MODELING OF ROAD TRAFFIC FATALITIES IN NAMIBIA: A GENERALIZED LINEAR MODEL APPROACH

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN APPLIED STATISTICS AND DEMOGRAPHY

Of

THE UNIVERSITY OF NAMIBIA

By

BERTHA BETTER THAN ALL NAMBAHU

200414127

2018

Main Supervisor: Dr Isak Neema

Co - Supervisor: Dr. Lillian Pazvakawambwa

Abstract

Fatalities caused by road traffic accidents have become a major concern worldwide and road traffic accidents happen almost every day on Namibian national roads claiming lives of many road users. It is predicted that by 2030, road crash fatalities would be the fifth most frequent cause of death worldwide. Many efforts have been made to prevent loss of lives and to curb down the MVA expenses on compensations due to road traffic accident; however, road traffic fatalities are still on the increase. Therefore, reliable statistical analysis is required to understand the main determinants of road fatal accidents. The main objective of this study was to identify the best model in modelling Road Traffic Fatalities to establish factors contributing to the number of road traffic fatalities in Namibia.The study was a quantitative cross – sectional study using secondary data of road accidents collected on daily basis by National Road Safety Council (NRSC) for the whole country from January 2012 to December 2013. Descriptive statistics were used to profile background characteristics of the population. The study explored various count models to assess factors influencing road traffic fatalities in Namibia. Six models namely; Poisson, Negative Binomial, Zero- Inflated Poisson, Zero Inflated Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial were explored and their performances were rated based on the AIC. The best model which was the Zero Inflated Negative Binomial (ZINB) was deployed for the estimation of parameters. Results showed that the region with highest number of road accidents from January 2012 to December 2013 with a frequency of (37 %) was Khomas region which is the region with the high population density and road traffic density. Therefore, intervention efforts should focus on more safety awareness campaigns in Khomas region. The ZINB has indicated that, the Road Traffic Fatalities in Namibia over a two calendar year under study were influenced by Khomas region; head on collision and side wipe collision.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

Praise the Lord, my soul; all my inmost being, praise his holy name (Psalm 103: 1, NIV). Above all, I want to thank the heavenly Father for all. I would also like to take this opportunity to express my profound gratitude and deep regards to my research supervisors' Dr Isak Neema and Dr Lillian Pazvakawambwa, as well as my colleague Mr Rodrigue Gnitchogna, for the patient guidance, enthusiastic encouragement and useful critiques of this research work. Many thanks to Hosianna prayers group, for the prayers and encouragement when things were getting tougher, I highly appreciate. To my sister Tangi Thimbopepi Shipombo, thank you so much my bestie for the moral support, love and care you have provided me with during my study. To all my friends, Mrs Paulina Kadhila, Dr Marius Hedimbi, Mrs Charlotte Amakutsi, Mrs Jambeko Shawelaka and Mrs Helana Nambili, you have been my peeler of strength, highly appreciated. To my late friend Laimi Shikongo, you were so determined to see me finishing this race, thank you so much for everything. To my MSc classmates, thank you so much buddies for your valuable contributions and selfless ideas shared. To all my friends, Hosianna Parish prayers group and family, I dearly appreciate your moral and spiritual support they gave me during study.

DEDICATION

I dedicate this work to my late friend Laimy Tumeniyeni Shikongo who passed on in a horrific car accident on 19th April 2013.

DECLARATION

I, Bertha Better Than All Nambahu, declare hereby that this study is a true reflection of my own research, and that this work, or part thereof has not been submitted for a degree in any other institution of higher education.

I, Bertha Better All Than Nambahu, grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit, for any person or institution requiring it for study and research; providing that The University of Namibia shall waive this right if the whole thesis has been or is being published in a manner satisfactory to the University.

Bertha B. Nambahu   Date

……………………………….   ……………………………..

LIST of ABBREVIATIONS

MVA	: Motor Vehicle Accident fund
NRSC	: National Road Safety Council
AIC	: Akaike Information Criterion
RTA	: Road Traffic Accident
CEO	: Chief Executive Officer
FARS	: Fatality Analysis Reporting System
RTIs	: Road Traffic Injuries
RAFs	: Road Accident Fatalities
WHO	: World Health Organization
NRSS	: Namibia Road Safety Strategy
OECD	: Organization for Economic Cooperation and Development
NEPRU	: Namibian Economic Policy Research Unit
GNP	: Gross National Product
MWT	: Ministry of Woks and Transport
GRSP	: Global Road Safety Partnership
INAR	: Integer – Valued Autoregressive
ANN	: Artificial Neural Networks
MNL	: Multinomial Logit
NL	: Nested Logit
OL	: Ordered Logit
OP	: Ordered Probit
HOL	: Heteroskedastic Ordered Logit
GOL	: Generalized Ordered Logit
GLM	: Generalized Linear Model
MXL	: Mixed Logit
GPR	: Generalized Poisson Regression

ZIP	: Zero Inflated Poisson
ZINB	: Zero Inflated Negative Binomial
NRAF	: Namibia Road Accident Forms
SPSS	: Statistical Package for Social Science
R	: Statistical software package
NB	: Negative Binomial

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background Information

Namibia is a southern African country whose western border is the Atlantic Ocean. It shares land borders with Angola and Zambia to the north, Botswana to the east and South Africa to the south. It is divided into 13 provinces and it`s capital city is Windhoek in the central part. It gained its independence on 21st March, 1990. Namibia`s population is about 2,113 million, with 43% in urban areas and 57% in rural areas (Namibia Census, 2011). The life expectancy at birth is 51.86 years for females and 52.47 for males (Namibia Census, 2012 est). Namibia is a lower middle-income country; its economic performance depends on agriculture, mining and manufacturing.

Transportation is one of the most important aspects in life. Transportation is an important question for any country`s development or economic progress, and it`s infrastructure are one of the measures of any country`s development or progress. Transportation not only includes moving people, but goods as well. Road traffic networks provide reliable, quick and flexible transportation system. The issue of traffic safety has been of great importance for many researchers, this is because Transportation is a mixed-blessing aspect (Lord & Mannering, 2010; Haleem, 2009).

No society can function without a good and reliable transportation infrastructure. Namibia is experiencing an increase in motorization while roads have deteriorated resulting in increased road traffic accidents. An efficient transportation system is not only critical but also essential for a country such as Namibia which is experiencing a moderately competitive economy. While the road safety situation in Namibia may not be the worst in the world, at least one person is killed and 16 others injured on our roads every day (MVA Fund Annual Report, 2010).

The Road Traffic Accidents (RTAs) can result in injury, property damage or death. Property damage refers to any harm or destruction of material goods such as vehicles, buildings and bridges etc. Injury is classified as serious or as slight injury. Serious injury means injury sustained in an accident that require hospitalization within 30 days after the occurrence of the accident while slight injury means minor cuts and bruises, sprains and light shocks sustained in an accident which may be treated at the scene of the accident or at home without being hospitalized (NRSC Action Plan 2009 – 2014, 2009).

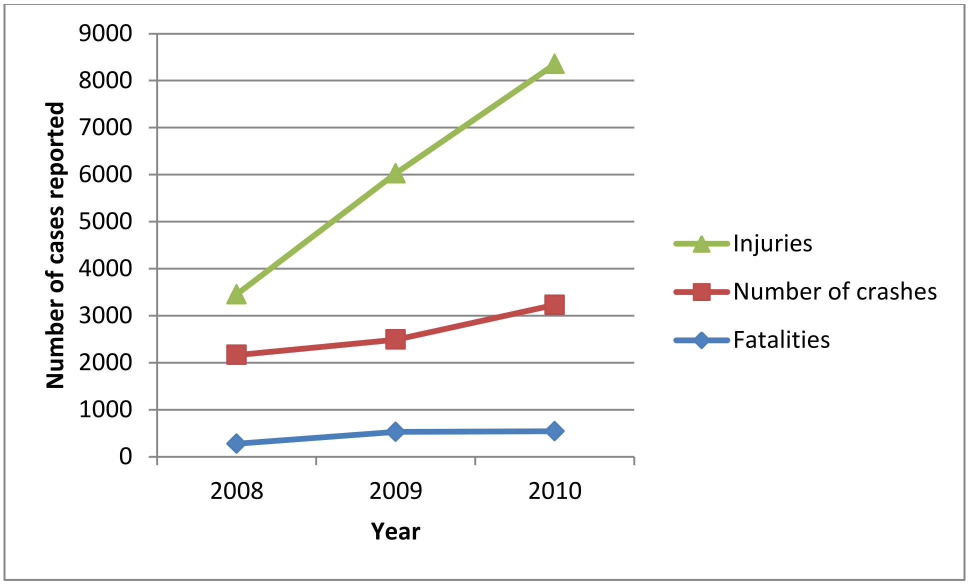
The National Road Safety Council (NRSC) defines a road traffic fatality as death of any person who dies outright or within 24 hours as a result of road traffic accident or collision or crash (NRSC, 2009). There are no mechanisms in place to follow up on the accident to be in line with the international definition of fatality. The definition used in the Fatality Analysis Reporting System (FARS) in United States of USA is “a person who dies within 30 days of crash on a US public road involving a vehicle with an engine, the death being the result of the crash.” All definitions single out that fatality is death of persons as a direct or indirect result of road traffic accidents.

Deaths caused by road traffic accidents have become a major concern worldwide. About 3000 people die from Road Traffic Injuries (RTIs) daily in the world. Eighty-five percent of the deaths and ninety percent of life lost due to RTIs are from the low income and middle income countries (Peden et al., 2004). One of the top three causes of deaths for 15 - 44 years’ age group in many countries is RTAs (World Bank, 2011). These have enormous implications for development in those countries. Some researchers like Ohakwe (2011), have indicated that RTAs result in the death of 1.2 million people worldwide each year and injuries of about 4 times this number (WHO, 2004). Further, Road Traffic Accidents (RTAs) are increasing with rapid pace and presently these are one of the leading causes of death in developing countries (Ohakwe, 2011).

The morbidity and mortality burden in developing countries is rising due to a combination of factors, including rapid motorization, poor road and traffic infrastructure as well as the behavior of road users (Nantulya & Reich, 2002). According to Manan (2011), in his study on accident prediction model at un-signalized intersections, he indicated that the road traffic accidents can be categorized in four common types. Where the driver collides with another vehicle or a roadside object, when the driver leaves his/her lane, rear-end collision and angle or side impacts, accidents involving pedestrians' cyclists and collision with animals.

In 2004, road traffic injuries were the ninth most frequent cause of death and WHO predicted that by 2030, they would become the fifth most frequent World Health Organization (2009). The Motor Vehicle Accident (MVA) fund is a statutory body established to design, develop, promote and implement motor vehicle accident and injury prevention measures. It also provides benefits to all people using Namibian roads. Motor Vehicle crashes are on the increase as recorded by MVA from 2008 to 2010, with 1889 reported cases in 2008, 1965 in 2009 and 2689 in 2010.

Figure 1.1: Road Traffic Accident trend in Namibia (2008 – 2010)



While the road safety situation in Namibia may not be worse in the world, current statistics indicate that every day at least one person is killed and 16 others are injured on Namibian roads. Given the size of Namibia's population (estimated at 2.1 million), this is unacceptably high. It is imperative therefore that, steps are taken to improve road safety in order to reduce the road carnage and create a safer and more forgiving road network. In the same vein, it is important to enhance public education activities in order to improve the understanding of road safety within all spheres of our road user communities (MVA Fund report, 2009/10). There have been several campaigns and plans that Namibia has put in place and currently running, in order to deal with the problem of road traffic accidents and ensure the safety of all road users in the country. Among those measures is the Xupifa Eemwenyo (Save lives) road safety project, which was established in 2005 with the aim of reducing fatalities, injuries and trauma on Namibian roads, create road safety awareness amongst Namibian road users. Other measures include Namibia

Road Safety Strategy (NRSS) and Khomas Region Transport Safety Plan, as well as, law enforcement of high traffic fines and the national drink-driving law.

1.2 Problem Statement

Road traffic networks are key economic drivers in these modern days. They provide a quick, reliable and flexible transportation system, for people, goods and services. Namibia is one of the developing Southern - African countries, where road traffic accidents happen almost every month claiming the lives of many Namibians. Throughout the world, the growths of transportation system have been and continue to be the backbone of economic development through import and export processes.

Increase in GDP (gross national product) is associated with greater movement of people goods and more investments in both vehicles and transport infrastructure. As a result, more vehicles are being imported to cater for transportation of people and goods and doing so enhancing development. Unfortunately, road traffic accidents happen every day – road injury was one of the top 10 causes of death in the world, according to the World Health Organization report (World Health Organization, 2011). For that reason, much research has been devoted to the analysis of road accidents (Brijs et al., 2007; Kim et al., 2007; Cafiso & Di Silvestro, 2011; Aguero – Valverde, 2013). It is of interest to analyse the number of fatalities that arise due to these accidents.

Financial implications due to fatalities and injuries caused by road traffic accidents (RTAs) have a tremendous impact on social well-being and socio economic development of every country. Road crashes are among the top leading causes of death among the people aged between 30 and 44 years according to the World Bank report of 2002. Furthermore, MVA Fund report of 2010 outlined that, the majority of the deceased in Namibia are young people within the productive age-group of 21 to 40 years. Therefore, it clearly shows that most of the victims are the youth. President O. R Tambo stated that “a nation that does not take care of its youth, it has no future” (National Youth Policy Report 2009 – 2014). Families are driven into poverty when bread-winners are killed in vehicle accidents. In Namibia, MVA fund spends a lot of money on funeral services and medical expenses on the road crashes victims. The question raised is until when and how much is the fund going to spend as fatalities are on the increase on our national roads? Substantial funneling used on awareness campaigns but road traffic fatalities are still on the increase.

The Regional Road Safety Plan (2011) estimated that accidents costs in 2006 amounted to about USD 25.1 million, an amount equivalent to 0.67% of the Gross National Product (GNP). One of the major obstacles in preventing the aforementioned problems is the lack of reliable statistical findings. The high number of role players and projects leads to misunderstandings as well as duplication of responsibilities.

Time is now to redouble efforts in order to prevent or reduce road accidents and associated health implications. According to Kononov and Janson (2002), one cannot improve safety without successfully relating accident frequency and severity to the causative variables. Most published reports about road traffic accidents in Namibia by MVA (Motor Vehicle Accident) fund and NRSC contain descriptive information only, and therefore there is a high need for in-depth studies. Factors such as crash cause, crash type, time of the day, weather condition, region where accidents are more likely to result in fatalities are quite useful to road users to decide when to travel. So there is a need to model and assess the effect of the above mentioned factors to road traffic fatalities in Namibia. Therefore, this study can be regarded as highly significant in road safety sector in the country. Our contribution in this study will be to identify the best model in modelling the number of road traffic fatalities in Namibia during the period of 2012 – 2013.

1.3 Research Objectives

The main objective of this study was to explore various count models and identify the best model in modelling the number of road traffic fatalities on Namibia roads in an effort to provide useful information to complement national prevention strategies and interventions aiming at minimizing road accidents fatalities. The following secondary objectives were targeted.

(i) To describe patterns and trends in the number of road accident fatalities

(ii) To model the number of road traffic fatalities and identify the best model that fit the number of traffic fatalities data on Namibian roads.

(iii) To identify factors influencing the number of road accident fatalities for the period of two years, 2012 – 2013.

1.4 Significance of the Study

It is essential to understand trends in collision rates and collision risk, especially when planning for road safety campaigns. This can be achieved by enabling road safety to be highlighted as a priority for action and developing robust arguments for the adoption of interventions (OECD, 2008). By comparing road crashes with other major causes of death such as other accidental deaths, HIV and Malaria, by evaluating the cost of road casualties in order to estimate medical costs, the burden of road accidents in terms of future disability, and the cost/benefit ratio for schemes to reduce casualties and efficiently deploying these interventions across areas of higher risk and/or where the greatest potential improvement can be achieved, such as in different road user groups, areas types (Urban/rural), or road types etc.

The study intended to understand major contributory factors to road fatal accidents in Namibia. The findings of the study will be useful Government particularly the to the Ministry of Works and Transport (MWT), through Motor Vehicle Accident (MVA) Fund and National Road Safety Council (NRSC) to assist on how to minimize the loss of lives due to road accidents in Namibia and contribute to all possible alternatives solutions to save lives. Other notable beneficiary will be the Namibian Police and road traffic units.

1.5 Limitation of the study

Secondary data was used for the study and if there were errors involved in data collection, they cannot be avoided. The data record had no information about the age and gender of the driver and the people who died in the accident as this would have given a valuable information. The researcher intended to make inferences on the age group that is mostly affected by the road accident fatalities as well as the gender.

1.6 Organizations of the chapters

The thesis is organized as follows: Chapter 1 is the introduction giving some background information, the problem statement, the research objectives and the significance of the study. Chapter 2 presents the literature review, which presents the state of research or the topic, what other researchers and book authors have done on topics similar to this research study. Chapter 3 consists of the methodology. A detailed explanation on how the study was carried out is given, the procedures that were adopted, the population of the research study and the analysis mode. Chapter 4, here we present the results of the analysis. Chapters 5 is the last chapter consisting of discussion, conclusion and recommendations made based on the results of the study.

CHAPTER 2: LITERATURE REVIEW

Road traffic accidents are major public health problem worldwide, accounting for almost 1.2 million deaths per year. In 2002, the overall global road traffic injury rate was 19/100 000 population, with 90% of cases in low-income and middle-income countries. Furthermore, according to the World Health Organization, the number of road traffic deaths in low-income and middle-income countries is expected to increase by 80% from the year 2000 to 2020 (World Health Organization, 2004).

