

Evaluation Of Minimum Sample Size Of Trip Production Model (Case Study: Middle Income Residential In Banjarmasin)

Rama Fauziah¹, Iphan F. Radam^{2,3}

¹ Graduate Student, Transportation Engineering and Management Program, Magister Study Program of Civil Engineering, Lambung Mangkurat University, Banjarmasin, Indonesia

² Professor, Magister Study Program of Civil Engineering, Lambung Mangkurat University, Banjarmasin, Indonesia

³ coresponden author: ifradam@ulm.ac.id

DOI: 10.29322/IJSRP.11.09.2021.p11732

<http://dx.doi.org/10.29322/IJSRP.11.09.2021.p11732>

Abstract- Residentials in Banjarmasin are dominated by middle-class families and are an area inhabited by residents with many activities, causing the emergence of trip production. The purpose of this study is to obtain the minimum sample size of the trip production model with the overall data using the multiple linear regression analysis methods. The combination of samples is analyzed using samples stratified random sampling technic with 50% data sample and simple random sampling with 60% data sample. Trip production regression equation model on overall data is $Y = 0.062 + 0.854 X_5 + 0.819 X_6 + 0.548 X_7 + 0.03321 X_8 + 0.063 X_1$ with the coefficient determination (R²) 0.916, where Y is the number of vehicles per hour, X₅ is car ownership, X₆ is motorcycle ownership, X₇ is bicycle ownership, X₈ is family head income and X₁ is the number of family members. Can be concluded that simple random sampling is the best with 60% data sample which is $Y = 0.041 + 0.940 X_5 + 0.878 X_6 + 0.579 X_7$.

Index Terms-, residential, sample size, trip production

I. INTRODUCTION

The location taken in this study case are Seven (7) Residentials which are *Komplek Perdagangan Permai 1*, *Komplek Perdagangan Permai 2*, *Komplek Griya Perkasa*, *Komplek Citra Permai*, *Komplek Murai*, *Komplek Margansa Residence*, and *Komplek Famili*. All these residentials are located separately, at a different distance with permanent and semi-permanent house types. The reason for determining the location of these seven residentials is because there are a lot of activities happened in these residentials such as trip to offices, schools, markets, and so on. These residentials have varying amounts and are inhomogeneous, have one access to go in and out of the residentials, and are chosen based on a different level of establishment area. These residentials are also chosen based on the ability of data retrieval, if collecting a huge scale of data, it will have a high risk because the data chosen is the overall sample of those residentials. A combination sample size is used in this study case along with making the comparison with a frequently used method such as Harry King Method and Slovin Method to obtain the minimum number of samples that can be used.

Accuracy and the success of a method hugely depend on the number of samples that can be used. The more samples taken will produce predictions (models) that are close to the reality as if we used all the data. The model formed is based on a zone-based model because the zone-based generation model is a better model than the household-based generation model [1]. Based on these things, the trip production model is used in several residentials with a variety of overall samples to discover the effect of the overall samples towards the trip production model.

II. REFERENCE LITERATURE

Transportation is the movement of goods and persons from place to place by using a mobile or motorized infrastructure; transportation is used to facilitate people in carrying out their daily routines [2]. A model is a representation of something designed for a specific purpose. The transportation model explains the relations between land use, traffic, and means of transportation [3]. The advantage in using a mathematical model in transportation planning is that during formulation making, the calibration and its use, the planners can acquire many things, through the experiments, about the behavior and internal mechanisms of the system that is being analyzed.

The population is the main factor that influences the transportation problems. In all areas of planning, the population can't be ignored [4]. Trip generation is the number of trips that occur in a unit of time in a land-use zone [5]. The population is the complete data

that is the main focus of a researcher in a predetermined scope and time [6]. A sample is a part of the whole characteristics that the population has [7].

The coefficient of determination (R^2) test is used to measure how big the percentage of the independent variable linear regression model is, in defining or influencing the dependent variable. One of the standards commonly used to see whether a suitable regression model is adequate is the coefficient of determination (R^2) [8]. The relations between one dependent variable with one or more independent variables can be described in multiple linear regressions [9].

