Australian Journal of Basic and Applied Sciences, 3(1): 194-200, 2009
ISSN 1991-8178

Determining the Likelihood Distribution of Average Household Trip Production (Case Study of Isfahan City)

[1]Ali Mansour Khaki, [2]Shahryar Afandizadeh, [3]Reza Moayedfar

[1]Associate professor of civil faculty, Iran University of Science & Technology
Email Address: mkhaki@iust.ac.ir
[2]Associate professor of civil faculty, Iran University of Science & Technology
Email Address: zargari@iust.ac.ir
[3]PHD candidate of transportation engineering, Iran University of Science & Technology
Phone: 00982144643158, Email Address: rmoayedfar@iust.ac.ir

Abstract: Background: at present study, the focus is on average of household trip production. Each municipal society is divided to household's categories based on socio-economic characteristics. Households in Each category produce the several numbers of trips. The purpose of this study is to determine and also predict the likelihood function of average trip production in each household's category. Methods: The city of Isfahan was selected as case study and the results of origin-destination survey conducted in this city were utilized. First, 15073 households were classified into 195 category based on their socio-economic characteristics. Then, based on "weighted mean" method, average of household trip production was calculated for each category. After that, likelihood distribution function was determined for average trip production of households. At last level, in order to insert households' socio-economic variables in the process, disaggregate trip generation model was calibrated. Results: χ^2 Test showed that average rate of household trip production follows gamma continuous distribution. Statistical indices and the validation of calibrated model was an evidence of its precision in predicting household trip production. Conclusions: with determined likelihood function, it would be possible to estimate at municipal society (like Isfahan city), what the probability of produced 1 trip averagely, 2 trips averagely and … is? Also it would be possible to predict average household trip production based on socio-economic characteristics (with disaggregate model) for each household's category.

Key word: disaggregate model, expectative frequency, gamma continuous distribution function, observed frequency, socio-economic characteristics.

INTRODUCTION

According to basic statistical concepts, each statistical society follows its own likelihood distribution function. One of these societies is the rate of trip production at municipal areas. There are several studies which have been done at this field. At mentioned studies the types of continuous probability functions has been utilized to determine the density function of average household trip production.

The probability distribution function of average household trip production can estimate likelihood of this parameter for each household's category. But there is no significant relation between this parameter and socio-economic characteristics. The purpose of this study is to estimate the likelihood of average household trip production based on socio-economic characteristics.

MATERIALS AND METHOD

The methods which have been utilized at present study, have divided to three sections. First, is the method of "weighted mean" to convert 15073 household data to 195 household categories at Isfahan city. Second, is the χ^2 test to determine probability distribution function of average household trip production And third method, is the utilizing of regression models to calibrate disaggregate model of average household trip production.

Corresponding Author: Ali Mansour Khaki, Associate professor of civil faculty, Iran University of Science & Technology;
Email Address: mkhaki@iust.ac.ir

The conducted survey in Isfahan and the prepared data:

Isfahan has 12 urban regions and 190 internal zones. In the origin-destination studies in Isfahan two major groups are recognized: first, trips of Isfahan residents and trips of non-residents in Isfahan. Isfahan citizens' trips make up the major portion of the trips in Isfahan. The information gathered about these trips is obtained through origin-destination survey and by direct interviews at people houses. In brief, in the origin-destination survey done in Isfahan, groups of male students of 3rd grade of junior high school were cooperating in filling out the related questionnaire designed to study the origin-destination. Each student was supposed to fill out two questionnaires, one for his own family and household and one for a neighbor. The questionnaire included some questions about the socio-economic characteristics of the household and its members and also some questions about the daily produced trips. The first page of the questionnaire included three questions including: 1. number of people in the household, age, gender, occupation and whether they did or did not have a driver's license, 2. second question was regarding the auto ownership per household considering the type of vehicle 3. the address of living location.