Global Road Safety Partnership (GRSP), a United Nation initiative, estimated that in 1999 alone, the cost of road crashes globally was in excess of US$ 500 billion, of which US$60 billion was contributed by developing countries including Namibia (NEPRU, 2006). Namibia, as a young country which recently gained its independence in 1990, is burdened with so many social problems and this situation is aggravated by road accidents which not only cause human suffering, but also carry a very high price tag. The Namibian government created the National Road Safety Council (NRSC) in 1996 to spearhead efforts to curb the increasing dangers of today`s road traffic system in an attempt to address the situation (NEPRU, 2006).

MVA fund report indicated that the number of reported crashes in Namibia was reduced by 8% from 2008 to 2009. Despite this reduction, fatalities increased by 68% and injuries also increased by 88% in the same period. This signals the challenge the country faces and justifies the need for a collective intervention by all stakeholders to tackle this growing health problem. A similar study conducted by Georgia department of transport in 2008 found that over six million people (an average of 2,394 each day) were involved in a motor crash either as a driver or a passenger or pedestrian from 2000 to 2006 in Georgia. Close to a million people were injured in motor vehicle crashes in Georgia over the same period. It was also reported that 11,435 people lost their lives with an average of 31 fatalities. Georgia department of transport also compared the number of motor vehicle fatalities which exceeded the number of murder; injuries from accidents, and far exceeded the number of aggravated assaults.

According to NRSC 2004 report, it has been observed that over the past years in Namibia, the highest number of fatalities was recorded during April, August and December months. Well, this could be because of public holidays or a festive season that falls in the above mentioned months in Namibia, Easter holiday, Heroes day and Christmas break respectively, which leads to higher traffic volumes on national roads in these periods in Namibia, putting road users at risk of road crashes. The design and conditions of roads and road networks, excess and inappropriate speed, young drivers, consumption of alcohol, no use of seat-belts, as well as the unavailability of air bags have all been reported as other determinants of road accidents including fatal events. However, data from transitional countries of Southeast Europe, including Albania, are sparse and there is a lack of research agenda on road safety in low-income and middle-income countries (Croat, 2008).

The human factor appears in the literature as being the most common determinants of road safety. As indicated by many researchers the strongest association of crash injury severity and fatalities are related to behavioral factors including: alcohol/drug use (Kuruc et al., 2009; Ponce et al., 2011); speeding (Dissanayake & Lu, 2002; Afukaar, 2003); failure to wear seat belts (Munk et al., 2008; Siskinda et al., 2011; Kashani et al., 2012); using mobile phone while driving (Violanti, 1998) and sleep/fatigue (Radun & Summala, 2004).

Speeding is a critical safety concern, especially for developing countries; where fatalities are more common among pedestrians and users of two and three wheelers (Mohan, 2002). Fatigue crashes are usually severe, as the driver makes no attempt to limit the consequence (Radum and Summala, 2004). Such types of crashes are linked with the nature of the road alignment; good condition of the road network make the driver task easy and monotonous, which demanded little effort as a result sleep/fatigue will be induced (Rossi et al., 2011). The most dominant factor that affects the severity of injury is the impact of speed. As reported by Krafft et al., (2009), 10% speed reduction before the impact can reduce fatality injury of road traffic crashes by 30%. It is further stated by Seyer et al. (2000) that, the impact of the speed increases the risk in injury.

However, actual impact speed is not readily available in most reports (Lee & Li, 2014). Therefore, variables such as the posted speed limits have been used as proxy variables for impact speed (Lee & Li, 2014). Moskal et al. (2012) found that the risk of injury is high for motorcycle riders than the moped riders due to difference in impact speed. This is due to the fact that mopeds have low speed limit and thus less impact. Road with high speed limit frequently causes fatal crashes; this is because vehicles` speeds increases with speed limits (Lee & Li, 2014).

Studies also suggested that road alignment (Hosseinpour et al., 2014), street light condition (Wanvik, 2009; Mogaka et al., 2011) and adverse weather conditions (Mondal et al., 2011) are important predictors that affect the severity that may lead to fatality of vehicle crashes. As stated by various researchers, rainy weather significantly affects crush severity that may lead to fatalities (Mogaka et al., 2011; Mondal et al., 2011). Wet roads reduce the tire traction efficiency, leading to poor braking performance (Cho et al., 2006). In addition, visibility obstruction due to fog results in higher severity crashes (Abdel – Aty et al., 2011). Commercial and public transportation operated in developing countries are influenced by the owners to work excessively long hours when exhausted (Mock et al., 1999).

On the other hand, in multi vehicle crashes, heavy trucks has better resistance to crash impact and protect its occupants that may results in more serious injuries leading to fatality to the other vehicle(s) involved in the crash (Kockelman & Kweon, 2002). Most of these facts on crash related injury severity and associated factors are obtained from developed countries. Evidence is rare from developing countries, African countries in particular (Lagarde, 2007). Therefore, this study will contribute to fill that gap, as it will assess and model the potential factors related to crash injury severity that may result in fatalities in Namibia.

Environmental factors such as lightning and road surface conditions are also found to be closely related to injury severity (Lee & Li, 2014). Khorashadi et al. (2005) found that crashes in the morning (5:31 – 8:00) are less likely to result in severe or fatal injury in rural and urban areas. Temperature and rainfall play a role too in severity injury of road traffic crashes. Some researchers found out that the driver`s condition have an effect on injury severity. Nassiri & Edrissi (2006), found that driver fatigue increases the likelihood of more severe injuries in truck crashes for two lane rural highways in Iran based on an Ordered Logit model. Similarly, Zhu and Srinivasan (2011) indicated that driver fatigue, illness, distraction and unfamiliarity with the vehicle significantly increase injury severity.

El-Basyouyn et.al (2014) stated that the development and application of models to investigate the relationship between weather and safety has been the subject of numerous studies. Qiu & Nixon (2008) conducted a meta-analysis and found that comparing with rates in non-adverse weather conditions; fatal crash rates increased by 9% in snow fall, and 8% when raining. Brijs et al.

(2008), performed an integer auto regressive model to study the effect of weather conditions on daily crash counts. They found that the highly significant factor is rainfall and that an increase of precipitation intensity leads to a higher number of crashes.

Several researchers analyzed the effects of weather elements on crash severity. Brijs et al. (2008) found a positive relationship between the number of hours of rainfall per day and the number of crashes. They further indicated that, if the duration of the precipitation increases by one unit (0.1 hours per day), then there is an expectation of an increase in the mean number of crashes by 0.27% and 0.33%. Similar findings were made by El-Basyouny & Kwon (2012). Additionally, El-Basyouny & Kwon established that for severe collisions, there was a significant positive relationship with mean temperature and a significant positive relationship with total snow fall and total precipitation. They further indicated that, for property-damage-only, there is a significant inverse relationship with mean temperature and a significant positive relationship with total and previous snow fall and total precipitation.

Previous research shows that various weather elements have significant effects on crash occurrence and risk; however, little is known about how these elements affect different crash type (El-Basyouyn et. al., 2014). El-Basyouny et al. (2014) investigated the effects of time and weather on crash types using full Bayesian multivariate Poisson lognormal models for seven crash types using five years of daily weather and crash data collected for the entire city. The overall results showed that temperature and snowfall were statistically significant with intuitive signs (crashes decrease with increasing temperature; crashes increase as snowfall intensity increases) for all crash types, while rainfall was mostly insignificant.

They further stated that, major snow or rainfall events following a dry weather condition were highly significant and positively related to three crash types: Follow-Too-Close, Stop-Sign-Violation, and Run-Off-Road crashes. The day-of-the-week dummy variables were statistically significant, indicating a possible weekly variation in exposure (El-Basyouny et al. 2014). Adverse weather and consequent conditions have a significant impact on road surface friction (due to precipitation), vehicle performance on the road (due to snow, ice or wind) and the driver (due to impaired visibility during inclement weather), which often increase the risk of crash occurrence (El-Basyouny et al. 2014).

The effect of weather elements has been assessed by several researchers, namely rainfall (Usman et al., 2012; Yu et al., 2013), temperature (Bergel et al., 2013; Brijs et al., 2008; El-Basyouny & Kwon, 2012), snow fall (Andrey, 2010; Eisenberg & Warner, 2005; Hermans et al., 2006; Khattak & Knap, 2001), on crash occurrence (Ahmed et al., 2012; Brijs et al., 2008; Usman et al., 2012; Yu et al., 2013), whereas other researchers investigated these effects on crash severity (Andrey, 2010; Bergel et al., 2013; Eisenberg, 2004; El-Basyouny & Know, 2012). The development and application of models to investigate the relationship between weather and crashes has been the subject of numerous studies. Brijs et.al, (2008) studied the effects of weather conditions on daily crash counts using a discrete time series model. They found out that the assumptions related to the effect of weather conditions on crash counts are found to be significant in the data and that if serial temporal correlation is not accounted for in the model, this may produce biased results.

Brijs et al., (2008) stated that, weekdays (Monday – Friday) are more dangerous than weekends (Sunday being the reference day). Interestingly, the difference between Saturday and Sunday is however not significant. Tuesday and Friday were the most dangerous days of the week in terms of the number of crashes. In fact, from the overall perspective, the variable "day of the week" is a highly significant variable in Integer – Valued Autoregressive (INAR) model (Brijs et al., 2008).

(Brijs et al., 2008) further found that day-of-the-week dummies can be seen as some kind of proxy variable for exposure when real traffic exposure information is missing. Martensen & Dupont, (2013), on their study of comparing single vehicle (sv) and multivehicle fatal road crashes analysis, found that crashes are more likely to be single vehicle crashes at night than during the day, in the darkness or twilight as compared to day light, at the weekend as compared to weekdays, and at weekend nights as compared to all other times (weekend-day, week-day, and week-night. They further found that using the joint model, the effect of daylight proved to be independent of the day light.

A substantial number of crashes occurred during the day but in the absence of daylight (120), or at night hours, but in daylight conditions (20). Each variable has its own effect anyway. At night and in the weekends, the share of single vehicles crashes is higher (Martensen & Dupont, 2013). The share of single vehicle crashes is higher in the darkness or twilight (Martensen & Dupont, 2013). Martensen & Dupont (2013), further indicated that, the only variable that did not become significant in the joint model for the time variables was weekend nights. They argued that this variable would be significant if there was an interaction between weekend and daytime, meaning, when the difference between day and night would be larger in the weekends than during the weekdays.

The effect of the time variables is related to that of the traffic flow (Martensen & Dupont; 2013). At night and in the weekend the traffic flow is lighter than during the weekdays. The joint analysis of time variables shows that the traffic density is more important for explanation of severity of crashes that other possible characteristic then may differ between weekday's one hand and nights or weekends on the other hand (Martensen & Dupont, 2013).

More research was done on the effects of weather on road crashes. Bergel-Hayat et al. (2013) studied the effects of weather on road accident risks in France, Netherland and Athens regions and made the following comments restricted to the parameters with a 95% significant level

- Rainfall is positively correlated with the number of injury accidents (100 mm of additional rainfall during a month increases the number of injury accidents in that month by 0.2 – 0.3%)
- Temperature is positively correlated with the number of injury accidents (1ºC of additional average temperature during that month increases the number of injury accidents in that month by 1 – 2%)
- Temperature is positively correlated with the number of injury accidents, but with different magnitudes on main roads and motorways in France (1ºC of additional average temperature during month increases injury accidents by 0.4% and about 2% on main roads and motorways respectively), it is also the case on rural roads and motorways in the Netherlands (1ºC of additional average temperature during a month increases injury accidents by about 1% and 2 – 3% on rural roads and motorways respectively).
- The occurrence of frost is negatively correlated with the number of injury accidents (1 additional day of frost during a month decreases the number of injury accidents in that month by 0.3 – 0.6%)

- Rainfall is positively correlated with the number of injury accidents both in France and the Netherlands (100 mm) of additional rainfall during a month increases injury accidents on motorways by about 0.5% on main roads in France, and 0.8% on motorways in the Netherlands Contrarily, rainfall is negatively correlated with the number of injury accidents in the region of Athens, where the network mainly consists of urban roads (100 mm additional rainfall during a month decreases injury accidents by approximately 5%)

Abbassi (2005) study on road accidents in Kuwait used an autoregressive integrated moving averages (Box Jenkins) model and compared it with the Artificial Neural Networks (ANN) analysis to predict accident fatalities in Kuwait. His study concluded that ANN was better in case of long term series without seasonal fluctuations of accidents or autocorrelation components. Brijs et al. (2007) provided a Bayesian Models for ranking hazardous road sites. They discussed the importance of identifying the sites that are more dangerous than others in order to help in better scheduling road safety policies.

According to Lee & Li (2014), in order to analyze the effects of various factors on injury severity and predict different levels of injury severity based on the factors, researchers have applied various statistical models including a Multinomial Logit (MNL) model, a nested logit (NL) model, an ordered logit (OL) or ordered probit (OP) model, a heteroskedastic ordered logit (HOL) model, a generalized ordered logit (GOL) model, a mixed logit (MXL) model, a Bayesian ordered probit model and a bivariate ordered probit model. The past studies identified the factors affecting injury severity using the above mentioned models. The factors are categorized into five groups as follows: (1) driver characteristics; (2) impact speed; (3) road geometric characteristics; (4) environmental characteristics and (5) vehicle characteristics.

Driver characteristics include driver demographic factors such as age and gender (Lee & Li, 2014). According to Zhang et al., (2000), older drivers (65+) are more likely to be killed or seriously injured in the traffic crashes than middle - age drivers. It is observed that drivers younger than 35 years old are more likely to have evasive actions that will prevent severe injury (Harb et al., 2009). Generally, females are more likely to face fatal injury than males (Srinivasan, 2002; Kockelman & Kweon, 2002).

According to Harb et al., (2009) on their road accident analysis study by investigating the influence of vehicle characteristics on injury severity, it was found that truck drivers are more likely to perform evasive actions to avoid crashes compared to passenger car drivers. Year model of the car is another important factor associated with injury severity as it indirectly reflects vehicle technology (Lee & Li, 2014). Lee & Li (2014) further stated that the effect of road geometry on injury has also been analyzed. Chung (2013) found out that injury severity can be increased by narrower median islands and fixed object in the median islands. Crashes on roadways with more number of lanes would result in less severe injury (Zhu & Srinivasan, 2011).

There has been considerable research conducted over the last 20 years focusing on predicting and modelling motor vehicle crashes on transportation facilities (Lord et al., 2004). The range of statistical models commonly used includes binomial, Poisson, Poisson – gamma (or Negative Binomial), Zero Inflated Poisson and Zero Inflated Negative Binomial. However, making an intelligent choice on which model to use is difficult each having its assumptions. Many researches have been conducted to investigate the suitability of the Poisson regression model to predict accident frequencies at intersections or roadways (Miaou & Lum; 1993, Jones et al., 1991).

Generalized Linear Models (GLM) as introduced by Nelder and Wedderburn (1972) is a powerful family of models that allow for a more flexible modeling of the response variable on any distribution from the exponential family. The strength of this is that there is no assumption about a normally distributed response but can take the forms of the counts, proportions, binary responses, positive continuous values, as well as Normal. Logistic, Poisson and negative binomial regression models are the three common GLM family members. The use of the particular models within the GLM family is dependent on the form of the data. The probability for the dependent component is linked to the independent variables through a link function.

Regression model is used to predict a count dependent variable affected by one or more independent variables. The model parameters are estimated by the method of maximum likelihood estimation (Famoye & Singh, 2003). Famoye and Singh (2003) stated that count data often exhibit substantial variation where the sample variance is larger than the sample mean (a case of over – dispersion) or the sample variance is smaller than the sample mean (a case of under – dispersion).

Poisson regression has been used to analyze count data where the sample mean and sample variance are almost equal (Famoye & Singh, 2003). The Poisson regression model is a generalization of the standard Poisson regression model (Famoye et.al, 2003). Poisson type regression models have been used to model count response variable affected by one or more covariates in recent years (Famoye et. al., 2003). However, according to (Winkelmann & Zimmermann, (1994), the Poisson regression model will not be appropriate when the data set exhibit over – dispersion or under – dispersion. In their same study, they have stated that if the variance is above the mean, it indicates over – dispersion, while if the variance is below the mean, it indicates under – dispersion. Winkelmann & Zimmermann, (1994), developed the generalized event count models based on the Poisson, negative binomial and the binomial distributions to describe the traffic accident phenomenon. In number of research findings, the accident data were found to be significantly overdispersed (Miaou, 1994, Shankar et al., 1995, Poch & Mannering, 1996, Barron, 1998).

The generalized Poisson Regression (GPR) Model has been used to model dispersed count data. It is a good competitor to Negative Binomial Regression model when the count data is over-dispersed. Based on the test for the dispersion parameter and the goodness – of – fit measure for the accident data, the GPR model performs as good as or better than other regression models (Famoye et. al., 2004). There are many studies in social sciences, such as traffic accident analysis, in which the event counts may be characterized by a large number of zero observations (Chin & Quddus, 2003). The most weakness of the Poisson model is the imposed equality of conditional mean and variance of the dependent variables. There are several ways to test for over dispersion as indicated in literature (Dean & Lawless, 1989, Lee 1986, Cameron & Trivendi, 1990, for example). Several researchers employed the negative binomial (NB) distribution instead of the Poisson distribution by relaxing the condition of the mean being equal to the variance, and the NB regression model was found to be more suitable in describing discrete and non-negative events with the aim of overcoming the over dispersion problem (Miaou, 1994, Shanker et al., 1995, Poch & Mannering, 1996, Barron, 1998).

