

Multinomial Logistic Regression Model to Analysis Traffic Accident on Indonesia's Regional Data

Agustini Tripena¹, Yhenis Apriliana², Agus Sugandha³, Agung Prabowo^{4,*}, Sukono⁵

^{1,2,3,4}Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto, Indonesia.

⁵Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, Indonesia

Corresponding Author: agung.prabowo@unsoed.ac.id

Abstract

Traffic accident is one of the highest causes of death in Indonesia following coronary disease and tuberculosis. Traffic accident is classified into three: mild, moderate and severe. The aim of this research was to determine the significant factors causing traffic accidents in Cilacap Regency using multinomial logistic regression. The data used were secondary data from the Resort Police of Cilacap Regency. The research's response variable was accident classification deemed as a nominal scale, while the predictor variables were day of occurrence, time of occurrence, accident type, accident location, situation, weather problem, number of vehicles involved, and number of victims of nominal, ordinal and ratio scales. The research results show that the accident type, accident location, situation, and number of victim variables significantly influence the three accident classifications.

Keyword: Cilacap, accident type, accident classification, accident location, data scale

INTRODUCTION

The introduction explains the urgency of this research, previous research and the research's objective. Traffic accident is an unexpected and incidental event occurring in public road and involving various types of vehicles and leading to death tolls and property damage. According to (Abdillah, 2020), traffic accident in Indonesia is the third biggest killer following coronary artery disease and tuberculosis. The factors causing traffic accident, according to



(Ambo, 2020) are human, vehicle and environment factors. Therefore, it is very important to research on driving behaviors. Furthermore, driving behaviors are observable through three accident classifications: mild, moderate and severe.

Based on the data acquired from Traffic Corps of the Indonesian National Police, in 2021 there were 103,645 cases of traffic accident in Indonesia with 117,913 having mild wound, 10,553 having severe wound and 25,266 death tolls. In the same year, in Cilacap Regency the traffic accident increased to 867 cases with mild wound, severe wound and death tolls respectively 1,056, 2 and 184.

In Indonesia context, research was conducted on the factor influencing the severity of traffic accident in Padang City, West Sumatera Province by Ambo, 2020) using an ordinal logistic regression analysis. The predictor variables used in the research were age, sex, type of accident, role of victim, and type of vehicle. Meanwhile, (Findasari, 2023) conducted similar research in Ambon City, Maluku Province using a multinomial logistic regression analysis with sex, age, education and type of vehicle as the influencing variables.

Statistic modeling related to traffic phenomena has, among others, been reviewed by (Abdillah, 2020), Ambo, 2020), and (Apriyani, 2021). The research by (Abdillah, 2020) using a multinomial logistic regression model concluded that there were different factors influencing crashes between non-interstate and interstates. In the research using a multinomial logit method, (Apriyani, 2021) found factors significantly influencing traffic violation. Meanwhile, (Abdillah, 2020) using a binary logit model identified factors influencing the mode of choice of transport in Johor Bahru City.

Logistic regression is a statistic method used to describe the relation between predictor variable and response variable with two or more categories. Data processing involving response variable in the form of nominal or ordinal data with binary or ternary data value cannot be conducted using simple or multiple linear regression. The appropriate method for such case is simple (binomial) or multiple (multinomial) logistic regression. Multinomial logistic regression is a logistic regression with response variable more than two categories (polychotomy) whose data are of nominal or ordinal scale (Apriyani, 2021). The research by (Ambo, 2020) using a multinomial logistic regression model showed that gender, regional origin, learning motivation and learning resources significantly influenced learning difficulty for Statistics course.

The binary and ternary logistic regressions methods can be used to determine the relation of a response variable with one or more predictor variables. The classifications of binary response variable for nominal data in logistic regression are 0 (fail) and 1 (success). Meanwhile, the classifications of ternary response variable for ordinal data are 0 (mild), 1 (moderate) and 2



(severe). Mohammed (2021) gave six assumptions to be satisfied for research using a multinomial logistic regression model.