The second page of the questionnaire regarded registering the trips made by the household. 15073 households were studied in Isfahan which includes 4.5 percent of the population of the city. Based on the conducted origin-destination studies in Isfahan, the required data for the present investigation has been prepared. These data include the family size, the number of drivers per household, auto ownership per household, and the household trip production rate. The related information about the 12 urban -municipal-regions in Isfahan has been presented in table[1].

Table 1: Trip production and socio-economic characteristics of the households in 12 urban regions in Isfahan.[1]

Municipal Area	Household size	Number of drivers per household	Auto ownership per household	Number of employee per household	Household Trip production rate
1	4.13	1.42	0.64	3.25	7.55
2	3.90	1.53	0.74	3.03	7.90
3	4.24	1.24	0.51	3.30	7.74
4	4.47	1.33	0.56	3.51	7.23
5	4.65	0.98	0.37	3.69	7.16
6	4.55	1.25	0.54	3.61	7.33
7	4.32	1.44	0.70	3.42	8.48
8	4.22	1.56	0.66	3.33	7.42
9	3.63	1.34	0.60	3.25	7.43
10	4.43	1.15	0.53	3.46	7.70
11	4.59	0.95	0.38	3.60	8.02
12	4.25	1.00	0.55	3.44	6.94

The method of categorizing the household's socio-economic characteristics and determining the average rate of trip production for each category:

In late 1960s, a new and innovative method titled Cross Classification, replaced the old Regression method in England. In this method, household trip production rate was defined as a function of socio-economic characteristics of households. The main assumption in this method is that the rate of trip production for each category is constant over time. The major problem in this method is predicting the number of households in each category in the horizon year of the plan.(Ortuzer, J.D. and G. Luis, 2001. Caldwell, L.C. *et al.*, 1978).

This categorization method can be mathematically presented as follows:

Relation [1]:

$t^p = (h)\ T^p\ (h)/H(h)$

In which,

$t^p\ (h)$: average household trip production with the purpose of p and household category h

$T^p\ (h)$: total number of produced trips by households in category h and with the purpose of p

$H(h)$: the number of households in category h

The advantages of this method include:(Ortuzer, J.D. and G. Luis, 2001).

1. It is independent from the traffic zones in the under-study area
2. There is no need to any pre-assumption regarding the quality and the how of relations between socio-economic characteristics of the households and the number of trips produced.
3. The quality and how of relation in each category can be different from other categories.

Determining the average trip production for each category:

The method of categorizing the household socio-economic characteristics in the present study is shown in table [2].

Table 2: the method of categorizing the household socio-economic characteristics.

Number	Family size	Auto ownership	Number of employee	Number of drivers
1	1	0	0	0
2	2	1	1	1
3	3	2 and more	2	2
4	4		3	3 and more
5	5		4 and more	
6	6 and more			

The above-mentioned categorization, [table 2], has been prepared based on the Isfahan data base and has included 15073 households. The maximum and minimum of each of the socio-economic characteristics have been presented. Based on the mentioned categorization, 195 household categories are determined in Isfahan. In order to determine the average household trip production rate for each category the following mathematical relation has been used:(Ortuzer, J.D. and G. Luis, 2001. stopher, P.R. and K.G. McDonald, 1983).

Relation 2:

$$t(h) = \frac{\sum_{I=1}^{n} (HS_i\ (h) * t_i\ (h))}{\sum_{I=1}^{n} HS_i\ (h)}$$

In which:

t *(h*): the average trip production rate for household belonging to category h

HSi (h): the size of household i in category h

ti (h) : the trip production rate of household i in category h

n : the number of households in category h

RESULTS AND DISCUSSION

Determining the likelihood distribution function for the household average trip production rate Considering section 2, the mean function in the composed likelihood distribution functions follows the principles of continuous functions of which the most likely would be gamma continuous distribution function. Now it is time to test whether the average rate of trip production possesses the gamma distribution or not.(Chang-Jen Lan and Patricia S. Hu, 2000. W.J., Conover, 1980).

