FACTORS INFLUENCING ROAD ACCIDENTS IN SRI LANKA: A LOGISTIC REGRESSION APPROACH

S.A.T. Dhananjaya[1], M.C. Alibuhtto[2]

[1]Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sri Lanka
[2]Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sri Lanka
satdhananjaya@gmail.commcabuhtto@seu.ac.lk

Abstract

Road accidents are a major socio-economic problem in Sri Lanka. This study aims to identify the factors that mainly contribute to accident severity in Sri Lanka and to identify the significance of the factors for formulating the model. For this purpose, road accident yearly data were collected for the period of 2010 to 2014 from the Police Traffic Headquarters in Colombo and 13 factors were considered. In this study, Binomial Logistic Regression has been used to analyze the data due to the dichotomous nature of the dependent variable.Based on the results, concluded that variables such as light condition, age of the driver, the validity of the license, urban / rural, weather, vehicle type and age of the vehicle have a decreasing effect on the probability of a fatal accident. Similarly, remaining variables such as location type, alcohol test and accident cause have an increasing effect on the probability of a fatal accident. Among them, Accident Cause is the most important variable in the model.

Keywords: Binary Logistic Regression Analysis, Drivers, Pedestrians, Road Traffic Accidents, Sri Lanka, Vehicles

Introduction

Road accidents are highly influenced to the public health in a country. And also increasing road accidents evolve social and economic problems due to loss of lives and damage possessions. Road accidents are really induced by interactions of the vehicles, road users and roadway conditions. Each of these basic components contains a number of sub components which are contributed to increase the risk of the road accidents like pavement characteristics, road characteristics, geometric features, traffic characteristics, design of vehicles, driver's characteristics, road user's behavior and environmental features. (Pakgohar et al., 2015)

Increase in number of vehicles in a country also generate immensely severe problems of road accidents. Specially, due to these retributions of road accidents such as injuries, impairments and fatalities are caused serious problems in developing countries. World Health Organization (WHO) has found more than 1.2 million people die each year on the world's roads and causing road traffic injuries a leading circumstance of death globally. And they found most of these deaths are in low and middle income countries. Sure enough, WHO indicated road traffic injuries are currently estimated to be the 9th leading cause of death across all age groups globally and predicted to become the 7th leading cause of death by 2030.Further, they show, this rise is driven by the low and middle income countries. (Toroyan et al., 2015) Sri Lanka is a developing country. Therefore still we have not good road system over the whole country. Because the road developing projects are going on many areas in the country. Therefore, this situation is affected to make huge problems in road accidents. Because the undeveloped road system leads to make many problems in the field.

Somasundaraswaran (2006) analyzed accident statistics of Sri Lanka during 1989-2005. The results of this study revealed that the main reason for the rapid increase of traffic accidents is due to the alarming rate of vehicle ownership together with inadequate road network development to support the demand.Renuraj et al. (2015) conducted a research on Factors Influencing Traffic Accidents in Jaffna. In this study, they used 692 accident cases for the analysis based on Jaffna police records during the period 2010-2013. They have used logistic regression approach for the analysis. Results from this study reveal that the fitted logistic regressionmodel can be used for the safety improvements against the traffic accidents in Jaffna. The conclusion of this research expressed that independent variables “Type of vehicle” and “Age” were identified as more influential variables influencing the accident severity.

Haadi, A. R (2012) conducted a case study on identification of factors that cause severity of road accidents in Ghana: Northern Region. The objective of the study was to identify the variables that mainly contribute to accident severity in the Northern Region and to describe the impact of these variables. In this study, the binary logistic regression has applied to a total of 398 accident data from 2007-2009 collected from motor transport and traffic unit (MTTU) Northern Region traffic-police records. The results of this study revealed that among the 398 records, 3.1% involved minor injuries. The rest were considered as accidents with fatal injuries (98.7%). The conclusion of the research expressed that most significantly associated with accident severity was overloading and obstruction.

Yordphol et al. (2005) conducted a research on traffic accidents in Thailand. Road traffic accidents in Thailand, 1983–2002 were used for this study. They found in this research, a higher number of vehicles, particularly motorcycles can be anticipated throughout the country which will result in more road casualties and tremendous economic losses, especially the extra health care costs for the accident victims and therefore, remains a challenging issue to all concerned parties to address this significant social problem and concurrently, to implement all the necessary measures promptly to fight this long and seemingly endless battle.

Al-Ghamdi (2002) conducted a research on using logistic regression to estimate the influence of accident factors on accident severity. Logistic regression was applied to accident-related data collected from traffic police records in order to examine the contribution of several variables to accident severity in Riyadh. The data set used in this study was derived from a sample of 560 subjects involved in serious accidents reported in traffic police records in Riyadh, the capital of Saudi Arabia. Only accidents occurring on urban roads in Riyadh were examined. The conclusion of this study expressed that logistic regression as used in this research is a promising tool in providing meaningful interpretations that can be used for future safety improvements in Riyadh.

Statement of the research problem

Traffic accidents are a serious public health problem and one of the leading causes of the death and injuries around the globe with ever rising trend. The magnitude of the problem of road traffic injuries in Sri Lanka significantly increased in the last decade.