The objective of this research was to determine the factors that significantly influence the occurrence of traffic accident in Cilacap Regency using the 2021 traffic accident data and compare the results with those of previous research. This research used one response variable with three categories and eight predictor variables with nominal, ordinal and ratio data scale (Table 1).

Using logistic regression requires information on the odds ratio capable of explaining the relation of ternary response variable with the predictor variable of the model. The parameter assessment in the research's logistic regression used a maximum likelihood (ML) method. Wald and G Statistics were used for partial and simultaneous and regression coefficient test, while Pearson chi-square test was used for model compatibility.

This article presents the introduction in Section 1, the theory about ARIMA and SARIMA in Section 2 and the methods in Section 3. Furthermore, Section 4 explains the results and discussion. This article ends with conclusion in Section 5.

The forecasting methods used for response variables of ration type include regression and time series. Multiple linear regression model connecting response variable Y with two or more predictor variables $\mathbf{x} = (x_1, x_2, \dots, x_p)$ in equation:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ is the regression parameter and ε is the model error assumed to be normally distributed with mean 0 and variance σ^2 . Since ε is normally distributed, response variable Y is also normally distributed, thus the range for Y value is real number set $\mathbb{R}$. Furthermore, since $E[\varepsilon] = 0$, the mean of Y given observation value $\mathbf{x} = (x_1, x_2, \dots, x_p)$ is equation:

$$E[Y|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2)$$

Forecasting using simple linear regression can use moment, ordinary least square (OLS) or maximum likelihood (ML) methods for parameter assessment. The use of simple linear regression with parameter assessment using OLS method can be found in (Apriyani, 2021). In case assumptions of simple linear regression are not satisfied, parameter assessment can be conducted using simple averaging method (Ambo, 2020) and improved simple averaging method (Hosmer, 2017). Meanwhile, in case the response variable is nominal or ordinal, simple linear regression cannot be used. The solution is logistic regression.

Research using logistic regression in its data modeling includes (Apriyani, 2021), (Findasari, 2023), (Hidayati, 2017), (Hosmer, 2013), and (Kholisatin, 2021). The logistic regression model in the three studies only used one or two predictor variables. The logistic regression model can be presented with equations

$$p(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (3)$$

$$p(x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}} \quad (4)$$

where $p(x_i)$ in equations (3) and (4) being the probability of success and β_i being the regression coefficient.

The multinomial logistic regression model in equation (4) only used two predictor variables. The multinomial logistic regression model influenced by P predictor variable $x_1, x_2, \dots, x_p$ can be expressed in equation:

$$E[Y|\mathbf{x}] = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (5)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_p)$ being the vector of observation for predictor variable $x_1, x_2, \dots, x_p$, while β_0 states the intercept variable of the model, and β_k states the regression parameter for predictor variable k , $k = 1, 2, \dots, p$.

Write conditional probability $P(Y=1|\mathbf{x}) = \pi(\mathbf{x})$ and $P(Y=0|\mathbf{x}) = 1 - \pi(\mathbf{x})$. Equation (5) can be written into equation:

$$\pi(\mathbf{x}) = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (6)$$

Further, equation (6) can be linearized using logit transformation and the result be stated as $g(\mathbf{x})$ in equation:

$$g(\mathbf{x}) = \ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

The result of equation (7) shows that the multinomial logistic regression equation with P predictor variable(s) can be stated as multiple linear regression in equation (2). This is made

through taking the exponential value from function $g(\mathbf{x})$ and $\ln(\pi(\mathbf{x})/1-\pi(\mathbf{x}))$ in the first and second segments of equation (7). Later, ratio $\pi(\mathbf{x})/1-\pi(\mathbf{x})$ in equation (7) is called odds ratio.