In order to do this, a table is produced having 20 categories in which the expectative frequencyineachcategory equals $\frac{195}{20}$ =9.75. This has clearly been shown in table [3].

Theχ^2 statisticequals $\sum_{i=1}^{20} \frac{(o_i-e_i)^2}{e_i}$ = 26.64103 which has a $\chi^2$with degree of freedom of 17. In other words, we have k groups in which two parameters have been assessed-(p=2), also, the degrees of freedom would be in the form of (k-1-p). As [p-value] equals 0.063547, so the hypothesis which mentions that the data have a gamma distributions is not rejected at the level of 0.95. That is, the data have parameters as follows:

- Shape parameter: 6.4088748
- Scale parameter: 0.95 103082

in which the parameters of location and scale has been gained through torque method. Accordingly the distribution of average trip production can be obtained through following formula:

Table 3: The analysis of χ^2test about gamma distribution.

group		observed frequency	expectative frequency	$\frac{(o_i - e_i)^2}{e_i}$
0.0000-	2.7100	7	9.75	0.775641
2.7100-	3.2460	7	9.75	0.775641
3.2460-	3 .647 1	4	9.75	3.391026
3.6471-	3.9895	13	9.75	1.083333
3.9895-	4.3005	6	9.75	1.442308
4.3005-	4.5936	5	9.75	2.314103
4.5936-	4.8771	7	9.75	0.775641
4.8771-	5.1569	7	9.75	0.775641
5.1569-	5.4378	8	9.75	0.314103
5.4378-	5.7240	8	9.75	0.314103
5 .7240-	6.0202	15	9.75	2.826923
6.0202-	6.3315	7	9.75	0.775641
6.3315-	6.6642	9	9.75	0.057692
6.6642-	7.0272	12	9.75	0.519231
7. 0272-	7.4332	11	9.75	0.160256
7.4332-	7.9031	11	9.75	0.160256
7.9031-	8.4748	15	9.75	2.826923
8.4748-	9.2313	16	9.75	4.00641
9.2313-	10.430 1	12	9.75	0.519231
10.4301-	And more	15	9.75	2.826923
total		195	195	26.64103

• e_i : expectative frequency
• o_i : observed frequency

[relation 3]:
λ_k ~ Gamma(shape, scale)

$$f(\lambda_k) = \frac{\lambda_k^{\text{shape } 1} (shape)^{\text{shape}}}{\Gamma\ (shape)} e^{-(shape)\lambda k} = \frac{\lambda_k^{6.41-1} (0.951)^{6.41}}{\Gamma(6.41)} e^{-(0.951)\lambda k}$$

The disaggregate household trip production model:

Before determining the final function, that is the composed likelihood distribution function, the mathematical relation between the average of household trip production and the socio-economic properties of the household will be determined. In order to achieve this mathematical relation between the average of household trip production (dependent variable) and household size, number of drivers, number of employees and auto ownership per household (independent variables), we will use a regression model. The purpose of determining the mentioned regression model, is obtaining some coefficients of the predicting variables along with having a meaningful effect on the target variable, the number of trips it should have an acceptable squared amount of the coefficient, or R-square. This amount is usually around 0.7.(Ortuzer, J.D. and G. Luis, Willumsen, 2001. Rose, G. and F.S. Koppelman, 1984).

In figure [1], the frequency distribution of household average trip rate (for 195 household category) has been illustrated.

According to the pattern presented in the graph, the average of household trip production can be divided into two major groups of low frequency and high frequency. In order to consider this issue, a new variable called trip frequency index enters the model.