In the case of excess zeros in the data that may cause over –dispersion in the results, Zero Inflated Poisson (ZIP) and Zero Inflated Negative (ZINB) regression models have been proposed (Famoye & Singh, 2006). The excess zeros can occur as a result of clustering and over – dispersion has the tendency to increase the proportion of zeros and whenever there are too many zeros relative to Poisson assumption, the negative binomial regression tends to improve the fit of data (Famoye & Singh, 2006). For a better fit, and over – dispersed model that incorporates excess zeros should serve as an alternate. Gurmu & Trivendi (1996) illustrated that the negative hurdles model which allows over – dispersion and the presence of excess zeros is more appropriate among the models they considered.

2.0 Road accidents in Africa

A generalized ordered logit model analysis/partial proportional odds model was used in Ethiopia to assess the effect of excessive speeding and falling asleep while driving on crash injury severity on the study that was done in 2014 using June (2012) to July (2013) data, on one of the main and busiest highway in Ethiopia (Abegaz et al., 2014). A total of 819 crashes were recorded during the study period. 424 (51.8%) were property damage only; 113 (13.8%) minor injury crashes; 106 (12.9%) serious injury crashes and 179 (21.5%) fatal crashes. Crash injury severity often categorized universally as discrete ordered categories such as: fatal injury, incapacitating injury, non- incapacitating injury, possible injury, and property damage only. In Ethiopia crash injury severity is classified into four categories: Fatal injury (death at the scene or up to one month following an incident); serious injury (victim hospitalized at least for 24h); minor injury (victim treated at an outpatient service or hospitalized for less than 24h), and property damage only (crash without any human injury).

Driver`s behavior: Excessive speeding, alcohol use and sleep/fatigue were identified as significant factors for crash injury severity. Of these factors, sleep/fatigue is one of the potential determinants; the result showed that, the divers who slept while driving ended up in a more serious injury category compared to alert drivers. Similarly, driving after drinking alcohol was significant cause; drunk drivers were more likely to sustain severe injuries comparing with the non-drunk drivers.

Speeding was identified to have varying coefficients for different injury levels, its highest effects on severe and fatal crashes, the first panel of coefficients (i.e., property damage Vs minor injury + serious injury + death), the second panel of coefficients (i.e., property damage + minor injury Vs serious injury + death), and the third panel of coefficients (i.e., property damage + minor injury + serious injury Vs Death). However, the results demonstrated that, those who drive on a congested road network (higher density of road that causes less spacing) were sustaining a less severe injury category compared with free flow traffic.

Environmental conditions: Among those environmental related conditions, rainfall and driving at night in the absence of street light showed a significant contribution to crash severity. Driving in a rainy situation was identified to be significant for crash severity comparing with driving on a clear or other weather conditions. Similarly, night time driving in the absence of street light resulted as a more serious injury outcome than day time driving and driving at night time in the presence of street lights.

Vehicle type: Vehicle types; minibus or Vans and two or three wheeled vehicles (bicycle, motorcycle and Bajaj taxi, were increasing the probability of higher crash severity. On the other hand, automobiles or station wagons or pickups were associated with less severe crashes. Heavy trucks traveled on a congested road network also showed protective effects on the severity, they were involved in less severe injury crashes.

2.1 Theoretical Review on GLMs

Modelling of count data such as accident crash data is a common task in social sciences and economics. An accident crash mostly results in two events, either resulting in a fatality (non-zero event) or non-fatality (zero events). If in many accidents that occurred, no fatality was recorded, that may result in excess zeros. The Poisson regression model is the simplest model in modeling count data limited by the assumption of equal mean and variance as count data may typically exhibit over-dispersion or excess zeros (zero-inflated data) in the data (Lambert, 1992, Yesilova et al., 2010). Therefore, this can be handled by using Negative Binomial regression which belongs to the family of generalized linear models (McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972). However, although NB can handle the over-dispersion rather well, in many applications it might not be sufficient for modeling excess zeros. Thus an increased interest in statistical models in zero-augmented models that address the issue of capturing excess zeros by a second model component was observed.

Some GLMs models that were recommended for capturing excess zeros are Hurdle models and Zero Inflated Models (Mullahy, 1986; Lambert, 1992, Rodout et al., 1998). The Zero-inflated models and Hurdle models are widely used in analysis of excess zeros (zero-inflated) in the data (Consul, 1989; Consul & Famoye, 1992). Hurdle models combine a left-truncated count component with a right – censored hurdle component (Mullahy, 1986), while Zero-Inflated models are mixture models that combine a count component and a point mass at zero (Lambert, 1992).

Hurdle and Zero-Inflated models are found in the pscl package in R software, by using the command hurdle () and zeroinfl () respectively, while the Poisson regression and Negative Binomial regression under the MASS package in R by using the command glm() and glm.nb() respectively(R manual). All GLMs use the same log linear function ($\log(\mu) = x^{\top}\beta$) but different assumptions on the remaining likelihood.

2.1.1 Poisson regression

Poisson regression is traditionally conceived of as the basic count model upon which a variety of other count models are based. The two most popular models for count data are the Poisson model and Negative Binomial model (Ehsan and Adnan, 2011). Mostly, modeling of counts, the starting point was the Poisson distribution. The Poisson Regression model can be written as

$$H_{Y_i}(y_i) = \frac{e^{\mu_i} \mu_i^{y_i}}{y_i!} \tag{2.1}$$

Where the logarithm of mean of Poisson distribution (μ_i) is assumed to be a linear function of the independent variable (x_i) given by

$$\log(\mu_i) = x_i \beta \qquad (2.2)$$

y_i denotes dependent variable having a Poisson distribution

x_i denotes the independent variables

Suppose the dependent variable (Y_i) is a count response variable that follows a Poisson distribution, Y_i can be modeled as follows with the probability of Y_i given by the equation below

$$f_i(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right) \frac{(1+\alpha y_i)^{y_i - 1}}{y_i!} \exp\left(\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right) \qquad (2.3)$$

Where $y_i = 0,1,2 \dots,;$

$\mu_i = \mu_i(X) = e^{XB}$, where X is a $(k-1)$ dimensional vector of covariates and B is a $k-$ dimensional vector of regression parameters. α is the dispersion parameter.

The parameter of dispersion is being observed on three cases (1) equi-dispersion; when $\alpha = 0$, and equation (2.3) reduce to PR, case (2) over-dispersion; when $\alpha > 0$, equation (2.3) will always sum to 1 and case (3) under-dispersion; when $\alpha < 0$, equation (2.3) may not sum to 1 as the equation get truncated. The variance and the mean of the response variable Y_i are given by

Variance: $$V(Y_i | x_i) = \mu_i(1+\alpha\mu_i)^2 \qquad (2.4)$$

Mean: $$(\mu_i) = E(Y_i|x + i) \quad (2.5)$$

2.1.2 Negative Binomial regression

The NB regression model is given by

$$f(y_i) = p(Y_i = y_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\, y_i}\left(\frac{\theta}{\theta + \mu_i}\right)^{\theta}\left(\frac{\mu_i}{\mu_i + \theta}\right)^{y_i}, y_i = 0,1,2 \quad (2.6)$$

Where $\theta = \frac{1}{\alpha}$, α is the dispersion parameter, Γ(.) is a gamma function, the dependent variables (Y_i) has a NB distribution with the two parameters $\mu_i \geq 0$ and $\theta \geq 0$, where the mean and variance are given by

Mean $$E(Y_i) = \theta\,\mu_i \quad (2.5)$$

And the

Variance $$\text{var}(Y_i) = E(Y_i)(1 + \mu_i) = \theta\,\mu_i(1 + \mu_i) \quad (2.6)$$

Generalized linear models (GLMs) are a set of models that are used when the response variable violates some of the important assumptions of linearity. In such a case, the response does not follow a normal distribution and efforts to minimize them have proven futile. If the response variable is not normally distributed, the GLMs which assume a link linear relationship based on a chosen link function may be utilized to complete the analysis. The GLMs are statistical models that can be written as

$$y = g(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + e \tag{2.9}$$

The GLMs are members of the exponential family that take the form

$$f(y;\theta;\phi) = \exp\left[\frac{(y\theta - b(\theta))}{a(\phi)} + c(y,\phi)\right] \tag{2.10}$$

Where, **Y** is a vector of response variable`s counts; the variables are the $x_i$ explanatory variables linearly associated covariates, β`s are the regression coefficients, e is the error variability that cannot be accountable for by the predictor (x_i) variables, $g(...)$ is a monotonic function relating the mean of the response variable to the linear predictors and some functions $a(.), b(.), c(.)$. The values of the regression coefficients (β`s) were obtained by Maximum Likelihood (ML) estimations as follows;

Here we wished to estimate parameters β which are related to the explanatory variables through

$$E(Y_i) = \mu_i \tag{2.11}$$

And

$$g(\mu_i) = X^T \beta$$

(2.12)

For each Y_i , the log-likelihood function is given by;

$$l_i = y_i\, b(\theta_i) + c(\theta_i) + d(y_i) \tag{2.13}$$

Where the functions b,c, and d are known functions. Also

$$E(Y_i) = \mu_i = -C'(\theta_i)/b'(\theta_i) \tag{2.14}$$

$$Var(Y_i) = \left[b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)\right] / \left[b'(\theta_i)\right]^3 \tag{2.15}$$

And $g(\mu_i) = X^T \beta = \eta_i$, where X is a vector with elements x_{ij}, $j = 1, \cdots p$.

The log-likelihood function for all the Y_i `s is

$$l = \sum_{i=1}^{N} l_i = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i) \tag{2.16}$$

Therefore to obtain the maximum likelihood estimator for the parameter β_j we need

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial \beta_j}\right] = \sum_{i}^{N}\left[\frac{\partial l_i}{\partial \theta_i} . \frac{\partial \theta_i}{\partial \mu_i} . \frac{\partial \mu_i}{\partial \beta_j}\right] \tag{2.17}$$

Using the chain rule for differentiation by considering each term on the right hand side and we obtain the following equation:

$$U_i = \sum_{i}^{N}\left[\frac{(y_i - \mu_i)}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)\right] \tag{2.18}$$

The variance-covariance matrix of the U_j `s has terms

$\Im_{jk} = E\left[U_j U_k\right]$. Therefore, the maximum likelihood estimation formula is given by

$$b^{(m)} = b^{(m-1)} + \left[\Im^{(m-1)}\right]^{-1} U^{(m-1)} \tag{2.19}$$

Where the difference between successive approximation $b^{(m-1)}$ and $b^{(m)}$ is sufficiently small.

2.1.3 Generalized Poisson Regression Model

For GLMs, the *glm* command will be used in the study for the estimation of the response variable, μ will be used in expressing the mean instead of λ which is used in many statistical literature for expressing the mean (Hilbe, 2011).The count data for the response variable and explanatory variables will be modeled with the Poisson regression as follows

$$H_{Y_i}(y_i) = \frac{e^{\mu_i} \mu_i^{\,y_i}}{y_i!}, \quad y_i = 0,1,2,\cdots; \mu_i > 0 \tag{2.20}$$

Where, $\mu_i = \exp(X' \beta)$ is the fitted mean of the model, X is a vector of explanatory variables) and Y_i is the counts of the response variable. The Poisson distribution assumes equal mean and variance therefore they are both equal to μ .

If the response variable (Y_i) is a count that follows a Poisson distribution, the response Y_i will be modeled as follows with the probability of Y_i given by the equation below

$$f_i(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha \mu_i} \right) \frac{(1+\alpha y_i)^{y_i - 1}}{y_i!} \exp\left(\frac{-\mu_i (1+\alpha y_i)}{1+\alpha \mu_i} \right) \tag{2.21}$$

Where $y_i = 0,1,2,...;$

- $\mu_i = \mu_i(X) = e^{XB}$ where X is a $(k-1)$ dimensional vector of covariates and B is a k – dimensional vector of regression parameters
- α is the dispersion parameter. The parameter of dispersion is being observed on three cases (1) equi-dispersion; when $\alpha = 0$, and equation (2.21) reduce to PR, case (2) over-dispersion; when $\alpha > 0$ equation (2.21) will always sum to 1 and case (3) under-

dispersion; when $\alpha < 0$ equation (2.21) may not sum to 1 as the equation get truncated.

The variance and the mean of the response variable Y_i are given by

Variance: $$V(Y_i \mid x_i) = \mu_i (1+\alpha\mu_i)^2 \tag{2.22}$$

Mean: $$\mu_i = E(Y_i \mid x + i) \tag{2.23}$$

Other models may be considered that take excess zeros into account if $\alpha > 0$ or $\alpha < 0$. Boucher et al., 2007 stated that since Poisson distribution has some severe drawback that limits its use, other distributions can be used such as Generalized Negative Binomial Regression Model (Famoye and Singh, 2003).

2.1.4 Generalized Negative Binomial Regression Model

The Poisson distribution has only one parameter, whereas the negative binomial distribution has two parameters. The negative binomial regression is more flexible than the Poisson regression model. The Negative Binomial Regression (NBR) models have been used for data analysis in many research areas (Hilbe, 2011). Count data often exhibit substantial variables when the variance is larger than the mean (case of over-dispersion) or when the variance is less than the mean (case of under-dispersion) as mentioned earlier. Negative Binomial Regression models have been recommended to be used for over – dispersed data (Famoye and Singh, 2003) rather than Poisson regression. The Negative Binomial Regression model is used to model the response count data (Y_i) data as follows

$$f(y_i) = p(Y_i = y_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta) y_i} \left(\frac{\theta}{\theta + \mu_i} \right)^{\theta} \left(\frac{\mu_i}{\mu_i + \theta} \right)^{y_i}, y_i = 0,1,2,\ldots,n, n < \infty \qquad (2.24)$$

Letting $v_i = \dfrac{\theta}{\theta + \mu_i}$

We get

$$f(y_i) = p(Y_i = y_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\, y_i!} v_i^{\theta} \left(-(v_i - 1)\right)^{y_i} \qquad (2.25)$$

Where $\theta = \dfrac{1}{\alpha}$, α is the dispersion parameter, $\Gamma(.)$ is a gamma function, response (Y_i) has a NB distribution with the two parameters $\mu_i \geq 0$ and $\theta \geq 0$, where the mean and variance are given by

$$\text{Mean} = E(y_i) = \theta \mu_i \qquad (2.26)$$

And the

$$\text{Variance} = Var(Y_i) = E(Y_i)(1 + \mu_i) = \theta \mu_i (1 + \mu_i) \qquad (2.27)$$

However, even though the Poisson and Negative Binomial Regression models are recommended to model the count models, in the case of excess zero, the Zero Inflated models are also suggested. Because of the excess zero`s, there can be an over representation of zero cases in the estimated model (Poisson and Negative Binomial). This can mistakenly regarded as over dispersion in the data due to these excess zero`s (Chin and Quddus, 2003). Over dispersed model that incorporates excess zeros is recommended as alternative for a better fit. Famoye and Singh

(2006) and Lambert (1992) proposed models that can handle the presence of excess zeros in the count data namely Zero Inflated and Hurdle models.

2.1.5 Zero **Inflated Poisson Regression Model**

Zero inflated regression models are used to model count data that having excess of zero counts. It is modeled by the command ***zeroinfl***() in the package **pscl** (Jackman, 2008) in R software. The R codes used can be seen in the appendix at the back of this study report. The response variable $(Y_i = 0)$ with the probability γ_i is assumed to follow a Poisson distribution with mean μ_i and probability $1-\gamma_i$, $i = 0,1,2,...,n$, will have the a distribution of two components, zero $(y_i = 0)$ and non-zero component $(y_i \neq 0)$ given by

Zero case

$$\Pr(Y_i = 0) = \gamma_i + (1-\gamma_i)^{-\mu_i} \tag{2.28}$$

Non - zero Case

$$\Pr(Y_i = r) = (1-\gamma_i)e^{\frac{(-\mu_i)(\mu_i)^r}{r!}}, r = 1,2,3,.... \tag{2.29}$$

With the

Mean: $$E(y_i \mid x_i, z_i) = \mu_i \theta \tag{2.30}$$

and

Variance: $$V(y_i \mid x_i, z_i) = \mu_i (1-\gamma_i)(1+\mu_i \gamma_i) \tag{2.31}$$

Where $i = 1,2,3,...n$

In addition, to assess the impact of the explanatory variables on the response variable in a ZIP, γ_i and μ_i can be expressed explicitly as a function of the covariates. The standard method to model the probability of excess zeros is by using the logistic regression model as follow

$$\text{logit}\,(\gamma_i) = XB \tag{2.32}$$

Where, X is the vector of covariates (x_i) and B is a vector of parameters (β`s). Furthermore, the effect of the covariate on the response variable excluding the excess zeros can be modeled through the Poisson distribution

$$\log(\mu_i) = Z\delta \tag{2.33}$$

Where Z and $\mathbf{X}$ are s- and w- dimensional vectors of the covariates while δ and B are corresponding vectors of the regression coefficients respectively.

