

**FACTORS AFFECTING THE SEVERITY OF ROAD
ACCIDENTS IN SRI LANKA: A LOGISTIC
REGRESSION APPROACH**

Navarathna Arrachchilage Manoja Rasanjali Senaviratna

158384M

Degree of Master of Science in Operations Research

Department of Mathematics

University of Moratuwa
Sri Lanka

November 2018

**FACTORS AFFECTING THE SEVERITY OF ROAD
ACCIDENTS IN SRI LANKA: A LOGISTIC
REGRESSION APPROACH**

Navarathna Arrachchilage Manoja Rasanjali Senaviratna

158384M

Thesis submitted in partial fulfillment of the requirements for the degree Master of
Science

Department of Mathematics

University of Moratuwa
Sri Lanka

November 2018

Declaration of the candidate and the supervisor

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 09.11.2018

The above candidate has carried out research for the Masters thesis under my supervision.

Signature of the supervisor:

Date:

Mr.T.M.J.A.Cooray

Senior Lecturer

Department of Mathematics

University of Moratuwa

Abstract

Road accidents have become a leading cause of death and injury as well as property damage worldwide. Ever increasing road accidents and traffic flow is a heavy burden to a developing country like Sri Lanka. In year 2016, 38915 accidents were reported where 7% of them are fatal contributing to 2824 deaths. Therefore, it is urgently needed to find solutions and reduce road accident deaths and injuries. The objective of this study is to identify the significant factors affecting for motorcycle and motor vehicle accidents in Sri Lanka. Secondary data used in this study between the period 2014 to 2016 were acquired from the police traffic headquarters, Colombo in Sri Lanka. A total number of 111457 road accidents where drivers at fault were included in the analysis. Among them 78531 were motor vehicle accidents and 32926 were motor cycle accidents. Motorcycle accidents are analyzed separately due to high accident rate of motorcycles.

Factors considered in the study were vehicle type, gender of driver, validity of license, accident cause, alcohol test, time of accident, weekday/weekend, road surface, weather condition, light condition, location and age of driver. Results revealed that male drivers (98%) have greater tendency to be involved in motorcycle and motor vehicle accidents rather than female drivers (2%). High number of motorcycle (75%) and motor vehicle (73%) accidents reported due to aggressive /negligent driving. Highest number of motor vehicle accidents (20.5%) reported by the drivers in between 29 - 34 years old. Highest number of motorcycle accidents (28.5%) reported by the drivers in between 19-24 years old. Majority of the accidents were occurred, while the vehicle was moving on a straight road. Among drivers and motorcyclists (7%) were found to have consumed alcohol. Most of motorcycle and motor vehicle accidents occurred in daytime under daylight on weekdays.

Binary logistic regression is applied motorcycle and motor vehicles accidents separately to evaluate the odds of grievous accidents compared to non-grievous accidents. For motor vehicle accidents vehicle type, validity of license, time, location, alcohol test, accident cause, age of driver and gender have a significant effect on the severity of accidents. Bend or junction location, aggressive/negligent driving, drive by male drivers, drive at daytime, driving light vehicle and drivers who use alcohol below legal limit or no alcohol, have a high chance to be a grievous accident. Moreover, the older drivers have less accident risk. For motorcycle accidents, location type, time, age of driver, accident cause and gender have a significant effect on the severity of accidents. Among them, location type, accident cause and gender have an increasing effect on the probability of a grievous accident. Time and age of driver have a decreasing effect on the probability of a grievous accident. Straight road, aggressive/negligent driving, drive by male motorcyclists, daytime have a high chance to be a grievous accident. Moreover, the older motorcyclists have less accident risk. These findings can aid modifying regulations and laws and establishing preventive and protective approaches and strategies.

Keywords: Road accidents, Logistic Regression, Accident severity, Motorcycle accidents, Motor vehicle accidents

Acknowledgement

I would like to acknowledge the efforts, support, guidance, cooperation and encouragement of numerous people who have made it possible for me to undertake this study.

I wish to express my sincere gratitude to my supervisor, Mr.T.M.J.A.Cooray, for his guidance, encouragement and support in shaping the outlook of this thesis. He provided invaluable insights that have guided my thinking and understanding.

I also would like to express gratitude to the Senior DIG of Police, Police Head Quarters for his permission to supply the data for this study. Besides all the police officers who support to get the data that needed for this research are also greatly appreciated.

I would like to gratefully acknowledge the staff of Department of Mathematics, The Open University of Sri Lanka for their support given to me.