So, this study statistically explored the significant factors influencing road accidents that are occurring in Sri Lankaand among them, break through the most influential factors on road accidents and attempt to fill the gaps by proposing solutions to the problem.

Research objectives

The objective of this study is to identify the factors that mainly contribute to accident severity in Sri Lanka and identify the significance of the factors for formulating the model.

Materials and Methods

This study is conducted to identify and analyze the factors influencing in fatal and non-fatal road accidents in Sri Lanka in between 2010 to 2014 time period. In this study, mainly focus on 13 factors influencing in road accidents such as road surface, light condition, location type, age of the driver, validity of license, alcohol test, accident cause, urban/rural, workday/holiday, weather, vehicle type, vehicle ownership and age of vehicle.

Data collection

Secondary data used in this research were acquired from the road accidents database obtained from the police traffic headquarters, Colombo in Sri Lanka. Data is received from 2010 to 2014 time period and those data has collected by the police officers. They have reported the related data according to the questionnaire which was prepared by traffic headquarters. Initial data were made over as MS Access database.

Data preparation

The initial database had 193,907 accidents. Initially, detected important 17 factors influencing in road accidents. In addition to the factors mentioned in above 2.1, pedestrian location, road pre-crash factor, vehicle pre-crash factor and accident type were drawn as influencing factors. But found some issues exists in this database. Basically, performed descriptive statistics and graphical analysis roughly. Then it leads to ascertain these 4 factors recorded more data (more than 100,000) under not known/not applicable level. So turned out to remove those factors from the analysis. Besides, some other factors also recorded as same as above.

Therefore, filtered those accident data from the database.Similarly, some factor had several factor levels, but some levels having few accident records. So such bulk of levels extracted as 'others' to vest significant percentage of accidents. In addition to that, alcohol level factor reported higher no of accidents under not tested level.

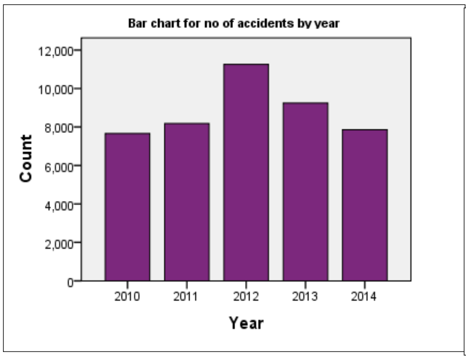
Because of these types of circumstances are not inappropriate for statistical analysis, filtered those data from the database with the help of the functions in MS excel and SPSS. Finally, prepared a database having 44,197 accidents and used it for further analysis.

Data Analysis

Data analyses in this study arrayed mainly under preliminary and fundamental analyses. In preliminary analysis included univariate analysis and bivariate analysis. Univariate analysis is performed to get a general understanding of the whole dataset and bivariate analyses is functioned to examine the relationships between the variables. Finally, due to the dichotomous nature of the dependent variable, carried out a binary logistic regression analysis as fundamental analysis to investigate the combined effect of the variables. These statistical data analysis was conducted by using MS Excel and SPSS software.

Results and Discussion

Univariate analysis



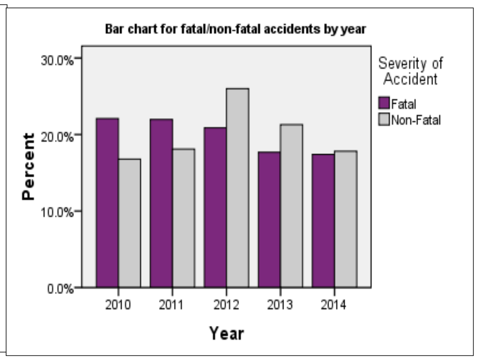


Figure 1: Bar chart for no of accidents by year

Figure 2: Bar chart fatal/non-fatal accidents by year

According to the above Figure 1 and Figure 2 while comparing the accident data for the years since 2010 to 2014, the majority of accidents occurred in 2012. Similarly highest no of non-fatal accidents occurred in 2012 and highest no of fatal accidents occurred in 2010.

Bivariate analysis

In this section, performed bivariate analysis by graphical analysis and Pearson chi-Square contingency table analysis. Among them, for only 2-level variables such as the validity of license, alcohol test and urban/rural described about the odds ratios.