2.1.6 Zero Inflated Negative Binomial

The Zero Inflated Negative Binomial is a combination of distributions assigning the mass of $(1-\gamma)$ to a NB distribution and a mass of γ to excess zeros, with $0 \le \gamma \le 1$. The ZINB distribution is given by

$$P(Y_i = r) = \begin{cases} \gamma + (1-\gamma)\left(\dfrac{\theta}{\theta+\mu}\right)^{\theta}, r = 0\ (zero) \\ (1-\gamma)\dfrac{\Gamma(\theta+r)}{r!\Gamma(\theta)}\left(\dfrac{\theta}{\theta+\mu}\right)^{\theta}\left(\dfrac{\mu}{\mu+\theta}\right)^{r}, r = 1,2,...,n\ (non-zero) \end{cases} \tag{2.34}$$

With the

Mean: $E(Y) = (1-\gamma)\mu$ (2.35)

and

Variance $Var(Y) = (1-\gamma)\mu\,(1+\gamma\mu+\frac{\mu}{\theta})$ (2.36)

When $\frac{1}{\theta} \approx 0$ and, $\mu \approx 0$ then equation (2.34) reduces to the Poisson distribution. Similarly, equation (3.34) approaches the ZIP when $\theta \to \infty$ and approaches the NB distribution when $\gamma \to 0$. The two parameters γ and μ will be related to the covariates by the ZINB regression model as follows

$$\log(\mu_i) = XB \qquad (2.37)$$

$$\log it(\gamma_i) = Z\delta \qquad (2.38)$$

Where Z and X are s- and w- dimensional vectors of the covariates while δ and β are corresponding vectors of the regression coefficients respectively.

2.1.7 Poisson Hurdle Model

The most important application of the hurdle model is the hurdle at zero. A generalized hurdle model is a flexible model that can handle either under or over dispersed count data. Suppose dependent variable is a positive dichotomous variable, zero case $(Y_i = 0)$ or non- zero $(Y_i \neq 0)$, then hurdle model is suitable. Suppose that $h_1(0)$ is the probability value when there is a zero and that, $h_2(k)$, $k \geq 1$, is a probability function when there is a non – zero count, The probability function of the hurdle at zero model is given by

$$P(Y_i = k) = \begin{cases} h_1(0) & k = 0 \\ \left(1 - h_1(0)\right) h_2(k), & k \geq 1 \end{cases} \tag{2.39}$$

Let $Y_i\ (i = 1, 2, ..., n)$ be a nonnegative integer – valued and suppose that the zero case $(Y_i = 0)$ is observed from the event under study with a frequency significantly higher that can be modeled by the usual model. The Poisson Hurdle model for the non-zero cases of the response variable $Y_i\ (i = 1, 2, ..., n)$ is given by

$$P(Y_i = 0) = 1 - q_i, \quad 0 \leq q_i \leq 1 \tag{2.40}$$

$$P(Y_i = r) = q_i \frac{\mu^r e^{-\mu}}{r!(1 - e^{-\mu})}, \quad r = 1, 2, 3, ...n; \quad 0 < n < \infty \tag{2.41}$$

Where q_i models all the zero cases, μ is the mean for the truncated Poisson distribution. The probability of the zero cases can also be modeled by the logistic regression model:

$\log it(q_i) = XB$

2.1.8 Hurdle Negative Binomial model

Alternative to the Poisson hurdle is the two-part model, Hurdle Negative Binomial that can also decompose the response (y_i) into two observed random components: $y_i > 0$ and $Y_i \mid y_i > 0$ further specify the appropriate regression component to each part. The HNB is given by

$$P(Y_i = 0) = 1 - q_i \,, \quad 0 \le q_i \le 1 \tag{2.42}$$

$$P(Y_i = r) = q_i \left(\frac{\Gamma(r+\theta)}{r!\Gamma(\theta)}\right)\left(1+\frac{\mu}{\theta}\right)^{-r}, r = 1,2,3,\ldots n, \quad 0 \le \mu < \infty \tag{2.43}$$

Where, μ is the mean parameter and θ is the over-dispersion parameter. Therefore, these models needed to be explored so that the best model can be used for this study.

CHAPTER 3 METHODOLOGY

3.1 Research Design

The study depended on secondary data of regional road accidents that were collected on daily basis by the Motor Vehicle Accident Fund (MVA) and the National Road Safety Council (NRSC), for regions of the whole country. The study was a quantitative analytical study focusing on the number of reported road fatalities for the 2012 – 2013.

3.2 Population of the study

The researcher used accident data on all fatality cases on Namibian roads for the two-year calendar years, January 2012 – December 2013. The population of the study was all road accidents that were reported on Namibian roads for a two calendar year period. There were 7195 road accident cases reported and recorded.

3.3 Sample of the study

There was no sampling undertaken as the study used the secondary data of the recorded road fatality cases for the two calendar years. This is because all the data were readily available to the researcher and it was of great benefit towards accurate estimation of the parameters of interest. Therefore, the entire data set of 7195 road accident cases was used for analysis.

3.4 Data collection and Management

Namibia Road Accident Forms (NRAF) were used to collect the data from the accident scenes by the Namibian police. The forms are administered by the police officers on daily duties who recorded the particulars of the person(s) involved in the accidents within the 24 hours of the accidents occurrence, time of the day, month of the year etc. The number of fatalities recorded on this forms are the number of people who died on the spot of the accident or the number of people who died within 24 hours in hospital as a result of road accident. The completed forms are then handed over to the offices of the National Road Safety Council for record keeping and data analysis.

The completeness of the data was checked and it was found that some variables had incomplete data. After data cleaning, the following variables were retained for analysis: region, time, day, month, week, type and the number of people injured. Since it is the accident data, there might be some missing information such as the cause of accident, the type of accident crash or the time that the accident if for instance the accident took place at night, if there were no witness when the accident happened.

3.5 Ethical Considerations

The format of the data used for the analysis is anonymous. Arrangements were made with the NSRT and MVA to ensure confidentiality, that nobody else had access to the data. A letter seeking permission to collect data from NRSC and MVA was sought from the Head of Department of Statistics, at the University of Namibia to allow the researcher to use road crash data as the study involves people`s information.

3.6 Methods for Data Analysis

The statistical Package for Social Sciences (IBM SPSS) version 23.0 and R (i386 3.2.1) package were used to analyze the data. SPSS was used to run descriptive statistics while the R software was used to analyze the data by fitting the generalized linear models. The specific generalized models fitted were discussed in details in the sections to follow. The data were coded accordingly for analysis in R and the analysis was based on five percent level of significance. Dummy variables for variables namely: Day, Time, Type of crash, Month, Week of the month, and region were created for the generalized linear models analysis in R, as 1 for each specific variable in consideration or otherwise as 0.

The dependent variable of interest in this study was the number of road accident fatalities (RAFs) in Namibia over the two calendar years (January 2012 to December 2013) and six qualitative explanatory variables were considered; Day, Time, Type of crash, Month, Week of the month, region and one quantitative variable; number of persons injured as categorized in the table 3.2 below. The data were merged for the two years and sorted according to the dates of the month.

Table 3.1 Key variables description

variables	**Description**
Qualitative variables	
Month	Month in which the accident took place {1 = January; 2 = February; 3= March; 4= April; 5 = May; 6 = June; 7 = July; 8 = August; 9 = September; 10 = October; 11 = November and 12 =

	December}
Week	The week of the month when the accident took place: {1 = 1st week of the Month, 2 = Second Week of the Month, 3 = Third Week of the Month, 4 = 4th Week of the Month, 5 = 5th Week of the Month}
Day	Whether the accident took place during the weekday or weekend day { 1= Week day: [Monday, Tuesday, Wednesday, Thursday & Friday]} and 2= Weekend day: [Saturday & Sunday]}
Time	Time of the day when the accident took place: {1 = Morning [5am -11:59 pm], 2 = Afternoon [12pm - 17:59 pm], 3 = Evening [18pm - 20:59pm] and 4 = Night [21pm - 4:59am]}.
Region	The region in which the road accident took place: { 1 = Erongo, 2 = Caprivi, 3 = Hardap, 4 = Karas, 5 = Kavango, 6 = Khomas, 7 = Ohangwena, 8 = Omaheke, 9 = Omusati, 10 = Oshana, 11 = Oshikoto, 12 = Otjozondjupa and 13 = Kunene}
Type	The accident crash type: {1 = Head on collision, 2 = Roll over, 3 = Head side or Side wipe collision with another vehicle, 4 = Collision with other object (Pedestrian,hit and run, head side with other vehicle,fixed object) and 5 = Uknown}
Quantitative variable	
Injured	The number of person(s) injured in the accident

3.6.1 Descriptive Analysis

The descriptive statistics were used to profile the number of road crashes by injuries sustained as a result thereof. Furthermore, a frequency table was used to display the proportion of the number of road accidents by region, month, day, week, time and the crash type of the accident. These analyses provided the basic information about the number of fatalities in Namibia for the study period.

3.6.2 Generalized Linear Models (GLMs)

Predicting the value of the number of RAFs say Y according to the time of the day say X under the assumption that the number of RAFs and time of the day are linear, a simple linear model can be used. But, what if the variable the number of RAFs and time of the day are not are not linearly related? As for this study, what if the number of road accident fatalities and the explanatory variables (Time, Month, Day, Type, Week and region) are not related linearly? Generalized linear models (GLMs) are the generalization of the linear models. If the number of RAFs data is not normally distributed, the GLMs which assume a link linear relationship based on a chosen link function may be utilized to complete the analysis. The GLMs take the form

$$Y = g(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + e \qquad (2.43)$$

Where, Y is a vector of the number of observed RAFs; the X variables (Time, Type, Day, Month, Region and Week) are linearly associated covariates, β`s are the regression coefficients, $e$ is the error variability that cannot be accountable for by the predictor (X) variables, and $g(...)$ is a monotonic function relating the mean of the number of RAFs to the linear predictors. The values of the regression coefficients (β`s) were obtained by Maximum Likelihood (ML) estimations.

The **pscl** and **MASS** packages were loaded in R to fit the generalized linear models of the data. The Poisson regression model was fitted by command $glm()$ that is found under **MASS** package while the Negative Binomial regression was fitted by the command $glm.nb()$ under the same package in R. The Zero Inflated models and Hurdle models were found under the **pscl** package, fitted by the command $zeroinfl()$ and $hurdle()$ respectively. Six Generalized Linear regression models namely: Generalized Poisson, Generalized Negative Binomial, Zero Inflated Poisson, Zero Inflated Negative Binomial, Poisson Hurdle and Hurdle Negative Binomial were explored and adjudicated based on their AIC.

3.7 Goodness of fit statistics

In this study, all six overviewed models were fitted. All six models were estimated for the number of RAFs using the road accident data. The Akaike`s Information Criterion (AIC) has been extensively used as goodness of fit in statistical studies on accident crash data (Basyouny & Sayed, 2009, Miaou 1996, AKaike 1974). The AIC is used as a tool for model selection by comparison of models, the model with the smallest values of AIC was considered as the best model.

This study aimed at identifying the best model among the six generalized candidate models. The smaller the value of AIC of the model, the best is the model (Cafiso et al., 2009). The value of the AIC for the six models fitted in this study was calculated using the following formula (Akaike, 1973):

$$AIC = -2\log L + 2p \tag{2.44}$$

Where $logL$ is the maximum log-likelihood of the fitted model while p is the number of estimated parameters in the fitted model.

CHAPTER 4 RESULTS

This chapter presents the results of the analysis of road accidents data on Namibian roads from NRSC for the two calendar year, 2012 – 2013.The results were divided into four parts, namely: descriptive analysis of potential explanatory variables and number of road accidents fatalities, Spearman`s rank correlation between the number road accidents fatalities and the potential explanatory variables, and generalized linear models modelling the effects of the explanatory variables on the number of road accidents fatalities, and finally the selection of the best model fit of the data. All the modeling efforts followed the three phases: exploration, validation and fitting.

The descriptive statistics section profiled the background characteristic of the data in terms of frequencies and graphs. This was followed by the Spearman`s rank correlation to identify the potential variables that are related to the number of road accidents fatalities. The third part comprised modelling of the number of road accident fatalities taking into account of the excess zeros in the data, and hence the zero inflated and hurdle models were considered in this regard. This was followed by assessing the models based on the AICs and the best model was identified among the six generalized linear models.

4.1 Descriptive statistics

The descriptive statistics on the number of RAFs results indicated that the number of accident fatalities ranges from 0 to 11 fatalities. On average, 15 persons per 100 persons died per road accident over the two-year calendar, with a standard deviation of 52 persons per 100. The variance (Std.deviation² = 0.268) is greater than the mean which implies over-dispersion in the number of fatalities. The skewness (7.27) indicated a lack of symmetry in number of RAFs data as evidenced by figure 4.1.

Table 4.1: Descriptive statistics of the number of fatalities

Measure	Minimum	Maximum	Mean	Std.deviation	95% CI	Skewness	Kurtosis
Fatality	0	11	0.15	0.518	[0.14 , 0.16]	7.27	95.039

Figure 4.1: Distribution road fatalities per road accident in Namibia

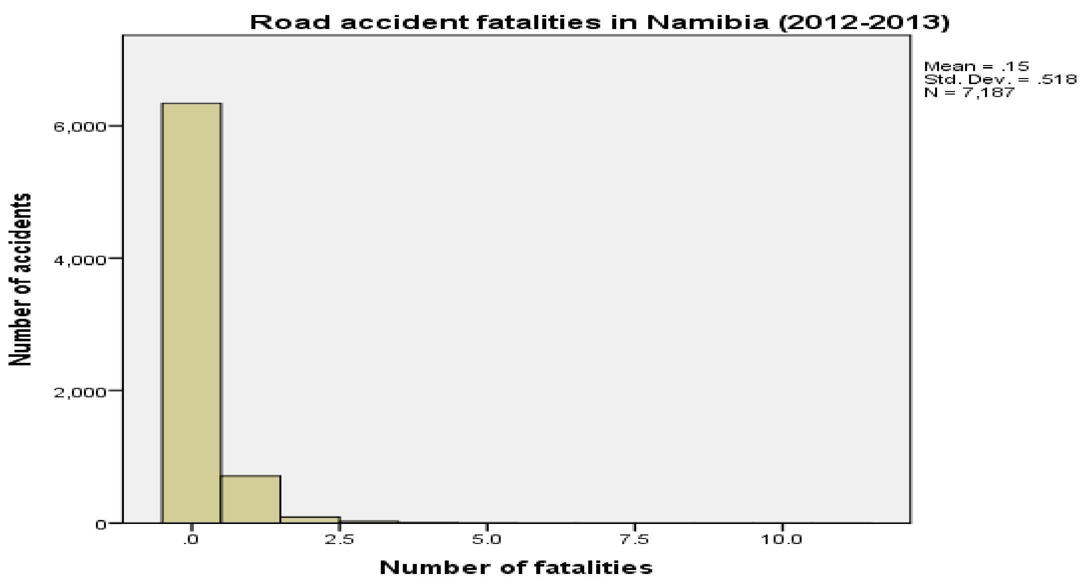


Table 4.2 below represents the frequency distribution of the total number of road accidents in Namibia per six independent variables; Week, month, region, Day, Type and Time of the day respectively. Most accidents happened in the fourth week of the month with 23.9 percent of the

total number of accidents that happen in a month. The high number of accidents on Namibian national roads from January 2012 to December 2013 happened in July and December which recorded 9.6 percent each.

Examining the question if road accidents issue could be approached from the regional level than national level, the regional frequencies of the number of accidents was assessed and the result are presented in figure 4.2. The region with the highest number of road accidents from January 2012 to December 2013 was the Khomas region, which accounts for 37 percent of the total national road accidents, followed by Erongo region which accounts for 11.1 percent. It was also of great interest to know which days of the week the road users' needs to be extra careful when driving. On average, 13 percent of the total accidents took place in a weekday while 17.5 percent on a weekend day, as a week have 5 days while the weekend constitute of only 2 days. The result shows that on average, most accidents happen during the weekend with 35 percent of the total accidents that occurred over the period of two years, 2012 to 2013. Lastly, with respect to the type of accident collision, collision with other object (pedestrian, hit and run, head side with other object...) which accounts for 53.5 percent while the side wipe collision was the least among the type of collision which accounts for 5.2 percent of the total known crash type.

