

Since the goal of the MLR is to develop a model with a good estimation capacity, a variable selection procedure intended to determine the explanatory variables (i.e., household size, number of employed persons in the household, number of students in the household) having significant contributions in explaining the dependent variable (Y or the number of daily trips) is essential. The backward elimination and forward selection models are suggested in selecting the independent variables to be included in the model. Being the simplest between the variable selection procedures and the easiest to implement, the technique employed to formulate the model was the backward elimination method.

This approach starts with all the independent inputs in the equation. Then, independent variables with a p-value greater than the level of significance (α_{crit}) are eliminated because these p-values indicate that it has the least significance among other variables. In this study, the confidence level used was set to 95%; therefore, an explanatory variable with a p-value larger than 0.05 was removed from the model. The next steps are refitting the model and reassessing the partial contributions of the remaining variables to the model. Several iterations may be made before arriving at a model that fits the data satisfactorily until p-values do not exceed the value of α_{crit} (i.e., 0.01, 0.05, and 0.10).

2.6 Statistical Tests

2.6.1 Test for Multicollinearity

One of the issues that could arise in regression analysis is multicollinearity. This problem occurs when two independent variables are highly correlated to each other, thus, affecting the interpretation of the relationship between the independent variables and the dependent variable. Two processes are being considered to detect multicollinearity. One is to perform collinearity diagnostics, which involves the determination of the tolerance value (TV) and the variance inflation factor (VIF) value.

These numbers depict the degree to which each explanatory variable is explained by other independent variables.

Furthermore, multicollinearity exists (Raagas and Sinco, 2013) if the TV is lesser than 0.10 and the VIF is greater than 5.0. Another way is to investigate ahead of the correlations between the independent variables. Parameters having a correlation coefficient greater than 0.70 indicate that an issue of multicollinearity exists (Raagas and Sinco, 2013). In order to solve the problem, one can either (1) submit the variables to a factor analysis which is a method used to create artificial variables from the original ones in such a way that the new variables are uncorrelated, or (2) eliminate one of the highly correlated variables having the lowest correlation with the dependent variable.

2.6.2 Coefficient of Determination

The goodness of fit of the multiple regression models can be evaluated through the R^2. Moreover, it is the proportion of variation in the dependent variable that can be explained by the explanatory variables. The coefficient of determination takes on values between 0 and 1. An R^2 value closer to 1 indicates that the model has a good fit; hence, the higher the R^2 value, the more useful the model becomes. However, this number should not be considered one of the criteria for selecting the MLR model since it always grows as the number of explanatory variables increases (Gupta, 2002), therefore, serving only as a summary measure of goodness of fit.

2.6.3 Overall Test of Significance

In the MLR, F-test measures whether the model is significant or not, hence, testing the overall significance of the equation. Moreover, it tests whether the R^2 is notably closer to 0. In testing the overall goodness of fit of the model, examining the following hypotheses is necessary: (a) Null hypothesis (H_o): $\beta_1 = 0$, and (b) Alternative hypothesis (H_1): $\beta_j \neq 0$, for at least one of the values of j. According to Gupta (2002), the acceptable F-values are as follows: (a) If F-value < 0.01, the model is significant at 99%; (b) If F-value < 0.05, the model is significant at 95%; (c) If F-value < 0.10, the model is significant at 90%. These values indicate whether the formulated model is acceptable or not. Typically, a model with an F-value greater than 0.05 signifies that it could not fit the data. Thus, the regression, as a whole, has failed.

2.6.4 Testing the Significance of the Explanatory Variables

The t-statistics measure the significance of the parameters included in the regression equation. The acceptable p-values for the t-test are the same as the ones specified in the F-test. In interpreting these numbers, three criteria are specified: (a) if p-value < 0.05, the coefficient estimate is assumed to be as true with a level of confidence of 95%; (b) if p-value > 0.10, the coefficient estimate is unreliable and statistically insignificant; and (c) if $0.05 <$ p-value < 0.10, it can only be asserted that the accuracy of the coefficient estimate with a 90% level of confidence (Gupta, 2002). Since the study used a confidence level of 95%, the variable's level of significance (p-value) considered in developing the regression model was 0.05 or lower.

ANOVA is a statistical tool used to determine the differences and relationships between group means. It also shows the variation among and between groups of variables.