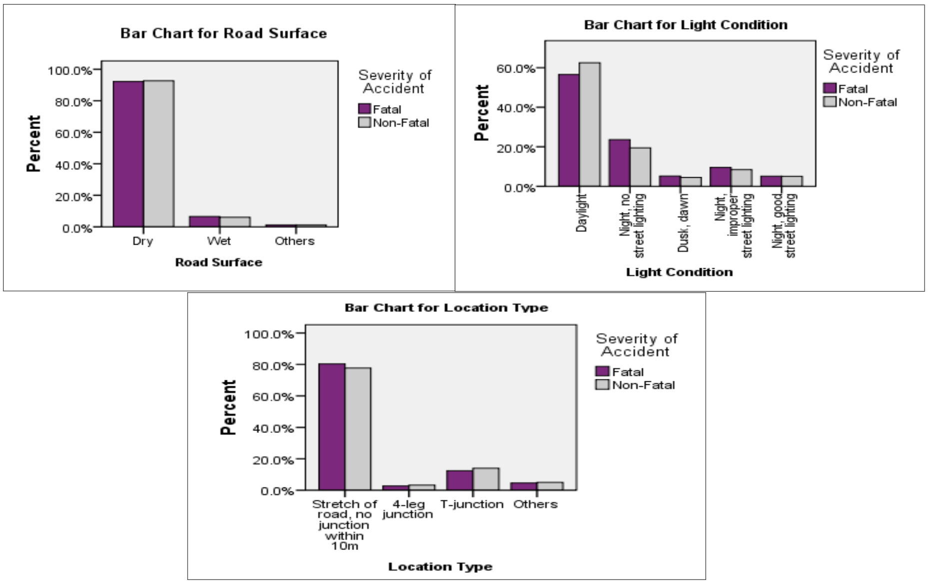


Figure 3: Bar charts for road characteristics

Besides, the results are discussed under 13 factors categorized into 4 characteristics such as road characteristics, human and accident characteristics, time and environmental characteristics, and vehicle characteristics with including percentages.