Table 4.2: Summary of the frequency distribution of the total number of road accidents per six potential variables

Variable	Frequencies	Percent (%)
WEEK of the month		
1st week of the week	1494	20.8

2nd week of the month	1537	21.4
3rd week of the month	1577	21.9
4th week of the month	1688	23.9
5th week of the month	899	12.5
	7195	100
Month		
January	490	6.8
February	538	7.5
March	619	8.6
April	554	7.7
May	576	8
June	622	8.6
July	691	9.6
August	650	9
September	613	8.5
October	597	8.3
November	555	7.7
December	690	9.6
Region		
Erongo	802	11.1

Caprivi	139	1.9
Hardap	376	5.2
Karas	233	3.2
Kavango	333	4.6
Khomas	2660	37
Ohangwena	350	4.9
Omaheke	170	2.4
Omusati	231	3.2
Oshana	641	8.9
Oshikoto	454	6.3
Otjozondjupa	638	8.9
Kunene	168	2.3
Day		
Weekend	2516	35.0
Weekday	4679	65.0
Type		
Head on collision	667	9.3
Roll over	1632	22.7
Side wipe collision	374	5.2
Collision with other object	3851	53.5

{Pedestrian, hit and run, head side with other vehicle, fixed object,..}		
Unknown	170	2.4
Total	6694	93.0
Missing System	501	7.0
Time		
Morning	2367	32.9
Afternoon	1392	19.3
Evening	1451	20.2
Night	1313	18.2
Missing system	672	9.3

Figure 4.2 Frequency distribution of the total number road accidents per region over a two calendar year of study

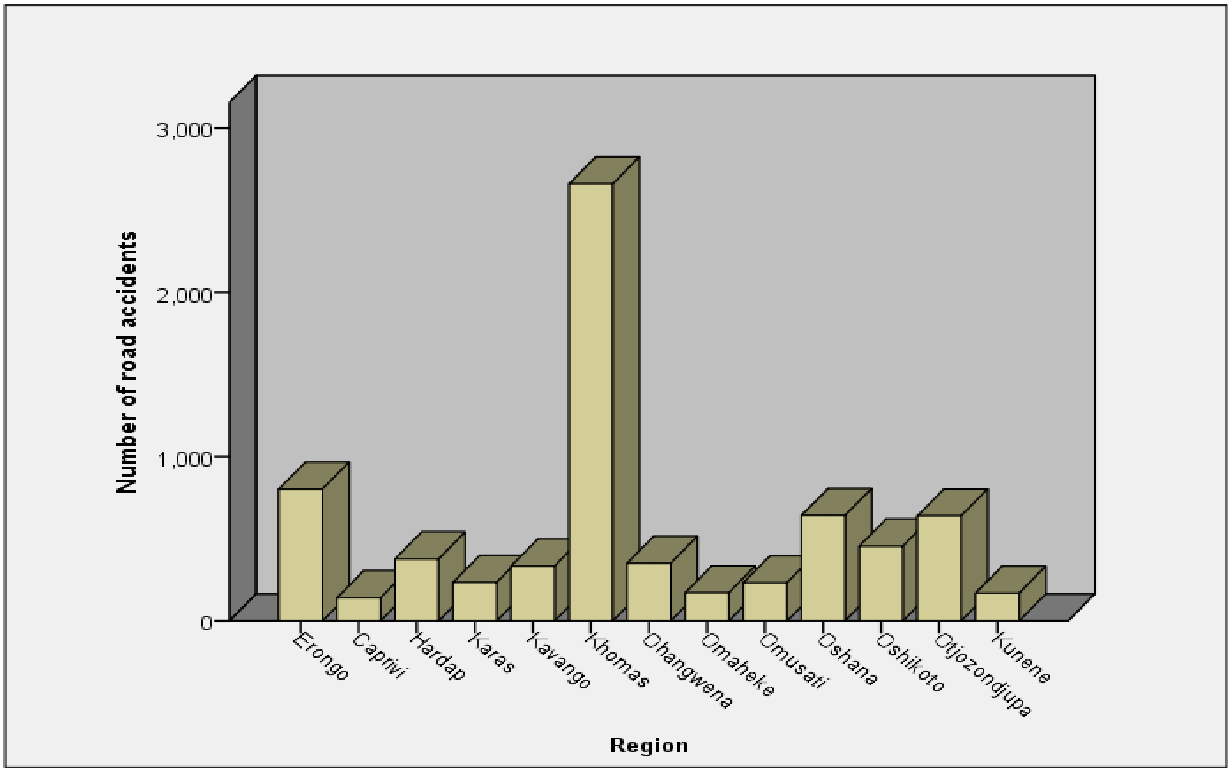
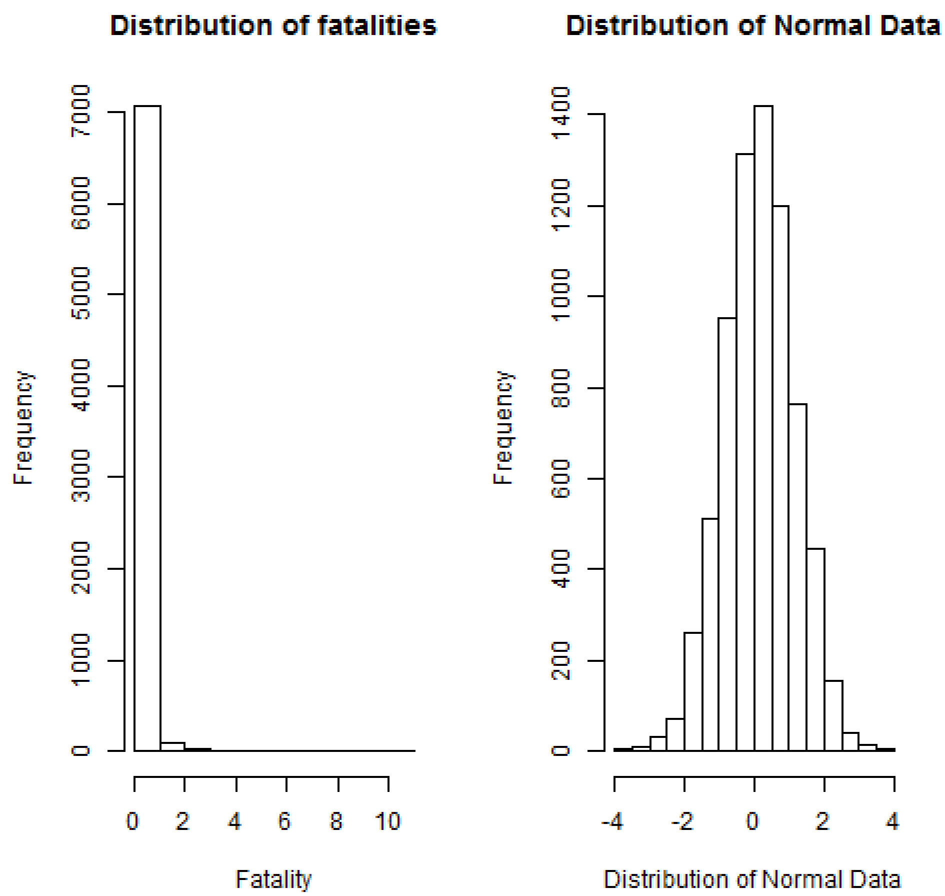
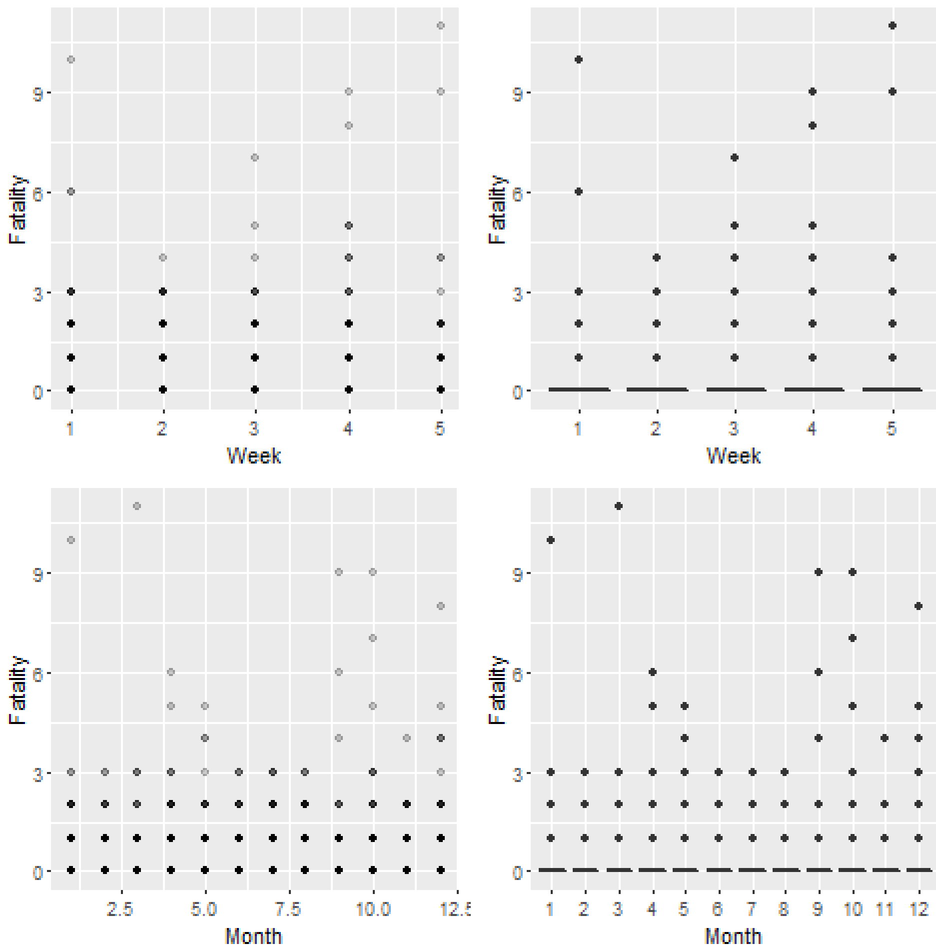


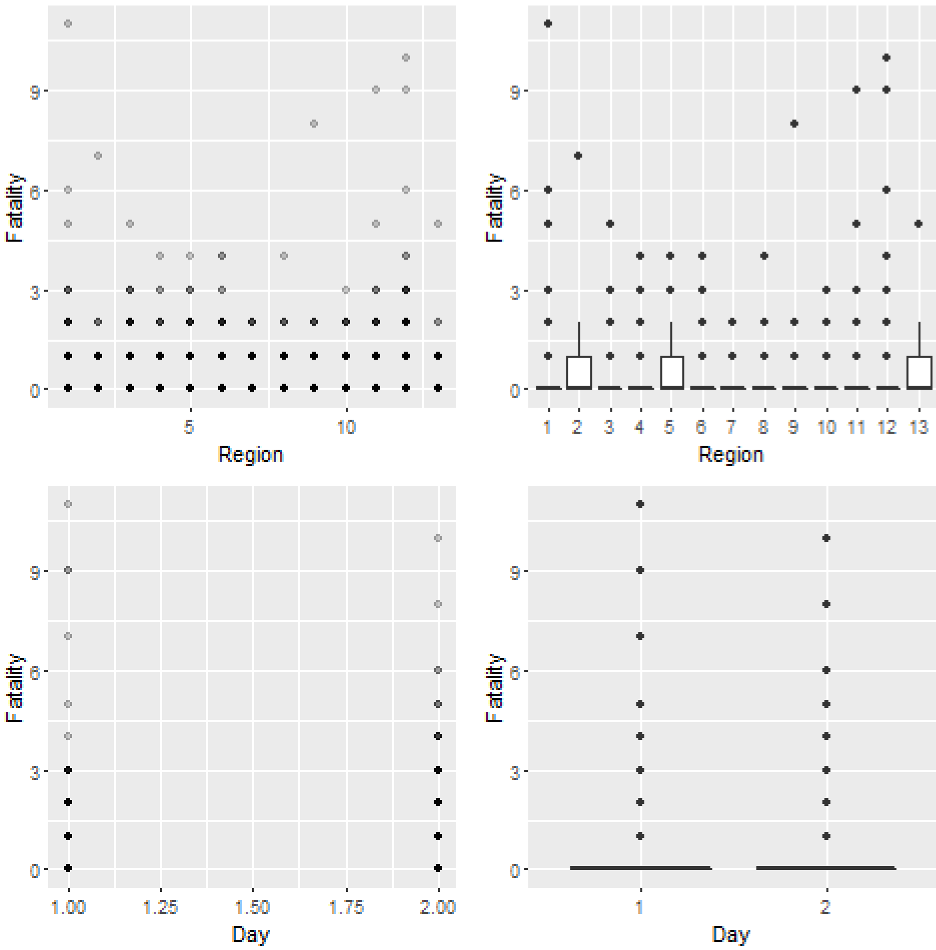
Figure 4.3 Frequency distribution of the number road fatalities versus distribution of normal data

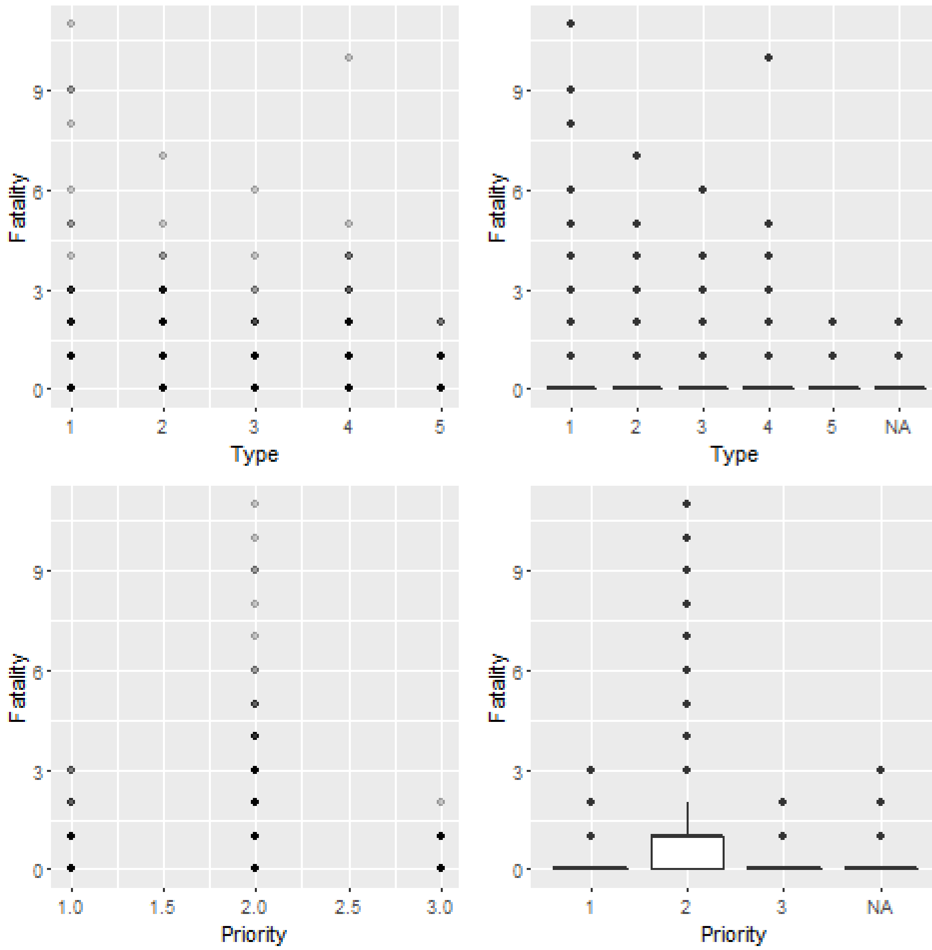


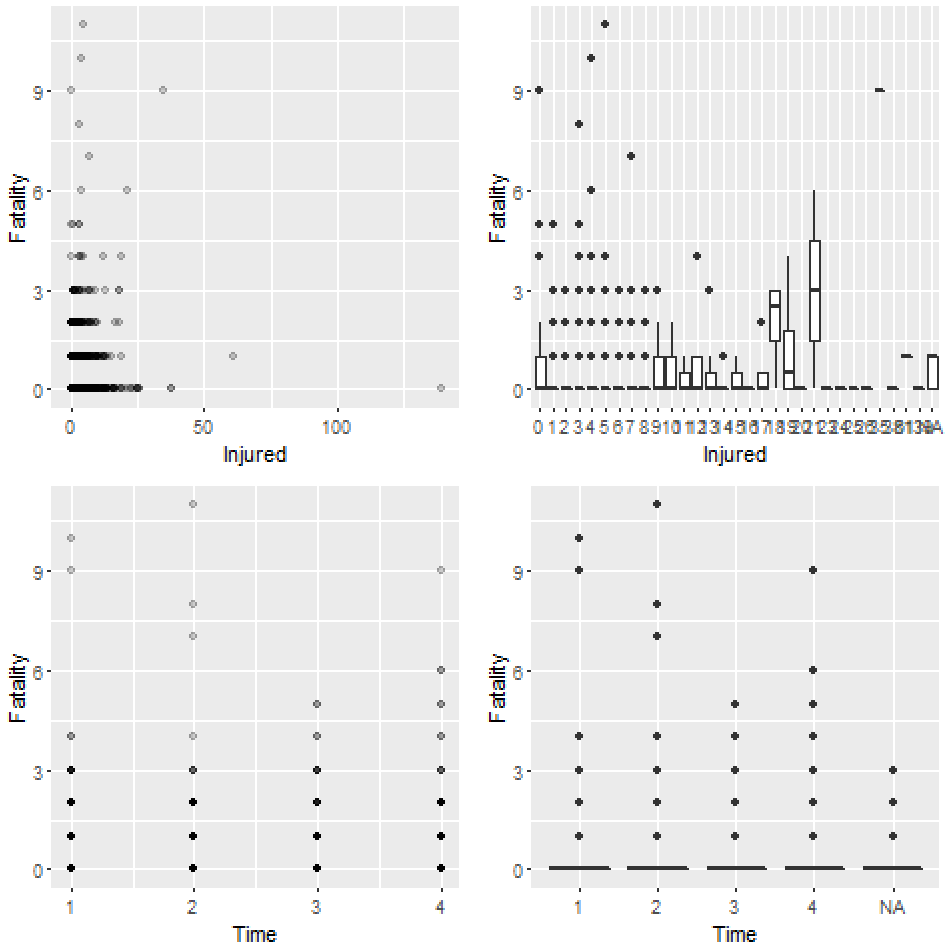
From figure 4.3 the skewed nature of the data makes it difficult to see any differences in the distributions of the number of fatalities. By examining the shape of this distribution the histogram is right skewed and doesn't look normal..