3. Results and Discussion

3.1 Interpretation of Regression Coefficients

With the use of cross-section data acquired from 445 respondents from Barangay 1 to Barangay 40, the general trip generation model was estimated used multiple linear regression (MLR) analysis. As presented in Table 2, four (4) of the thirteen (13) independent variables (i.e. gender distribution (S), monthly income categories (I), type of dwelling (D), and house ownership(O)) are categorical and thus must be subdivided into dummy variables prior to the development of trip generation model. The remaining nine (9) non-dummy variables used in the formulation of the model are as well presented in Table 1.

In the final model, all regression coefficients (β_1, β_2, β_3, β_5, β_9, β_{12}, and β_{17}) have a positive sign indicating a positive association between the predictor variables and the number of trips generated. This means that an increase in the household size (X_1), number of employed (X_2), number of students (X_3), number of persons between 17 years to 30 years of age (X_5), and the number of vehicles owned (X_9))entails an increase in the number of trips generated. Considering the dummy variables (i.e. I_1, I_2, I_3, D_1, D_2, D_3, and O_1), only households whose average monthly income is between Php10,001.00 and Php15,000.00 (I_1) and whose dwelling type is a boarding house (D_3) shows a positive relationship to the number of trips generated.

In view of the fact that household members themselves are the performers of the trips, a greater household size shall correspond to a higher number of daily trips. Similarly, the number of employed household members and students affects work and education trips, which in turn contributes to the total number of daily trips. The age range of 17 years old to 30 years old are the ages when members of the household receive tertiary education and for some, start to be a part of the labor force generating work trips. The ownership of vehicles also provides access to the generation of trips raising the total number of daily trips. The dummy variables retained in the final general trip generation model confirm the particular household income group and the type of dwelling which gives the highest probability to generate trips. Constituting the final trip generation model, households having an average monthly income of Php10,000.00 to P15,000.00 are more likely to travel compared to any other households belonging to other income groups. Finally, households whose dwelling type is a boarding house are the most likely to generate trips.

3.2 Testing Individual Coefficients

For the testing of individual coefficients, only variables with a 95% level of significance or higher were retained in the final general trip generation model. Shown in Tables 3 and 4 are the t-values and significance levels of each explanatory variable. The study used the following hypothesis: H_{O-1}: The explanatory variables (such as household size and number of employed) do not significantly influence the number of daily trips per household. H_{a-1}: The explanatory variables (such as household size and number of employed) significantly influence the number of daily trips per household.

The t-value for the coefficient of X_1 (household size) is 3.466 (at p=0.001). This value would lead us to reject the null hypothesis (H_{O-1}). Thus, the household size has a significant effect on the total number of daily trips per household. Similarly, the coefficient of X_2(number of employed) has a t-value of 13.171 (at p=0.000). This value would lead us to reject the null hypothesis (H_{O-1}). Thus, the number of employed persons in the household has a significant effect on the total number of daily trips per household. The t-statistic for the hypothesis that the number of students (X_3) has no effect on the number of trips is 16.779 (at p=0.000). This value would lead us to reject the null hypothesis (H_{O-1}). Thus, the number of students in the household has a significant effect on the total number of daily trips per household. The coefficient of X_5 (number of persons between 17 years to 30 years of age) has a t-value of 2.831 (at p=0.005). This value would lead us to reject the null hypothesis (H_{O-1}). Thus, the number of persons between 17 years to 30 years of age in the household has a significant effect on the total number of daily trips per household. The t-statistic for the hypothesis that the number of vehicles owned (X_9) has no effect on the number of trips is 2.615 (at p=0.009). This value would lead us to reject the null hypothesis (H_{O-1}). Thus, the number of vehicles owned by the household has a significant effect on the total number of daily trips per household. Finally, the hypotheses that have a monthly income of Php10,000.00 to Php15,000.00 (I_1) and a dwelling type of boarding house (D_3) have no effects on the number of trips have t-values of 2.315 (at p=0.021) and 3.326 (at p=0.001), respectively. These values would lead us to reject the null hypothesis (H_{O-1}). Thus, having an average monthly income of Php10,000.00 to Php15,000.00 and a dwelling type of boarding house have significant effects on the total number of daily trips per household. In summary, all of the independent variables (X_1, X_2, X_3, X_5, X_9, I_1, $and D_3$) in the model are significant at 95% level of significance. This means that each of the independent variable has a significant effect on the number of daily trips.

