Abusini & Kilicman, *Cogent Mathematics* (2017), 4: 1379675
https://doi.org/10.1080/23311835.2017.1379675

cogent mathematics

Received: 15 May 2017
Accepted: 06 September 2017
First Published: 15 September 2017

*Corresponding author: Adem Kilicman, Department of Mathematics, University Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
E-mail: akilic@upm.edu.my

Reviewing editor:
Hiroshi Shiraishi, Keio university, Japan

Additional information is available at the end of the article

STATISTICS | RESEARCH ARTICLE

Modelling motorcycle accident on the road section by using general linear model: Case studies in Batu City

Sobri Abusini[1] and Adem Kilicman[2*]

Abstract: In the present study, we consider modelling the transportation and traffic problems in Batu City of East Java Province. Based on the accident data, the number of traffic accidents in Batu City is considered too high. Thus, the objective of this research is to develop a suitable model for motorcycle accidents around Batu City in East Java Province. Using some statistical analysis it is found that the best-fit motorcycle accident model is computed by : $Acc = 7.019 \times 10^{-6} F^{1.702} \exp(-0.2884 SW + 0.091 S)$ where : Acc = number of accident, F = Flow, pcu/h, SW = shoulder width, m, S = speed, km/h. This model shows that the affecting factors are flow, shoulder width and speed, therefore local government should improve some related factors (flow, shoulder width and speed) that can reduce the number of motorcycle accidents at crossing road in Batu City.

Subjects: Applied Mathematics; Statistics & Probability; Industry & Industrial Studies

Keywords: model; accident; motorcycle; general linearized model; Batu City

AMS subject classifications: 90B20; 90B06

1. Preliminary

Transportation is an important element in the national development and also considered as growth engine for economies. If the transportation network is not efficient then it will be detrimental to the public and leads to the congestion/delays, accidents, impaired mobility etc. The high cost of

Adem Kilicman

ABOUT THE AUTHORS

Sobri Abusini is a lecturer at Department of Mathematics, Brawijaya University, Indonesia. Sobri,PhD, graduated from Padjadjaran University, Bandung Indonesia majoring in Mathematical Sciences in 1987 and obtained Master from Institute Technology Bandung, Bandung, Indonesia. Sobri received his PhD Airlangga University, Surabaya, Indonesia in 2010. His research area includes modelling the transportation problems.

Adem Kilicman is a full professor at the Department of Mathematics, Faculty of Science, University Putra Malaysia. He received his BSc and MSc degrees from Department of Mathematics, Hacettepe University, Turkey and PhD from Leicester University, UK. He has joined University Putra Malaysia in 1997 since then working with Faculty of Science and He is also an active member of Institute for Mathematical Research, University Putra Malaysia. His research areas include Functional Analysis and Topology.

PUBLIC INTEREST STATEMENT

Road safety is a global issue and every day many people are killed or injured on roads, the cost of the damages during the accidents are economically very costly. Thus, it is an important topic to study the modelling the cases. In the present study, we consider modelling the transportation such as traffic problems and motorcycles accidents in Batu City of East Java Province. Based on the data, the number of traffic accidents in Batu City is considered very high. Thus, the objective of this research is to develop a suitable model for the motorcycle accidents around Batu City in East Java Province.

transportation for low-income people can be routine daily work and cost the 30–40% of total revenues, which reasonable cost should only amount to 10% of revenues (Sulistio, 2007).

On the other side, the transportation safety is a global problem and not solely a problem of transport, but has become a social problem in the society. WHO's concern for the safety of road transport is realized by setting the World Health Day 2004 with the theme: Road Safety is no Accident in order to reduce the toll of deaths and injuries. According to WHO, road transport accident rate in the world has reached 1.2 million deaths and over 30 million injured/disabled per year. As many as 85% of the victims who died in traffic crashes occur in developing countries, the number of vehicles is only 32% of the total number of vehicles in the whole world.

From the description in Table 1 we can see the number of traffic accidents at this time is quite high, especially, a motorcycle accident and a serious threat to traffic safety.