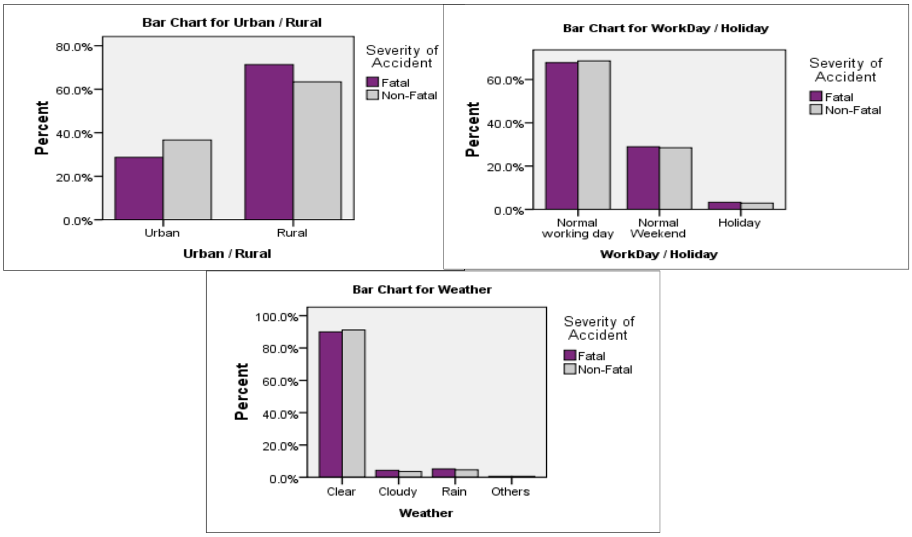


Figure 4: Bar charts for time and environmental characteristics

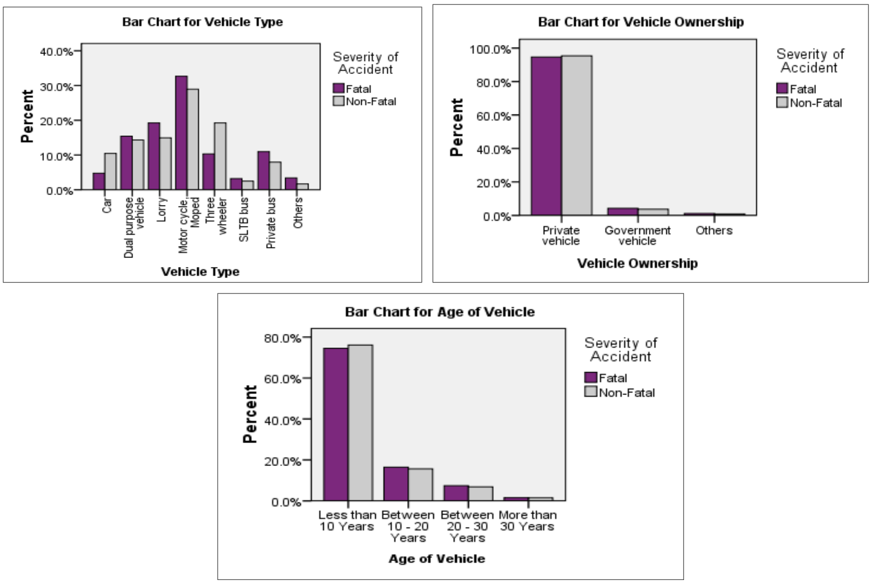


Figure 5: Bar charts for vehicle characteristics

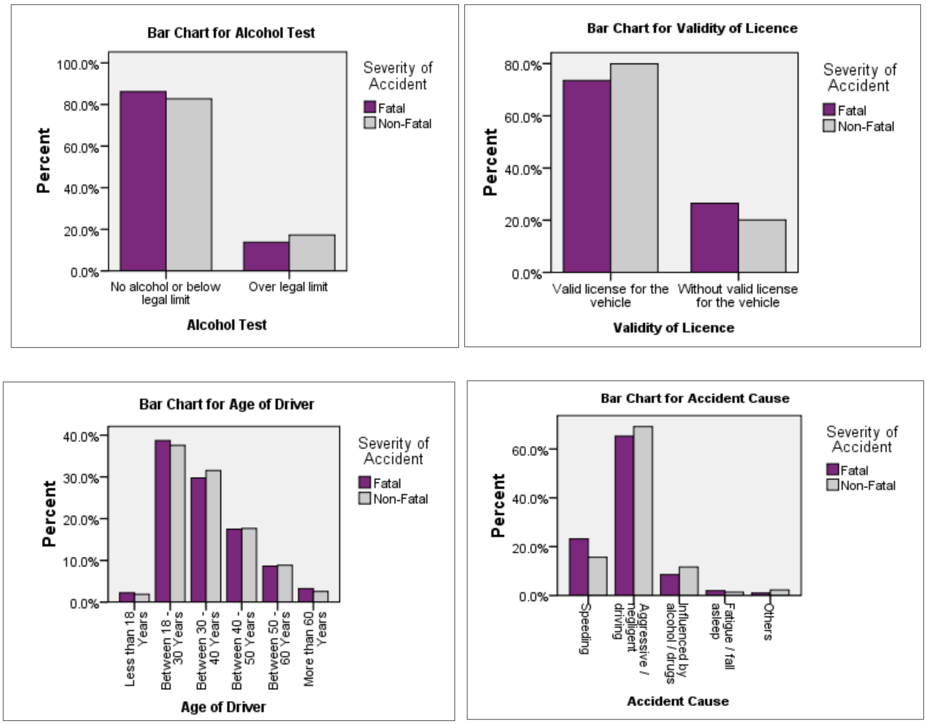


Figure 6: Bar charts for human and accident characteristics

Table 1: Frequencies of fatal and non-fatal accidents by road characteristics

Variable	Levels of Factor	Severity of Accident			
		Fatal		Nonfatal	
		Count	Percentage	Count	Percentage
Road Surface	Dry	4293	92.2	36634	92.7
	Wet	307	6.6	2406	6.1
	Others	57	1.2	500	1.3
Light Condition	Daylight	2632	56.5	24708	62.5
	Night, no street lighting	1099	23.6	7694	19.5
	Dusk, dawn	241	5.2	1776	4.5
	Night, improper street lighting	446	9.6	3359	8.5
	Night, good street lighting	239	5.1	2003	5.1
Location Type	Stretch of road, no junction within 10m	3739	80.3	30726	77.7
	4-leg junction	127	2.7	1306	3.3
	T-junction	578	12.4	5525	14.0

	Others	213	4.6	1983	5

Table 2: Frequencies of fatal and non-fatal accidents of time and environmental characteristics

Variable	Levels of Factor	Severity of Accident			
		Fatal		Nonfatal	
		Count	%	Count	%
Urban / Rural	Urban	1336	28.7	14490	36.6
	Rural	3321	71.3	25050	63.4
Work Day / Holiday	Normal working day	3158	67.8	27135	68.6
	Normal Weekend	1347	28.9	11270	28.5
	Holiday	152	3.3	1135	2.9
Weather	Clear	4189	90.0	36048	91.2
	Cloudy	201	4.3	1430	3.6
	Rain	241	5.2	1841	4.7
	Others	26	0.6	221	0.6

Table 3: Frequencies of fatal and non-fatal accidents by vehicle characteristics

Variable	Levels of Factor	Severity of Accident			
		Fatal		Nonfatal	
		Count	Percentage	Count	Percentage
Vehicle Type	Car	222	4.8	4136	10.5
	Dual purpose vehicle	718	15.4	5670	14.3
	Lorry	896	19.2	5906	14.9
	Motorcycle, Moped	1522	32.7	11444	28.9
	Three wheeler	480	10.3	7598	19.2
	SLTB bus	149	3.2	986	2.5
	Private bus	511	11.0	3141	7.9
	Others	159	3.4	659	1.7
Vehicle Ownership	Private vehicle	4406	94.6	37714	95.4
	Government vehicle	198	4.3	1476	3.7
	Others	53	1.1	350	0.9
Age of Vehicle	Less than 10 Years	3471	74.5	30086	76.1
	Between 10 - 20 Years	767	16.5	6175	15.6
	Between 20 - 30 Years	346	7.4	2697	6.8
	More than 30 Years	73	1.6	582	1.5

Moreover, if quantified the result by the odds ratio, the odds of an accident occurred by no alcohol are 1.303 times more likely to be odds of an accident occurred by alcohol. Similarly, if quantify this result of the relative risk, a fatal accident occurred by no alcohol is 1.27 times more likely to be a fatal accident occurred by alcohol and a non-fatal accident occurred by no alcohol is 0.975 times less likely to be a non-fatal accident occurred by alcohol.