Table 3. Regression coefficients for the independent variables of the general trip generation model 1 (Number of daily trips per household, Y)

Model	**Unstandardized Coefficient**		**Standard Coefficient**	**T**
	Beta	*Standard Error*	*Beta*	
α	0.278	0.114		2.437
X_1	0.120	0.034	0.1501	3.466
X_2	0.573	0.043	0.4110	13.17
X_3	0.799	0.047	0.5480	16.77
X_5	0.112	0.039	0.0861	2.831
X_9	0.147	0.056	0.0620	2.615
I_1	0.266	0.115	0.0545	2.315
D_3	1.283	0.385	0.0778	3.326

Table 4. Regression coefficients for the independent variables of the general trip generation model 2 (Number of daily trips per household, Y)

Model	Significance	Collinearity Statistics	
		Tolerance	VIF
α	0.015		
X_1	0.000	0.290	3.44
X_2	0.000	0.558	1.79
X_3	0.000	0.509	1.96
X_5	0.004	0.588	1.70
X_9	0.009	0.968	1.03
I_1	0.021	0.979	1.02
D_3	0.001	0.994	1.00

3.3 Testing for Multicollinearity

Another problem in MLR technique involves the situation in which the independent variables are highly correlated with each other. This is referred to as multicollinearity. To test for multicollinearity, the correlation matrix is usually obtained. In the correlation matrix, the entries on the main diagonal give the correlation of one variable with itself, which is always 1 by definition. The entries off the main diagonal are the pair-wise correlations among the explanatory variables. Referring to the Pearson correlation matrix shown in Table 5, all of the pair-wise correlation coefficients are less than 0.70, suggesting that there is no severe multicollinearity problem. Included in Table 5 are the results of multicollinearity diagnostic tests quantified by the variance inflation factors (VIFs) and tolerance values of each explanatory variable involved in the model. Since the tolerance values of all the variables are greater than 0.10 and the VIF values are less than 5, the problem of multicollinearity does not exist in the final general trip generation model.

Table 5. Pearson Correlation Matrix for Independent Variables

	X_1	X_2	X_3	X_5	X_9	I_1	D_3
X_1	1	0.560	0.608	0.603	0.092	0.083	-0.030
X_2		1	0.085	0.473	0.164	0.084	0.001
X_3			1	0.209	0.022	0.110	0.002
X_5				1	0.050	0.049	-0.025
X_9					1	-0.013	-0.047
I_1		.				1	0.001
D_3							1

3.4 Testing the Goodness of Fit: Coefficient of Determination

The coefficient of determination (R^2) is a measure of the goodness of fit of the model and is interpreted as the proportion of the variance in the dependent variable (Y) that is predictable from the independent variables (X_1, $X_2, \dots X_n$)included in the regression model. Otherwise known as the R-squared value, it is the square of the correlation coefficient (R) between the predicted Y values and actual Y values. Table A (Appendix) is an Analysis of Variance (ANOVA) which shows the total variance of the general trip generation model. This variance is divided into the regression sum of squares (RSS) and the total sum of squares (TSS), the ratio of which is the coefficient of determination (R^2) of the trip generation model. For this study, an R^2 value of 0.8113 was obtained. This value indicates that 81.13% of the variance in the total daily trips (Y) is explained by the household size (X_1), number of employed (X_2), number of students (X_3), number of persons between 17 to 30 years of age (X_5), and the number of vehicles owned (X_9). Moreover, having an average monthly income of Php10,000.00 to Php15,000.00 (I_1) and a dwelling type of boarding house (D_3) increases the likelihood for households to generate trips. The remaining 18.9% is explained by other factors not included in the model. In the case of cross section data, such an R^2 value is considered reasonable.

3.5 Overall test of Significance of the Model

From the ANOVA, the Significance of F value (Sig.) indicates the probability that the final general trip generation model was merely obtained by chance. A small Significance of F affirms the validity of the regression model. Since the obtained significance of F (Table A, Appendix) value is 1.62×10^{-121}, which is almost equal to 0, it means that there is almost 0% chance that the regression model was merely a chance occurrence, thus increasing the overall accuracy and significance of the formulated model. Since this is a highly significant value, the hypothesis that all the independent variables entered into the regression model (X_1, X_2, X_3, X_5, X_9, I_1, $and D_3$) have no impact on the number of daily trips is rejected at approximately 100% level of significance in favor of the alternative hypothesis that these variables jointly affect the total daily trips.

