

# Model Relationship Severity due to Traffic Accidents by Using Loglinear Method

Hasmar Halim<sup>1</sup>, Sakti Adji Adisasmita<sup>2</sup>, Muh. Isran Ramli<sup>3</sup>, Sumarni Hamid Aly<sup>4</sup>

<sup>1</sup>Doctoral Student, Hasanuddin University, Faculty of Engineering Department of Civil Engineering, Jalan Poros Malino Km.6, Gowa, South of Sulawesi, Indonesia

<sup>2</sup>Professor, Hasanuddin University, Faculty of Engineering Department of Civil Engineering, Jalan Poros Malino Km.6, Gowa, South of Sulawesi, Indonesia

<sup>3</sup>Associate Professor, Hasanuddin University, Faculty of Engineering Department of Civil Engineering, Jalan Poros Malino Km.6, Gowa, South of Sulawesi, Indonesia

<sup>4</sup>Associate Professor, Hasanuddin University, Faculty of Engineering Department of Civil Engineering, Jalan Poros Malino Km.6, Gowa, South of Sulawesi, Indonesia

**Abstract:** *The Makassar city as the capital of the province in Indonesia has a fairly high accident rate. In 2015 recorded 810 events with a total of as many as 1,223 people victims. These accidents can be caused by a variety of variables that interact. It is necessary for the analysis of the relationship between variables to get the best relationship model. Loglinear analysis can be used to analyze this type of categorical data and a qualitative one. This study aimed to create a pattern of relationships between variables observation, analyzing models of the best relationships between variables observation that occurred in the city of Makassar. In this study using secondary data obtained from the Traffic Accident Unit (TAU), Police of Makassar and analyzed using loglinear methods to determine the best model of the relationship between variable severity, sex, time and number of wheels of vehicles. A total of 1108 samples accident year 2015. And the period of analysis performed using a hierarchical model testing loglinear produced that all second-order model. A partial test of the highest order by 2 ( $K=2$ ) known to have a level of acceptance of the model unless the relationship between gender and the vehicle does not have a significant relationship.*

**Keyword:** Severity, Traffic Accidents, Log-linear

## 1. Introduction

Traffic accidents are a major indicator of the level of road safety. In developed countries the attention and efforts on the issue continue to be developed in order to minimize the quantity and quality of the accident. Underlying it, the World Health Organization (WHO) and the World Bank issued a report entitled World Report on Road Traffic Injury Prevention. The report mentions, among others, every day at least 3,000 people died in traffic accidents. Of that total at least 85 percent occur in countries with low and moderate incomes. Traffic accidents have also been the cause of 90 percent of disability adjusted life years. In Southeast Asia, the World Health Organization (WHO) notes that, every hour there are 34 people died in an accident on the highway. In 2001 there are 354 000 people died in an accident on the road and around 6.2 million people were hospitalized. The costs resulting from accidents in Southeast Asia reached 14 billion US dollars. WHO estimates that, in the next two decades, this number is expected to rise 144 percent, the highest in the region WHO.[1]

Traffic accident in Indonesia had entered the stage apprehensive. The traffic accident rate was the cause of death largest in Indonesia. The number of victims large enough will give of the effects of economic (material losses and social that were significant. According to data from Indonesian police in 2013, total traffic accident reached 101.037 cases with person died, 25.157 people. With material losses reached IDR. 254.608.956.399, --. Indonesian police data states that, in 2013 happened 12 cases per hour

accident by person died about three person/hours, With potential harm socioeconomic around IDR 203 trillion -- IDR 217 trillion per year (2.9 % to 3.1 % of gross domestic /GDP Indonesia).

City Makassar as the gateway eastern Indonesia having the growth rates traffic accident high. That crash traffic accident unit (TAU) city Makassar noted over a period of years 2015 has happened, traffic accident as many as 810 scenes caused the downfall of the victims as many as 117 people dying, 56 people suffered severe injuries, 917 people had severe wound light and than 133 others only suffered material loss. By 810 the scene of an accident 100% account of your forgetting humans (*human error*).

According to [2], but the human factor the other thing that could led to disaster is a factor vehicles and factors of road/environment .In other research suggested that is not simply of of the three this factor will accident but can be caused by the interplay of third factors are .The interaction can be caused by the human factor by a factor of vehicles , the human factor by a factor of road / environment and factors vehicles and factors of road/environment as well as interplay of third factors are. [3]

The relationship between these variables will form a relationship with a high-dimensional models for the interaction between variables is strong enough. Measurement of the relationship between variables can be done by various methods. To obtain the results of the study are expected to be used in accordance with the analysis techniques corresponding to the data type and scale of measurement. In

general, is one of the linear regression data analysis techniques are used to analyze the data with interval or ratio scale to rely on some classical assumptions.