This research used a multinomial logistic regression model with eight predictor variables and one response variable with three categories $Y=0, Y=1, Y=2$. One of the three categories was taken as reference in this case category $Y=2$. The multiple regression model obtained was equation that involved four predictor variables:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \quad (8)$$

The logistic regression model corresponding to multiple linear regression equation (8) was:

$$p(x) = \frac{\exp(g_j(x))}{\sum_{j=0}^{J-1} \exp(g_j(x))} \quad (9)$$

where J stating the number of predictor variables. Further, $g_0(0)=0$ applied to $j=0$, thus equation (9) could be expressed as:

$$p(x) = \frac{\exp(g_j(x))}{1 + \sum_{j=1}^{J-1} \exp(g_j(x))}$$

The parameter assessment in equation (9) was performed using ML method numerically settled using Newton-Raphson iteration with assistance of SPSS 26.0 software. Significance test on the coefficient of logistic regression equation (9) was conducted using G ratio likelihood test that was distributed chi-square with degree of freedom p that was equation:

$$G = -2 \ln \left[\frac{\left(\frac{1}{n} \sum_{i=1}^n y_{0i} \right)^{\sum_{i=1}^n y_{0i}} \left(\frac{1}{n} \sum_{i=1}^n y_{1i} \right)^{\sum_{i=1}^n y_{1i}} \dots \left(\frac{1}{n} \sum_{i=1}^n y_{(J-1)i} \right)^{\sum_{i=1}^n y_{(J-1)i}}}{\prod_{i=1}^n \hat{\pi}_{0i}^{y_{0i}} \hat{\pi}_{1i}^{y_{1i}} \dots \hat{\pi}_{(J-1)i}^{y_{(J-1)i}}} \right]$$

where $\hat{\pi}_{ji} = \hat{\pi}_j(\mathbf{x}_i)$, $j=0,1,2,\dots,J-1$. Numerator and denominator in equation (9) were respectively the likelihood of model without predictor variable and the estimated likelihood of model with predictor variable. Hypothesis H_0 was rejected if $G > \chi^2_{(\alpha;p)}$. If H_0 were rejected, it was to conclude that at least one predictor variable influenced response variable.

A partial significance test was conducted using Wald test to test H_0 that the predictor variables partially influenced the response variable. The test statistics used was the equation

$$W_{jk} = \left[\frac{\hat{\beta}_{jk}}{SE(\hat{\beta}_{jk})} \right]^2, \quad (11)$$

where the test was distributed chi-square with degree of freedom 1. The hypothesis H_0 was rejected if $W_{jk} > \chi^2_{(\alpha;1)}$. If H_0 were rejected, it was to conclude that predictor variable k influenced logit function j . Symbol $\hat{\beta}_{jk}$ in equation (11) states the estimated parameter value k and $SE(\hat{\beta}_{jk})$ states standard deviation of $\hat{\beta}_{jk}$, $j=1,2,\dots,J-1$ and $k=1,2,\dots,p$.

The model fit test was conducted using Pearson Chi-square test with test statistics in equation

$$\hat{C} = \sum_{g=1}^G \sum_{j=1}^{J-1} \frac{(O_{gj} - \hat{E}_{gj})^2}{\hat{E}_{gj}}, \quad (12)$$

that was distributed Chi-square with degree of freedom $(G-2)(J-1)$. The symbols in equation (12) were O_{gj} , stating the frequency of appearance of category j in group g , and $\hat{E}_{gj}$ stating estimated expectation value of appearance of category j in group g , $g=1,2,\dots,G$. Hypothesis zero was rejected if $\hat{C} > \chi^2_{(\alpha;(G-2)(J-2))}$. Hypothesis H_0 rejection concluded that the model was incompatible with the data.

The coefficient could be interpreted based on the parameters of logistic regression using odds ratio. Odds mean the probability of an event occurring compared to the probability of non-occurrence of such event. If response variable $Y=0$ were assumed to be reference category, the odds ratio for response variable $Y=j$ to $Y=0$ calculate on two values (e.g. $x=a$ and $x=b$) was equation:

$$OR_j(a,b) = \frac{\exp(g_j(x=a))}{\exp(g_j(x=b))}. \quad (13)$$

The value $OR=1$ shows that $x=a$ had the same tendency with $x=b$ to generate $Y=j$. The value $1 < OR < \infty$ means $x=a$ had the tendency of bigger than $x=b$ to generate $Y=j$ and vice versa for $0 < OR < 1$.

METHOD

The research methodology explained the data used and the research measures. The data used were divided into two variables: a response variable of nominal scale and eight predictor variables of nominal, ordinal and ratio scales.