Figure 4.4 Relationship between each covariate and the number of fatalities




Fatality
0
3
6
9
5
10
Region
Fatality
1 2 3 4 5 6 7 8 9 10 11 12 13
Region
Fatality
1.00
1.25
1.50
1.75
2.00
Day
Fatality
1
2
Day


Fatality
Type
Fatality
Type
NA
Fatality
Priority
Fatality
Priority
NA



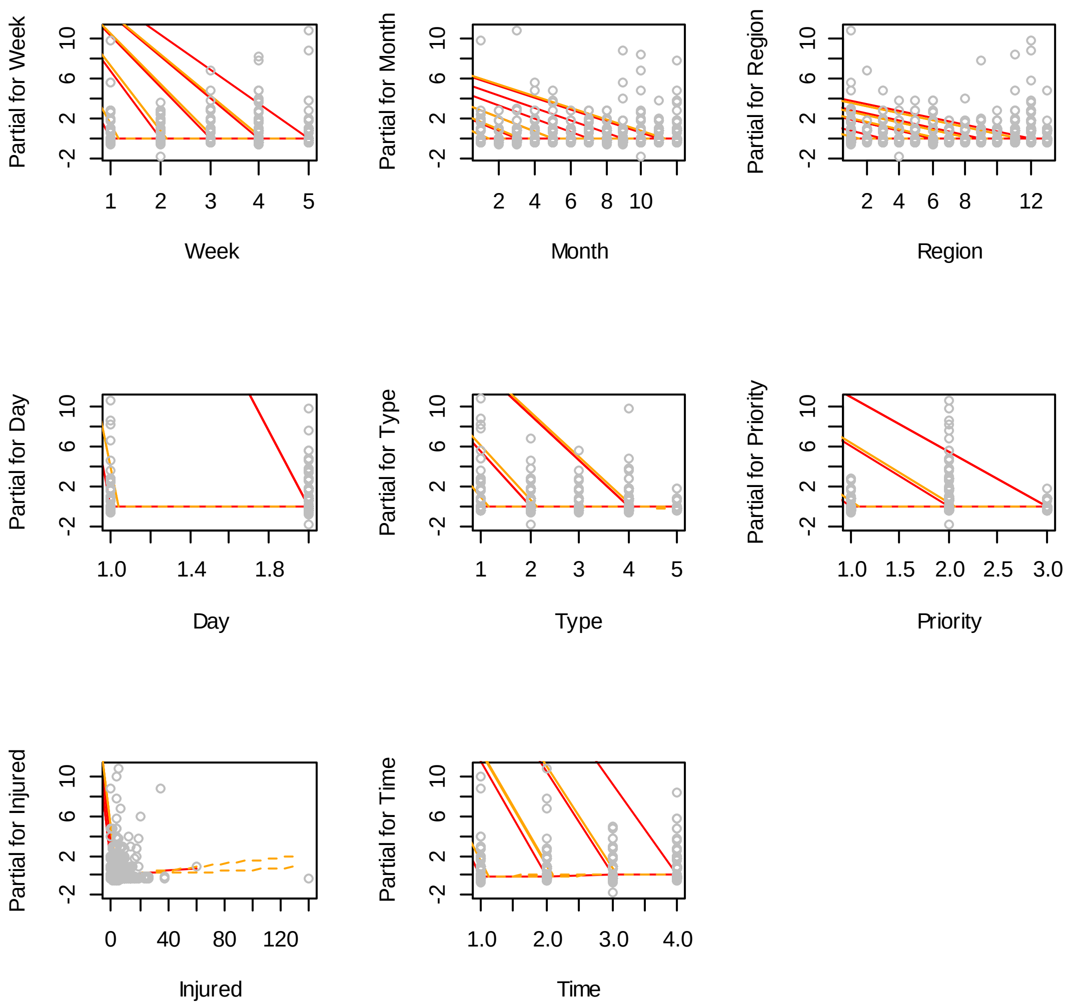
The data was further plotted using the boxplot in ggplot to assess the relationship between the covariates and the number of fatalities. The results showed that the number of zeros in the fatalities heavily influenced all the plots. The highest number of fatalities was recorded for March and October as shown in figure 4.4 above. In terms of the weeks of the month, the fourth and fifth week of the month have the biggest number of fatalities. Figure 4.4 further revealed the highest number of fatalities in Erongo, Omusati, Oshikoto and Otjozondjupa region. It further showed that there was high number of fatalities during the weekday as well as weekend days during morning and night hours. However, days, priority and the number of injured persons don't show a very strong relationship with fatality but time, type, day and priority show possible non-linear relationships with the number of fatalities. This was further asses with the linear regression models, simple and multiple linear regressions.

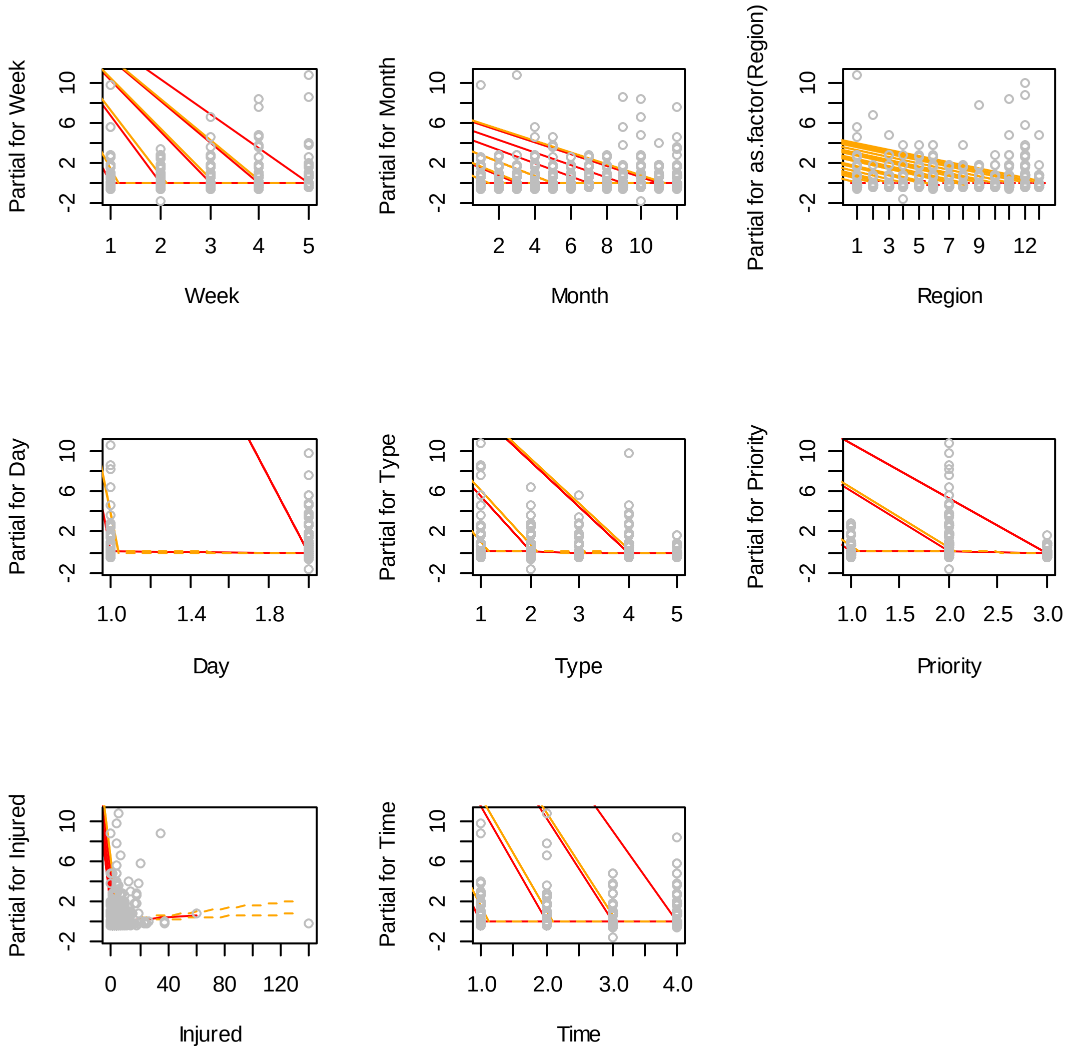
4.2 Linear Regression assumption checking

The four model assumptions namely; linearity, constant error variance, residual independence and normality of residuals were checked to see if the conclusions of the models fitted are to be believed.

- **Assessing linearity**. Partial residual plots for the full model and the model 3 with Region as a factor were plotted respectively below to assess if the continuous covariates in the model are appropriately modelled as linear. Owing to the size of the partial residuals, it is difficult to assess whether the linearity relationship is reasonable for the continuous covariates. However, Injured and Month seems to be showing some linearity

Figure 4.5 Partial Residual plot for linearity assessment for a full model





Constant Error Variance assumption testing:

Full Model: $Y_{it} \ = \ \beta_0 + \beta_1 Week + \beta_2 Month + \beta_3 Day + \beta_4 Type + \beta_5 Time + \beta_6 Region$

H_0 = Constant error variance

H_a = Not all variances are equal

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 2895.16   Df = 1   **p = 0**

α **= 0.05**

Since the p-value (p = 0) is less than 0.05, we reject the null hypothesis and therefore homoscedasticity assumption is violated.

- **Residual Independence**

H_0 = Residuals are independent

H_a = Residuals are not independent

Durbin Watson test

lag	Autocorrelation	D-W Statistic	**p-value**
1	0.6882123	0.619086	**0**

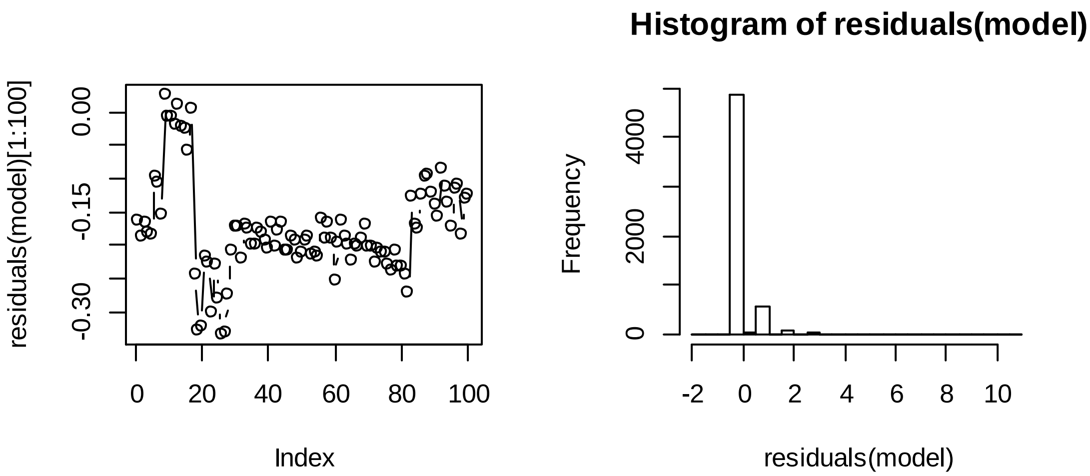
Alternative hypothesis: rho != 0

α **= 0.05**

Since the p-value (p = 0) is less than 0.05, we reject the null hypothesis. Therefore, the independence of residuals assumption is violated.

- **Normality of Errors**

Figure 4.6 Distribution of normality of errors for the full model



The histogram of residuals of the full model above shows that, the residual is right skewed (tail on the right), therefore the errors are not normally distributed. The assumption of normality of errors is also violated.

Model 3: $Y_{it} = \beta_0 + \beta_1 Week + \beta_2 Month + \beta_3 Day + \beta_4 Type + \beta_5 Time + \beta_6 Region(as\ factor)$

Constant Error Variance assumption testing:

H_0 = Constant error variance

H_a = Not all variances are equal

Breusch-Pagen test

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 3193.253   Df = 1   **p = 0**

α = 0.05

Since the p-value (p = 0) is less than 0.05, we reject the null hypothesis and therefore homoscedasticity assumption is violated

durbinWatsonTest(model3)

lag	Autocorrelation	D-W Statistic	p-value
1	0.6696777	0.6561877	0

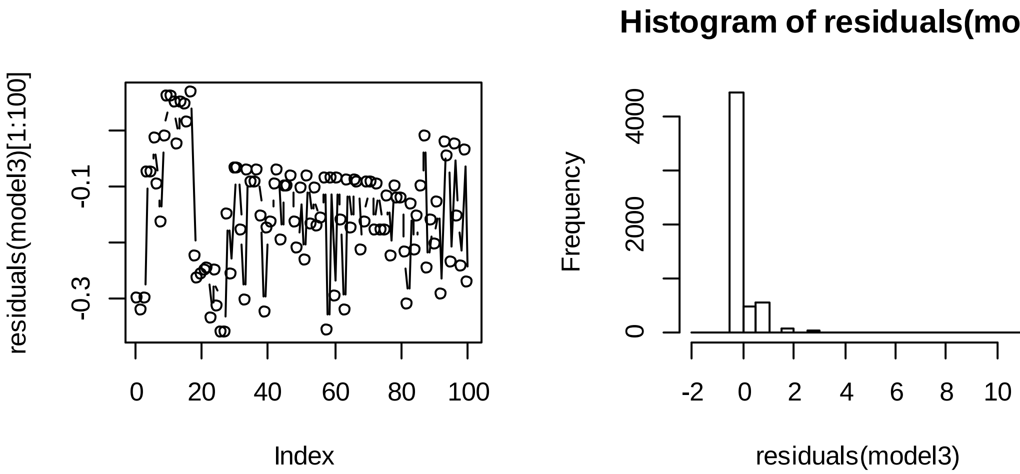
Alternative hypothesis: rho != 0

α = 0.05

Since the p-value (p = 0) is less than 0.05, we reject the null hypothesis. Therefore, the independence of residuals assumption is violated.

- **Normality of Errors**

Figure 4.7 Distribution of normality of errors for model 3 (with region as a factor)



The histogram of residuals of model 3 above shows that, the residuals are right skewed (tail on the right), therefore the errors are not normally distributed. The assumption of normality of errors is also violated.

Therefore, both simple and multiple linear regression models that could have been considered to be best in explaining the relationship between the number of fatalities and the covariates did not satisfy the homoscedasticity, independence of residuals and normality of errors assumptions of linear regression. We cannot therefore validate conclusions inferred from linear models, and this justifies a need of generalized linear models.

4.3 Generalized Linear Models

Predicting the value of the number of RAFs (Y) by changing the values for the explanatory variable (X), under the assumption that the number of RAFs and potential explanatory variable are linear in future, generalized linear models (GLMs) are the generalization of the linear models. Since all the assumptions of the linear models were violated by RAFs data, the GLMs which assume a link linear relationship based on a chosen link function were considered to complete the analysis.

The six generalized linear models, PR, NB, ZIP, ZINB, HP, and HNB were performed on the data in R. The AIC for each model was also calculated for comparison and identifies the best model. The ZINB came out as the best model in fitting the data and therefore the outputs for ZINB were considered for interpretations. The R commands that were used for the analysis and outputs are attached in the appendix. The table 4.4 below shows the AIC values of the six GLMs:

Table 4.3 Akaike Information Criterion value of the six generalized models

Model	**AIC**
Poisson	5778.6
Negative Binomial	5501.2
Zero Inflated Poisson	5481.4
Zero Inflated Negative Binomial	5438.9
Hurdle Poisson	5446.5
Hurdle Negative Binomial	5447.5

Based on the Akaike Information Criteria's (AIC) of the six GLMs that was performed on the data, ZINB was identified as the best model in fitting the fatality data, with the smaller value of the AIC (5438.9) compared to Poisson Regression (AIC = 5778.6), Negative Binomial

Regression (AIC = 5501.2), Zero-Inflated Poisson (AIC = 5481.4), Hurdle Poisson regression (AIC = 5446.5) and Hurdle Negative Binomial regression (AIC = 5447.5) value.