Table 4: Frequencies of fatal and non-fatal accidents by human and accident characteristics

Variable	Levels of Factor	Severity of Accident			
		Fatal		Nonfatal	
		Count	%	Count	%
Alcohol Test	No alcohol or below legal limit	4014	86.2	32710	82.7
	Over legal limit	643	13.8	6830	17.3
Validity of License	Valid license for the vehicle	3424	73.5	31600	79.9
	Without valid license for the vehicle	1233	26.5	7940	20.1
Age of Driver	Less than 18 Years	103	2.2	742	1.9
	Between 18 – 30 Years	1803	38.7	14852	37.6
	Between 30 - 40 Years	1386	29.8	12471	31.5
	Between 40 - 50 Years	813	17.5	6978	17.6
	Between 50 - 60 Years	402	8.6	3491	8.8
	More than 60 Years	150	3.2	1006	2.5
Accident Cause	Speeding	1079	23.2	6182	15.6
	Aggressive / negligent driving	3041	65.3	27343	69.2
	Influenced by alcohol / drugs	397	8.5	4604	11.6
	Fatigue / fall asleep	92	2.0	528	1.3
	Others	48	1	883	2.2

Table 5: Odds ratio for alcohol test

Odds Ratio		
Alcohol Test	For cohort Severity of Accident = Fatal	For cohort Severity of Accident = nonfatal
1.303	1.270	0.975

Table 6: Odds ratio for validity of license

Odds Ratio		
Validity of License	For cohort Severity of Accident = Fatal	For cohort Severity of Accident = nonfatal
0.698	0.727	1.042

Similarly, according to the results of the odds ratio, the odds of an accident occurred by drivers who, having the valid license for the vehicle are 0.698 times less likely to be odds of an accident occurred by drivers who without a valid license for the vehicle.

Similarly, if quantify this result of the relative risk, a fatal accident occurred by drivers who, having the valid license for the vehicle is 0.727 times less likely to be a fatal accident occurred by alcohol and a non-fatal accident occurred by drivers who, having the valid license for the vehicle is 1.042 times more likely to be a non-fatal accident occurred by drivers who without a valid license for the vehicle.

Table 7: Odds ratio for urban / rural

Odds Ratio		
Urban / Rural	For cohort Severity of Accident = Fatal	For cohort Severity of Accident = nonfatal
0.695	0.721	1.037

Similarly, according to the result by the odds ratio, the odds of an accident occurred in an urban area is 0.695 times less likely to be odds of an accident occurred in a rural area. Similarly, if quantify this result of the relative risk, a fatal accident occurred in an urban area is 1.037 times more likely to be a fatal accident occurred in a rural area and a non-fatal accident occurred in an urban area is 0.695 times less likely to be a non-fatal accident occurred in a rural area.

Pearson Chi-Square contingency table nalysis

Pearson chi-square contingency table analysis, which performed to check whether exist or not a significant relationship between the independent variables and a dependent variable. Similarly, therewith Phi & Cramer's V also discussed for all variables. According to this analysis, following table describes the association between the each factor and the severity.

Table 8: Chi-square test results for association between each contributory factor and the severity

Analyzed Variables	Pearson' s χ^2(df), Phi & Cramer's V, p	Significance
Road characteristics		
Road surface	χ^2 (2) = 1.899, crv (4) = 0.007, p= 0.387	Not Significance
Light condition	χ^2 (4) = 69.874, crv (4) = 0.04, p= 0.000	Significance
Location type	χ^2 (3) = 16.809, crv (3) = 0.02, p= 0.001	Significance
Human and accident characteristics		
Alcohol Test	χ^2 (1) = 35.634, crv (1) = 0.028, p= 0.000	Significance
Validity of license	χ^2 (1) = 0.01036, crv (1) = 0.048, p= 0.000	Significance
Age of driver	χ^2 (5) = 15.687, crv (5) = 0.019, p= 0.008	Significance
Accident cause	χ^2 (4) = 0.02296, crv (4) = 0.072, p= 0.000	Significance
Time and environmental characteristics		
Urban / Rural	χ^2 (1) = 0.01148, crv (1) = 0.051, p= 0.000	Significance
Workday/holiday	χ^2 (2) = 2.877, crv (2) = 0.008, p= 0.237	Not Significance
Weather	χ^2 (3) = 8.585, crv (3) = 0.014, p= 0.035	Significance
Vehicle characteristics		
Vehicle type	χ^2 (7) = 0.05143, crv (7) = 0.108, p= 0.000	Significance
Vehicle ownership	χ^2 (2) = 6.142, crv (2) = 0.012, p= 0.046	Significance

Age of vehicle	χ^2 (3) = 5.758, crv (3) = 0.013, p= 0.124	Not Significance

According to the results of Table 8, we can denote light condition, location type, alcohol test, validity of license, age of driver, accident cause, urban / rural, weather, vehicle type and vehicle ownership are statistically significantly associated with the severity. Only three variables such as road surface, workday/holiday and age of vehicle are statistically not significantly associated with the severity.
As well as according to the Cramer's V values it appears that the association between severity and the variables such as road surface, workday/holiday, vehicle ownership and age of the vehicle were shown to be in the strongest weak association type.

The remaining variables were shown fairly weak association with the severity according to their Cramer's V values.