The research's response variable was nominal data composed of three types of traffic accident: mild, moderate, and severe. Research using similar response variable had been conducted by (Apriyani, 2021), where response variable was classified into three categories, victims with mild, moderate and severe wound, as type of data of nominal scale.

Traffic accident according to law No. 22 Year 2009 (Abdillah, 2020) is classified into three: (1) mild if the accident only causes damage to vehicle and or goods, (2) moderate if it causes mild wound and damage to vehicle and or goods, and (3) severe if it causes victim severe wound or death.

The definition of mild, moderate and severe wounds refers to Government Regulation No. 17 Year 2013 (Ambo, 2020) on Road Facilities and Traffic. Traffic accident victim severity is classified into three: (1) mild wound if victim does not need hospitalization or needs hospitalization for less than 30 days, (2) severe wound if the wound causes permanent handicap or hospitalization for more than 30 days, and (3) death if the victim is confirmed dead as the consequence of the traffic accident within 30 days from the accident.

Meanwhile, the predictor variables used was day of occurrence, time of occurrence, accident type, accident location, situation, weather problem, number of vehicles involved, and number of victims. The response and predictor variables are explained in Table 1.

The order of research measures was as follows: (1) descriptive statistical analysis on traffic accident data; (2) multicollinearity test between predictor variables; (3) formulating multinomial logistic regression model; (4) estimating model parameter using maximum likelihood method; (5) model significance test composed of simultaneous test and partial test; (6) model compatibility test; and (7) interpreting odds ratio value. The order of measures (8) presented analysis on the research results in relation to some previous research results.

RESULTS AND DISCUSSION

Descriptive Statistics of Response and Predictor Variables

There were 867 accidents in 2021 in Cilacap Regency, Central Java Province. Based on the predetermined criteria, there were three accident classifications: mild, moderate and severe. Table 2 presents the number and percentage of each accident classification. Based on Table 2, despite only 1 mild accident, referring to the definition of types of accident set forth in Law No. 22 Year 2009, the data cannot be integrated into moderate accident, or removed from the analysis.

Table 3 summarizes the number of accidents occurring by predictor variable used in this research. Referring to Table 3, we may conclude that:



1. the frequency of accident on business days tends to be dominating compared to that on holidays for each type of accident;
2. the frequency of accident in the morning and afternoon shows only slight difference;
3. the frequency of front-side accident tends to be dominating each traffic accident;
4. the frequency of straight-location accidents tend to be dominating each traffic accident, and crossing locations are also points prone to moderate accident;
5. crowded situation is quite dominant for accident, and despite moderate level accident, but the number is quite high; and
6. there were totally 834 accidents in sunny weather out of the 867 accident cases, and moderate accident occurrence in sunny weather was very high, 675 out of 867 accident cases.

The next predictor variable is number of vehicles involved in each type of accident, symbolized with X_7 . The observation data show only 1 vehicle having mild accident. There were 1,086 and 313 vehicles having moderate and severe accidents respectively. Thus, there were totally 1400 vehicles having accident in Cilacap Regency in 2021.

Meanwhile, the number of victims of accident X_8 in Cilacap Regency in 2021 was totally 1240 people. Almost half had moderate accident of 955 people. Meanwhile, 285 people had severe accident.

Multicollinearity Test

The multicollinearity test in multinomial logistic regression was conducted to test linear correlation between the eight predictor variables. Based on Table 4, there was no VIF value more than 10, which means there was no multicollinearity between the predictor variables. Thus, all of the predictor variables could be used for further test.

Early Model of Multinomial Logistic Regression

Multinomial logistic regression was formulated involving eight predictor variables. Regression coefficient was determined with assistance of SPSS whose output was presented in Table 5. Each calculation made final category as reference. For example, for the day of occurrence variable, the Holiday category was made reference, so that in Table 5 only parameter estimates are available for the Working Days category. In addition, this calculation also uses the classification of serious accidents as a reference. As a result, in Table 5 there are only response variables for mild and moderate classification.