Linear regression is one of data analysis techniques used to analyze data with scales at least intervals with, dependent on some the assumption classical. Linear regression not can apply to analyze data in cases where variables bound represent the data nominal scale and ordinal or are the type category. Log-linear models can be applied in cases where a bound variable data which type of categorical, while the independent variables can use the existing four measurement scale. In this study, log-linear models were used to analyze the data type of categorical presented in contingency tables to create a pattern of relationships between variables observation, analyzing models of the best relationships between variables observation that occurred in the city of Makassar.

### 1.2. Severity of Traffic Accident

Accidents do not happen by chance, but there is. Therefore, there is cause, the accident must be analyzed and discovered, so that corrective actions for the cause of it can be done as well with the preventive measures further accidents can be prevented. According D.A. Colling cited by [4] accidents can be defined as any event that is not planned and controlled to be caused by humans, circumstances, environmental factors, or combinations of these things that disrupt the work process and can cause injury or not, morbidity, mortality, the damage of property or other undesirable events.

Several other definitions suggests that a traffic accident is an event in the unanticipated and unintended involve vehicles with or without other road users, which resulted in human casualties and/or loss of property. [5]

While understanding the accident, according to [6], Article 93 states that a traffic accident is an event in the way that unexpected and unintentional involving a moving vehicle with or without road users other, resulting in human casualties and loss of property.

As the previous definition that accidents result in losses. These losses are measured by indicators of severity caused by traffic accidents. The severity can be either 1) the victim to death (fatal), 2) serious injury, and 3) slight injury. Accidents not involving other road users is called a single accident. In addition there are types of traffic accidents without casualties, the accident with property damage only (PDO). [7]

### 1.3. The factors that affect accident

Several previous studies have revealed the relationship of the variables that can cause variable severity of traffic accidents. In a study conducted by [8], the severity of the dichotomy into two categories, fatal and non-fatal. The severity as the dependent variable and is supported by nine independent variables using logistic regression concluded that the variable location and the cause of the accident the most significant variables associated with the severity of the accident.

Other research suggests that the severity is affected by vehicle type, season and time of the accident. Estimates of the level of traffic accidents with fatalities with greater conditions occur in the rainy season and greater in the two-wheeled vehicle that occurred at the time of the light. The high accident rate per 1000 incident occurred when the light on the 2-wheeled vehicles in the dry season. [9].

Another case done by [10], suggests that the horizontal alignment of a road condition is strongly associated with the severity of traffic. In addition, the traffic signal conditions seem to play an important role in reducing fatal accidents. From several studies and researches in the field can be concluded that traffic accidents can be influenced by the human factor, vehicle and road environment and the interaction and combination of two or more factors mentioned above [11]. Figure 1 below shows the interaction between these factors shows the interaction between the factors that cause traffic accidents.

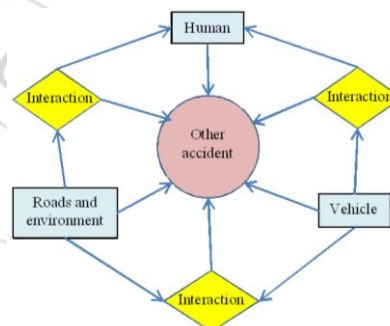


Figure 1: Accident Cause Factors and their Interaction

### 1.4. Log-Linear Model on Contingency Tables

Log-linear regression was applied a linear modeling approach that can be used for data that is spread Poisson. Log-linear model is a development of two-way analysis of cross tables or where there is a relationship between two or more variables were analyzed using the natural logarithm of the content of cells in the table. Log-linear models are widely used in analyzing the data of the type categorical. Type categorical a transformation function of the four measurement scale, the scale of nominal, ordinal, ratio, and interval.