Table 4.4 Zero Inflated Negative Binomial output in R

			Zero inflated negative binomial regression				
Count Model coefficients					Zero-Inflation model coefficients		
	β	Pr(>|z|)			β	Pr(>|z|)	
Intercept	-1.46536	6.28E-02	**		-3.0772	0.13601	
Weekend	0.01291	0.92263			-0.3865	0.31479	
Erongo	-0.50492	1.29E-01			1.4814	0.20406	
Caprivi	0.60093	0.06418			4.4637	0.2247	
Hardap	-0.20209	0.49954			0.9104	0.44281	
Karas	-0.17882	0.60262			0.8664	0.53513	
Kavango	0.44159	0.09484			1.0804	0.31557	
Khomas	-0.95211	0.0204	*		2.4583	0.02726	*
Ohangwena	-0.32634	0.26447			0.9762	0.35007	
Omaheke	-0.54745	0.09271			-0.4094	0.78061	
Omusati	-0.04046	0.89785			0.9487	0.37783	
Oshana	-0.2389	0.040913			1.514	0.15174	
Oshikoto	-0.19607	0.53248			1.0512	0.31553	
Otjozondjupa	-0.07319	0.80292			0.9311	0.37815	
Morning	-0.15097	3.86E-01			1.2826	0.00148	**
Afternoon	-0.05431	0.73044			0.9482	0.0223	*
Evening	-0.22133	2.62E-01			1.2845	0.00485	**
Head on Collision	2.1394	2.77E-05	***		2.4861	0.14965	
Roll Over	0.74873	1.21E-01			0.4804	0.77052	
Side wipe collision	1.52958	0.00795	**		2.0666	0.23946	
Collision with other object{Pedestrian..}	0.12768	7.98E-01			-0.8195	0.65799	
log(theta)	0.36604	2.10E-01					

Count model coefficients

The Zero-Inflated Negative Binomial regression model revealed that, a value of β = -0.50492 indicating a magnitude decrease change of 0.60 in Erongo, a value of β = -0.20209 indicating a magnitude decrease change of 0.82 in Hardap, a value of β = -0.17882 indicating a magnitude decrease change of 0.84 in Karas, a value of β = -0.95211 suggests a magnitude decrease change of 0.39 in Khomas, a value of β = -0.32634 suggests a magnitude decrease change of 0.72 in Ohangwena, a value of β = -0.54745 suggests a magnitude decrease change of 0.58 in Omaheke, a value of β = -0.04046 suggests a magnitude decrease change of 0.96 in Omusati, a value of β = -0.2389 suggests a magnitude decrease change of 0.79 in Oshana, a value of β = -0.19607 suggests a magnitude decrease change of 0.82 in Oshikoto and a value of β = -0.07319 suggests a magnitude decrease change of 0.93 say one fatality for a unit change in the covariate compared to Kunene region. However, a value of β = 0.60093; 0.44159 suggests a magnitude increase change of 1.82 and 1.56 in Caprivi and Kavango respectively say one fatality for a unit change in the covariate. It was also found that the magnitude increase in all types of collision with the head on and head side collisions being more significant contributors. If the accident crash is a head on collision, a value of β = 2.1394 suggests a magnitude increase change of 8.49 while a value of β = 1.52958 suggests a magnitude increase change of 4.61 say one fatality for a unit change in the covariate. With the roll over and collision with other object **(**Pedestrian, hit and run, head side with other vehicle, fixed object..,), one fatality for a unit change in the covariate, a value of β = 0.74873 suggests a magnitude increase change of 2.11 and a value of β = 0.12768 suggests a magnitude increase change of 1.14 compared to unknown collisions respectively.

With time of the day, the ZINB results shows that a value of β = -0.15097 suggests a magnitude decrease change of 0.86 in the morning {5:00am – 11:59am}, a value of $\beta = 0.05431$indicating a magnitude decrease change of 0.95 in the afternoon {12:00pm – 17:59pm}, while a value of β = -0.22133 suggests a magnitude decrease change of 0.80 in the evening {18:00pm – 20:59pm} compared to night hours. The expected number of fatalities during the weekend is 1.01299 times the expected number of fatalities during weekdays holding all other factors constant.

Zero Inflation model components

The Zero-Inflated Negative Binomial component results revealed that, the baseline odds of accidents with no fatalities is 0.05. The odds of number of fatalities decreasing by one is 0.68 if the accident happens during the weekend, 0.66 if the accident happens in Omaheke region and with 0.44 if the accident occurred as a result of collision with other object. If the accident took place in other regions, the odds of number of fatalities increasing by one is 4.41 in Erongo; 86.8 in Caprivi; 2.49 in Hardap; 2.38 in Karas; 2.95 in Kavango, 11.68 in khomas; 2.65 in Ohangwena; 2.58 in Omusati; 4.54 in Oshana; 2.86 in Oshikoto and 2.54 in Otjozondjupa however it is worth noticing the only significant coefficient is that of Khomas region. Moreover with the variable time of the day taken in consideration, the odds that the number of fatalities would increase by one are: 3.61 during morning and evening hours while 2.58 during afternoon hours and all those statistics are significant. Both Head on collision, Roll over and Side wipe collision increase the odds of zero fatality but they are non-significant.

CHAPTER 5 DISCUSSIONS OF RESULTS AND CONCLUSIONS

This chapter discusses the findings of the study as presented in the previous chapter. It is divided into three sections namely; the descriptive statistics, linear regression models and the Generalized Linear models. The chapter relates the outcome of the study to the outcomes of the other related studies as discussed in the Literature Review.

5.1 Descriptive Statistics

The result showed that most accidents happen in the fourth week of the month. This can be due to month end and most people are paid towards the end of the month, hence the high number of traffic movement since people are travelling from place to place or town to town for shopping and other purposes. Another reason could be the use of alcohol which always increased on month ends because of pay days. These findings agreed with that of Farmer & Williams (2005) in their study of temporal factors in motor vehicle crash deaths in the United states, that one of the factor that affects the distribution of fatalities is alcohol and that, day 1 of January and day 31 of October were found with the highest number of pedestrian crash deaths over the 16 years, 1986 – 2002 with the average of 24 deaths per day. Insurance Institute for Highway Safety Highway Loss Data Institute report (2014) also found out that July 4, January 1 and August 2 were among the top five days with most Motor vehicle crash deaths over a period of years 2010 to 2014 in America. A further study is recommended for the contributory factors to high number of accidents on the Namibian national roads on the fourth week of the month.

The result also showed that high number of accidents on Namibian national roads in January 2012 to December 2013 happened in July and December that appeared to be equal with 9.6 %. Well, it is not surprising about December as most people travel long distances to celebrate the festive season with their families which results in high traffic density on national roads. Furthermore, December falls in a rainy season in Namibia therefore poor visibility the less friction on car tires since the roads are wet due to rainfall. It is somehow surprising about the high number of accidents happened in July with 9.6 %. There is no specific event in July in the country, no school holidays nor public holiday. These findings correspond with that of Farmer & Williams (2004) and El-Basyouny et.al. (2014) that most fatalities (averaging 120 to 132 deaths per day) were the summer and fall months, June through December. Fewer accidents happened in January over the same period with 6.8 % which agree with the finds of Farmer & Williams (2004) who found that January and February averaged the lowest number of deaths per day. This is worrisome as well as one expected high number of accidents in January as most people travels back to work and schools starts within those months in the country.

The region with highest number of road accidents from January 2012 to December 2013 with a tremendous frequency of 2660 (37 %) was Khomas region, followed by Erongo region with the number of accident of 802 (11.1 %). Khomas region has the highest population density in Namibia as per 2011 census and with the capital city of Namibia where the traffic flow is high. It could be due to this reason that the number of accidents is high in Khomas region. The region with the lowest number of accidents that happened over the same period of two years was Caprivi region with 139 (1.9 %) of the total accidents.

Furthermore, most accidents happen during the weekdays with 65 % of the total accidents that occurred over the period of two years, 2012 to 2013. This may result as the number of weekdays is five days compared to weekend days which are just two days. Therefore, on average, the number of accidents that occur per weekday is 936 (13%) while per weekend day is 1259 (17.5) accidents per weekend day. Therefore, on average, more accidents happen during the weekend than weekdays. These findings are in agreement with that of Nangombe & Neema (2012), in their study of the Statistical analysis of road traffic fatalities in Namibia from 2007 to 2009. On the effect of the day of the week on fatality in a vehicle collision in Namibia, they found out that, Sunday and Saturday was recording the highest number of fatalities with 147 and 146 respectively.

The results also show that 50.2 % of the total accidents that happened during January 2012 to December 2013 were severe while 15.9 % are slight injuries. This is not a good indication as severe injuries may lead to fatality or disabilities that affect the productivity of a human being in his/her family and country at large. The impact that results between vehicles leaving the road and solid roadside objects such as trees, poles and road signs is a major road safety problem worldwide. The type of crash that resulted in the high number of fatalities is collision with other object (pedestrian, hit and run, head side with other object...) with the frequency of 3851 (53.5 %) while the side wipe collision was the least among the type of collision with 374 (5.2 %) of the known crash type. This may due to the fact that, when the car is in motion, the collision on any object will result in high impact that may result in more severe injury. The study also found that most road accidents happens during the mornings hours {05:00 am – 11:59 am}, followed by evening hours {18:00 pm – 20:59 pm}. The results correspond with findings of the previous years on road fatalities in Namibia by Nangombe & Neema (2012), who found that evening hours contributes to the highest number of fatalities every day. These results however partly correspond with that of Farmer & Williams (2005) and Insurance Institute for Highway Safety

Highway Loss Data Institute report (2014) who found out that most crash deaths were high during afternoon and evenings hours.

5.2 Modeling of the number of RAFs by various GLMs

The main objective of this study was to assess and model factors that contribute to Road Accident Fatalities on Namibian roads in an effort to provide useful information that can be used for national prevention strategies and interventions to minimize road accidents fatalities. This study considered several models to quantify factors associated with the number of RAFs in Namibia based on the model`s AIC. Figure 4.1 shows that there were excess zeros in the accident data. The results of the study found out that best model in modeling the number of fatalities in Namibia is the Zero Inflated Negative Binomial. ZINB models have been suggested by numerous researches, namely Chipeta et al. (2014) in their study of Zero adjusted models with applications to analyzing helminths count data found out that ZINB was one of the best models in analysing count data with excess zeros. Furthermore, Hu et al. (2012) in their study of Zero-Inflated and Hurdle Models of Count Data with extra zeros on HIV –risk reduction intervention trial founds out that Zero-Inflated negative binomial model provided the best fit. Sharma & Landge (2013) also proposed ZINB as the best model that established an empirical relationship between heavy traffic accidents, highway geometric and traffic parameters

5.4 CONCLUSION

The main objective of this study was to assess and model factors that contribute to road accident fatalities on Namibia roads. The results show that the ZINB is the best model to fit the factors contributing to the high number of Road Accident Fatalities in Namibia. The results further showed that, the number of Road Accident Fatalities in Namibia increased towards month ends and head on crashes contributed to more fatalities. Most accidents took place in Khomas region which houses the capital city followed by Erongo region. Furthermore, weekends contributed to high number of fatalities than the week days. The riskiest time of the day to travel is morning {5:00 am -11:59 am} and evening hours {12 pm - 17:59 pm} as those hours of the day contributed with of total percentage of 53.1 to the number of road accidents.

Recommendations

There is a need of further in-depth studies on each of the contributing factors to the number of road accident fatalities in Namibia, especially in Khomas region, weekend days and mornings hours of the day. However, drivers are advised to be extra careful during pick up hours more especial morning hours and weekend days. More serious actions need to be taken, either by introducing frequent road safety campaigns in Khomas region on road safety campaigns and probably many cameras around the region. With regards to the limitation encountered in the study, training or induction courses on how to complete NRAFs forms and the importance of fully completed of these forms must be provided to all police officer joining the force as there are a lot of missing cases in the data for two years (2012 – 2013). More research in this area is needed as the data in this study has some major limitations, no information on age and gender given, which is critical to understanding of road traffic fatalities. The researcher further recommend spatial analysis which would further improve the understanding of road traffic fatalities in Nmaibia. Lastly, let road safety be a responsibility of all individuals.

REFERENCES

Abbassi, A. H. (2005) Statistical methods for predicting fatalities of Road Traffic Accidents in Kuwait. *Egyptian Population and Family Planning Review*, 2005: 2 – 19.

Abdel-Aty, M., Ekram, A – A., Huang, H., & Choi, k. (2011). A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis and Prevention*, 43: 1730 – 1737.

Abdel-Aty, M., Ekram, A-A., Huang, H., & Choi, K. (2011). A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis Prevention*, 43: 170 – 1737.

Abegaz, T., Berhane, Y., Worku, A., Assrat, A., & Assefa, A. Effects of excessive speeding and falling asleep while driving on crash injury severity in Ethiopia: a generalized ordered logit model analysis.

Afukaar, F. (2003). Speed control in developing countries: issues challenges and opportunities in reducing road traffic injuries. *Injury Control and Safety Promotion*: 10: 77 – 81.

Aguero- Valverde, J. (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis and Prevention*, 50: 289 – 297.

Ahmed, M. M., Abdel-Aty, M., & Yu, R. (2012). Assessment of Interaction of Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather and Traffic Data. *In Transportation Research Record: Journal of the Transportation Research Board*, No. 2280, Transportation Research Board of the National Academics, Washington D.C: 51 – 59.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. Second International Symposium on Information Theory, 1:267-281.

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on Automatic Control, 19(6): 716 – 723.

Andrey, J. (2010). Long Term Trends in Weather – Related Crash Risks. *Journal of Transport Geography*, 18: 247 – 256.

Barron, D. N. (1998). The Analysis of Count Data: Over dispersion and Autocorrelation. *Sociological Methodology,* 22: 179-219.

Basyouny, E. K., & Sayed, T. (2009). Accident prediction models with random corridor parameter. *Accident Analysis & Prevention,* 41(5): 1118-1123.

Bergel, R., Debbarh, M., Antoniou, C., & Yannis, G. (2013). Explaining the Road Accident Risk: Weather Effects. *Accident Analysis Prevention*, 60: 456 – 465.

Brijs, T., Karlis, D., & Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis Prevention,* 40: 1180 – 1190.

Brijs, T., Karlis, D., Van den Bossche, F., & Wets, G. (2007). A Bayesian model for ranking hazardous road sites. *Journal of the Royal Statistics Society*: Series A (Statistics in Society), 170(4): 1001 – 1017.

Cafiso, S., & Di Silvestro, G. (2011). Performance of safety indicators in edentification of black spots on two-lane rural roads. *Transportation Research Record*, 2237: 78 – 87.

Cafiso, S., Di Graziano, A., Di Silvestro, G., & Persaud, B. (2009). Development of comprehensive accident models for two – lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis and Prevention.*

Cameron, A., & Trivendi, P.K. (1990). Regression – based tests for over dispersion in the Poisson model. *Econometrics*, 46: 347 – 364.

Chin, H.C., & Quddus, M. A. (2003). Modelling Count Data with Excess Zeros. *An Empirical Application to Traffic Accidents,* 32(1): 90-116

Chipeta, M. G., Ngwira, B. M., Simoonga, C., & Kazembe, L. N. (2014). Zero adjusted models with applications to analysing helminths count data. BMC Research Notes. http://www.biomedcentral.com/1756-0500/7/856

Cho, J., Lee, H., & Yoo, W. (2006). A wet-road braking distance estimate utilizing the hydroplaning analysis of patterned tire. *Int. J. Numer. Methods Eng* 69, 1423 – 1445.

Chung, Y.-S. (2013). Factor complexity of crash occurrence: An empirical demonstration classification trees and random forests. *Accident Analysis and Prevention*, 41: 98 -

Consul, P. C. (1989). Generalized Poisson Distributions: Properties and Applications. Maker Dekker Inc., New York.

Consul, P. C., & Famoye, F. (1992). Generalized Poisson Regression Model. *Communication in Statistics – Theory and Methods*, 21(1): 89 – 209.

Dean, C., & Lawless, J.F. (1989). Tests for detecting over dispersion in Poisson regression models. *American Statistical Association*, 84: 467 – 472.

Dissanayake, S., & Lu, J.J. (2002). Factors influential in making an injury severity difference to older drivers involved in fixed object-passenger car crashes. *Accident Analysis and Prevention,* 34: 609 – 618.

Ehsan, S., & Adnan, R. (2011). Zero-Inflated Negative Binomial Regression Model with Right Censoring Count Data.

Eisenberg, D. (2004). The Mixed Effects of Precipitation on Traffic Crashes. *Accident Analysis and Prevention,* 36: 637 – 647.

Eisenberg, D., & Warner, K.E. (2005). Effects on Motor Vehicle Collisions, Injuries, and Fatalities. *American Journal of Public Health*, 95(1): 120 – 124.

El- Basyouny, K., Islam, T. M., Barua, S., & Li, R. (2014). Assessing the effect of weather states on crash severity and Types by use of full Bayesian Multivariate Safety Models.Transportation Reserch Record Jounal of the Transportation Research Board

El-Basyouny, K., & Kwon, D-W. (2012). Assessing Time and Weather Effects on Collision Frequency by Severity in Edmonton Using Multivariate Safety Performance Functions. Washington D.C., Presented at the 91st TRB Annual Meeting, Transportation Research Board, National Research Council.

Famoye, F., & Singh, K. R, (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science,* 4: 117 – 130.

Famoye, F., & Singh, P.K. (2006). Zero-Inflated Generalized Poisson Regression Model with an application to Domestic Violence Data. Journal of Data Science, 4: 117 – 130.

Famoye, F., Wulu, T. J., & Singh, K. P. (2003) Generalized Poisson Regression with an application to Accident Data. *Journal of Data Science*, 2(2004): 287 – 295.

Farmer, C.M., & Williams, A.F. (2005). Temporal factors in motor vehicle crash deaths. Injury Prevention, 11: 18-23.

Gurmu, S., & Trivedi, P. K. (1996). Excess Zeros in Count Models for Recreational Trips. *Business and Economic Statistics,* 14: 469 – 477.

Haleem, K. M. (2009). Comprehensive Analytical Investigation of the Safety of Unsignalized Intersections. A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Civil, Environmental & Construction Engineering in the College of Engineering and Computer Science at the University of Central Florida Orlando, Florida.

Harb, R., Yan, X., Radwan, E., & Su, X. (2009). Exploring precrash maneuvers using classifications trees and random forests. *Accident Analysis and Prevention,* 41(1): 98 – 107.

Hermans, E., Wets, G., Van Den Bossche., F. (2006). Frequency and Severity of Belgian road traffic accidents studied by state-space methods. *Jounal of Transportation and Statistics,* 99(1): 63 – 76.

Hilbe, J. M. (2011). Negative Binomial Regression. 2nd edition

Hosseinpour, M., Yahaya, A.S., Sadullah, A.F. (2014). Exploring the effects off roadway characteristics on the frequency and severity of head on crashes: Case studies from Malaysian federal roads. *Accident Analysis and Prevention,* 62: 209 – 222.