Binary logistic regression analysis
According to the methodology, main dataset (44,197 accidents) was divided into two portions; 60% (26,540 accidents) was used to develop the model, and the remaining 40% (17,657) was used to validate the model. (Rana et.al, 2010).

Baseline model
The baseline model exists a predictive power of 11.3%, which shows the overall percentage of correctly classified cases when there are no explanatory variables in the model. The log likelihood value of the base model is 18698.721. This value is used to select a best model.

Table 9: Variables in the baseline model

	B	S.E.	Wald	Df	Sig.	Exp (B)
Constant	2.063	0.019	1.130E4	1	0.000	7.867

Initial -2 Log Likelihood: 18698.721

Table 9 shows the coefficient for the constant of the baseline model. According to this table, it can be annotate the model with just the constant is a statistically significant predictor of the outcome ($p < 0.05$).

Developed model
For developing the binary logistic model, used the "Backward Elimination (Likelihood Ratio)" method. Allvariables were entered into the analysis and by extracting insignificant ones, model iteration occurred up to four steps. The analysis was performed on P value = 0.05 significance level to formulate the model.

Table 10: Omnibus tests of developed model coefficients

	Chi-square	**df**	**Sig.**
Step 4	-2.517	1	0.113
Block	413.886	10	0.000
Model	413.886	10	0.000

Table 10, indicates the chi-square values for the block and the model are highly significant (chi-square=413.886, $p<0.05$). Therefore, the developed model is significantly better than the baseline model.

That means the accuracy of the model improved when added the explanatory variables. In this case, added all 10 explanatory variables in one block and therefore the chi-square values are same for the block and model. But the Step 4 shows insignificant. It is obtained due to the model iteration stopped on this fourth step.

Variance explanation

Table 11: Developed model summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
18284.835	0.015	0.031

The results of the Table 11 is used to check that the developed model which is with the explanatory variables is an improvement over the baseline model. As the log likelihood results of this table, describes that there is a significant difference between the log likelihoods (specifically the -2LLs) of the baseline model and the developed model.

According to the Table 11, the developed model has a significantly reduced log likelihood value (18284.835) compared to the baseline model. Then it's revealed that the developed model is explaining more of the variance in the outcome and it is an improvement over the baseline model. Thus, it can be concluded that the developed model is better at predicting the severity of the accidents than the baseline model where no predictor variables were added.

In addition to that, Table 11 contains the Cox & Snell R Square and Nagelkerke R Square values, which are used to calculate the explained variation. These values are sometimes referred to as pseudo R^2 values. According to these both values, the explained variation in the dependent variable based on the model ranges from 1.5% to 3.1% respectively.

Hosmer and lemeshow test

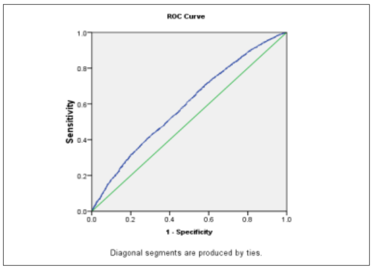
Table 12: Hosmer and Lemeshow test results

Chi-square	df	Sig.
13.198	8	0.105

In binary logistic regression analysis, Hosmer and Lemeshow test use to represent that data fit the model satisfactorily. As the results shown in the Table 12, Hosmer & Lemeshow test of the goodness of fit suggests the model is a good fit to the data as $p=0.105$ (>0.05).

ROC curve

Table 13: Area under the curve



Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.592	0.004	0.000	0.583	0.600

Figure 7: ROC curve

This curve is called the receiver operating characteristic (ROC) curve. According to the above Table 13 and Figure 7, the area under the curve is 0.592 with 95% confidence interval (0.583, 0.600). Also, the area under the curve is significantly different from 0.05 since the p - value is 0.000. That means, the logistic regression classifies the group significantly better than by chance.

Category prediction

Table 14: Developed model classification table

Observed		Predicted					
		Model Development Set			Validation Set		
		SOA		% Correct	SOA		% Correct
		F	NF		F	NF	
SOA	F	1705	1288	57.0	806	858	48.4
	NF	10072	13475	57.2	6656	9337	58.4
Overall Percentage				57.2			57.4

When closely observed the process of this analysis, the model at the fourth step was the best of all for predicting the severity of accidents. That's the prediction power is estimated at 57.2%, which is greater than to the predictive power of the baseline model. (See the Table 14).

Similarly, it was found that the same model correctly predicted 57.4% of the validation data. That means the developed model more accurately predicts the severity of accidents than the prediction in baseline model.

Developed model interpretation

Table 15: Variables not in the developed model

Variable	Score	df	Sig.
Road Surface	0.716	1	0.398
Workday/Holiday	0.467	1	0.495
Vehicle Ownership	2.613	1	0.106
Overall Statistics	3.880	3	0.275

According to the above Table 15, describes that road surface, workday/holiday and vehicle ownership variables are removed due to statistically not significantlyassociated with the severity of accidents. (P values of 0.398, 0.495, 0.106 >0.05 respectively).