Among the accidents, motorcycle accidents have the largest percentage, about 63% of all traffic accidents. This is due to lack of seriousness in handling the road safety. Therefore, the objective of this study is to develop a model of a traffic accident involving a motorcycle in the town of Batu.

2. Model accident

Model as a tool or media that can be used to reflect and simplify a reality (the real world) is scalable. The models might display the properties of the complexity to be understood by everyone who translated into written form or image (Bhat, 2015).

While the transportation model is an mathematical expression and is used to illustrate the relationship among the parameters (Comi & Russo, 2004; Ortuzar & Wilumsen, 1994; Tamin, 1999).

Modelling the accidents on roads (*X*-roads) by using Generalized Linear Model done in (Swan et al., 1994), and the format of the model is as follows:

Table 1. Data on accidents in Batu

Type of accident	Batu city	
	2014	2015
Death	16	25
Serious wound	10	3
Minor injuries	132	128
Total amount	158	156

Source: Police Batu.

Table 2. Data input in Batu City

No	Road	*Acc*	*NL*	*LW*	*SW*	Median	Gradien	80% speed	Flow
1	Pattimura street	23	1	3.5	1.5	0	1	28.84	1,739
2	Sudirman street	27	1	5	2.2	0	1	47.36	858
3	Dewi Sartika street	3	1	5.9	1.6	0	1	28.43	625
4	Surapati street	6	1	2.85	1.15	0	1	34.34	705
5	Raya Mojorejo street	32	1	4.05	1.4	0	1	27.77	2,401
6	Raya Payung street	6	1	2.9	1	0	1	35.1	519
7	Kalilanang street	3	1	4.35	2.15	0	1	34.07	627
8	Raya Punten street	5	1	3.4	1	0	1	40.06	323

Source: Survey.

$$A = kQ^{\alpha}$$

$$A = kQ^{\alpha} e^{b_1 g_1 + b_2 g_2 + \cdots} \quad (1)$$

where A = average motorcycle accident, Q = traffic flow, $g_1, g_2, \ldots$ = variables of geometry, $\alpha, k, b_1, b_2, \ldots$ = parameters for estimation

To develop a model of traffic accident, in particular to motorcycle accident, should be carried out to minimize the number of accidents by using multiple distribution in statistics (Aitkin, Anderson, Francis, & Hindle, 1989). Based on the several related literature, we conduct a study on traffic accidents, especially motorcycle accidents on several roads in the town of Batu. In the process of research we apply Geometric distribution. The data input in Batu City can be seen in Table 2.

Input data were provided by 15 of data (15 roads), then the data are processed by performing several iterations. Similarly, we develop a model based on eight roads' location data models that have shown the best and most significant parameters (most suitable). For crash analysis SPSS version 16 (Ghozali, 2006) and Software Easyfit 5.2 Standard version of the data were used and can be seen in Table 2.

3. Materials and methods

3.1. Test distribution

Before we set up a model of a motorcycle accident on the town of Batu, we use the input data which is necessary to check the suitability of the distribution of each variable (Conover, 1999). Thus, we need to determine the shape of the distribution of the data in the study and to determine which distribution is most appropriate. Testing is only done on the variable distribution of the number of accidents (*Acc*) as the dependent variable or response variable (Zuherman, 2014).

To determine the suitability of the distribution, Kolmogorov–Smirnov test was done on Standard Software Easyfit 5.2 and calculations are carried out directly using the formula:

$$D_{\max} = \sup \left[\left| F_n(x) - F_0(x) \right| \right]. \quad (2)$$

The calculation of the distribution tests were conducted on roads with eight data involving one variable response and five explanatory variables. Among them are: the number of accidents (*Acc*) as the response variable, the variable number of lanes (*NL*), lane width (*LW*), road shoulder width (*SW*), a variable average speed (*S*) and variable traffic volume (*F*) (Tables 3 and 4).