This attributes 1 to high number of trips and 0 to low number of trips. Due to the linear dependence between the predicting variables in the model, new variables are produced based on the primary linear combinations which are linear independence to each other.

z1 = 0.634 * *hs* +0.259* *ao* +0.646 **ne*+0.337**nd*
z2 = - .340* *hs* + 0.7 16 * *ao* - 0.254 **ne*+0.558**nd*
z3 = 0.075**hs* - 0.621**ao*- .215 **ne*+0.750**nd*
z4 = 0.691**hs* +0.182**ao* -.690**ne* -.116**nd ne* -.116**nd*
In which:
hs: household size
ao: auto ownership
ne: number of employees nd: number of drivers

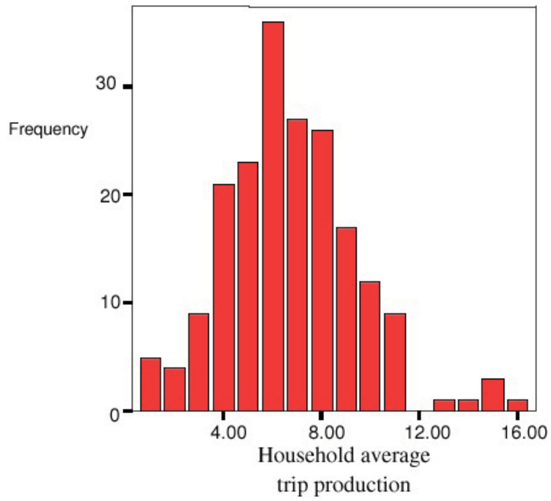

Fig 1: the frequency distribution of household average trip production.

Using the new variables, the trend of would be as follows via software SPSS 11.5: determining the most suitable calibrated model

Table 4: The quantities of correlation coefficient in the model

Model	R	R Square	Adjusted RSquare	Std. Error of the Estimate
1	.698(a)	.488	.485	30.11669
2	.878(b)	.770	.768	20.22845
3	.885(c)	.783	.780	19.70425

a indicators: Trip Frequency, Constant
b indicators: Z1, Trip Frequency, Constant
c indicators: Z1,Z2, Trip Frequency, Constant
d Response Variable: The Square of number of trips

Table 5: Analysis of Variances of the New Model

Model		Total of square	Degree of Freedom	Mean of square	F test	Level of Significan
1	Regression	166660.239	1	166660.239	183.746	.000(a)
	Residuals	175053.855	193	907.015		
	Total	341714.094	194			
2	Regression	263149.598	2	131574.799	321.549	.000(b)
	residuals	78564.496	192	409.190		
	Total	341714.094	194			
3	Regression	267556.922	3	89185.641	229.707	.000(c)
	residuals	74157.172	191	388.257		
	Total	341714.094	194			

a indicators: Trip Frequency, Constant
b Indicators: Z1, Trip Frequency, Constant
c Indicators: Z1, Z2, Trip Frequency, Constant
d Response Variable: The Square of number of trips

Table 6: Determining the calibration coefficient in the new model

Model		Model Coefficients		Standardized Model Coefficients	Statistic t	Level of significance	Level of significance of 0.95 for B	
		Coefficient	coefficient Standard	Beta			Low boundary Deviation	High boundary
1	(Constant)	216.374	12.295		17.598	.000	192.124	240.624
	Trip frequency index	- 169.288	12.489	-.698	-13.555	.000	-193.920	-144.656
2	(Constant)	145.468	9.461		15.375	.000	126.806	164.130
	Trip frequency index	-156.873	8.427	-.647	-18.615	.000	-173.494	-140.251
	Z1	12.417	.809	.534	15.356	.000	10.822	14.012
3	(Constant)	140.643	9.327		15.079	.000	122.246	159.040
	Trip frequency index	-154.171	8.248	-.636	-18.692	.000	-170.440	-137.903
	Z1	13. 322	.832	.573	16.008	.000	11.681	14.964
	Z2	4.557	1.352	.120	3.369	.001	1.889	7.224

The final model in which all of the regression pre-assumptions are followed is as follows:

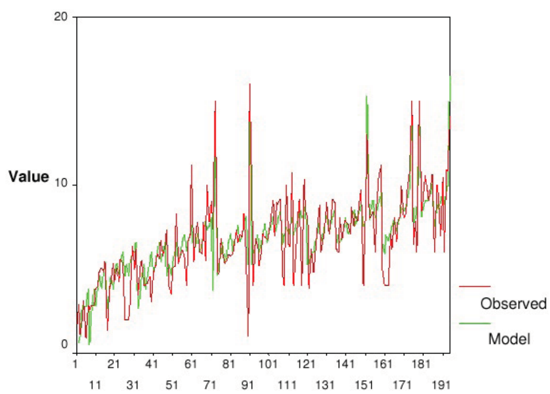

Fig. 2: The comparrison of model results and observations.