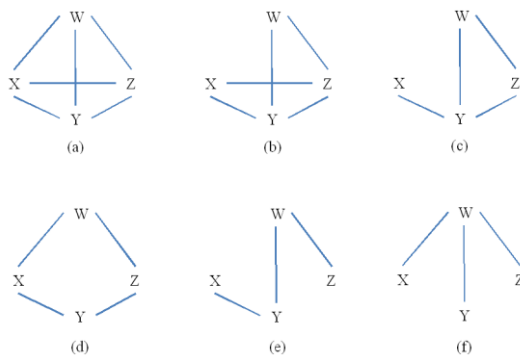
Besides can be used to analyze relations between two variable categories, this model can also be used to analyze table contingency multivariate involving three or more using test goodness-of-fit statistics. Model logs linear, multivariate is the expansion of the logs linear. Trivariate namely model logs-linear having three variations. Model logs-linear, multivariate happens if classifications, cross done with three variables category or more.

### 1.5. Log-Linear Model Four Dimensions

According to [12] theoretical model for the four-dimensional log linear model is shown in Figure 2. This figure shows an illustration of six kinds of theoretical models that occurs between the four dimensions, where a line connecting a pair dimension would argue that both dimensions associated

substantially included the possibility of entry into force of a causal relationship.

Based on the four dimensions there are six pairs maximum possible dimensions of association, the most comprehensive theoretical model that includes six pairs of the association presented graphically on the model (a). Model (b) contains only 5 pairs of variables associated. Model (c) and Model (d) contains 4 pairs dimensions with different patterns of association. While the model (e) and model (f) contains three pairs of variables with different patterns of association and partnership dimensions that are not related to the line will state that these dimensions are not associated with substance, or one dimension does not have a direct impact on other dimensions.



**Figure 2:** Illustration of six kinds of theoretical models that occurs between the four dimensions

**1.6. Model hierarchy**

According to [12] log linear model hierarchy of the four dimensions category will include four main variables, example W, X, Y, Z and subsets (or most) of the set of all the variables of interaction has a) Six interaction of two variables b) Four interaction 3 variables and c) An interaction 4 variables. From these variables there are 23 models it is possible to log linear model analysis of four dimensions. As an illustration of a log linear model for a table multivariate involving three or more variables, given a table of four dimensions measuring  $i \times j \times k \times l$  with the variables W, X, Y, Z where total surveillance is no and the observation of the cell  $(i, j, k, l)$  with  $n_{ijkl}$  while  $m_{ijkl}$  denoted as an expectation value for the cell  $(i, j, k, l)$  under the assumption of several parametric models.

Log linear model is the simplest to the table four independent dimensions of the complete model / mutual independence for all the variables to the equation: [13]

$$\text{Log } m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z \quad (1)$$

Four-dimensional contingency tables, log linear models of complex containing  $\binom{4}{2} = 6$ , the possibility of a form of interaction of two variables such as  $\{\lambda_{ij}^{WX}\}$ ,  $\binom{4}{3} = 4$  the possibility of such variables forms of interaction  $3\{\lambda_{ijk}^{WXY}\}$  and one form of interaction of the four variables  $\{\lambda_{ijkl}^{WXYZ}\}$ . This model is symbolized by (WXYZ) and have similarities. [14]

$$\text{Log } m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} + \lambda_{ijk}^{WXY} + \lambda_{ijl}^{WXZ} + \lambda_{jkl}^{XYZ} + \lambda_{ikl}^{WYZ} + \lambda_{ijkl}^{WXYZ} \quad (2)$$

Model without interaction three factors , denoted or symbolized with (WX , WY , WZ , XY , XZ , YZ) have form model linear.

$$\text{Log } m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} \quad (3)$$

Analogous to table three dimensions , it was reflected from a model for the table four dimensions restricted to the principle of a hierarchy , where if enroll a form of interaction , then can also have to cover all forms of the interaction that lower grade. For example, if the quote  $\{\lambda_{ijl}^{WXZ}\}$  in a model, then it must include  $\lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{jk}^{XY}$ .

**2. Methodology**

The data used in this research is secondary data obtained from the Traffic Accident Unit, Polrestabes of Makassar in 2015. The population is all that involves two wheels, three wheels, four wheels, six wheels and wheel  $\geq 10$  who suffered a traffic accident in the City Makassar period in 2015. the sample is part of the population from accidents resulting in severity from minor injuries to death.

The variability of this study is the first is the severity of traffic casualties. In this variable are divided into three categories, namely death, serious injuries and minor injuries. The second variable is the gender variable. This variable consists of categories of male and female. The third variable is the time of the accident that dichotomy into daylight and night. Daylight defined as accidents that occur between the hours of 6:00 to 18:00., while the night is an accident that occurred between 18:00 to 24:00. The last variable is a variable vehicle, are divided into five categories depending on the number of tires on the vehicle involved in the accident. The category consists of two wheels, three wheels, four wheels, six wheels and wheel  $\geq 10$ .