3.6 Model Validation

For the general trip generation model to gain credibility, it is essential that the results be validated by comparing the predicted number of daily trips to the observed number of daily trips. The predicted daily trips are the resulting Y values using the final model upon the substitution of the independent variables from 20% of randomly selected cases from the 445 respondents. This comparison tests the ability of the trip generation model to predict future behavior. For each randomly selected observation, the results demonstrate that the model is capable of generating satisfactory estimates as validated by the good match between the predicted and actual daily trips per household. The root-mean-squared error (RMSE) of 0.8061 indicates that there is an average discrepancy of ±0.8061 number of trips between the predicted and actual number of trips in each case that the regression model is employed. Considering the overall predicted number of daily trips from 90 randomly selected observations, 0.213% of the total is in excess or deficit of the overall actual number of daily trips. Generally, the model is successfully validated by comparing the predicted results with the actual data. Thus, the results support the final general trip generation model as a valid model for predicting number of trips per household.

3.7 General Trip Generation Model

Subsequent to the treatment of data by outlier detection testing, multicollinearity testing, and dimension reduction operations was the backward elimination of insignificant variables in each iteration of the modeling process. The criterion for the elimination of insignificant variables is to retain variables having a p-value of 0.05 or lower attaining a 95% level of significance. The summary in Table 3 and 4 are the regression results for the estimated general trip generation model. The final estimated trip generation model is:

$$Y = 0.278 + 0.121X_1 + 0.574X_2 + 0.799X_3 + 0.113X_5 + 0.147X_9 + 0.267I_1 + 1.283D_3$$

4. Conclusion

This study established the following factors to be considered in the data collection procedure in developing a trip generation model: (1) the selection of the study area, (2) the determination of the specific land use, (3) the definition of the unit of analysis, (4) the statistical determination of the sample size, and (5) the formulation of the survey tool. This study also identified that the significant variables affecting the trip generation (Y) of the downtown residential area of Cagayan de Oro City are: the household size (X_1), the number of employed persons in the household (X_2), the number of students in the household (X_3), the number of household members and their corresponding age range (X_5), the number of vehicles owned by the household (X_9), the average monthly income (I_1), and the dwelling type (D_3). Using the multiple linear regression analysis, the final estimated general trip generation model is:

$$Y = 0.278 + 0.121X_1 + 0.574X_2 + 0.799X_3 + 0.113X_5 + 0.147X_9 + 0.267I_1 + 1.283D_3$$

Based on the derived model, it was concluded that the number of trips per household corresponds with the increase in the number of members (or size), the number of vehicles, the income level, and the dwelling type per household. Moreover, in a model-specific scale, an increase in household size will increase the plausible number of employed members, the number of students, and the number of persons with ages between 17 years old to 30 years old. The degree of change in the independent variables in the trip generation model should be taken into careful consideration in order to understand the variations in the number of trips produced by the household. It is highly recommended that the study and the developed model will be employed as a reference and as a tool for regional and urban planners, particularly in the formulation of policies and recommendations that will advance the land use and transportation setting of the City.

Acknowledgment

We would like to acknowledge Engr. Kirt Mufty S. Isnani, Engr. Theodore Nikko G. Lauron, and Engr. Anna Louise V. Saballa for contributing significant inputs in the overall conduct of the study. We would also like to convey our sincerest gratitude to the Kinaadman Research Center (KRC) for funding our research.

References

Aloc, D., and Amar, J. (2013). Trip Generation Modelling of Lipa City. Retrieved July 23, 2015, from https://www.academia.edu/7601857/Trip_ Generation_Modelling_of_ Lipa_City

Daniel, E., and Ituen, U. (2013). Inter-Urban Trip Generation Models for the Urban Centers in Akwa Ibom State, Nigeria. Civil and Environmental Research, 3(4), 55-62. Retrieved July 18, 2015, from http://www.iiste.org/Journals/index.php/ CER/article/view/5048/5522

Dooden, A. (2014). Developing Trip Generation Models Utilizing Linear Regression Analysis: Jericho City as a Case Study. Retrieved July 16, 2015, from http://scholar.najah.edu/sites/default/ files /Alaa Dodeen.pdf

Ewing, R., Greenwald, M., Zhang, M., Walters, J., Feldman, M., Cervero, R., and Thomas, J. (2011). Traffic Generated by Mixed-Use Developments – Six-Region Study Using Consistent Built Environmental Measures. Journal of Urban Planning and Development, 137, 248-261.

Fillone, A. M. (2015). Evaluating Proposed Transportation Infrastructure Projects in Metro Manila Using the Transport Co-Benefit Analysis. Journal of Eastern Asia Society for Transportation Studies, 11, 189-208. Retrieved January 16, 2016, from https://www.researchgate.net/publication/288824278_Evaluating_Proposed_Transportation_Infrastructure_Projects_in_Metro_Manila_Using_the_Transport_Co-Benefit_Analysis.