$(tp)^2$ =- 140.643154.171**tf* +13.322*z1 +4.557*z2

In which *tf* is trip frequency index and $(tp)^2$ is the average of trip production squared.
In other words, the final model would be as relation 6 demonstrates.

Relation 4:

$(tp)^2 = 140.643 - 154.171*tf + 6.897*hs + 6.713*ao + 7.490*ne + 7.032*nd$

In which:

- *tf* is trip frequency index
- *hs* is household size
- *ao* is auto ownership
- *ne* is the number of employees
- *nd* is the number of drivers

$(tp)^2$ is the square of average trip production

Since the above mentioned model can be attributed by negative quantities, it would be modified into the following: $\max\{0, (tp)^2\}$

This way, the model does not predict negative values; in addition, the value of the modeling coefficient increases.

In this section, the values of the predicting variables have been put into the model and the changes in the independent variable; household trip production has been calculated out of the model. Finally, the obtained results from the model and the values acquired through observation in the city of Isfahan have been compared and contrasted, as it is presented in graph (Yari, Gholamhussain, 1375).

The results show that the obtained results of the model have a good concordance with the results of the observations; hence, the calibrated model has the required precision.

Conclusion:

The χ^2 test in terms of likelihood distribution of average household trip production (section 5-1) demonstrated that the related parameter is conforming to and following the gamma continuous distribution. the statistical indicators and assessment of the calibration coefficients for disaggregate model was demonstrative of its high accuracy and precision in predicting the average trip production rate.

The outputs of section 5 show that it is possible to find significant relation between the likelihood of average produced trips of each household's category and socio-economic characteristics. So it would be possible to predict this parameter (by inserting the socio-economic characteristics in the disaggregate model) and then determine the likelihood of it.

REFERENCES

The Comprehensive Studies of Transportation in Isfahan, 1379, Sharif Transportation Research Center, pages, 1-250.

Yari, Gholamhussain, 1375, Statistics in Engineering, Volume1-2. Iran University of Science and Technology Publication, pages, 1-545.

Agresti, A., "An Introduction to Categorical Data Analysis", University of Florida, John Wiley & Sons, 1996, pages, 71-174.

Caldwell, L.C., III, and M. J. Demetsky, 1978. "An Evaluation of the Transferability of Cross Classification Trip Generation Models", VHTRC 78-R39. Virginia Highway and Transportation Research Council, Charlottesville, pages, 74-81.

Chang-Jen Lan and Patricia S. Hu, 2000. "Mixed Generalized Linear Model for Estimating Household Trip Production", Transportation Research Record 1718, Paper., 61-67: 00-1282.

Ortuzer, J.D. and G. Luis, Willumsen, 2001. "Modelling Transport", John Wiley & Sons, Third Edition, pages, 33-160.

Rose, G. and F.S. Koppelman, 1984. "Transferability of disaggregate trip generation models", In J. Volmuller and R. Hamerslag. Proceeding of the Ninth International Symposium on Transportation and Traffic Theory. VNU Science Press, Utrecht, pages, 140-147.

Stopher, P.R. and K.G. McDonald, 1983. "Tripgenerationbycross- classification:analternative methodology", Transportation Research Record, 944, pages, 84-91.

Conover, W.J., 1980. "Practical Non-parametric Statistics", Second Edition, JOHN WILEY & SONS, pages, 1-105.