Last but not least I would like to thank my parents for their endless support throughout my journey. Without them I wouldn't be able to accomplish any of the achievements that I have attained so far.

TABLE OF CONTENTS

| | |
|-----------------------------------------------------------------------------------|-----|
| Declaration of the candidate and the supervisor..... | i |
| Abstract | ii |
| Acknowledgement | iii |
| Table of contents..... | iv |
| List of figures..... | ix |
| List of tables | x |
| List of abbreviations..... | xii |
| 1.INTRODUCTION | 1 |
| 1.1 Overview | 1 |
| 1.2 Background of the study | 1 |
| 1.3 Statement of the problem | 2 |
| 1.4 Objectives..... | 2 |
| 1.5 Significance of the study | 3 |
| 1.6 Source of Data | 3 |
| 1.7 Limitations of the study | 3 |
| 1.8 Outline of the report..... | 3 |
| 2.LITERATURE REVIEW | 4 |
| 2.1 Overview | 4 |
| 2.2 Previous Studies related to Motor Vehicle Accidents in Sri Lanka | 4 |
| 2.2.1 Descriptive Analysis | 4 |
| 2.2.2 Risk Factors Influencing Motor Vehicle Accidents | 4 |
| 2.3 Previous Studies related to Motor Vehicles Accidents in Other Countries | 6 |
| 2.3.1 Descriptive Analysis | 6 |
| 2.3.2 Risk Factors Influencing Motor Vehicle Accidents | 8 |
| 2.4 Previous Studies related to Motorcycle Accidents in Other Countries..... | 10 |

| | | |
|--------|-----------------------------------------------------------------------------|----|
| 2.4.1 | Descriptive Analysis | 10 |
| 2.4.2 | Risk Factors Influencing Motorcycle Accidents Using Logistic Regression ... | 11 |
| 3. | METHODOLOGY | 12 |
| 3.1 | Overview | 12 |
| 3.2 | Description of Data | 12 |
| 3.3 | Explanatory Variables..... | 12 |
| 3.4 | Research Methodology..... | 12 |
| 4. | ANALYSIS OF MOTOR VEHICLE ACCIDENTS..... | 14 |
| 4.1 | Overview | 14 |
| 4.2 | Frequency of accidents..... | 14 |
| 4.3 | Severity of accidents by year..... | 14 |
| 4.4 | Road Accidents by Vehicle Types..... | 15 |
| 4.5 | Results of Analysis of Motor Vehicle Accidents | 15 |
| 4.5.1 | Severity of Accidents by Motor Vehicles | 15 |
| 4.5.2 | Severity of Accidents by Validity of License..... | 17 |
| 4.5.3 | Severity of Accidents by Accident Cause | 17 |
| 4.5.4 | Severity of Accidents by Alcohol Test | 18 |
| 4.5.5 | Severity of accidents by Time of Accident..... | 18 |
| 4.5.6 | Severity of accidents by Weekday/ Weekend | 19 |
| 4.5.7 | Severity of Accidents by Road Surface | 20 |
| 4.5.8 | Severity of Accidents by Weather Condition..... | 20 |
| 4.5.9 | Severity of Accidents by Light Condition..... | 21 |
| 4.5.10 | Severity of Accidents by Location of the Accident..... | 21 |
| 4.5.11 | Descriptive Statistics of Age of Driver | 22 |
| 4.6 | Reduction of Levels of Factors..... | 23 |
| 4.7 | Results of Pearson Chi-Square Test..... | 25 |

| | | |
|--------|-----------------------------------------------------------------------|----|
| 4.8 | Detecting Multicollinearity in Binary Logistic Regression | 25 |
| 4.8.1 | Correlation Coefficients of Explanatory Variables | 26 |
| 4.8.2 | Detection of Multicollinearity based on Collinearity Statistics | 26 |
| 4.9 | Solutions to Multicollinearity | 27 |
| 4.10 | Selection of Number of Variables | 28 |
| 4.11 | Checking Linearity Assumption | 29 |
| 4.12 | Binary logistic regression analysis | 30 |
| 4.12.1 | Baseline Model | 30 |
| 4.12.2 | Developed model | 31 |
| 4.12.3 | Developed model interpretation | 31 |
| 4.12.4 | Logit Model | 33 |
| 4.12.5 | Variables not in the Developed Model | 33 |
| 4.12.6 | Importance of Variables in the Model | 33 |
| 4.13 | Measures of Goodness of Fit | 34 |
| 4.13.1 | Test of Model coefficients | 34 |
| 4.13.2 | Model Summary | 35 |
| 4.13.3 | Predictive Accuracy of Developed Model | 35 |
| 4.13.4 | Hosmer and Lemeshow test | 36 |
| 4.13.5 | Receiver Operating Characteristic Curve | 36 |
| 4.14 | Model Diagnostics | 37 |
| 4.14.1 | Testing Residuals | 38 |
| 4.14.2 | Checking Normality in Residuals | 38 |
| 4.14.3 | Testing Heteroskedasticity and Correlation in Residuals | 39 |
| 4.14.4 | Detecting Influential Observations | 39 |
| 4.15 | Validation of the Model | 40 |
| 4.16 | Summary | 40 |