Table 5. Output Estimation for Model Parameter with All Predictor Variables

Predictor Variables		Response Variable	
		Parameter Estimation for Accident Classification	
		Mild	Moderate
Intercept	-	11.715	0.503
Day of occurrence	Working days	0.241	-0.425
Time of occurrence	Early morning	-2.176	-0.116
	Morning	-3.013	0.176
	Noon	-3.036	0.167
Accident type	Single	4.581	1.845
	Front-Front	-0.780	-0.935
	Front-Side	-1.403	-0.762
	Front-Back	-1.434	-0.806
	Sideways	-1.022	-0.583
	Streak	-1.198	-0.679
	Hit pedestrian	-1.232	-0.503
	Hit animal	2.837	-1.920
Accident location	Go strike	1.127	1.144
	Bend	-0.355	1.192
	Intersection	1.132	1.169
Situation	Lonely	-3.953	0.265
Weather problem	Rain	-1.208	0.435
Number of vehicle involved	-	0.230	-0.306
Number of victims	-	-27.679	-0.283

Simultaneous significance test with G ratio likelihood statistical test using equation (10) concluded that there was at least one predictor variable influencing response variable. The test is carried out by comparing the p-value with a significance level of 0.05. Therefore, it is necessary

to proceed with the partial test. SPSS output results for simultaneous tests are available in Table 6. Meanwhile, partial test using Wald statistical test with equation (11) suggested that further analysis, only four predictor variable were to be used: accident type, accident location, road situation and number of victims (Table 7).

Table 6. Output Simultaneous Model Significance Test

Model	Log Likelihood	Likelihood Ratio Test		
		Value of Statistic Test	Degree of Freedom	<i>p-value</i>
No Variable	535.619			
With Variable	423.212	112.407	38	0.000

Based on the output of Table 7, in minor accidents there is not a single category of predictor variables that affect the response variable. Whereas in moderate accidents, there are four influential predictor variables, namely the type of accident, accident location, situation and number of victims with each influential category having a *p-value* lower than the 0.05 significance level (yellow color). These results suggest that in the subsequent analysis only four predictor variables were used, namely the type of accident, the location of the accident, the situation and the number of victims

Table 7. Output Partial Model Significance Test

Predictor Variable		Response Variable					
		Mild Accident			Moderate Accident		
		Parameter Estimation	Wald Value	<i>p-value</i>	Parameter Estimation	Wald Value	<i>p-value</i>
Day of occurrence	Working days	0.241	0.000	1.000	-0.425	3,438	0.064
Time of occurrence	Early morning	-2.176	0.000	0.999	-0.116	0.108	0.742
	Morning	-3.013	0.000	0.998	0.176	0.424	0.515
	Noon	-3.036	0.000	0.998	0.167	0.379	0.538

Accident type	Single	4.581	0.000	0.998	1.845	9.327	0.002
	Front-Front	-0.780	0.000	1.000	-0.935	1.184	0.277
	Front-Side	-1.403	0.000	1.000	-0.762	0.823	0.364
	Front-Back	-1.434	0.000	1.000	-0.806	0.873	0.350
	Sideways	-1.022	0.000	1.000	-0.583	0.366	0.545
	Streak	-1.198	0.000	1.000	-0.679	0.170	0.680
	Hit pedestrian	-1.232	0.000	1.000	-0.503	1.080	0.299
	Hit animal	2.837	0.000	1.000	-1.920	1.694	0.193
Accident location	Go strike	1.127	0.000	1.000	1.144	5.552	0.018
	Bend	-0.355	0.000	1.000	1.192	4.438	0.035
	Intersection	1.132	0.000	1.000	1.169	6.854	0.009
Situation	Lonely	-3.953	0.000	0.997	0.265	5.067	0.024
Weather problem	Rain	-1.208	0.000	1.000	0.435	1.095	0.295
Number of vehicle involved	-	0.230	0.000	1.000	-0.306	0.219	0.640
Number of victims	-	-27.679	0.001	0.982	-0.283	6.737	0.009

Final Model of Multinomial Logistic Regression

The parameters were estimated for predictor variables that were proven significantly influencing response variable. Logistic regression formulation in this stage only involved four predictor variables, and estimation for each category of the four predictor variables is presented in Table 8.