Table 3. Fitting results of KS test

No	Distribution	Parameters
1	Geometric	$p = 0.0708$
2	Logarithmic	$\theta = 0.98118$
3	Neg. binomial	$n = 1$ $p = 0.10305$
4	Poisson	$\lambda = 13.125$

Table 4. Goodness of fit—Summary

No	Distribution	Kolmogorov Smirnov	
		Statistic	Rank
1	Geometric	0.25451	1
2	Logaritmic	0.44737	3
3	Neg. binomial	0.35276	2
4	Poisson	0.60094	4

Table 5. Goodness of fit—Details					
Geometric [No. 2]					
Kolmogorov–Smirnov					
Sample size	8				
Statistic	0.25451				
p-value	0.5921				
Rank	1				
α	0.2	0.1	0.05	0.02	0.01
Critical value	0.35831	0.40962	0.45427	0.50654	0.54179
Reject?	No	No	No	No	No

3.2. Calculation Kolmogorov–Smirnov test

Based on standard output above 5.2 easyfit indicates that the distribution of data on the number of motorcycle accidents (*Acc*) is to follow the Geometric distribution (first order) with the value of KS statistic which is 0.25451, Negative Binomial distribution (second order) with the value of KS statistic 0.35276 and distribution Logarithmic ranks third with KS statistical value of 0.44737.

Furthermore, the Kolmogorov–Smirnov hypothesis is details in order to determine a distribution is not denied or rejected from a distribution of empirical data by using Goodness of Fit—Details (Table 5).

where H_0 = Data *Acc* follows distribution Geometric distribution and H_1 = Data *Acc* does not follow the distribution Geometric distribution

Statistical Test Kolmogorov–Smirnov or D_{max} = 0.25451 Critical Value Kolmogorov–Smirnov 5% or D (5%) = 0.45427.

Decision: because $D_{max} < D$ (5%). Then, H_0 is not rejected.

Conclusion: With an error rate of 5%, it can be said that the accident data (Acc) following distribution Geometric distribution, with $p = 0.0708$.

Distribution of test results and graphs P–P plot above shows that the distribution that best fits the data field distribution is Geometric distribution, so it can be concluded that the data motorcycle accident on roads in the town of Batu are following the Geometric distribution.

3.3. Distribution of geometric on Kolmogorov–Smirnov test statistics

For more specifically, on the following steps performed statistical calculations Kolmogorov–Smirnov test against Geometric distribution based on the value of the maximum absolute deviation

(D_{max}) defined as follows:

$$D_{\max} = \sup\left[\left|F_n(x) - F_0(x)\right|\right]$$

D_{max}: the maximum absolute deviation value Between $F_n(x)$ and $F_0(x)$

$F_n(x)$: the function of the cumulative odds observed, $n = 8$ (number of observations)

$$F_n(x) = \frac{1}{n}\left[\text{number of observation} \le x\right]$$

$F_0(x)$: the cumulative distribution function distribution Geometric

$$F_0(x) = 1 - (1-p)^{x+1} = 1 - (1 - 0.0708)^x + 1$$

cogent • mathematics

Table 6. Geometric distribution calculation in Kolmogorov–Smirnov test

x	Frek	Frek Kum	$S(x)$	$F_o(x)$	$\|S(x) - F_o(x)\|$	D_{max}	$\|S(x(i-1)) - F_o(x)\|$	D_{max}
3	2	2	0.25	0.2545186	0.00452	0.22309	0.25452	0.25452
5	1	3	0.375	0.3563419	0.01866		0.10634	
6	2	5	0.625	0.4019129	0.22309		0.02691	
23	1	6	0.75	0.8283591	0.07836		0.20336	
27	1	7	0.875	0.8720449	0.00296		0.12204	
32	1	8	1	0.9113653	0.08863		0.03637	

Source: Results of analysis.

From Table 6 we observe that the values $D_{max} = 0.25452$. This value is almost equal to the value obtained using software Dmaks easyfit 5.2 standard.

We can conclude that the calculation of the suitability of the distribution of Kolmogorov–Smirnov test is the same as the distribution of the test performed by using the statistical program Easyfit version 5.2 standard.