Table 16 explains the variables in the developed model used to predict the severity of accidents. When exploring results of this table, light condition, location type, age of the driver, validity of license, alcohol test, accident cause, urban/ rural, weather and vehicle type have a significant effect on the severity of accidents.

By observing the B coefficients, it reveals that variables such as light condition, age of the driver, validity of the license, urban / rural, weather, vehicle type and age of the vehicle have a decreasing effect on the probability of a fatal accident. Then the rest of variables such as location type, alcohol test and accident cause have an increasing effect on the probability of a fatal accident.

Table 16: Variables in the developed model

Variable	B	S.E.	Wald	df	Sig.	Exp (B)	95% C.I. For Exp (B)	
							Lower	Upper
Light Condition	-0.081	0.016	24.119	1	0.000	0.922	0.893	0.953
Location Type	0.054	0.022	5.982	1	0.014	1.055	1.011	1.102
Age of Driver	-0.037	0.018	4.175	1	0.041	0.963	0.929	0.998
Validity of License	-0.442	0.046	91.673	1	0.000	0.643	0.587	0.704
Alcohol Test	0.202	0.061	11.049	1	0.001	1.224	1.086	1.379
Accident Cause	0.304	0.032	90.783	1	0.000	1.356	1.274	1.443
Urban / Rural	-0.400	0.044	84.489	1	0.000	0.670	0.615	0.730
Weather	-0.077	0.037	4.472	1	0.034	0.926	0.862	0.994
Vehicle Type	-0.085	0.012	53.601	1	0.000	0.919	0.898	0.940
Age of Vehicle	-0.101	0.029	12.191	1	0.000	0.904	0.854	0.957
Constant	3.193	0.156	417.392	1	0.000	24.371		

Table 17: Relative importance of variables in the developed model

Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Light Condition	-9154.119	23.402	1	0.000

Location Type	-9145.475	6.115	1	0.013
Age of Driver	-9144.490	4.144	1	0.042
Validity of License	-9186.406	87.977	1	0.000
Alcohol Test	-9148.103	11.370	1	0.001
Accident Cause	-9191.977	99.118	1	0.000
Urban / Rural	-9186.314	87.793	1	0.000
Weather	-9144.586	4.336	1	0.037
Vehicle Type	-9168.876	52.917	1	0.000
Age of Vehicle	-9148.335	11.834	1	0.001

Table 17 presents the information how the model is affected if an explanatory variable is removed from the model. In other words, which variable is important for the model? In answering this problem, used above results to examine the importance of a variable in the model.

So, according to that results, the removal of Accident Cause from the model makes the biggest change in the model's log likelihood value. Therefore, Accident Cause is the most important variable in this model. It is followed by the validity of the license, urban / rural, vehicle type, light condition, age of vehicle, alcohol test, location type, weather and age of driver respectively.

Conclusion and Recommendation

The study presented in this thesis was conducted to identify and analyze the factors associated with the fatal and non-fatal road accidents in Sri Lanka in between 2010 to 2014. The aim of this study is to provide some realization findings of traffic accidents in Sri Lanka. These research results are categorized mainly as preliminary and fundamental analyses. Under preliminary analysis, univariate analysis was performed to collaborate frequency distributions of the factors. Then graphical analysis and Pearson chi-square contingency table analyses were conducted simultaneously to establish associations between the severity of road accidents and indicated factors to determine the significance.

Finally, binary logistic regression was carried out to predict future outcomes in terms of significant influencing factors. The conclusions achieve from this research are summarized as below.

In the section of univariate analysis, the results concluded many information. According to the descriptive statistics, it revealed that, there is no big variation in the accident counts. But majority of road accidents occurred in 2012. Similarly, among them, highest number of fatal accidents occurred in 2010 and highest no of non-fatal accidents occurred in 2012.

Besides, the most of accidents were occurring in dry road surface with clear weather condition.Motorcycles are found to have a higher probability of causing traffic accidents in Sri Lanka. In addition to that, most of accidents recorded by newly registered vehicles (age is less than 10 years). According to the statistic reports by the Ministry of transport & civil aviation, it indicated that there is a continuous rapid increase in new vehicle registrations in each year. Therefore, this fact caused to increase the vehicle population in Sri Lanka and it has mainly affected to increase the traffic accidents.

One of important exposure in descriptive statistics is the high number of traffic accidents reported due to aggressive / negligent driving. So this is a great teaser of drivers in Sri Lanka. The next leading cause is speeding. Similarly, highest number of traffic accidents reported by the drivers in between 18 – 30 years old. It is convinced that young drivers are most influenced in traffic accidents. It may be due to most of young drivers have not satisfactory experiencing in driving and lack of relevant knowledge. Therefore, since these facts, they drive often coolly and involve increasing the traffic accidents. Furthermore, other consequential revelations are a large number of accidents occurred in rural area and by private vehicles.

Then, according to the Pearson chi-square contingency table analysis, it was concluded that the severity of the accident is in statistically significant association with the factors such as light condition, location type, alcohol test, validity of license, age of driver, accident cause, urban / rural, weather, vehicle type and vehicle ownership. Therefore, only three variables such as road surface, workday/holiday and age of vehicle are statistically not significantly associated with the severity of the accident.