Table 8. Output Final Estimation for Model Parameter with Significance Predictor Variables

Predictor Variable		Response Variable					
		Mild Accident			Moderate Accident		
		Parameter Estimation	Wald Value	P-value	Parameter Estimation	Wald Value	P-value
Intercept	-	10.921	0.000	0.999	0.037	0.001	0.975
Accident type X_3	Single x_{30}	4.945	0.000	0.999	1.791	9.061	0.003
	Front-Front	0.034	0.000	1.000	-1.345	9.492	0.002

	X_{31}						
	Front-Side X_{32}	-0.794	0.000	1.000	-1.084	6.691	0.010
	Front-Back X_{33}	-0.828	0.000	1.000	-1.227	7.606	0.006
	Sideways X_{34}	-0.348	0.000	1.000	-0.931	2.452	0.117
	Streak X_{35}	0.337	0.000	1.000	-1.474	5.099	0.024
	Hit pedestrian X_{36}	-0.823	0.000	1.000	-0.557	1.398	0.237
	Hit animal X_{37}	1.761	0.000	1.000	-1.955	1.768	0.184
Accident location X_4	Go strike X_{40}	2.753	0.000	1.000	2.648	5.500	0.019
	Bent X_{41}	1.056	0.000	1.000	2.490	4.459	0.035
	Intersection X_{42}	2.551	0.000	1.000	3.028	6.874	0.009
Situation X_5	Lonley X_{50}	-3.279	0.000	0.998	-0.693	8.687	0.003
Number of victims X_8	-	-30.976	0.001	0.982	-0.256	5.651	0.017

Based on Table 8, we could formulate two logit functions with predictor variables X_3 , X_4 , X_5 , and X_8 for accident type category with p-value lower than 0.05, and with response variable category $Y = 2$ (severe accident) as reference variable (equation (14)). Logit functions mild accidents were equation (15).

$$\hat{g}_1(\mathbf{x}) = 10.921 + 4.945x_{30} + 0.034x_{31} - 0.794x_{32} - 0.828x_{33} + 0.337x_{35} + 2.753x_{40} + 1.056x_{41} + 2.551x_{42} - 3.279x_{50} - 30.976x_8.$$

$$\hat{g}_2(\mathbf{x}) = 0.037 + 1.791x_{30} - 1.345x_{31} - 1.084x_{32} - 1.227x_{33} - 1.474x_{35} + 2.648x_{40} + 2.490x_{41} + 3.028x_{42} - 0.693x_{50} - 0.256x_8.$$

Referring to equation (8), the final models of multinomial logistic regression with predictor variable significantly influencing response variable were equation (16), (17) and (18).

$$\hat{\pi}_0(\mathbf{x}) = \frac{\exp(\hat{g}_1(\mathbf{x}))}{1 + \exp(\hat{g}_1(\mathbf{x})) + \exp(\hat{g}_2(\mathbf{x}))}$$

$$\hat{\pi}_1(\mathbf{x}) = \frac{\exp(\hat{g}_2(\mathbf{x}))}{1 + \exp(\hat{g}_1(\mathbf{x})) + \exp(\hat{g}_2(\mathbf{x}))}$$

$$\hat{\pi}_2(\mathbf{x}) = 1 - \hat{\pi}_0(\mathbf{x}) - \hat{\pi}_1(\mathbf{x}).$$

where $\hat{\pi}_0(\mathbf{x})$ being the probability of mild traffic accident, $\hat{\pi}_1(\mathbf{x})$ the probability of moderate traffic accident, and $\hat{\pi}_2(\mathbf{x})$ the probability of severe traffic accident.

Parameter significance test for final model concluded that there was at least one predictor variable influencing response variable, since p-value in Table 6 is lower than 0.05. Meanwhile, based on the output of partial early model significance test in Table 8, in moderate accident, there were four predictor variables influencing response variable, including X_3 : accident type, X_4 accident location, X_5 : situation, and X_8 : number of victims. Therefore, we could match the model for equations (16), (17), and (18).