III. RESEARCH METHODOLOGY

A. Data Processing

In this study, a model is used with overall sample sizes then is analyzed with a different number of samples with the purpose to determine the sample combination against the model along with the trip production result that occurred. Furthermore, every sample size variation is analyzed using multiple analysis regression and classic assumption test to discover to what extent the sample limits are acceptable.

B. Data Analysis

The following are several explanations of data analysis stages:

1. Analyzing the data using the entire sample and the resulting data that can be used to multiple linear regression analysis;
2. Combining the samples with Stratified Random Sampling with 4 data trials, the data is retrieved based on reducing the number of families;
3. Combining the samples with Simple Random Sampling with 4 data trials, the data is retrieved based on Mersenne Twister Method;
4. Analyzing classic assumption test and goodness of fit;
5. Analyzing the number of samples compared with Slovin formula and Harry King Nomogram as follows:

Slovin Formula:

$$n = \frac{N}{1 + Ne^2}$$

With:

n = Sample Sizes

N = Population

e = Percentage of disengagement looseness because sampling error is still wanted

How to use *Harry King Nomogram Technic*

The way to use this technic is by drawing a straight line from the right line that is the population outline, passing through the centerline that is the desired error rate line, then it will reach the left line that shows sample size percentage, when the sample size is known furthermore is to multiply sample percentage with the total population and multiplier.

6. Analyzing the acceptable sample size limit in the same way as the previous analysis.

IV. ANALYSIS RESULTS

Trip production model is analyzed using SPSS program with stepwise method and is made by multiple models with different sample sizes approaches in order to get the best model. The results of the analysis above explained that a model only insert the chosen independent variable. The best considered model is $Y = 0.062 + 0.854 X5 + 0.819 X6 + 0.548 X7 + 0.03321 X8 + 0.063 X1$, because R^2 is bigger and has more independent variable. In multiple regression analysis, the result is obtained in the Table 1 to Table 6.

Table 1. Multicollinearity Test Analysis Results

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
5 (Constant)	0.062	0.071		0.873	0.383		
X5	0.854	0.035	0.541	24.097	0.000	0.542	1.846
X6	0.819	0.028	0.573	28.829	0.000	0.691	1.448
X7	0.548	0.041	0.233	13.509	0.000	0.918	1.089
X8	0.03321	0.000	0.078	3.629	0.000	0.592	1.689
X1	0.063	0.019	0.072	3.386	0.001	0.606	1.651

Table 2. Autocorrelation Test Analysis Results

Model	R	R ²	Adjusted R Square	SEE	DW
1	.708 ^a	0.501	0.500	0.855	
2	.920 ^b	0.847	0.846	0.474	
3	.954 ^c	0.910	0.909	0.365	
4	.955 ^d	0.913	0.912	0.359	
5	.957 ^e	0.916	0.914	0.354	1.870

Table 3. Heteroscedasticity Test Analysis Results

Variable	Sig.	Conclusion
X ₅	0.474	✓
X ₆	0.134	✓
X ₇	0.405	✓
X ₈	0.271	✓
X ₁	0.910	✓

Table 4. Normality Test Analysis Results

Unstandardized Residual		
N		315
Most	Absolute	0.076
Extreme	Positive	0.076
Differences	Negative	-0.046
Test Statistic		0.076
Exact Sig. (2-tailed)		0.051

Table 5. Model Reliability Test Analysis Results (F Test)

Model	Sum of Squares	df	Mean Square	F	Sig.
5 Regression	420.055	5	84.011	671.985	0.000^f
Residual	38.631	309	0.125		
Total	458.686	314			

Table 6. Regression Coefficient Test Results (t Test)

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
5 (Constant)	0.062	0.071		0.873	0.383		
X5	0.854	0.035	0.541	24.097	0.000	0.542	1.846
X6	0.819	0.028	0.573	28.829	0.000	0.691	1.448
X7	0.548	0.041	0.233	13.509	0.000	0.918	1.089
X8	0.03321	0.000	0.078	3.629	0.000	0.592	1.689
X1	0.063	0.019	0.072	3.386	0.001	0.606	1.651

Based on results of the table above can be concluded that all classic assumption test with the overall data is meet the requirements of the multiple regression analysis.