Based on the binary logistic regression analysis results road surface, workday/holiday and vehicle ownership variables are statistically not significantly associated with the severity of accidents. So remaining factors such as light condition, location type, age of the driver, validity of license, alcohol test, accident cause, urban / rural, weather and vehicle type have found a significant effect on the severity of accidents.

Besides, according to B coefficients, it concludes that variables such as light condition, age of the driver, validity of the license, urban / rural, weather, vehicle type and age of the vehicle have a decreasing effect on the probability of a fatal accident. Then the rest of variables such as location type, alcohol test and accident cause have an increasing effect on the probability of a fatal accident. That means, according to the research outcome, these three variables are the most influential factors in this study. Among them, finally concluded that Accident Cause is the most important variable in the model.

According to the inferential statistical test results under binary logistic analysis of the traffic police accident data, the majority of the contributing factors for the occurrence of road traffic accidents in Sri Lanka is mainly due to the factor of “Accident Cause”. It is described by the drivers’ faults such as speeding, aggressive/negligent driving, influenced by alcohol/drugs, fatigue/falls asleep. This is an issue which needs high level attention from drivers and high commitment by traffic police.So, not only the government of Sri Lanka, but also the drivers is reflected a great responsibility to reduce road accidents, and control this ambience.

For achieving this objective of reduce traffic accidents, government should make programs for educating all stakeholders, especially drivers and pedestrians as a whole about road safety using media (TVs, Radios, Newspapers, magazines, etc.) or arrange those in formal organizations in schools and other governmental and non-governmental organizations or in religious institutions.In addition to that, setting a national road safety policies, laws and regulations, raising road traffic fines, renewingdriving license frequently and driver's health, driving skills must be rechecked during the renewal of the license and consecration to oblige the laws and regulations and performing infrastructural strategies for traffic police to facilitate their duties.

The database which is used for the analysis in this study had 4 levels of severity of accidents such as fatal, grievous, non-grievous, damage only etc. For this study, extracted last 3 levels as non-fatal. So binary logistic regression is used for this study due to the dichotomous nature of the dependent variable. But, recommend, it is better to perform a research on the same field by considering above all four levels using Multinomial Logistic Regression Analysis.

Besides, recommend to perform a research by considering the whole database into distinct Models such as "Factors influencing by pedestrians", "Factors influencing by drivers", "Factors influencing by vehicles" and so on.

Acknowledgements

The authors would like to express gratitude to the Senior DIG Traffic Administration and Road safety for his permission to supply the data for this study. And all the police officers who support to get the data that needed for this research are also greatly appreciated.

References

1. Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention,34*(6), 729-741.

2. Bicalho, M. A. H., Sukys-Claudino, L., Guarnieri, R., Lin, K., & Walz, R. (2012). Socio-demographic and clinical characteristics of Brazilian patients with epilepsy who drive and their association with traffic accidents. *Epilepsy & Behavior, 24*(2), 216-220.

3. Dissanayake, S., & Roy, U. (2014). Crash severity analysis of single vehicle run-off-road crashes. *Journal of Transportation Technologies*, *4*(01), 1.

4. Dozza, M., Flannagan, C. A., & Sayer, J. R. (2015). Real-world effects of using a phone while driving on lateral and longitudinal control of vehicles.*Journal of safety research*, *55*, 81-87.

5. Ernstberger, A., Joeris, A., Daigl, M., Kiss, M., Angerpointner, K., Nerlich, M., & Schmucker, U. (2015). Decrease of morbidity in road traffic accidents in a high income country–an analysis of 24,405 accidents in a 21 year period. *Injury*, *46*, S135-S143.

6. Haadi, A. R. (2012). *Identification of Factors that Cause Severity of Road Accidents in Ghana: A Case Study of the Northern Region* (Doctoral dissertation).

7. Kavade, H. (2009). A logistic regression model to predict incident severity using the human factors analysis and classification system.

8. Kazan, E. E. (2013). Analysis Of Fatal And Nonfatal Accidents Involving Earthmoving Equipment Operators And On-Foot Workers.

9. Khalili, M., & Pakgohar, A. (2013). Logistic regression approach in road defects impact on accident severity. *Journal of Emerging Technologies in Web Intelligence*, *5*(2), 132-135.

10.Pakgohar, A., & Kazemi, M. (2015). An examination of accident severity differences between male and female drivers, Using Logistic Regression Model. *Civil engineering journal*, *1*(1), 31-36.

11. Renuraj, S., Varathan, N., & Satkunananthan, N. (2015). Factors Influencing Traffic Accidents in Jaffna. *Sri Lankan Journal of Applied Statistics*, *16*(2).
12. Somasundaraswaran, A. K. (2006). accident statistics in Sri Lanka. *IATSS research*, *30*(1), 115-117.

13. Toroyan, T., Peden, M. M., & Iaych, K. (2013). WHO launches second global status report on road safety. *Injury prevention*, *19*(2), 150-150.

14. Tanaboriboon, Y., & Satiennam, T. (2005). Traffic accidents in Thailand. *IATSS research*, *29*(1), 88-100.