Table 9. Output of Final Model Significance Test by Simultaneous

Model	Log Likelihood	Likelihood Ratio Test		
		Value of Statistic Test	Degree of Freedom	p-value
No Variable	299.288			
With Variable	192.785	106.503	26	0.000

Model compatibility test used Pearson chi-square test on equation (12). Based on the output of model compatibility test in Table 7, p-value = 1.000 higher than the significance level 0.05 thus H_0 was not rejected. We may conclude that the model conforms to the data. Therefore, this research's final models are equations (16), (17), and (18).

Table 10. Output of Fit Test Model

	Value of Statistic Test	Degree of Freedom	p-value
Pearson	83.941	164	1.000

Model Parameter Interpretation using Odds Ratio

The final measure was model parameter interpretation using odds ratio to examine the meaning of estimated parameter value with predictor variable. Odds ratio was the measure of association predicting the probability of predictor variables against the response variable (Hosmer et al., 2013:52).

Table 11. Odds Ratio Value

Response Variable	Predictor Variable		Odds Ratio
Moderate accident	Accident type	Single	5.993
		Front-Front	0.260
		Front-Side	0.338
		Front-Back	0.293
		Streak	0.229
	Accident location	Go strike	14.124
		Bend	12.066
		Intersection	20.652
	Situation	Lonely	0.500
	Number of victims	-	0.774

Below is one interpretation of odds ratio based on Table 8. Single accident had the tendency of 5.993 times moderate traffic accident compared to hit-and-run accident.

Analysis on Research Results Based on Previous Researches

The research by (Hidayati 2017) concluded that there were different factors influencing the occurrence of crashes between non-interstate and interstates. The age, road condition, driver's restraint and negotiating a curve factors influenced traffic barrier crashes on non-interstate highways. Meanwhile driver's condition, type of barriers, speed limit compliance and citation record influenced interstate barrier crashes.

Based on this research result, the four factors only influenced moderate accident (Table 5). Meanwhile, for mild accident there was no factor with significant influence although the modeling was included into the four significant factors on moderate accident for model for mild accident. Furthermore, this research result could only prove that for each accident category, the

influencing factors were the same. Meanwhile, the research by (Findasari, 2023) successfully proved different factors for two different categories of crashes: non-interstate and interstates.

By the results of research by (Apriyani, 2021), we may summarize that the mode choice of transport in Johor Bahru City was statistically correlated with users' age, income, vehicle ownership, cars comfort, reliability of bus service, affective motives and instrumental motives. The results of research by (Apriyani, 2021) may serve suggestions related to administering safe, comfortable and secure transportation. Expectantly, the implementation of results of research by (Apriyani, 2021) may shift the mode (frequency) of accident towards accident of mild category.

This research and the research by (Findasari, 2023) used the same sample data, Indonesia, but different in terms of level of locality. Miranti (2020) used Maluku Province in the Maluku Islands, while this research used Central Java Province in Java Island. The two studies used response variable of nominal scale. The difference was that (Findasari, 2023) focused on victim classification, while this research widened the scope into accident classification. Miranti (2020) found four predictor variables significantly influencing victim classification: sex, age, education and type of vehicle. Meanwhile, this research produced the accident type, accident location, road situation, and number of victim variables as the four predictor variables significantly influencing accident classification.

In their research, (Findasari, 2023) found six factors influencing traffic violation: traffic mark and sign violation, traffic light violation, illegal parking, wrong-way driving, speeding, and not wearing a seat belt. The research by (Findasari, 2023) focused on traffic violation. The six factors influencing the traffic violation are greatly relevant to Indonesia conditions.

Kholisatin and Choiruddin (2021) argued that factors related to the cause of accident were human, vehicle, and environment factors. This research concluded that accident classification was influenced by four factors: accident type, accident location, situation, and number of victims. Therefore, the results of this research are in line with (Findasari, 2023).