Hu, M. C., Pavlicova, M., & Nunes. E.V. (2011). Zero-Inflated and hurdle models of count data with extra zeros: examples from and HIV-risk reduction intervention trial. *The American Journal of drug and alcohol abuse, Pubmed,* 37(5): 365 – 75.

Insurance Institute for Highway Safety Highway Loss Data Institute Report

Jackman, S. (2008). Pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory. Stanford University. R package version 0.95.URL http://CRAN.R-project.org/package=pscl.

Jones, B., Janssen, L., & Mannering, F. (1991). Analyis of the frequency and Duration of Freeway Accidnts in Seattle. *Accident Analysis and Prevention*, 23(4): 239 – 255.

Kashani, A.T., Shariat-Mohaymany, A., Ranjbari, A. (2012). Analysis of factors associated with traffic injury severity on rural roads in Iran. *Journal of Injury and Violence Research*, 4(1): 36 – 41.

Khattak, A. J., & Knapp, K.K. (2001). Snow Event Effects on Interstate Highway Crashes. *Journal of Cold Regions Engineering*, 15(4): 219 – 229.

Khorashadi, A., Niemeier, D., Shanker, V., & Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory

Kim, J. K., Kim, S., Ulfarsson, C. F., Porrello, L.A. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis and Prevention*, 39:238 – 251.

Kockelman, K.M., & Kweon, Y.J. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis and Prevention,* 34: 312 – 321.

Kononov, J., & Janson, B.N. (2002). Diagnostic Methodology for the Detection of Safety Problems at Intersections. Jounal Transortation Research Record, 1784(2002): 51 – 56.

Krafft, M., Kullgren, A., Lie, A., & Tingvall, C. G., (2009). The effect of Automatic Emergency Braking on fatal and serious injuries.

Kuruc, R., Sidlo, J., Valent, D., Mlynar, J., & Murarikova, B. (2009). The responsibility of alcohol impaired road users in fatal road traffic accidents. *Brastisl Lek Listy*, 110(12): 802 – 806.

Lagarde, E. (2007) Road traffic injury is an escalating burden in Africa and deserves Proportionate Research Efforts. *ProS Med,* 4(3): 170.

Lambert, D. (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics*, 34:1-14.

Lee, C., & Li, X. (2014). Analysis of injury severity of drivers involved in single- and two-vehicle crashes on highways in Ontario. *Accident Analysis and Prevention,* 71: 286 – 295.

Lee, L. F. (1986). Specification tests for Poisson regression models. *International Economic Review*, 27: 687 – 706.

Lord, D., & Mannering, F. (2010). The statistical analysis of Crash –Frequency Data. *Forthcoming Transportation Report Part A,* 44: 12-24.

Lord, D., & Mannering, F. (2010). The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A,* 44: p. 291-305.

Lord, D., Washington, S. P., & Ivan, J. N. (2004). Statistical Challenges with Modelling Motor Vehicle Crashes: Understanding the Implications of Alternative Approches. Research Report, Center for Transportation Safety. Texas Transportation Institute. Texas A&M University System. 3135 TAMU, College Station, TX, 77843 – 3135.

Manan, W. N. B. (2011). Accident prediction model at un-signalized intersections using multiple regression method. A project report submitted is partial fulfillment of the requirement for the award of the degree of Master of Civil Engineering.

Martensen, H., & Dupon, E. (2013). Comparing single vehicle and multivehicle fatal road crashes: A joint analysis of road conditions, time variables and driver characteristics. *Accident Analysis and Prevention,* 60: 466 – 471.

McCullagh, P., & Nelder, J. (1989). Generalized Linear models. Chapman and Hall, London, 2nd edition.

Miaou, S. P. (1994). The Relationship Between Truck Accidents and Geometric Design of Road Section: Poisson Versus Negative Binomial Regression. *Accident Analysis and Prevention,* 26(4): 471-82.

Miaou, S. P. (1996). Measuring goodness of fit of accident prediction models. (*The National Academies of Sciences, Engineering, and Sciences:* 130.

Miaou, S.P., & Lum, H. (1993). Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accident Analysis and Prevention*, 25(6): 689-709.

Mock, C., Amegashi, J., 7 Darteh, K. (1999). Role of commercial drivers in motor vehicle related injuries in Ghana. *Injury Prevention,* 5: 268 – 271.

Mogaka, E. O., Ng`ang`a, Z., OundoJ., Omolo, J., & Luman, E. (2011). Factors associated with severity of road traffic injuries, Thika, Kenya. *Pan Afri.Med.J.* 8(20).

Mohan, D. (2002). Road Safety in less motorized environments: Future conerns. *International Journal of Epidemiology*, 31(3): 527 – 532.

Mondal, P., Sharma, N., Kumar, A., Vijay, P., Bhangale, U.D., & Tyagi, D. (2011). Are road accidents affected by rainfall? A case study from a large Indian metropolitan city. *Applied Science Technology*, 1(2): 16 – 26.

Moskal, A., Martin, j., & Laumon, B. (2012). Risk factors for injury accidents among moped and motrorcycle riders.

Motor Vehicle Accidents fund (2009/10). MVA Fund Crash and Claims report. MVA publications. Windhoek.

Mullahy, S.P. (1986). Specifications and testing of modified count data models. *Journal of Econometrics*, 33: 341 – 365.

Munk, M.D., Carboneau, D.M., Hardan, M., Ali, F.M. (2008). Seatbelt use in Qatar in association with severe injuries and death in the pre-hospital setting. Prehospital Disaster Med. 23(6): 547 – 552.

Namibia 2011 population & housing Census. Main Report. Namibia Statistics Agency. Windhoek, Namibia.

Namibia Road Safety Council. Strategy & Action Plan, 2009 – 2014. http://www.nrsc.org.na/.../NSRC%20Long%20Term%20Strategy_web.pdf. *Access date: 17/03/2014*

Nangombe, A. N.K., & Neema, I. (2012). Statistical Analysis of Road Traffic Fatalities In Namibia From 2007 – 2009. Research Project in partial fulfilment of Bachelor of Science Honors Degree in Statistics.

Nantulya, V. M., & Reich, M. R. (2002). The Neglected Epidemic: *Road Traffic Injuries in developing countries*, 324: 1139 – 1141. PubMed |Direct Link|

Nassiri, H., & Edrissi, A. (2006). Modeling Truck Accident Severity on Two-Lane Rural Highways. *Scientia Iranica*, 13(2): 193 – 200.

Nelder, J.A., & Wedderburn R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society A*, 135: 370 – 384.

Ohakwe, J., Iwueze, I. S., & Chikezie, D. C. (2011). Analysis of road traffic accidents in Nigeria: *A case study of Obinze/Nikede/Ihengwa road in Imo State, Sourtheastern, Nigeria.* Asian J. Applied Sci. 4: 166 – 175.

Ohakwe, J., Iwueze, I. S., & Chikezie, D. C. (2011). Analysis of Road Traffic Accidents in Nigeria: A case Study of Obinze/Nekede/Iheagwa Road in Imo State, Southeastern, Nigeria. *Asian Journal of Applied Sciences*, 4: 166-175.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., & Jawaren, E. (2004). World report on road traffic injury prevention.Geneva: World Health Organization.

Poch, M., & Mannering, F. L. (1996). Negative Binomial Analysis of Intersection Accident Frequencies. *Journal of Transportation Engineering***,** 122(2): 105-13.

Ponce, J.D.C., Mu`Ñoz, D.R., Andreuccetti, G., Carvalho, D.G.D., Leyton, V. (2011). Alcohol-related traffic accidents with fatal outcomes in the city of Sao Paulo. *Accident Analysis and Prevention*, 43: 782 – 787.

Qiu L., &Nixon, W.A. (2008). Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis. *In Transportation Research Record: Journal of the Transportation Research Board*, No.2055, Transportation Research Board of the National Academics, Washington D.C: 139 – 146.

Qui, L., & Nixon, W. A. (2008). Effects of Adverse Weather on Traffic Crashes. In *Transportation Research Record: Journal of Transportation Research Board*, No 2055. Transportation Research Board, National Research Council, Washington D.C., 139 – 146.

Radun, I., & Summala, H. (2004). Sleep-related fatal vehicle accidents: Characteristics of decisions made by multidisciplinary investigation teams. *Sleep*, 27(2): 224 – 227.

Rodout, M., Demétrio, C. G. B., & Hinde, J. (1998). Models for count data with many zeros. International Biometric Conference.

Rossi, R., Gastaldi, M., Gecchee, G. (2011). Analysis of driver task-related fatigue using driving simulator experiments. *Procedia-Social and Behavioral Sciences*, 20: 666 – 675.

Seyer, K., Newland, C., Terell, M., & Dalmotas, D. (2000). The effect of mass, stiffness and geometry on injury outcome in side impacts – parametric study. *PubMed*, 44: 1- 11.

Shankar, V. N., Mannering, F., & Barfield, W. (1995). Effect of Roadway Geometric and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention*, 27 (3): 371-89.

Sharma, A. K., Landge, V.S., & Deshpande, N.V. (2013). Modeling Motorcycle Accident on Rulal Highway. *International Journal of Chemical, Environmental & Biological Science,* 1(2): 2320-4087.

Siskinda, V., Steinhardta, D., Sheehana, M., O`connorb, T., & Hanksb, H. (2011). Risk factors for fatal crashes in rural Austraria. *Accident Analysis and Prevention*, 43: 1082 – 1088.

Srinivasan, K. K. (2002). Injury severity analysis with variable and correlated thresholds. *Transportation Research Record,* 1784: 132-142.

Usman, T., Fu, L., & Miranda-Moreno, L. F. (2012). A Disaggregate Model for Quantifying the Safety Effects of Winter Road Maintenance Activities at an Operational Level. *Accident Analysis and Prevention*, 48: 368 – 378.

Violanti, J. M. (1998). Cellular phones and fatal traffic collisions. *Accident Analysis and Prevention,* 30: 519 – 524.

Wanvik, P.O. (2009). Effects of road lighting: Analysis based on Dutch accident statistics 1987-2006. *Accident Analysis prevention*, 4: 123 – 128.

WHO, (2004). World Report on Road Traffic Injury Prevention: Summary. World Health Organization, Geneva, Switzerland.

Winkelmann, R., & Zimmermann, K. F. (1994). Count data models for demographic data. *Mathematical Population Studies*, 4(3): 205-221

World Health Organization, 2009. Global Status Report On Road Safety 2013, Geneva, Switzerland: World Health Organization.

Yesilova, A., Kaya, Y., Kaki, B., & Kasapi, I. (2010). Analysis of Plant Protection studies with excess zeros using zero-inflated and negative binomial hurdles mode. *Gazi University journal sciences*, 23(2): 131 – 136.

Yu, R., Abdel-Aty, M., & Ahmed, M. (2013). Bayesian Random Effect Models Incorporating Real – Time Weather and Traffic Data to Investigate Mountainous Freeway Hazardous Factors. *Accident Analysis and Prevention*, 50: 371 – 376.

Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y. (2000). Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. *Accident Analysis and Prevention*, 32: 117 – 125.

Zhu, X., & Srinivasan, S. (2011). A comprehensive analysis of factors influencing the injury severity of large-truck crashes. *Accident Analysis and Prevention*, 43: 49-57.

ANNEXURE

ANNEXURE A: R codes

This is a freely available statistical package and is the software used for the explanatory analysis of this thesis. The R software can be freely downloaded from the site (http://www.r-project.org/.)

Below is some of the R syntax used for the explanatory analysis for the thesis:

Loading data and packages installation

```
> library(foreign)
> BERTHA<-read.table("clipboard",header=T)
> attach(BERTHA)
> install.packages(c("binom", "effects", "gplots", "MuMIn", "car", "gridExtra",
+ "faraway", "fields", "ggplot2", "Hmisc", "lmtest", "dplyr",
+ "nlme", "plyr", "proto", "reshape2", "spam", "zoo",
+ "geepack"))
```

Frequency distribution of the number road fatalities versus distribution of normal data

```
> par(mfrow=c(1,2))
> hist(Fatality, main="Distribution of fatalities")
> hist(rnorm(nrow(BERTHA), mean(Fatality)), main='Distribution of Normal Data',
+      xlab='Distribution of Normal Data')
```

qplot

```
> require(ggplot2)
Loading required package: ggplot2
> qplot(Fatality, log="x", data=BERTHA, xlab="Fatality shown with a logged x-axis")
```

Relationship between each covariate and the number of fatalities

```
> require(ggplot2)
> require(gridExtra)
> p1<-qplot(Week, Fatality, xlab='Week', ylab='Fatality',alpha=I(1/5))
> p2<-qplot(as.factor(Week), Fatality, xlab='Week', ylab='Fatality', geom = 'boxplot')
> p3<-qplot(Month, Fatality, xlab='Month', ylab='Fatality',alpha=I(1/5))
> p4<-qplot(as.factor(Month), Fatality, xlab='Month', ylab='Fatality', geom = 'boxplot')
> grid.arrange(p1, p2, p3, p4, nrow=2, ncol=2)
> p5<-qplot(Region, Fatality, xlab='Region', ylab='Fatality',alpha=I(1/5))
> p6<-qplot(as.factor(Region), Fatality, xlab='Region', ylab='Fatality', geom = 'boxplot')
> p7<-qplot(Day, Fatality, xlab='Day', ylab='Fatality',alpha=I(1/5))
> p8<-qplot(as.factor(Day), Fatality, xlab='Day', ylab='Fatality', geom = 'boxplot')
> grid.arrange(p5, p6, p7, p8, nrow=2, ncol=2)
> p9<-qplot(Type, Fatality, xlab='Type', ylab='Fatality',alpha=I(1/5))
> p10<-qplot(as.factor(Type), Fatality, xlab='Type', ylab='Fatality', geom = 'boxplot')
> p11<-qplot(Priority, Fatality, xlab='Priority', ylab='Fatality',alpha=I(1/5))
> p12<-qplot(as.factor(Priority), Fatality, xlab='Priority', ylab='Fatality', geom = 'boxplot')
> grid.arrange(p9, p10, p11, p12, nrow=2, ncol=2)
> p13<-qplot(Injured, Fatality, xlab='Injured', ylab='Fatality',alpha=I(1/5))
> p14<-qplot(as.factor(Injured), Fatality, xlab='Injured', ylab='Fatality', geom = 'boxplot')
> p15<-qplot(Time, Fatality, xlab='Time', ylab='Fatality',alpha=I(1/5))
```

> p16<-qplot(as.factor(Time), Fatality, xlab='Time', ylab='Fatality', geom = 'boxplot')

> grid.arrange(p13, p14, p15, p16, nrow=2, ncol=2)

Fitting Linear models

> model<-lm(Fatality~Week+Month+Region+Day+Type+Priority+Injured+Time,data=BERTHA)

> summary(model)

> model1<-
lm(Fatality~as.factor(Week)+Month+Region+Day+Type+Priority+Injured+Time,data=BERTHA)

> summary(model1)

> model2<-
lm(Fatality~Week+as.factor(Month)+Region+Day+Type+Priority+Injured+Time,data=BERTHA)

> summary(model2)

> model3<-
lm(Fatality~Week+Month+as.factor(Region)+Day+Type+Priority+Injured+Time,data=BERTHA)

> summary(model3)

> model4<-
lm(Fatality~Week+Month+Region+as.factor(Day)+Type+Priority+Injured+Time,data=BERTHA)

> summary(model4)

> model5<-
lm(Fatality~Week+Month+Region+Day+as.factor(Type)+Priority+Injured+Time,data=BERTHA)

> summary(model5)

> model6<-lm(Fatality~Week+Month+Region+Type+as.factor(Priority)+Injured+Time,data=BERTHA)

> summary(model6)

> model7<-lm(Fatality~Week+Month+Region+Type+Priority+as.factor(Injured)+Time,data=BERTHA)

> summary(model7)

> model8<-lm(Fatality~Week+Month+Region+Type+Priority+Injured+as.factor(Time),data=BERTHA)

> summary(model8)

Assumption checking

```
> par(mfrow=c(3,2))

> termplot(model,se=T,partial.resid=T)

> require(car)

> ncvTest(model3)

> durbinWatsonTest(model3)

 > plot(residuals(model3)[1:100], type='b')

> abline(h=0)

> hist(residuals(model3), nclass = 20)
```

GLMS

```
> library(MASS)

> library(pscl)

> library(foreign)

> BB<-read.spss(file.choose())

> summary(BB)

> formula1<-
(Fatality~Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+
Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4)

> GLMP<-glm(formula1,data=BB,family=poisson)

> summary(GLMP)

> AIC(GLMP

> GLMNB<-glm.nb(formula1,data=BB)

> summary(GLMNB)

> AIC(GLMNB)

> ZIP<-
zeroinfl(Fatality~Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Re
```

```
gion9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4|Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4,data=BB,dist="poisson")

> summary(ZIP)

> ZINB<-
zeroinfl(Fatality~Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4|Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4,data=BB,dist="negbin")

> summary(ZINB)

> AIC(ZIP)

> AIC(ZINB)

> HP<-
hurdle(Fatality~Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4,data=BB,zero.dist="poisson",link="logit")

> summary(HP)

> AIC(HP)

>  HNB<-
hurdle(Fatality~Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4,data=BB,zero.dist="negbin",link="logit")

> summary(HNB)

> AIC(HNB)

ZINB<-
zeroinfl(Fatality~Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4|Day1+Region1+Region2+Region3+Region4+Region5+Region6+Region7+Region8+Region9+Region10+Region11+Region12+Time1+Time2+Time3+Type1+Type2+Type3+Type4,data=BB,dist="negbin")
```