Gadepalli, R., Jahed, M., Ramachandra Rao, K., and Tiwari, G. (2014). "Multiple Classification Analysis for Trip Production Models Using Household Data: Case Study of Patna, India." Journal of Urban Planning Development, 140(1), 05013001.

Garrick, N., Marshall, W., Jackson, E. (2008). Value Pricing in Connecticut. The University of Connecticut. Connecticut Transportation Institute.

Garber, N., &Hoel, L. (2009). Traffic and highway engineering (4th ed., pp. 549-603). Toronto: Cengage Learning.

Gupta, V. (2002). SPSS for Beginners. Retrieved December 15, 2015, from http://instructional1.calstatela.edu/ikim/mkt446/spss/SPSSforBeginners.pdf

Huntsinger, L., Rouphail, N., and Bloomfield, P. (2013). "Trip Generation Models Using Cumulative Logistic Regression." Journal of Urban Planning Development, 139(3), 176–184.

Lohani, B. (2005). Asian Urbanization, Transport Development, and Environmental Sustainability: Is Sustainable Transport still Possible for Asian Cities? Retrieved July 31, 2015, from cleanairasia.org/portal/system/files/presentations/58966_paper.doc

Muromachi, Y., Lim, I., Wicaksono, A., Vergel, K. N., Choocharukul, K., Tan, V., . . . Yai, T. (2015). A Comparative Study on Road-based Urban Public Transport Policies in Six Asian Countries from the Viewpoint of Governance, Urban Planning, and Financial Aspects. Journal of Eastern Asia Society for Transportation Studies, 11, 1433-1450. Retrieved January 16, 2016, from http://www.dynamicglobalsoft.com/easts2015/program/pdf_files/1383.pdf

Ortuzar, J., and Willumsen, L. (2011). Modelling Transport (4th ed.). John Wiley and Sons.

Padmini, A.K, Abdul K.V., and Paneckel, L. (2013). Forecasting Trip Production Based on Residential Land Use Characteristics. International Journal of Innovative Technology and Exploring Engineering, 3(2), 55-60. Retrieved July 31, 2015, from http://www.ijitee.org/attachments/File/v3i2/B1024073213.pdf

Poumanyvong, P., Kaneko, S., and Dhakal, S. (2012). Impacts of Urbanization on National Transport and Road Energy Use: Evidence from low-, middle- and high-income countries.

Retrieved July 31, 2015, from https://ideas.repec.org/p/hir/idecdp/2-2.html

Raagas, E. L., and Sinco A. L. (2013). Advanced Data Analysis Manual: A Simplified Approach. Cagayan de Oro City. DATStat Analysis Center.

Rhee, J. (2003). Improvement of Trip Generation Forecast with Category Analysis in Seoul Metropolitan Area. Journal of the Eastern Asia Society for Transportation Studies, 4, 818-831 Retrieved July 31, 2015, from http://www.easts.info/2003proceedings/papers/0818.pdf

Sanghai, S. (2014). Development of Trip Generation Model Using Activity Based Approach. Retrieved September 2, 2015, from http://www.researchgate.net/profile/Sanket_Sanghai/publication/262419972_Development_of_Trip_Generation_model_using_Activity_based_Approach/links/00463537b40ad51552000000.pdf

Sekhar, S., Anand, S., and Karim, M. (1996). Comparison of Regression Model and Category Analysis (A Case Study). Journal of the Eastern Asia Society for Transportation Studies, 2(3), 917-929. Retrieved July 23, 2015, from http://easts.info/on-line/journal/vol2no3/23018.pdf

Sterling, D. L., (2003). A Comparison of Spatial Interpolation Techniques for the Determining of Shaling Rates of the Atlantic Ocean Channel. Virginia Polytechnic Institute and State University. Retrieved September 2, 2015 from https://theses.lib.vt.edu/theses/available/etd-09162004-094346/unrestricted/Full_Thesis_Doc_9_28_04.pdf

Soltani, A., and Ivaki, Y. E. (2011). The Influence of Urban Physical Form on Trip Generatiom, Evidence from Metropolitan Shiraz, Iran, Indian. Journal of Science and Technology, 4(9), 1168-1174. http://www.indjst.org/index.php/indjst/article/viewFile/30250/26185

Weisberg, S. (2005). Applied Linear Regression. (3rd ed.). Hoboken, New Jersey: John Wiley and Sons, Inc.

APPENDIX

Table A. ANOVA for the General Trip Generation Model

	Sum of Squares	d_f	Mean Square	F	Sig.
Regression	1089.016	7	155.574	213.128	.0000
Residual	253.294	347	.730		
Total	1342.310	354			