3.4. Accident prediction model

3.4.1. Initial parameter estimation

In the initial parameter estiomation, we use three stages in the case of a motorcycle accident in the town of Batu and we use the software SPSS 16.0 for Windows. The calculations can be seen as follows:

Based on Table 7, it can be explained that the coefficient of determination of the model is indicated by the adjusted R^2 that is equal to 0.943 which means the variability motorcycle accident (*Acc*) as the response variable can be explained by the shoulder width, speed and flow as a predictor or independent variable was 94, 3% and the remaining 5.7% is explained by other variables not included in the model.

Simultaneous effect (*F* test) in Table 8 Variance Analysis above are used to determine whether the independent variables or predictor variables together or simultaneously affect the response variable or the dependent variable. This can be seen in the value of the *F* test of 39.526 and 0.002 significant at 0.002 or 0.050, which means that the independent variables (predictors) flow, shoulder width and speed simultaneously influence the variable speed motorcycle accident (*Acc*).

Table 9 can determine the effect of each independent variable (predictor) and dependent variable (response). Of the three independent variables included in the model, all significant at 0.05. These variables are: flow, speed, and shoulder width significantly at 0.05 for all grades assymtotic

Table 7. Model summary

Model	R	R^2	Adjusted R^2	Std. error of the estimate
1	0.984[a]	0.967	0.943	0.2350804

Note: Predictors = (Constant), flow, shoulder width, speed.

Table 8. Analysis of variance

Model		Sum of squares	*df*	Mean square	F	Sig.
1	Regression	6.553	3	2.184	39.526	.002[a]
	Residual	0.221	4	0.055		
	Total	6.774	7			

[a]Predictors = (constant), flow, shoulder width, speed.

[b]Dependent variable = Acc.

Table 9. Coefficient

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	β		
1	(Constant)	−12.929	1.429		−9.047	0.001
	Shoulder width	−0.673	0.214	−0.323	−3.148	0.035
	Speed	0.111	0.017	0.754	6.611	0.003
	Flow	1.835	0.170	1.213	10.762	0.000

Note: Dependent variable = Acc

significantly much smaller than 0.05, then from here we can conclude that a motorcycle accident variable (*Acc*) is influenced by factors flow, speed, and road shoulder (shoulder width).

3.4.2. Model accident

Model accidents on roads (*X*-roads) by using the approach of Generalized Linear Model (GLM), using Equation (1) and paying attention to Table 9 in the above equation maths can be formed, where the value of the coefficients can be put into the equation following (Aitkin et al., 1989):

$$Acc = k\,\text{Flow}^{\beta_0}\ \exp(\beta_1\ \text{shoulder width} + \beta_2\ \text{speed})$$

If it is defined by:

$Acc = A$ (the number of motorcycle accidents)

F = Flow (pcu/h)

pcu = pessenger car uni

SW = Shoulder width (m)

S = Speed (km/h)

3.4.3. Non linear regression analysis

To predict non-linear regression model equation can be used non-linear regression analysis (non-linear regression analysis) on SPSS version 16.

Having regard to the model of a motorcycle accident above, it can result in the following calculation:

From Table 10, estimation Parameters (Parameter Estimate) above can be seen the estimated values of the parameters to form a non-linear regression model equation which is a model of motorcycle accidents on roads in the town of Batu, as the following equation:

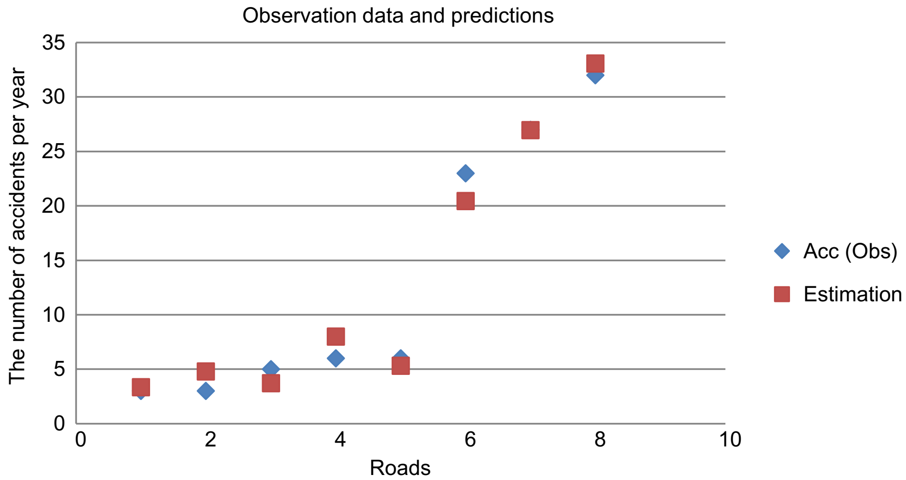
$$Acc = 7.019 \times 10^{-6} F^{1.702}\ \exp(-0.2884SW + 0.091S).$$