The next step log-linear model analysis of four dimensions with the help of IBM-SPPS 21 application. The analysis measures include:

- a) Determining the Minimal Self Statistics and Likelihood Function  
 Minimal sufficient statistic for each model is the number of observed frequencies corresponding to the index of the high order of the parameters that exist in the model. Therefore, the value of statistical adequacy is also statistically quite minimal, then how to find statistically quite minimal, assuming a model of simple samples  $\{n_{ijkl}\}$  for cross-classification of variables random poisson (W, X, Y, Z) are independent with expected value  $m_{ijkl}$ .
- b) Estimated Frequency Hope  
 Estimates of the expected frequency is an estimate of the value that occurs each categorical on each variable. How to find hope for the frequency estimation log linear model four-dimensional analog of frequency estimation expectations for the log-linear model of three dimensions. Estimates of the expected frequency is used to calculate

the residual value.

c) Test *Goodness Of Fit*

Goodness of Fit Test is used to determine whether a log linear model used matches the actual circumstances. Goodness of Fit Test can use two test statistics are statistics Chi-Square or Likelihood Ratio Square. Statistics Chi-Square also be used to determine whether there is a significant relationship between the variables measured.

hypothesis:

$H_0$  : log linear model obtained in accordance with the actual state.

$H_1$  : log linear model was not obtained in accordance with the actual state.

Chi-Square statistical formula is:

$$\chi^2 = \sum_{i,j,k,l} \frac{(O_{ijkl} - E_{ijkl})^2}{E_{ijkl}} \quad (4)$$

If  $\chi^2_{\text{calculate}} \leq \chi^2_{\text{tabel}}$  and p-value > significance level  $\alpha = 0.05$  then log linear model used in accordance with the actual situation.

Likelihood Ratio Test statistical formula Square is:

$$G^2 = 2 \sum_i \sum_j \sum_k \sum_l O_{ijkl} \log \left( \frac{O_{ijkl}}{E_{ijkl}} \right) \quad (5)$$

where:

$O_{ijkl}$  = observation on variable i, j, k, l

$E_{ijkl}$  = the expected frequency

Degree of freedom for the test statistic Goodness of Fit table, four dimensions are determined using a way analogous to the three-dimensional table Formulas degrees of freedom for the model with all the forms of interaction of three factors at table four dimensions.

d) Selection of Models

Multivariate contingency table can have one among many that might be a log linear model. Selection of a suitable log linear model is very difficult, there should be a strategy. The strategies that can be done is with gradual strategy procedure. It starts with choosing a significant level and then proceed with testing the Goodness of Fit by selecting criteria  $G^2$  likelihood ratio statistic values are relatively smaller (less than the value of Chi-Square table) and the p-value is relatively large (more than a significance level  $\alpha = 0.05$ ). Once the model is selected, do further analysis to choose the best model that is using the partition Chi-Square.

### 3. Results and Discussion

#### 3.1. Multivariate contingency table Four Dimensions Using Log-Linear Model

Secondary data from the traffic accidents unit Police of Makassar, then calculated the amount of each category and multivariate contingency table included in four dimensions. Four-dimensional multivariate contingency tables are presented in Table 1 below:

**Table 1:** Multivariate contingency table Four Dimensions

Severity	sex	Time	Vehicle (number of wheels)					Total
			R2	R3	R4	R6	>=R10	
death	male	daylight	37	1	12	12	1	63
		night	43	1	18	6	1	69
	female	daylight	8	0	6	1	0	15
		night	4	0	2	0	0	6
Serious injury	male	daylight	21	1	5	4	0	31
		night	13	0	5	2	0	20
	female	daylight	5	0	6	0	0	11
		night	6	0	1	2	0	9
Slight injury	male	daylight	253	9	88	17	2	369
		night	121	2	81	12	0	216
	female	daylight	147	6	54	3	1	211
		night	64	0	20	3	1	88
<b>Total</b>			722	20	298	62	6	1108

#### 3.2. Model Relationships with Hierarchical Log Linear

Based on the relationship patterns theoretically described earlier, and then do the analysis of the log-linear hierarchy to determine the statistical significance level models. Results of testing the model with log-linear hierarchy as illustrated in Table 2. This output provides information on the effects of variables that have been selected to construct a model.