This analysis stage is equal to the previous analysis and the sample used are 85% - 25%, sample is taken based on criteria I executed like taking samples with more family members. Trip production analysis results with multiple overall sample combination can be found in Table 7.

Table 7. Trip Production Model Results with Multiple Overall Samples

Sampling	Random Sampling	R ²	F	t	Classic Assumption			
					Multicollinearity	Heteroscedasticity	Autocorrelation	Normality
100%		0.916	✓	✓	✓	✓	✓	✓
85%	Stratified	0.927	✓	✓	✓	✓	✓	✓
	Simple 1	0.915	✓	✓	✓	✓	✓	✓
	Simple 2	0.914	✓	✓	✓	✓	✓	✓
	Simple 3	0.913	✓	✓	✓	✓	✓	✓
	Simple 4	0.911	✓	✓	✓	✓	✓	✓
75%	Stratified	0.872	✓	✓	✓	✓	✓	✓
	Simple 1	0.925	✓	✓	✓	✓	✓	✓
	Simple 2	0.922	✓	✓	✓	✓	✓	✓
	Simple 3	0.918	✓	✓	✓	✓	✓	✓
	Simple 4	0.91	✓	✓	✓	✓	✓	✓
50%	Stratified	0.91	✓	✓	✓	✓	✓	✓
	Simple 1	0.932	✓	✓	✓	✓	✓	☒
	Simple 2	0.93	✓	✓	✓	☒	✓	☒
	Simple 3	0.909	✓	✓	✓	☒	✓	☒
	Simple 4	0.904	✓	✓	✓	✓	✓	☒
25%	Stratified	0.902	✓	✓	✓	☒	✓	☒
	Simple 1	0.924	✓	✓	✓	☒	✓	☒
	Simple 2	0.916	✓	☒	✓	☒	✓	☒
	Simple 3	0.892	✓	✓	✓	☒	✓	☒
	Simple 4	0.885	✓	☒	✓	☒	✓	☒

At 50% sample with *stratified random sampling* showed that a model can still be used, meanwhile with *simple random sampling* showed that data usage can only be used up to a sample of 75%. Therefore, the minimum number of samples is tested with the same analysis as the previous analysis between 75% and 50% which are 70%, 65%, 60%, and 55% sample then the comparison is tested using Slovin Method and Harry King Nomogram Method which can be seen as follows:

1. Slovin Method

$$n = \frac{N}{1 + Ne^2}$$

Obtained:

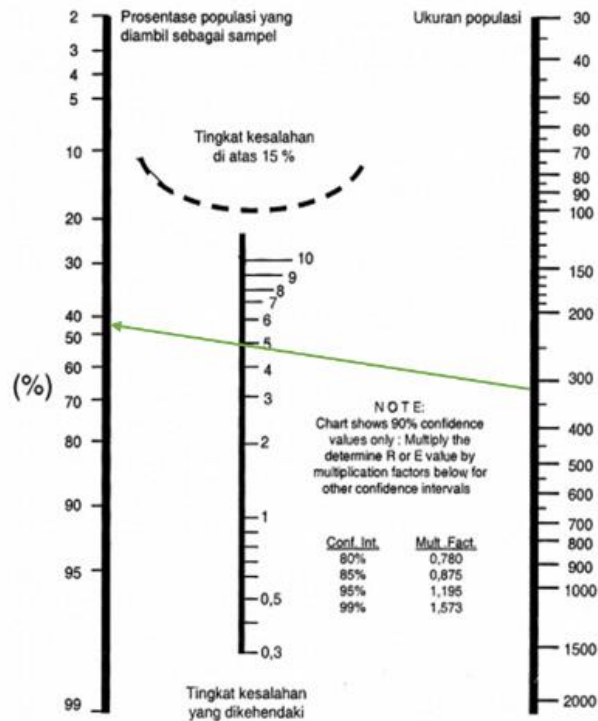
$$n = 315 / (1 + (315 \times 0.05^2))$$

$$n = 315 / (1 + 0.7875)$$

$$n = 315 / 1.7875$$

$$n = 176.223 \approx 176 \text{ samples (56%).}$$

2. Harry King Nomogram Method



Picture 1. Harry King Nomogram

$$N = 315 \times (45\%) \times 1.195$$

$$= 169.39 \approx 169 \text{ sample (54\%).}$$

On sample, 70% - 55%, sample 56% (Slovin Method), and sample 54% (Harry King Nomogram Method) analysis is executed 4 times to obtain the best sample size limit that can be used. The number or samples comparison results can be seen in Table 8.

Table 8. Number of Samples Comparison

No	Sampling	Sample Percentages	Classic Assumption Test (4 times trials)
1	Slovin Method	56% (176 sample)	25% fulfilled
2	Harry King Nomogram Method	54% (169 sample)	Unfulfilled
3	Analysis		
	- Stratified Random Sampling	50% (157 sample)	100% fulfilled
	- Simple Random Sampling	70% (220 sample)	100% fulfilled
		65% (204 sample)	100% fulfilled
		60% (189 sample)	100% fulfilled
		55% (173 sample)	Unfulfilled

Based on the results above can be concluded that Slovin Method with sample 56% are analyzed 4 times randomly, only 25% of the results meet the eligibility requirements for multiple analysis test. For Harry King Nomogram Method with sample 54%, showed that data doesn't meet the eligibility requirements for multiple regression analysis tests and on *simple random sampling* 60% is analyzed with several times trial data that showed fulfilled model, it can be concluded that the used sample limit is 60%.

V. CONCLUSION

The best model for overall trip production residentials in Banjarmasin is model $Y = 0.062 + 0.0854 X5 + 0.819 X6 + 0.548 X7 + 0.03321 X8 + 0.063 X1$ consists of variable X5 (number of private passenger car ownership), X6 (number of motorcycle ownership), X7 (number of bicycle ownership), X8 (total of family income in Rupiah) and X1 (number of family members) with $R^2 = 0.916$.

On sample 50% with stratified random sampling showed that a model can still be used, while with simple random sampling showed that usage data can only be used up to sample 60%. The results of this analysis with Slovin Method and Harry King Nomogram Method showed that classic assumption test isn't fulfilled, hence attempted to obtain the minimum number of samples in order that the minimum number of samples that can be used is 60% with the best considered model is $Y = 0.041 + 0.940 X_5 + 0.878 X_6 + 0.579 X_7$.

The sampling method is expected to be able to use a different method such as Jacob Cohen Formula Method, and Isaac and Michael Method. Residential numbers also need to be added and more spread out so that the trip model can represent maximally and equally to all trip production in Banjarmasin city and more variety of sample sizes so that changes of effect are more visible with scaled-up sample sizes.

REFERENCES

- [1] Petruswandie Eddy, and Radam, Iphan F. 2018. Analysis of the Impact of Generation of Housing on the Performance of Soekarno-Hatta Street in Kasongan City. *International Journal of Advanced Engineering Research and Science*. 5 (8). 257-263
- [2] Morlok, E.K., 1998, *Introduction to Transportation Engineering and Planning*, Jakarta: Erlangga.
- [3] Bruton, M. J. 1985. *Introduction to Transportation Planning*. Melbourne: Hutchinson.
- [4] Warpani, S. 1990. *Planning and Transportation System*. Bandung: ITB.
- [5] Hobbs, F.D, 1995, *Traffic Planning and Engineering*, Yogyakarta: Gadjah Mada University Press.
- [6] Margono, 2004. *Educational Research Methodology*, Jakarta: Rineka Cipta.
- [7] Sugiyono. 2008. *Qualitative and Quantitative Research Methods and R & D*. Bandung: ALFABETA.
- [8] Walpole, Ronald E., 1995, *Introduction to Statistics*, 3rd Edition, Jakarta: Gramedia Pustaka Utama.
- [9] Draper, N. Smith, H. 1992. *Applied Regression Analysis*. Jakarta: Gramedia Pustaka Utama.