Table 10. Parameter estimation

Parameter	Estimate	Std. error	95% confidence interval	
			Lower bound	Upper bound
k	7.02E-06	0	−2.66E-05	4.06E-05
F	1.702	0.196	1.158	2.246
SW	−0.288	0.349	−1.257	0.681
S	0.091	0.021	0.032	0.15

Source: Results of analysis.

Figure 1. Data observation and prediction model accident in Batu.



4. Results and discussion

4.1. Observation data comparison

Model A, good model, is expected to give a prediction or estimate the value of close observation data, for it is necessary to check the model of motorcycle accidents on these roads. Table 11 and Figure 1 show that the estimation of the model is approaching the observation data (95% confidence level), so that the resulting model can be considered in accordance with existing conditions.

4.2. Pairwise test

Further testing Paired Sample *t*_Test to the data of observation and data estimation results, to see how relations between the two data produced from SPSS can be seen as follows:

In Table 12 Paired Sample Statistics show that the observational data have increased very little to the data of the average estimate of 13.12 into 13.21.

Table 13 Paired Samples Correlation analyses whether there is a relationship between the data and the observation data estimation. Here, we can see that the correlation with the estimated observation data is very strong (0.992). When viewed value Sig (0.000) $< \alpha$ it can be concluded very significant correlation.

Table 11. Data observation and estimation in Batu

Roads	*Acc* (obs)	Estimation	Residu	Residu^2
1	3	3.35	−0.3548	0.1259
2	3	4.8	−1.8009	3.2431
3	5	3.72	1.2773	1.6315
4	6	8.01	−2.0119	4.0475
5	6	5.32	0.6806	0.4632
6	23	20.45	2.5479	6.4918
7	27	26.96	0.0386	0.0015
8	32	33.08	−1.0754	1.1566
			MSE	2.1451

Source: Results of analysis.

Table 12. Paired samples statistics

		Mean	*N*	Std. deviation	Std. error mean
Pair 1	Acc	13.12	8	12.065	4.265
	Predicted values	13.2123	8	11.85215	4.19037

Source: Results of analysis.

Table 13. Paired sample correlation

		N	Correlation	Sig.
Pair 1	Acc and Predicted values	8	0.992	0

Source: Results of analysis.

Table 14. Paired samples test

		Paired differences			*t*	*df*	Sig. (2-tailed)
		Mean	Std. deviation	Std. error mean			
Pair 1	Acc—Predicted values	−0.08733	1.56297	0.55259	−0.158	7	0.879

Source: Results of analysis.

Hypothesis:

H_0: $\mu_{Acc} = \mu_{\text{estimates}}$

H_1: $\mu_{Acc} \neq \mu_{\text{estimates}}$

From Table 14 Paired Samples Test, Sig column (two-tailed) is a probability value to reach *t*_statistics. To *t*_arithmetic (−0.158) < *t*_table (8; 0.025) is 2.306 so H_0 is not rejected so as not significant. Besides using *t*_arithmetic comparison with *t*_table, can also do the comparison Sig (2-tailed) with α.

Sig (2-tailed) 0,879 > α (0,025), H_0 not rejected, so it can be concluded that the observed data and the estimation is no different.

4.3. Contributions variable accident

Setelah Having obtained the model of a motorcycle accident and by taking into account the coefficient of each variable, it will show the contribution of each of these variables that influence the occurrence of motorcycle accidents.