**Table 2:** Testing Hierarchy log linear model with K-High Order

	K	df	Likelihood Ratio		Pearson	
			Chi-Square	Sig.	Chi-Square	Sig.
K-way and Higher Order Effects <sup>a</sup>	1	59	2991,131	,000	5787,018	,000
	2	51	119,676	,000	124,835	,000
	3	30	34,723	,253	35,662	,219
	4	8	4,270	,832	3,194	,922
K-way Effects <sup>b</sup>	1	8	2871,455	,000	5662,183	,000
	2	21	84,953	,000	89,173	,000
	3	22	30,453	,108	32,468	,070
	4	8	4,270	,832	3,194	,922

In the log-linear analysis of higher order (K = 4), which contains all models (saturated) to the conclusion that the model was not significant ( $\alpha > 0.05$ ), so use lower order. The next analysis is the 3rd order model, in this analysis have the same result with the previous analysis of the resulting models are not real ( $\alpha > 0.05$ ). Because of the two previous analyzes have not produced tangible model then continued on the next analysis is the order of the 2nd. In this order, lead to the conclusion that the real model ( $\alpha < 0.05$ ). Based on the analysis, it is known that the model loglinear used by a minimum must contain two possible order. Next is to analyze the variables with the first order and second order which have a significant relationship with the use of partial analysis models loglinear.

Partial analysis log linear model with a second order shown in Table 3. This analysis is used to infer the level of significance of these variables partially relationship.

**Table 3:** Partial testing Associations log linear with K-High Order

Effect	df	Partial Chi-Square	Sig.
TK*JK	2	22,747	,000
TK*W	2	9,454	,009
JK*W	1	6,245	,012
TK*K	8	18,043	,021
JK*K	4	6,659	,155
W*K	4	10,111	,039
TK	2	1039,208	,000
JK	1	169,707	,000
W	1	77,870	,000
K	4	1584,670	,000

TK : Severity  
 JK : Sex  
 W : Time  
 K : Vehicle  
 Df : Degree of freedom  
 Sig. : Signifikan

From the analysis of the partial with the highest order of 2 as illustrated in Table 3, it can be concluded that there is a relationship or a very strong association between sex, time and vehicle to the severity of traffic accidents. This is evident from the significant value ( $\alpha < 0,05$ ). This is according to research conducted by [15] suggests that the difference between the sexes did not differ in the involvement of traffic accidents. In other studies also suggest that the risk of death was found to be associated with age and gender of victims, as well as the type of vehicle involved. [16]. In Table 3 also provides an illustration of the lack of a significant relationship between the variables gender and vehicle variables. This can be seen with significant values ( $\alpha > 0.05$ ). This is contrary to the opinion [2] and [17], which reported an interaction between the variables of human, vehicle and road/environment.

The estimated number of traffic accidents in each association in question can be seen by the expected value. Table 4 portrays the expectation value for each association studied. Estimates of the number of accidents, according to Table 4 identifies that the tendency of men to have a greater chance of dying in the same conditions that caused or were using motorcycles than women at the same time conditions, namely when the daylight conditions.

**Table 4:** Parameter estimation of a log linear model with variable severity, sex, time and vehicle

severity	sex	time	vehicle				
			2-wheels	3-wheels	4-wheels	6-wheels	≥10-wheels
death	male	daylight	41,265	1,224	12,979	9,517	1,089
		night	36,652	0,53	20,217	8,03	0,498
	female	daylight	7,979	0,246	3,043	0,497	0,162
		night	6,104	0	1,761	0,956	0,252
Serious injury	male	daylight	18,767	0,552	5,716	4,13	0
		night	12,423	0,178	6,635	2,598	0
	female	daylight	8,795	0,27	3,247	0,522	0
		night	5,015	0	1,401	0,75	0
Slight injury	male	daylight	250,968	9,224	86,305	19,353	1,911
		night	127,926	2,292	77,147	9,372	0,502
	female	daylight	143,225	5,484	59,711	2,981	0,838
		night	62,881	0	19,838	3,294	0,748

Estimates of the highest severity level that caused death in the amount of 41.265 on the motorcycle while the lowest in three-wheeled vehicles. The highest percentage of severity

that caused the death of 62.45% of motorcycles compared followed by 4-wheel vehicles (light vehicles) reached 19.64%. The tendency of expectation was higher by 3.17 times compared to lighter vehicles. While based on time (light and dark), the tendency of an accident which would cause death occurs in conditions when the light except 4-wheel vehicle inclination occurs in a dark time.