CONCLUSION

Based on the analysis results and discussion, we may conclude that factors influencing the response variable were X_3 : accident type, X_4 : accident location, X_5 : situation, and X_8 : number of victims. The multinomial logistic regression model for traffic accident in Cilacap Regency was as follows:

$$\hat{\pi}_0(\mathbf{x}) = \frac{\exp(\hat{g}_1(\mathbf{x}))}{1 + \exp(\hat{g}_1(\mathbf{x})) + \exp(\hat{g}_2(\mathbf{x}))}$$

$$\hat{\pi}_1(\mathbf{x}) = \frac{\exp(\hat{g}_2(\mathbf{x}))}{1 + \exp(\hat{g}_1(\mathbf{x})) + \exp(\hat{g}_2(\mathbf{x}))}$$

$$\hat{\pi}_2(\mathbf{x}) = 1 - \hat{\pi}_0(\mathbf{x}) - \hat{\pi}_1(\mathbf{x}).$$

with the logit function:

$$\begin{aligned}\hat{g}_1(\mathbf{x}) &= 10.921 + 4.945x_{30} + 0.034x_{31} - 0.794x_{32} - 0.828x_{33} + \\ &0.337x_{35} + 2.753x_{40} + 1.056x_{41} + 2.551x_{42} - 3.279x_{50} - \\ &30.976x_8.\end{aligned}$$

$$\begin{aligned}\hat{g}_2(\mathbf{x}) &= 0.037 + 1.791x_{30} - 1.345x_{31} - 1.084x_{32} - 1.227x_{33} - \\ &1.474x_{35} + 2.648x_{40} + 2.490x_{41} + 3.028x_{42} - 0.693x_{50} - \\ &0.256x_8.\end{aligned}$$

ACKNOWLEDGEMENTS

We would like to express our gratitude to LPPM of Universitas Jenderal Soedirman for funding this research and publication of this research's results. This research was conducted as part of Basic Research of Unsoed (RDU) of LPPM Unsoed for funding year 2023, under Contract Number: 27.258/UN23.37/PT.01.03/II/2023

REFERENCE

- Abdillah, A., Sutisna, A., Tarijah, I., Fitria, D., and Widiyanto, T. (2020). Application of Multinomial Logistic Regression to Analyze Learning Difficulties in Statistics Courses. *Journal of Physics: Conference Series*, 1490 (2020) 012012. <https://doi.org/10.1088/1472-6596/1490/1/012012>.
- Ambo, T.B., Ma, J., and Fu, C. (2020). Investigating Influence Factors of Traffic Violation using Multinomial Logit Method. *International Journal of Injury Control and Safety Promotion*, 28(1): 78-85. <https://doi.org/10.1080/17457300.2020.1843499>.
- Apriyani, M., Jajang, and Tripena, A. (2021). Implementasi Model Regresi Logistik Multinomial pada Pengelompokan Penyakit Tuberkulosis [Implementation of a Multinomial Logistic Regression Model in Tuberculosis Disease Groupings]. *Jurnal Matematika dan Pendidikan Matematika*, 13(1), 27-38. <https://doi.org/10.20884/1.jmp.2021.13.1.3612>.
- Findasari and Himayati, A.I.A. (2023). Analisis Regresi Logistik Biner pada Faktor Resiko Kejadian Tuberkulosis [Binary Logistic Regression Analysis on Risk Factors for Tuberculosis]. *Jurnal Matematika, Sains dan Teknologi*, 24(1): 1-14. <https://doi.org/10.33830/jmst.v24i1.4666.2023>.





- Hidayati, A., and Hendrati, L. Y. (2017). Traffic Accident Risk Analysis by Knowledge, the Use of Traffic Lane, and Speed. *Jurnal Berkala Epidemiologi*, 4(2), 275–287. <https://doi.org/10.20473/jbe.V4i22016.275-287>.
- Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression: Third Edition*. New York: John Wiley dan Sons Inc.
- Kholisatin, N., and Choiruddin, A. (2021). Analisis Risiko Kecelakaan Lalu Lintas di Jalan Raya Kabupaten Nganjuk Menggunakan Poisson Point Process on a Linear Network [Traffic Accident Risk Analysis on Nganjuk Highway Using Poisson Point Process on a Linear Network]. *Inferensi*, 4(2), 99-108. <http://dx.doi.org/10.12962/j27213862.v4i2.10924>.