(a) The influence of the volume of vehicles (Flow) to the number of accidents.

From the result of the contribution model can be seen that the volume of the vehicle greatly affects the occurrence of a motorcycle accident, if there is no vehicle volume (flow) then the accident would not have happened. And the number of motorcycle accidents will increase along with the increase in the volume of vehicles on the roads in the town of Batu.

(b) The influence of the shoulder (shoulder width) to the number of accidents.

The coefficient of the road shoulder of −0.288 states that for each additional 1-m shoulders (shoulder width), it will be less accidents as $e^{-0.288}$ multiplication or multiplication by 0.7498.

(c) Effect of velocity (speed) to the number of vehicle accidents

The coefficient of velocity (speed) of +0.091 states that for any increase in the value of 1 from the value of velocity (speed), a motorcycle accident will increase by multiplication $e^{0,091}$. or multiplication 1.0953. The higher speed will increase the number of traffic accidents, especially motorcycle accidents.

5. Conclusions and recommendations

5.1. Conclusion

From the observation and data analysis of motorcycle accidents on roads in the city of Batu several conclusions can be made :

(a) The results of modelling a motorcycle accident on the city area Batu is formulated as:

$$\text{Acc} = 7.019 \times 10^{-6} F^{1.702} \exp(-0.2884SW + 0.091S).$$

(b) The main factor which is the infrastructre that influences the geometry of road and costs the motorcycle accident on a road in the city of Batu is the road shoulder width (SW)

(c) As another factor is also known as traffic factor that affects motorcycle accident on a road in the town of Batu that includes the speed and volume of traffic.

(d) From the test results of the distribution of the above models, the empirical data of this study are in accordance with the Geometric distribution.

5.2. Recommendations

(1) The number of samples used in predictive modelling should be more of data related to accidents, in order to obtain a better model.

(2) The collection of the date system should be improved, especially with regard to the identification of the location and cause of traffic accidents, making it easier to analyse the accident and in the decision-making as well as handling of the accident.

(3) It needs also further research based on the results of this study, for example, the causes of the accident from the human aspect, the vehicles and the environment that is infrastructure.

Funding
The authors received no direct funding for this research.

Author details
Sobri Abusini[1]
E-mail: sobri@ub.ac.id
Adem Kilicman[2]
E-mail: akilic@upm.edu.my
ORCID ID: http://orcid.org/0000-0002-1217-963X
[1] Department of Mathematics, University Brawijaya, Malang, Indonesia.
[2] Department of Mathematics, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

Citation information
Cite this article as: Modelling motorcycle accident on the road section by using general linear model: Case studies in Batu City, Sobri Abusini & Adem Kilicman, *Cogent Mathematics* (2017), 4: 1379675.

References

Aitkin, M., Anderson, D., Francis, B., & Hindle, J. (1989). *Statistical modelling in GLIM*. Oxpord: Clarendon Press.

Bhat, U. N. (2015). *An introduction to queueing theory modeling and analysis in aplications*. New York, NY: Copyright Holder Springer Science+Business Media. https://doi.org/10.1007/978-0-8176-8421-1

Comi, A., & Russo, F. (2004). *A modelling system to link end-consumers and distribution logistics* (pp. 6–19).

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York, NY: John Wiley & Sons, Inc./Texas Tech University.

Ghozali, I. (2006). *Applications multivariate analysis with SPSS program*. Semarang: The Agency Publisher Diponegoro University.

Ortuzar, J. D., & Wilumsen, L. G. (1994). *Modelling transport*. New York, NY: John Wiley & Sons.

Sulistio, H. (2007). *In the role of regional road transport safety program*. Malang: University of Brawijaya.

Swan, T., Gilchrist, R., Bradley, M., Clarke, M., Green, P., … O'Brien, C. (1994). *The glim system, release 4 manual*. New York, NY: Oxford University Press.

Tamin, O. Z. (1999). *Planning and transport modeling*. Bandung: Publisher ITB.

Zuherman, M. (2014). *Introduction to queueing theory and stochastic teletraffic model*. Retrieved from https://arxiv.org/pdf/1307.2968v16.pdf.