#### 4. Conclusion and Recommendation

Loglinear method can determine the model of the relationship between variables, severity, gender, time and vehicle. Hierarchical Model of testing loglinear with K-High Order containing all models (saturated) note that the real models that can be accepted is the order of the 2nd. It is identified that the minimum must include a significant variable. Partial test of the highest order by 2 ( $K = 2$ ) note that all variables have other variables association unless the association between the variables sex with a vehicle that is not significant. The study also recommends that further research with different variables and methods.

#### 5. Acknowledgements

A big thank you to windha and Silvia who have helped in the process of inputting the data of traffic accidents.

#### References

- [1] Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. (2004). *World report on road traffic injury prevention*. World Health Organization. Retrieved from [http://www.who.int/violence\\_injury\\_prevention/publications/road\\_traffic/world\\_report/en/](http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/)
- [2] Vogel, L.; Bester, C. J. (2005). A Relationship Between Accident Types and Causes. *The 24th Annual Southern African Transport Conference*, (July), 233–241.
- [3] Lum, Harry; Reagan, J. A. (1995). *Interactive Highway Safety Design Model: Accident Predictive Module. Public Roads*.
- [4] Manurung, J. R. H. (2012). *Hubungan Faktor-Faktor Penyebab Dan Akibat Kecelakaan Lalu Lintas Pada Pengendara Sepeda Motor Di Kota Medan Tahun 2008 – 2010*. Universitas Sumatera Utara.
- [5] Undang-Undang Republik Indonesia, Pub. L. No. 22, Tentang Lalu Lintas Dan Angkutan Jalan (2009). Indonesia.
- [6] Peraturan Pemerintah Republik Indonesia, Pub. L. No. 43, Tentang Prasarana Dan Lalu Lintas Jalan (1993). Indonesia.
- [7] Soemitro, R. A.,;Bahat, Y. S. (2005). Accident Analysis Assessment To The Accident Influence Factors On Traffic Safety Improvement. EASTS.
- [8] Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis and Prevention*, 34(6), 729–741. [http://doi.org/10.1016/S0001-4575\(01\)00073-2](http://doi.org/10.1016/S0001-4575(01)00073-2).
- [9] Indriani, Diah; Indawati, R. (2006). Model Hubungan dan Estimasi Tingkat Keparahan Kecelakaan Lalu Lintas. *Berita Kedokteran Masyarakat*, 22, 100–106.
- [10] Sohn, So Young, S. H. W. (2001). Traffic Accident

- Severity. In *Proceedings of the Eastern Asia Society for Transportation Studies* (Vol. 3, pp. 173–184).
- [11] Departemen, & Perhubungan. (2006). *Rencana Umum Keselamatan Transportasi Darat*.
- [12] Agung, I. G. N. (1989). *Analisis Data Kategorik Multivariat: Pemakaian Model Log-Linier* (Pertama). Yogyakarta: Pusat Penelitian Kependudukan Universitas Gadjah Mada.
- [13] Agresti, A. (2007). Loglinear Models for Contingency Tables. In *An Introduction to Categorical Data Analysis* (pp. 204–243). <http://doi.org/10.1002/9780470114759.ch7>
- [14] Maronna, R. (2005). Analyzing Categorical data. *Other*, 46(3), 473–474. <http://doi.org/10.1007/978-0-387-21727-7>
- [15] Lourens, P. F., Vissers, J. A. M. M., & Jessurun, M. (1999). Annual mileage, driving violations, and accident involvement in relation to drivers' sex, age, and level of education. *Accident Analysis and Prevention*, 31(5), 593–597. [http://doi.org/10.1016/S0001-4575\(99\)00015-9](http://doi.org/10.1016/S0001-4575(99)00015-9).
- [16] Jones, A. P., & Jørgensen, S. H. (2003). The use of multilevel models for the prediction of road accident outcomes. *Accident Analysis and Prevention*, 35(1), 59–69. [http://doi.org/10.1016/S0001-4575\(01\)00086-0](http://doi.org/10.1016/S0001-4575(01)00086-0)
- [17] Zhang, G., Yau, K. K. W., & Chen, G. (2013). Risk factors associated with traffic violations and accident severity in China. *Accident Analysis and Prevention*, 59, 18–25. <http://doi.org/10.1016/j.aap.2013.05.004>.