| | |
|-----------------------------------------------------------------------------|----|
| 5. ANALYSIS OF MOTOR CYCLE ACCIDENTS..... | 41 |
| 5.1 Overview | 41 |
| 5.2 Severity of Motorcycle Accidents | 41 |
| 5.1 Factors Associated with Severity of Accidents | 42 |
| 5.2 Distribution of Age of Driver | 43 |
| 5.3 Reduction of Levels of Factors..... | 44 |
| 5.4 Results of Pearson Chi-Square Test..... | 46 |
| 5.7 Detecting Multicollinearity in Binary Logistic Regression | 47 |
| 5.7.1 Correlation Coefficients of Explanatory Variables | 47 |
| 5.7.2 Detection of Multicollinearity based on Collinearity Statistics | 48 |
| 5.8 Solutions to Multicollinearity..... | 49 |
| 5.9 Checking Linearity Assumption..... | 50 |
| 5.10 Binary logistic regression analysis | 50 |
| 5.10.1 Baseline Model..... | 50 |
| 5.10.2 Developed model..... | 51 |
| 5.10.3 Developed model interpretation | 51 |
| 5.10.4 Logit Model..... | 52 |
| 5.10.5 Variables not in the Developed Model | 52 |
| 5.10.6 Importance of Variables in the Model | 53 |
| 5.11 Measures of Goodness of Fit..... | 53 |
| 5.11.1 Test of Model coefficients..... | 53 |
| 5.11.2 Model Summary | 54 |
| 5.11.3 Predictive Accuracy of Developed Model | 55 |
| 5.11.4 Hosmer and Lemeshow test | 55 |
| 5.11.5 Receiver Operating Characteristic Curve..... | 55 |
| 5.12 Model Diagnostics | 56 |

| | | |
|--------------------------------|---------------------------------------------------------------|----|
| 5.12.1 | Testing Residuals..... | 56 |
| 5.12.2 | Checking Normality in Residuals..... | 57 |
| 5.12.3 | Testing Heteroskedasticity and Correlation in Residuals | 57 |
| 5.12.4 | Detecting Influential Observations | 58 |
| 5.12.5 | Validation of the Model | 58 |
| 5.13 | Summary | 59 |
| CONCLUSION AND DISCUSSION..... | | 60 |
| 6.1 | Overview | 60 |
| 6.2 | Conclusion and Discussion | 60 |
| 6.3 | Recommendations..... | 61 |
| REFERENCE LIST | | 63 |
| APPENDIX | | 67 |

LIST OF FIGURES

| | Page |
|---------------------------------------------------------------------|------|
| Figure 1.1: Number of accidents by year | 1 |
| Figure 4.1: Severity of accidents by year | 14 |
| Figure 4.2: Road accidents by vehicle types | 15 |
| Figure 4.3: Main types of motor vehicles..... | 16 |
| Figure 4.4: Severity of accidents by validity of license | 17 |
| Figure 4.5: Severity of accidents by alcohol test | 18 |
| Figure 4.6: Severity of accidents by time of accident..... | 19 |
| Figure 4.7: Severity of accidents by weekday/ weekend | 19 |
| Figure 4.8: Severity of accidents by road surface..... | 20 |
| Figure 4.9: Severity of accidents by weather condition | 20 |
| Figure 4.10: Severity of accidents by location of the accident..... | 22 |
| Figure 4.11: Histogram of age of driver..... | 23 |
| Figure 4.12: Log odds of severity vs. age of driver | 30 |
| Figure 4.13: ROC Curve | 37 |
| Figure 4.14: Normal probability plot of residuals | 38 |
| Figure 4.15: Plot of Cook's Distance..... | 39 |
| Figure 5.1: Severity of accidents vs. year | 41 |
| Figure 5.2: Distribution of age of driver | 43 |
| Figure 5.3: Log odds of severity Vs. Age of driver | 50 |
| Figure 5.4: ROC Curve | 56 |
| Figure 5.5: Normal probability plot of residuals | 57 |
| Figure 5.6: Plot of Cook's Distanc | 58 |

LIST OF TABLES

| | Page |
|----------------------------------------------------------------------------|------|
| Table 3.1: Description of Factors..... | 13 |
| Table 4.1: Accident frequency by year | 14 |
| Table 4.2: Accident Severity by Motor Vehicles | 15 |
| Table 4.3: Accident severity by accident cause..... | 17 |
| Table 4.4: Accident severity by light condition..... | 21 |
| Table 4.5: Descriptive statistics of age of driver | 22 |
| Table 4.6: Summary statistics for Proportion Test | 24 |
| Table 4.7: Chi-square test results..... | 25 |
| Table 4.8: Pearson Correlation matrix between 2 explanatory variables..... | 26 |
| Table 4.9: Collinearity statistics | 27 |
| Table 4.10: Collinearity statistics for remained variables..... | 28 |
| Table 4.11: Results of best subsets | 29 |
| Table 4.12: Results of baseline model | 31 |
| Table 4.13: Variables in the model | 32 |
| Table 4.14: Variables not in the model | 33 |
| Table 4.15: Importance of variables in the model | 34 |
| Table 4.16: Test of developed model coefficients..... | 34 |
| Table 4.17: Model Summary | 35 |
| Table 4.18: Classification Table | 36 |
| Table 4.19: Results of Hosmer and Lemeshow test..... | 36 |
| Table 4.20: Area Under the Curve..... | 37 |
| Table 4.21: Testing Residuals..... | 38 |
| Table 4.22: DW and LM test of residuals | 39 |
| Table 4.23: Category prediction | 40 |

| | |
|----------------------------------------------------------------------------|----|
| Table 5.1: Frequency Table for accident severity and risk factors | 42 |
| Table 5.2: Descriptive statistics of age of driver | 43 |
| Table 5.3: Summary statistics for Proportion Test | 45 |
| Table 5.4: Chi-square test results..... | 46 |
| Table 5.5: Pearson Correlation matrix between 2 explanatory variables..... | 47 |
| Table 5.6: Collinearity statistics | 48 |
| Table 5.7: Collinearity statistics of remained variables | 49 |
| Table 5.8: Results of baseline model | 51 |
| Table 5.9: Variables in the model | 51 |
| Table 5.10: Variables not in the Model..... | 52 |
| Table 5.11: Importance of variables in the model | 53 |
| Table 5.12: Test of developed model coefficients..... | 54 |
| Table 5.13: Model Summary | 54 |
| Table 5.14: Classification Table | 55 |
| Table 5.15: Hosmer and Lemeshow Test..... | 55 |
| Table 5.16: Area Under the Curve..... | 56 |
| Table 5.17: Testing Residuals..... | 57 |
| Table 5.18: DW and LM test of residuals | 58 |
| Table 5.19: Category prediction | 59 |

LIST OF ABBREVIATION

| Abbreviation | Description |
|--------------|------------------------------------------|
| B | Coefficient of logistic regression model |
| BJ | Bend/Junction |
| BL | No alcohol/below legal limit |
| Cause1 | Speeding |
| Cause2 | Aggressive/ negligent driving |
| Cause3 | Influenced by alcohol/drugs |
| Cause4 | Fatigue/fall asleep |
| Cause5 | Others |
| CI | Confidence Interval |
| CL | Clear |
| D | Dry |
| DD | Dusk/Dawn |
| Df | Degrees of freedom |
| DL | Daylight |
| DT | Day Time |
| DW | Durbin Watson Test |
| F | Female |
| GSL | Night, Good street lighting |
| HV | Heavy Vehicle |
| LM | Lagrange Multiplier Test |
| LV | Light Vehicle |
| M | Male |
| ML | Maximum Likelihood |
| NSL | Night, no street lighting |

| | |
|--------|------------------------|
| NT | Night Time |
| OL | Over Legal Limit |
| OR | Odds Ratio |
| Others | OT |
| RA | Rainy |
| RD | Road |
| RTA | Road Traffic Accidents |
| S.E | Standard error |
| Sig. | Significant |
| W | Wet |
| WD | Weekday |
| WE | Weekend |
| WL | With Valid License |
| WO | Others |
| WOL | Without Valid License |